Development of new approaches to study the role of chromatin in dna damage response

Muhammad Shoaib

UNIVERSITY OF PARIS-SUD XI
FACULTY OF MEDICINE

Graduate School : **Oncology: Biology, Medicine, Health**
Specialty : **Biochemistry, Cell & Molecular Biology**

Thesis submitted to obtain the degree of
DOCTOR OF THE UNIVERSITY OF PARIS-SUD XI

Presented by
Muhammad SHOAIB

Defended on **6^{th} of October 2011**

DEVELOPMENT OF NEW APPROACHES TO STUDY THE ROLE OF CHROMATIN IN DNA DAMAGE RESPONSE

Thesis Supervisor : **Dr. Vasily OGRYZKO**

Thesis Jury :

Dr. Myriam FERRO	CEA, Grenoble	Reviewer
Dr. Carl MANN	CEA, Saclay	Reviewer
Dr. Thierry GRANGE	Institut Jacques Monod	Examiner
Dr. François DAUTRY	Institut André Lwoff	Examiner
Prof. Christian AUCLAIR	ENS, Cachan	President
Dr. Vasily OGRYZKO	Institut Gustave Roussy	Thesis Supervisor

UNIVERSITÉ PARIS-SUD XI
FACULTÉ DE MÉDECINE

École Doctorale : **Cancérologie: Biologie, Médecine, Santé**
Discipline : **Biochimie, Biologie Cellulaire et Moléculaire**

Thèse pour l'obtention du grade de
DOCTEUR DE L'UNIVERSITÉ PARIS-SUD XI

Présentée et soutenue publiquement par
Muhammad SHOAIB

le **6 octobre 2011**

DÉVELOPPEMENT DE NOUVELLES APPROCHES POUR ÉTUDIER LE RÔLE DE LA CHROMATINE EN RÉPONSE AUX DOMMAGES DE L'ADN

Directeur de Thèse : **Dr. Vasily OGRYZKO**

Devant le jury composé de :

Dr. Myriam FERRO	CEA/Grenoble	Rapporteur
Dr. Carl MANN	CEA/Saclay	Rapporteur
Dr. Thierry GRANGE	Institut Jacques Monod	Examinateur
Dr. François DAUTRY	Institut André Lwoff	Examinateur
Prof. Christian AUCLAIR	ENS, Cachan	Président
Dr. Vasily OGRYZKO	Institut Gustave Roussy	Directeur de Thèse

dedicated to my parents, Shahida SALAM & Abdul SALAM
for they raised me, supported me, taught me, and loved me.

Acknowledgments

Foremost, I would like to express my sincere gratitude to my supervisor and mentor Dr. Vasily Ogryzko for accepting me in his group and for the confidence he entrusted to me knowing that I had zero experience in molecular biology. I am grateful to him for the time he spent and the efforts he put to teach me the very basics of molecular biology starting from holding a micro-pipette to the complex proteomics experiments on anything, anytime basis. His continuous support, enthusiasm, availability and motivation not only nurtured my scientific curiosity but also provided me a strong methodological insight towards answering many outstanding questions in molecular biology. I have certainly benefited from his intelligent ideas, problem solving skills and his eagerness to explain me science.

I am extremely thankful to Dr. Carl Mann and Dr. Myriam Ferro, Engineer-Researcher from CEA Saclay and CEA Grenoble, respectively, for reviewing this thesis and helping me to ameliorate the quality of thesis. I am also thankful to Prof. Christian Auclair, Professor at ENS Cachan, Dr. François Dautry, Researcher at Institute André Lwoff, and Dr. Thierry Grange, Researcher at Institute Jacques Monod, for evaluating this work by participating as examiners.

I am extremely grateful to Dr. Marc Lipinski, the then Director of UMR8126, for his efforts in getting me admitted in the Masters program in oncology and his continuous support and guidance throughout my stay in this lab. I would also like to appreciate Dr. Joëlle Wiels, the current Director of UMR8126, for her help on various scientific as well as administrative fronts. I am thankful to them for their positive criticism on my work during lab seminars, which certainly allowed me to evaluate and eventually improve my work. I am also thankful to Mme. Muriel Nicoletti, the secretary of Graduate School of Oncology, for her patience and assistance that she provided during the last six years.

I would like to express my deep gratitude to Chloé Robin, Meriem Hasmim, Yosra Messai and Zaeem Noman for their strong support, encouragement and caring, especially during difficult times, and for providing me a helping hand whenever needed. Their advices have always been a valuable asset to me and my stay in IGR would have been much less fun without their significant presence. I would also like to thank Arman Kulyyassov, who was a post-doc in our group, for his help and support on many projects. I would like to take this opportunity to thank all the current and previous members of our group including Emilie Cochet, Ulykbek Kairov, Evelyne Saade, Damien Vertut and Martine Comisso for their help and guidance during all this time.

I would also like to extend my thanks to many internees that have worked with me during last six years in this lab and provided me the opportunity to pass on knowledge that I have acquired. I would especially thank Kinga Winczura, with whom I enjoyed working the most and became very good friends. Besides, while I was busy writing my thesis, she obtained some very interesting results related to my project. Many thanks to her for all the hard work she has done and the motivation she showed on my project.

During this work I have collaborated with many colleagues for whom I have great regard,

and I wish to extend my warmest thanks to Dr. Filippo Rosselli, Dr. Valeria Naim and Dr. Philippe Dessen. I have learned a lot during long discussions with them and the ideas generated during these discussions have certainly helped me achieve my goals.

I would also like to express my appreciation to Syed Qasim Raza, Shoaib Ahmad Malik, Abdul Qadir, Adnan Arshad and Afaq Ali for their time, friendship and never-ending amusing discussions during lunch and tea time in IGR, which were not only entertaining but also refreshing for all of us. I appreciate all my acquaintances in IGR for their time and help they have offered to me during last six years including M'barka Mokrani, Marie Boutet, Intissar Akalay, Laëitia Thomas, Charle-Henri Gattoliat and Petr Dmitriev.

I cannot forget to mention my friends outside IGR for their support, encouragement and friendship during my stay in Paris. Special mention to Kashif Saeed, Aamer Baqai, Husain Parvez, Shehzad Hanif, Khurram Khurshid and Kashif Zahoor with whom I have learned and enjoyed a great deal during traveling, eating, endless hours of table tennis and discussions on a wide variety of subjects. I would also like to thank Kashif Mehmood, Abdul Rauf, Shehab Aamer, Sheraz Khan, Imran Siddiqui, and Masood Maqbool for their kind support. My stay in France would have been much less exciting and pleasurable if it were not the company of these friends.

My mother and father have always been a consistent and perpetual source of motivation and encouragement throughout my studies and personal life. I owe them all my achievements till today for their profound devotion, unconditional love and endless faith they have in me. This thesis is as much an achievement of theirs as it is mine. I am highly indebted to them for their strong affection towards me and their earnest attachment to my career goals. I am also thankful to my kind sister Sobia Salam and my loving brothers Muhammad Atif and Muhammad Jawad for their emotional support and encouragement over the last six years.

In the end, I would like to humbly acknowledge the Higher Education Commission of Pakistan, Société Française d'Hématologie and my PhD. supervisor for the financial support during the last six years.

Abstract

In eukaryotic cells, the genome is packed into chromatin, a hierarchically organized complex composed of DNA and histone and non-histone proteins. In this thesis we have addressed the role of chromatin in cellular response to DNA damage (DDR) using various methodologies encompassing functional genomics and proteomics. First, we analyzed histone post-translational modifications (PTM) in the context of specific kind of DNA lesions (ICL-Interstrand Crosslinks) in Fanconi anemia using quantitative proteomics methodology, SILAC (Stable Isotope Labeling of Amino acids during Cell Culture). Using mass spectrometry (MS), we have successfully identified and quantified a number of histone PTM marks in histone H3 and H4, mainly acetylations and methylations, which have shown dependence upon functional FA-pathway. As a next step, we applied a functional genomics approach to study DDR in FA cells. In this analysis we first monitored the expression profile of histone modifying enzymes related to histone acetylations and methylations. Our results suggest some correlations between histone PTMs and gene expression of histone modifying enzymes, although conclusive evidence warrants further investigations. Next, we analyzed the total transcriptome after DNA damage induction in FA mutant and wild type cells. We also included in this analysis IR irradiation, in an attempt to dissociate more generic DDR from more specific changes that are associated with the role of FA pathway to the DNA ICLs. By performing a factorial interaction analysis, we were able to isolate the part of transcriptional response to DNA damage that was requiring functional FA pathway, as well as the genes that were sensitized to DNA damage by the inactivation of FA pathway. In the final part of the thesis, we attempted to solve one of the limitations that we encountered in the histone PTM analysis. The current approaches used to study histone PTMs from particular loci involves classical chromatin immunoprecipitation, which due to involvement of formaldehyde crosslinking render the protein part mostly unavailable for MS-based proteomics. We have proposed a novel methodology, which is based upon the biotin tagging of histones proximal to a protein of interest and subsequent purification of nucleosomes carrying the tagged histone. This methodology does not involve any crosslinking, enabling us to purify histones from specific loci, and subject them to large scale MS-based histone PTM analysis. A time dimension can also be added to our approach, as we can follow the modification status of particular fraction of histones once they get biotinylated. Another advantage is the use of alternate variant histones, which allows us to study the PTM profile of different functional states of chromatin.

KEYWORDS: chromatin, histone PTM,DNAdamage response, transcriptome analysis, SILAC, biotinylation, native chromatin immunoprecipitation.

Résumé

Le génôme des cellules eucaryotes est condensé au sein d'une structure complexe hiérarchiquement organisée : la chromatine. La chromatine est composée d'ADN, de protéines histone et non-histone. Cette thèse a pour but d'étudier le rôle de la chromatine dans la réponse cellulaire aux dommages de l'ADN (DDR) par les méthodologies de génomique fonctionnelle et de protéomique. Nous avons tout d'abord analysé les modifications post-traductionnelles (PTM) des histones dans le cadre des pontages inter-brins (ou "Interstrand Crosslinks", ICL), type particulier de lésions de l'ADN, en choisissant le modèle de l'Anémie de Fanconi (FA). Ceci a été réalisé grâce aux techniques de protéomique quantitative SILAC (Stable Isotope Labeling of Amino acid during Cell culture) et de spectrométrie de masse (MS). Nous avons ainsi réussi à identifier et à quantifier de nombreuses PTMs dans les histones H3 et H4, et à démontrer que certaines de ces PTM sont dépendantes d'une voie fonctionnelle de la signalisation de FA. Nous avons également approfondi l'étude des DDR dans les cellules de FA par une approche de génomique fonctionnelle. Pour cela, nous avons analysé le profil d'expression d'enzymes associées à l'acétylation et à la méthylation des histones. Nos résultats suggèrent l'existence de corrélations entre le profil d'expression de ces enzymes et les PTMs des histones. Des études complémentaires sont nécessaires en vue de confirmer ces corrélations. Nous avons également comparé le transcriptome de deux lignées cellulaires de FA (mutée en FANCC et corrigée en FANCC) après induction de dommages à l'ADN. Afin de différencier les changements spécifiquement associés à la voie de signalisation de FA en réponse aux ICL de l'ADN des réponses plus générales aux dommages de l'ADN, nous avons inclus des cellules traitées par rayonnement ionisant. En réalisant une analyse d'interactions factorielles, nous avons pu identifier une réponse transcriptionnelle aux dommages de l'ADN nécessitant une voie fonctionnelle de la signalisation de FA. Nous avons également tenté de pallier aux limitations rencontrées dans l'analyse des PTMs des histones. En effet, les PTMs des histones que nous avons identifiées représentent l'ensemble des modifications, c'est-à-dire les PTMs concernant les histones se trouvant immédiatement à proximité du site du dommage et en relation directe avec celui-ci, et les PTMs se trouvant à distance du dommage et pouvant ne pas être en relation directe avec celui-ci. Les approches courantes pour identifier les PTMs se trouvant à des loci particuliers sont basées sur l'immunoprécipitation classique de la chromatine où l'utilisation de formaldéhyde altère les protéines, ce qui en rend impossible l'analyse par MS. Nous avons proposé une nouvelle méthodologie basée sur la biotinylation expérimentale d'histones situées à proximité d'une protéine particulière, suivie de la purification des nucléosomes contenant ces histones biotinylées. Contrairement à l'immunoprécipiatation classique de la chromatine, cette méthode n'induit pas d'altération des protéines, permettant ainsi de purifier les histones à partir d'un locus spécifique et d'analyser à grande échelle leurs PTMs par MS. Cette approche permet aussi de suivre dans le temps les PTMs d'une fraction des histones juste

après leur biotinylation. Enfin, elle présente l'avantage de pouvoir étudier le profil des PTMs de différents états fonctionnels de la chromatine grâce à l'utilisation de variants d'histones.

MOTS CLÉS : chromatine, modification post-traductionnelle des histones (PTM), réponse aux dommages de l'ADN (DDR), analyse du transcriptome, SILAC, biotinylation, immuno-précipitation de chromatine native.

Contents

List of Figures

List of Tables

Nomenclature

ADP	Adenosine Diphosphate
ASF-1	Anti Silencing Factor-1
ATM	Ataxia Telangiectasia Mutated
ATP	Adenosine Triphosphate
ATR	Ataxia Teleangiestasia and Rad3 Related
Au-PAGE	Acid-Urea Polyacrylamide Gel Electrophoresis
BAP	Biotin Acceptor Peptide
BBAP	B-lymphoma and BAL-associated protein
BER	Base Excision Repair
BirA	Biotin Ligase A
BRCA1	Breast Cancer Associated-1
BRCT	BRCA1 C-terminus
BSA	Bovine Serum Albumin
C-terminal	Carboxy-terminal
CAF-1	Chromatin Assembly Factor-1
CBP	Creb-Binding Protein
CDK	Cyclin Dependent Kinase
CE-ESI-MS	Capillary Electrophoresis-Electrospray Ionization-Mass Spectrometry
CenpA	Centromere protein A
CenpA	Ubiquitin Conjugating Enzyme-13
CHD1	Chromodomain Helicase DNA Binding Protein-1
ChIP	Chromatin Immunoprecipitation

CID	Collision Induced Dissociation
CK2	Casein Kinase-2
CMV	Cytomegalo Virus
Co-REST	Corepressor to RE1 Silencing Transcription Factor/Neural Restrictive Silencing Complexes
COMPASS	Complex Proteins Associated with Set1
Cy3	Cyanine3
DamID	DNA Adenine Methyltransferase Identification
DDR	DNA Damage Response
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic Acid
DNA-PK	DNA Dependent Protein Kinase
DNMT1	DNA Methyltransferase-1
DNMT3L	[DNA Methyltransferase-3L
DSB	Double Strand Break
DTT	Dithiothreitol
ECD	Electron Capture Dissociation
EHMT1	Euchromatic Histone-Lysine N-methyltransferase-1
EIC	Extracted Ion Chromatogram
emPAI	Exponentially Modified Protein Abundance Index
ESI	Electrospray Ionization
ETD	Electron Transfer Dissociation
EYA1	Eyes Absent-1
EZH2	Enhancer of Zest-2
FA	Fanconi Anemia
FACT	(Facilitates Chromatin Transcription
FHA	forkhead-associated
FOSL1	Fos-Related Antigen-1

FT-ICR	Fourier Transform - Ion Cyclotron Resonance
GeLCMS	Gel Enhanced Liquid Chromatography Mass Spectrometry
GFP	Green Fluorescent Protein
GNAT	GCN5-related N-acetyltransferase
GO	Gene Ontology
H/L	Heavy/Light Ratio
H2A-Bbd	H2A-Bar Body deficient
HA	Hydroxyapatite
HA-Tag	Hemagglutinin
HAT	histone acetyltransferase
HCl	Hydrochloric Acid
HDAC	Histone Deacetylase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HFD	Histone Fold Domain
HILIC	Hydrophilic Interaction Chromatography
HIRA	HIR Histone Cell Cycle Regulation Defective Homolog A
HKDM	Histone Lysine Demethylase
HKMT	Histone Lysine Methyltransferase
HMR	Hidden MAT Right
HMT	histone methyltransferase
HP1	Heterochromatin Protein-1
HPCE	High Performance Capillary electrophoresis
HPLC-MS/MS	High Performance Liquid Chromatography-Tandem Mass Spectrometry
HPMC	Hydropropylmethylcellulose
HR	Homologous Recombination
HRP	Horse Raddish Peroxidase
HSB	High Salt Buffer
ICL	Interstrand Crosslink

ICPL	Isotope Coded Protein Label
IEX	Ion Exchange Chromatography
IMAC	Immobilized Metal Affinity Chromatography
ING	Inhibitor of Growth
INHAT	[Inhibitor of Histone Acetyltransferase
IR	Ionizing Radiation
IRIF	Ionizing Radiation Induced Foci
iTRAQ	Isotope Tags for Relative and Absolute Quantification
JMJD6	Jumonji Domain Containing-6
KDM	Lysine Demthylase
KDM2A	Lysine Demethylase-2A
LC	Liquid Chromatography
LDS	Lithium Dodecyl Sulphate
LSB	Low Salt Buffer
LSD1	Lysine Specific Demthylase-1
m/z	mass/charge
MALDI	Matrix Assisted Laser Desorption Ionization
MDC1	Mediator of DNA Damage Checkpoint Protein-1
MLL	Mixed Lineage Leukemia
MMC	Mitomycin C
MMR	Mismatch Repair
MNase	Micrococcal Nuclease
MOF	Males Absent on the First
MOP	Methoxypsoralen
MRM	Multiple Reaction Monitoring
MRN	Mre11/Rad50/Nbs1
mRNA	Messenger RNA
MS	Mass Spectrometry

MSK1/2	Mitogen and Stress-Activated Protein Kinase
MYST	MOZ, YBF2/SAS3, SAS2 and Tip60
N-CoR/SMRT	Nuclear Receptor-Corepressor/Silencing Mediator for Retinoid and Thyroid Hormone Receptors
N-terminal	Amino-terminal
NAP1L1	Nucleosome Assembly Protein 1-Like 1
NChIP	Native Chromatin Immunoprecipitation
NER	Nucleotide Excision Repair
NHEJ	Non-Homologous End Joining
NuRD	Nucleosome Remodeling and Histone Deacetylation
P-TEFb	Positive Transcription Elongation Factor-b
p53BP1	p53 Binding Protein-1
PBS	Phosphate Buffered saline
PCAF	P300/CBP-Associated Factor
PCNA	Proliferating Cell Nuclear Antigen
PHD	Plant Homeodomain
PHF8	PHD Finger Protein-8
PI3K	Phosphatidyl Inositol 3' Kinase
PIKK	Phosphatidyl Inositol 3' Kinase-Related Kinases
PIM1	Proto-oncogene Serine/Threonine-Protein Kinase Pim-1
PKD2	Polycystic Kidney Disease-2
PMSF	Phenylmethanesulfonylfluoride
PP2A	Protein Phosphatase-2A
PP4C	Protein Phosphatase-4C
PPP	Protein-Protein Proximity
PRDM2	PR-Domain Zinc Finger Protein-2
PRMT	Protein Arginine Methyltransferase
PTM	Post-translational Modification

PUB-MS	Proximity Utilizing Biotinylation - Mass Spectrometry
PVDF	Polyvinylidene Fluoride
rDNA	Replication Independent
rDNA	Ribosomal Deoxyribonucleic Acid
RNA	Ribonucleic Acid
RP	Reverse Phase
RSC	Remodels the Structure of Chromatin Complex
SAM	S-Adenosylmethionine
SCX	Strong Cation Exchange Chromatography
SDS-PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
SET	SU(VAR)39, Enhancer of Zest, Trithorax
SH2	Src Homology 2
SILAC	Stable Isotope Labeling of Aminoacids during Cell Cuture
SMYD1	SET and MYND Domain Containing-1
SNF2	Sucrose Non Fermentable-2
SWI/SNF	SWItch/Sucrose NonFermentable
TCR	T Cell Receptor
TIC	Total Ion Chromatogram
Tip60	Tat Interacting Protein-60
TMT	Tandem Mass Tags
TOF	Time of Flight
TRRAP	Transformation-Domain Associated Protein
TSA	Trichostatin A
TSS	Transcription Start Site
UBZ	Ubiquitin Binding Zinc Finger Domain
UHRF1	Ubiquitin-like, containing PHD and RING Finger Domains-1
UV	Ultra Violet
UVA	Ultra Violet-A

UVC	Ultra Violet-C
V(D)J	Variable(Diverse)Joining
WCX-HILIC	Weak Cation Exchange-HILIC
WICH	WSTF-ISWI ATP-dependent chromatin-remodeling complex
WSTF	WilliamsBeuren Syndrome Transcription Factor
XPC	Xeroderma Pigmentosum Protein C

1 Introduction

1.1 Epigenetics

1.1.1 The Central Dogma of Molecular Biology

Epigenetic phenomena have recently gained a renewed interest and represent a topic of active research in current biology, owing in part due to going "against the grain" of the Central Dogma of Molecular Biology. If understood in an over-simplified way, the Central Dogma professes a "genocentric" view of a biological system, according to which all information necessary to define the state of an organism (up to the reaction norm due to the effects of environment) is contained in the sequence of its genome. Although it adequately describes the situation in simple single-cellular organisms, such as bacteria, the "genocentric" view certainly runs into problem when one start to deal with the multicellular organisms that exhibit the phenomenon of cellular differentiation. Different cell types from the same organism can be taken from an organism and propagated in a cell culture, in identical environmental conditions. The genome sequence of these different cell types is identical (as demonstrated by the phenomenon of somatic cloning), nevertheless they exhibit very different but stable phenotypes. This observation suggests that there should be an additional information responsible for the maintenance of the stable differentiated phenotypes. Understanding the nature and the mechanisms of processing and propagation of this information represent one of the major topics in the epigenetic studies.

In fact, the Central Dogma, as it was first formulated by Francis Crick in 1958 and re-stated in 1970 [Crick, 1970] asserts:

> "The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred from protein to either protein or nucleic acid".

A careful reading of this definition shows that it leaves the door open for other types of information that could be required to specify the state of the organism (e.g. stored in macromolecular conformations, interactions and post-translational modifications) and might propagate independently from the DNA sequence. Thus, while the existence of epigenetic phenomena conflicts with the 'genocentric' view of biological systems, it is compatible with the Central Dogma of Molecular Biology as it was formulated by Crick. The importance of epigenetics has also been renewed more recently by James Watson, while celebrating the 50th anniversary of the double helix, in 2003. He gave the following comment while revisiting the Central Dogma:

> "The major problem, I think, is chromatin.... you can inherit something beyond the DNA sequence. That's where the real excitement of genetics is now" [Watson, 2003].

1.1.2 The Changing Definitions of Epigenetics

Historically, the term epigenetics was used to describe various biological phenomena that could not be explained by genetic principles. Towards the middle of the twentieth century the field of genetics was rapidly progressing, however, the embryologists and developmental biologists were not yet enlightened by the cause and effects of gene regulation. There were a few leading scientists who realized that genetics and developmental biology were indeed related and should eventually come together in a common discipline. Conrad Waddington (1905-1975), who was an expert in both fields of research, took the Greek word epigenesis, a theory of development which proposed that the early embryo was undifferentiated, and transformed it to epigenetics. Although Waddington was given credit for coining the term epigenetics, his initial understanding of epigenetics was very broad and not much different from embryology. According to him, all those events which lead to the unfolding of the genetic program for development can be grouped under the term epigenetics. However, in 1942, Waddington defined epigenetics more precisely as "the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being" [Waddington, 1942]. Later, in 1957, he proposed the idea of an epigenetic landscape to represent the process of cellular decision making during various phases of development in a multicellular organisms [Waddington, 1957] [Figure1.1].

Over the years, after its birth, the term epigenetics was very scarcely used and often confused with the developmental biology. In the decades of 70's and 80's very few people used the term in the same manner as Waddigton did. However, in late 1980's and early 1990's, the term moved into a new era where it began to be considered as a distinct field of biology instead of rather a concept. In 1987, Robin Holliday described the inheritance of epigenetic

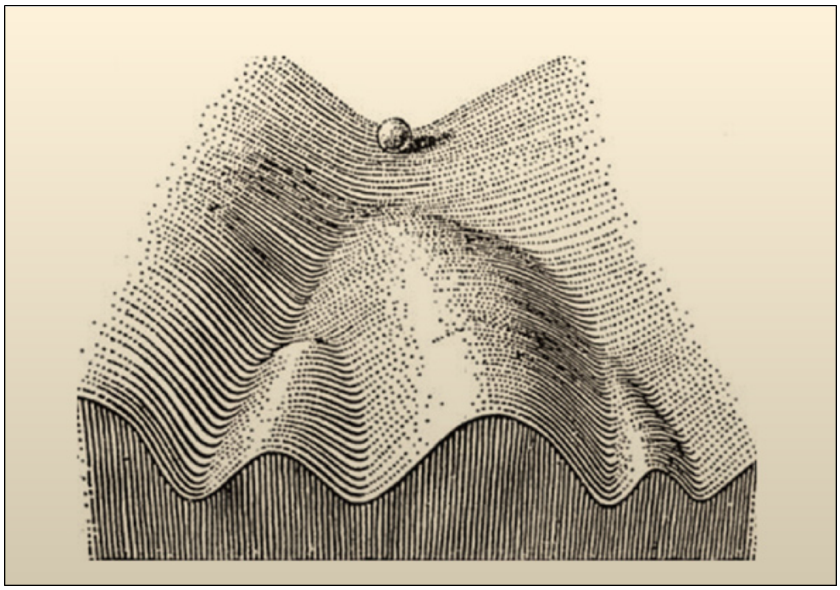


Figure 1.1: Waddington's Epigenetic Landscape that defines different stable trajectories that a cell can take during development and differentiation.

defects where he revisited the term epigenetics and applied it to a scenario where he described that changes in DNA methylation also changed the gene activity, in an attempt to explain epigenetics in a more specific manner [Holliday, 1987]. At that time the word epimutation was introduced to describe heritable changes in genes, which were not due to changes in DNA sequence. This publication is considered as the critical paper that laid the foundations for the explosion in the use of the term "epigenetics" in 1990s.

Later, in 1992, Hall described epigenetics as "the sum of the genetic and non genetic factors acting upon cells to selectively control the gene expression that produces increasing phenotypic complexity during development" [Hall, 1992]. This decisive change in the way how the term epigenetics is defined came about as the molecular mechanisms controlling gene activity and the cellular inheritance of phenotypes began to be unraveled. Holliday's work on cell memory paved way for further narrowing the field as the decade progressed. He wrote that "epigenetics can be defined as the study of the mechanisms of temporal and spatial control of gene activity during the development of complex organisms". He also added that, "Mechanisms of epigenetic control must include the inheritance of a particular spectrum of gene activities in each specialized cell. In addition to the classical DNA code, it is necessary to envisage the superimposition of an additional layer of information which comprises part of the hereditary material, and in many cases this is very stable. The term epigenetic inheritance has been introduced to describe this situation" [Holliday, 1990]. A few years later Holliday redefine epigenetics as 1- The study of the changes in gene expression, which occur in organisms with differentiated cells, and the mitotic inheritance of given patterns of gene expression, and 2- Nuclear inheritance which is not based on differences in DNA sequence [Holliday, 1994]. In 1996, a book entitled "Epigenetic Mechanisms of Gene Regulation" defined the term as "The study of mitotically and/or meiotically heritable changes in

gene function that cannot be explained by changes in DNA sequence".

By the end of twentieth century, epigenetics had already been recognized as a sub discipline of biology owing to the unraveling of the molecular mechanisms controlling gene activity and the inheritance of cell phenotypes. Epigenetics, in a broad sense, is considered to be a bridge between genotype and phenotype, a phenomenon that changes the final outcome of a locus or chromosome without changing the underlying DNA sequence. For instance, despite the identical genotype of vast majority of cells in a multicellular organism, organismal development generates a diversity of cell types with disparate, yet stable, profiles of gene expression and distinct cellular functions. Thus, cellular differentiation may be considered an epigenetic phenomenon, largely governed by changes in what Waddington described as the "epigenetic landscape" rather than alterations in genetic inheritance [Figure1.1].

Although the distinction between DNA and non-DNA inheritance can be easily made, there are no simple criteria for distinguishing between genetic and epigenetic phenomena. In general, genetics today deals with the transmission and processing of information in DNA, whereas epigenetics deals with its interpretation and integration with information from other sources and reflect this information into a phenotype. Epigenetics is therefore concerned with the systems of interactions that lead to predictable and usually functional phenotypic outcomes. Thus in today's modern terms, epigenetics can be mechanistically defined as "The sum of the alterations to the chromatin template that collectively establish and propagate different patterns of gene expression (transcription) and silencing from the same genome" [Allis et al., 2007].

1.1.3 Epigenetic information: Heritability vs Stability

The above discussion suggests an explanation of how the meaning of the current term 'epigenetic' came about. However, it appears that the requirement of 'heritability', currently emphasized as the criterion for a particular phenomenon to fall under the scope of epigenetic research, might be too restrictive.

Terminally differentiated cells such as neurons and muscle cells live for decades, maintaining their distinct phenotypic differences in spite of environmental stresses, thermal noise and DNA damage/repair (Green and Almouzni, 2002; Ikura and Ogryzko, 2003). These cells do not proliferate, thus the term 'heritable' does not apply to their distinct and stable traits. On the other hand, it is reasonable to expect that maintenance and propagation of information responsible for these traits employs mechanisms that are similar to those that are utilized in the replicating cells. Thus, although one cannot use the criteria of 'heritability', there are good reasons to qualify these variations (and the underlying mechanisms) as epigenetic. The term "epigenetic stability" refers to a broader phenomenon that encompasses maintenance of phenotypic traits in both replicating and non-replicating cells independent from DNA sequence [Ogryzko, 2008]; [Viens et al., 2006].

1.1.4 Role of Epigenetic Information

Why life needs epigenetic information? As discussed above, one of its roles is in facilitating maintenance of differentiated cell phenotypes during development of multicellular organisms, that is in supporting alternative interpretations of genetic information. However, one might argue that from evolutionary perspective, a more fundamental and primary role could have been protection of genetic information. Cells have developed systems to distinguish between their own DNA and foreign DNA, as a form of defense mechanism to protect themselves from invading DNA. One example of such mechanisms is seen in bacteria, where a restriction endonuclease and DNA methylase enzymes target the same DNA sequence, depending on its methylated status. A foreign DNA invading the cell is not methylated and thus becomes a target of the endonuclease. However, the same target sequence present originally in the cell is fully methylated, and the methylated state is maintained after replication due to the action of methylase. Thus, cellular machinery that helps cell to protect its own genetic information is based on recognition of epigenetic information associated with DNA.

A similar mechanism of replication of methylated status of DNA serves additional protective purpose, by helping to prevent errors during replication of genetic information. In order to correctly remove the erroneously incorporated nucleotide, mismatch repair mechanisms need to recognize the parental strand of the newly duplicated DNA. While the newly replicated DNA is in the hemi-methylated state, the parental strand is labeled with methyl, which allows the mismatch repair machinery to correctly identify which base of a mismatch pair has to be removed. Importantly, this mechanism works only if there is a time window when the hemi-methylated DNA has not yet been converted to the fully methylated state. Thus, the increased accuracy of replication of genetic information is ensured by the difference between the rates of replication of genetic and epigenetic information, making mismatch repair an example of a general kinetic proofreading scheme [Hopfield, 1974].

In the above examples, epigenetic information acts as a part of protection system serving to differentiate the 'self' from the 'foreign' and the 'old' from the 'new'. Thus, at least two distinct roles of epigenetic information can be distinguished: interpretation of genetic information and its protection. What could be the evolutionary relation between these two functions? We can speculate that the protection of genetic information might have emerged first in evolution, since it would be beneficial for the single cellular organisms. Only later, epigenetic mechanisms could be recruited to play a role in differentiation. Supporting this idea is the evidence that many epigenetic mechanisms appear to be related to mechanisms of suppression of parasitic genetic elements. Only after the establishment of the machinery that allows recognition of different epigenetic marks and channeling the signals encoded in these marks along appropriate response pathways, these mechanisms of processing of epigenetic information could be recruited for other purposes, such as for stabilizing different alternative states of the same organism.

1.1.5 Epigenetic Templating

After having discussed the role of epigenetic information and its evolving nature, we next ask this question, how epigenetic information can be propagated and maintained? Studies of DNA methylation (an earliest recognized epigenetic mark) provided an early clue on one mechanism, which is based on a different behavior of the so called maintenance methylase enzyme towards un-methylated versus hemi-methylated DNA. The semi-conservative DNA replication of a methylated double strand gives rise to hemi-methylated bases comprising a methylated parental strand and an unmethylated newly synthesized one. Methyltransferase enzymes bind to the hemi-methylated sites where they methylate the new strand. The modifying machinery is not recruited to non methylated strands.

Later on, this model motivated Francis Crick to propose a similar mechanism as the molecular basis for neurobiological memory, essentially an epigenetic phenomenon [Crick, 1984]. He postulated that i) the strength of a synapse is determined by phosphorylation of a protein molecule, ii) this protein can form dimers (or oligomers) and iii) the kinase responsible for the modification will only modify monomer in a dimer that has second monomer already phosphorylated.

Consistent with the recognized role of chromatin as a principal carrier of epigenetic information, very similar mechanisms have been suggested for perpetuation of post-translational histone modifications. Some histone acetyltransferase (HAT) complexes could be preferentially recruited to acetylated chromatin due to the presence of bromodomains, recognizing acetylated lysines, whereas the chromodomain containing histone methyl transferase (HMT) complexes could be likewise recruited to methylated chromatin [Owen et al., 2000]; [Ogryzko, 2001]; [Jenuwein and Allis, 2001]. Also, domain structure of some protein kinases suggests that similar mechanism could operate not only on the level of chromatin, but in controlling the organization of cytoplasm as well. Tyrosine kinases often contain SH2-domains , which recognize phosphotyrosine, whereas some serine/threonine kinases contain FHA domains that recognize phosphoserine/phosphothreonine. If some of the targets of these kinases form dimers (or oligomers), their phosphorylation status can be perpetuated according to the Crick's original proposal.

For the absence of a accepted terminology, Vasily Ogryzko's group used the term 'epigenetic templating' [Ogryzko, 2008]; [Viens et al., 2006] to refer to a mechanism of perpetuation of epigenetic information that is based on the preferential activity of enzymes that deposit a particular epigenetic mark on macromolecular complexes already containing the same mark [Figure 1.2]. They tested whether this model can also apply to variant histones, putative epigenetic marks on chromatin different from post-translational histone modifications.

1.2 The Chromatin Template

In eukaryotic cells, the genetic material is organized into a complex structure that is known as chromatin (from the Greek "khroma" meaning colored), which was first detected with

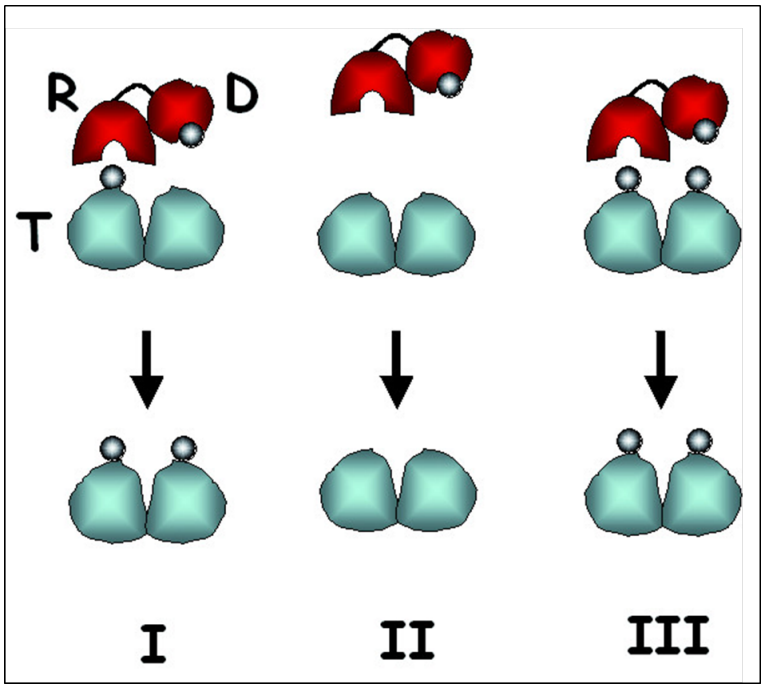


Figure 1.2: Scheme of epigenetic templating. In case I, the modifying machinery is recruited to sites exhibiting a particular mark and therefore it amplifies the information. In case II, if the mark is absent, the machinery is not recruited. In case III, the recruitment occurs but no additional modification takes place because target sites are already modified. Adapted from [Ogryzko, 2008].

basic dyes as a stainable substance in the nucleus, towards the end of the nineteenth century. Currently, it is considered as a principal carrier of epigenetic information. It is an hierarchically organized complex of DNA, histones and non-histone proteins. Histones, being the most abundant and notable protein components of chromatin, have two properties that would allow them to carry epigenetic information. First, they are very tightly associated with DNA, as demonstrated by biochemical methods and also by direct measurements of their mobility in vivo by photobleaching experiments [Kimura and Cook, 2001]. This tight association can persist through mitosis [Gerlich et al., 2003], thus histones can provide a stable mark on a particular locus of genome. Second, they are known to carry many post-translational modifications, which can dictate specific biological readouts and stably transmitted across generations (Discussed later).

1.2.1 The Nucleosome

Packaging of chromatin is a distinctive feature of eukaryotic genomes. The fundamental packaging unit of chromatin is the nucleosome that consists of a core histone octamer composed of two copies of each core histone protein H2A, H2B, H3 and H4, with 145-147 base pairs of DNA wrapped around this core. This nucleosome core particle represents the basic repeating unit in chromatin and exist in the form of arrays that forms basis for higher-order chromatin structure [Kornberg, 1974]. As the principal packaging element of DNA within

Figure 1.3: Nucleosome core particle. Ribbon traces for the 146-bp DNA phosphodiester backbones (brown and turquoise) and eight histone protein main chains (blue: H3; green: H4; yellow: H2A; red: H2B. The views are down the DNA super helix axis for the left particle and perpendicular to it for the right particle. Adapted from [Luger et al., 1997].

the nucleus, it is the primary determinant of DNA accessibility during fundamental cellular processes such as recombination, replication, mitotic condensation, transcription and DNA damage sensing and repair.

The basic structure of the nucleosome core particle was resolved by X-ray crystallography at 2.8 A°, showing in atomic detail how the histone protein octamer is assembled and how DNA base pairs are organized into a super helix around it [Luger et al., 1997]. It showed a disc shaped structure resulting from the flat 1.65 left-handed super helical DNA structure, corresponding to 145-147 base-pair stretch of genomic DNA. It also helped to understand the secondary structure of the core histones [Figure1.3]. Nucleosomes are connected by a linker DNA of variable length (10-80 base pairs [Widom, 1992]) that forms a 10-nm beads-on-a-string array [Giannasca et al., 1993].

Ever since the discovery of the nucleosome in 1974, there has been evidence, appearing in the literature from time to time, that in-vivo structure of nucleosome is subject to both static and dynamic variations in structure and composition. Scientists have discovered discrete particles in which DNA is wrapped around histone complexes of different stoichiometries: octasomes, hexasomes, tetrasomes, "split" half-nucleosomes, and, recently, hemisomes (reviewed in [Lavelle and Prunell, 2007] and [Zlatanova et al., 2009]). These observations have completely changed the way how scientists looked at the nucleosome as it they can no longer be viewed as a single static entity: rather, it is a family of particles differing in their structural and dynamic properties. Such variations in nucleosome structure have important functional implications during essential cellular processes like transcription, replication, repair and re-

Table 1.1: Nucleosome core particles with suggested traditional names. Adapted from [Zlatanova et al., 2009].

Suggested Name	Traditional Name	Histone Composition/Stoichiometry
Nucleosome (L-octasome,L-nucleosome)	Nucleosome	H2A/H2B (H3/H4)2 H2A/H2B
R-nucleosome (R-octasome)	Reversome	H2A/H2B (H3/H4)2 H2A/H2B
Hexasome	Hexasome	H2A/H2B (H3/H4)2 or (H3/H4)2 H2A/H2B
L-tetrasome	Tetrasome	(H3/H4)2
R-tetrasome	Right-handed tetrasome	(H3/H4)2
Split nucleosome	Split nucleosome	H2A/H2B/H3/H4 + H3/H4/H2A/H2B
Hemisome	Half-nucleosome	H2A/H2B/H3/H4 (could be histone variant specific)

combination etc., imparting versatility to this basic organizational unit of chromatin. Consequently, the "nucleosome" must now be considered not as a unique entity, but as a dynamic family of particles that differ in either the extent of wrapping of the DNA around the core histone proteins, or in the composition and/or stoichiometry of the internal histone core, or both [Table1.1].

1.2.2 Structure of Core Histones

The core histones are small proteins with a molecular weight between 11 and 20 kDa, containing highly basic amino acid residues. Each core histone displays a characteristic and highly conserved structured domain, which is called Histone Fold Domain (HFD) , formed by three α-helixes linked by two short loops L1 and L2. This HFD forms "handshake" arrangements between histones H2A and H2B, and histones H3 and H4 [Figure1.4]. The protein octamer is divided into four dimers defined by H2A-H2B and H3-H4 pairs. The two H3-H4 dimers interact through a four α-helix bundle formed by the two H3 histone folds to define the H3-H4 tetramer. Then, each H2A-H2B dimer interacts with the tetramer through a second four α-helix bundle between H2B and H4 histone-folds [Arents and Moudrianakis, 1995]. The HFD's fold together in anti-parallel pairs: H3 with H4 and H2A with H2B. This dimeric structure of HFD's is an ancient building block that gives rise to dynamic variation in composition and structural conformation of nucleosome particles [Zlatanova et al., 2009].

All the core histones have a flexible relatively unstructured N-terminal charged tails. In addition, H2A and H2B also display C-terminal tails. These tails represent up to 28% of the mass of the core histones. They show no secondary structures and extend radially beyond the disc-shaped surface of the nucleosome, in position to associate with linker DNA or adjacent nucleosomes. These tails, although unstructured, hold important clues to nucleosomal variability as many of the residues on these tails form ideal targets for an extensive pattern of post-translational modifications (discussed later). Moreover, the N-terminal H4 histone tail

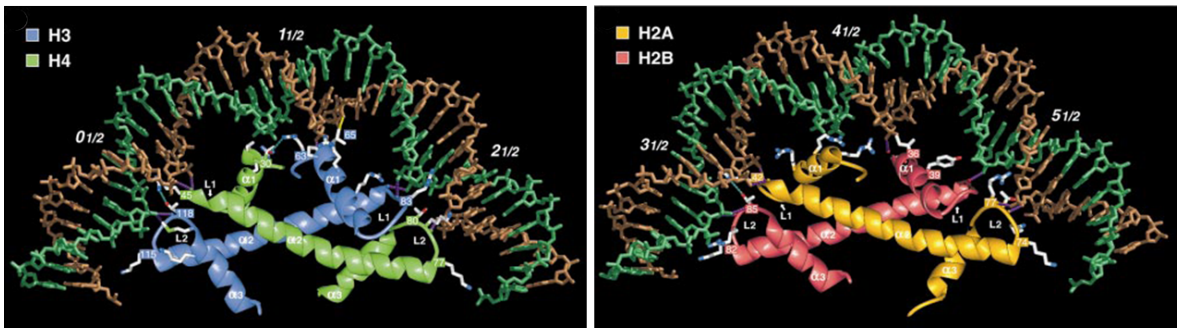


Figure 1.4: Ribbon traces of the H3-H4 and H2A-H2B histone-fold pairs. The α1-L1-α2-L2-α3 structure is shown. Adapted from [Luger et al., 1997].

has internucleosomal interactions with histone H2A (α-helix 2) that could stabilize higher order structures [Luger et al., 1997].

Structural analysis of nucleosomes containing core histones from different species have revealed that the overall nucleosome structure is surprisingly resistant to structural alterations and highly conserved from yeast to humans. It is believed that sequence differences in strategic regions of the histone fold create subtle structural changes at histone-histone interfaces, which may have significant functional implications. The high degree of evolutionary conservation of the major type core histones may therefore have a purpose above and beyond architectural preservation of nucleosomes [Chakravarthy et al., 2005]. Although histones have traditionally been viewed as slowly evolving architectural proteins that lack diversification beyond their abundant tail modifications, recent studies have revealed that variant histones have evolved for diverse functions (discussed later). This diversity in form and function of histones act as a switch to dictate when genetic information is turned on or off, it's faithfully propagation through cell division and it's repair.

1.2.3 Molecular Genetics of Core Histones

Histone protein synthesis is essential for cell proliferation and required for the packaging of DNA into chromatin. The genes encoding the core histones and histone H1 exist as multiple clusters on several chromosomes in mammalian cells. Each cluster includes several copies of the gene. Furthermore, the genes show a specific structure lacking introns [Tripputi et al., 1986]. These genes are mostly expressed during S phase of the cell cycle, with a 10 to 50 fold increase compared to the level of expression observed in quiescent cells, and with the exception of alternative histone variants (discussed later). They produce mRNAs displaying short 5′ and 3′ untranslated regions and a stem-loop structure at their 3′ end instead of a polyadenylated tail [Marzluff and Pandey, 1988]. Core histone gene expression has been shown to be evolutionarily conserved from yeast to humans [Marino-Ramirez et al., 2006].

1.3 Higher Order Chromatin Organization

The first observations of eukaryotic interphasic nuclei have shown the presence of two forms of chromatin on the basis of differential compaction; a condensed form named heterochromatin, and a more relaxed one named euchromatin [Heitz, 1928]. Euchromatin has been reported to exist under two states of condensation. The more relaxed one, representing only ten percent of euchromatin, seems to be transcriptionally active, whereas the more condensed form, does not display any transcriptional activity [Lamond and Earnshaw, 1998]; [Spector, 2001]. The euchromatin consists largely of coding sequences that accounts for only less than 4% of mammalian genomes. It represents a complex and dynamic state of chromatin that harbours and interacts with a dedicated cellular machinery (transcription machinery, chromatin modifying and remodelling complexes etc.) to bring about the transcription of functional RNAs. The exact nature of these interactions between transcription apparatus and chromatin "processing" machines is currently an exciting area of research.

Heterochromatin is generally described as transcriptionally inactive, unlike euchromatin, which tends to be associated with the active part of the genome. However, all inactive genes and non-transcribed regions of the genome do not appear as heterochromatin. Heterochromatin can also exist in two forms i.e. "constitutitve" and "facultative" heterochromatin. Constitutive heterochromatin is the permanently silent chromatin and contains a high density of repetitive DNA elements such as clusters of satellite sequences and transposable elements [Birchler et al., 2000]. These regions remain condensed throughout the cell cycle. Facultative heterochromatin, on the other hand, is found at the developmentally regulated loci, where the chromatin state can change in response to cellular signals and gene activity. It represents genomic regions in the nucleus of a eukaryotic cell that have the opportunity to adopt open or compact conformations within temporal and spatial contexts [Trojer and Reinberg, 2007]. In general, heterochromatin serves important genome maintenance functions by imparting stability to the genome.

In vitro reconstitution of chromatin in low ionic strength conditions shows that the nucleosomes are spaced by 10 to 90 bp of linker DNA and organized on a continuous DNA helix, in a linear string [Giannasca et al., 1993]. These arrays of regularly separated nucleosomes give rise to the classical 11-nm "beads on a string" template [Figure 1.5], which is also called the primary structure of chromatin. The addition of divalent cations results in the further compaction of the arrays of nucleosomes into a more compact 30-nm fibre that is typically regarded as the secondary structure of chromatin [Robinson and Rhodes, 2006]; [Robinson et al., 2006]. Binding of the linker histone, H1/H5, to the linker DNA, between the two adjacent nucleosomes, stabilizes the structure of the nucleosome, and determines the trajectory of the entering and exiting DNA [Sivolob and Prunell, 2003]. Consequently, the linker histone is likely to direct the relative positioning of successive nucleosomes and the pattern of nucleosome-nucleosome contacts.

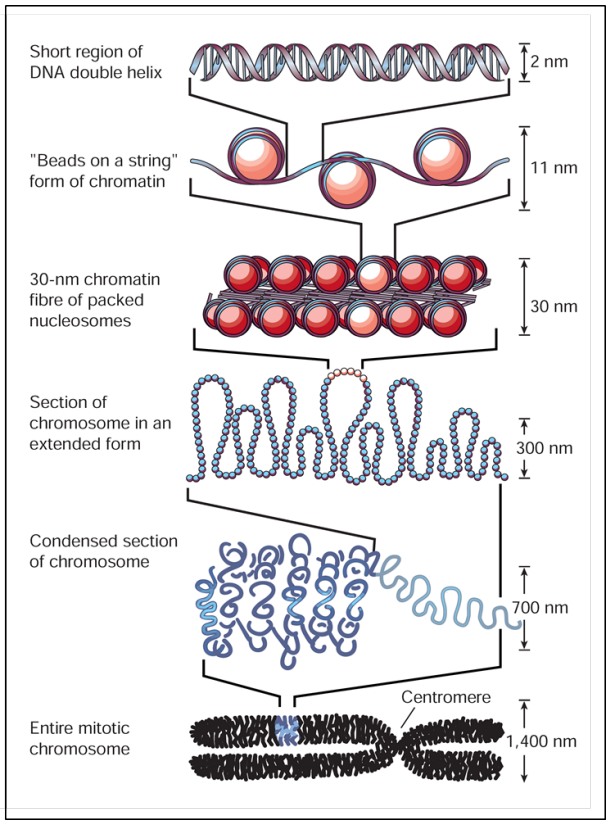


Figure 1.5: Higher order structuring of chromatin.

1.3.1 *Cis* VS *Trans*-Effects

Interestingly, the histone tails are essential for intra-molecular folding as well as fiber-fiber interaction in vitro. These observations indicate that the histone tails are indeed mediators of key nucleosome interactions in vivo. The covalent modifications of histones along with incorporation of variant histones may lead to the rearrangement of 11-nm fiber by *cis*-effects or *trans*-effects. The *cis*-effects are brought about by alteration in the physical properties of modified histone tails, such as modulation in the electrostatic charge or tail structure that, in turn, alters inter-nucleosomal contacts. A widely accepted example of this kind of effect is histone acetylation that neutralizes positive charges of highly basic histone tails, thereby generating a local expansion of chromatin fiber, increasing access to various cellular machineries to the DNA double helix [Gorisch et al., 2005]. Histone modifications may also lead to the *trans*-effect by recruiting modification-binding proteins to the chromatin. This is kind of a "reading" phenomena where certain binding proteins, present in much larger complex of proteins, recognizes or have preferences for certain type of histone modifications, leading to "docking" of this complex onto the chromatin. The well known examples of this kind of modulation include the bromodomain and chromodomain containing proteins that read acetylated and methylated histone residues respectively. Histone modifications of both HFD and tail region can also recruit ATP-dependent remodeling complexes to the 11-nm

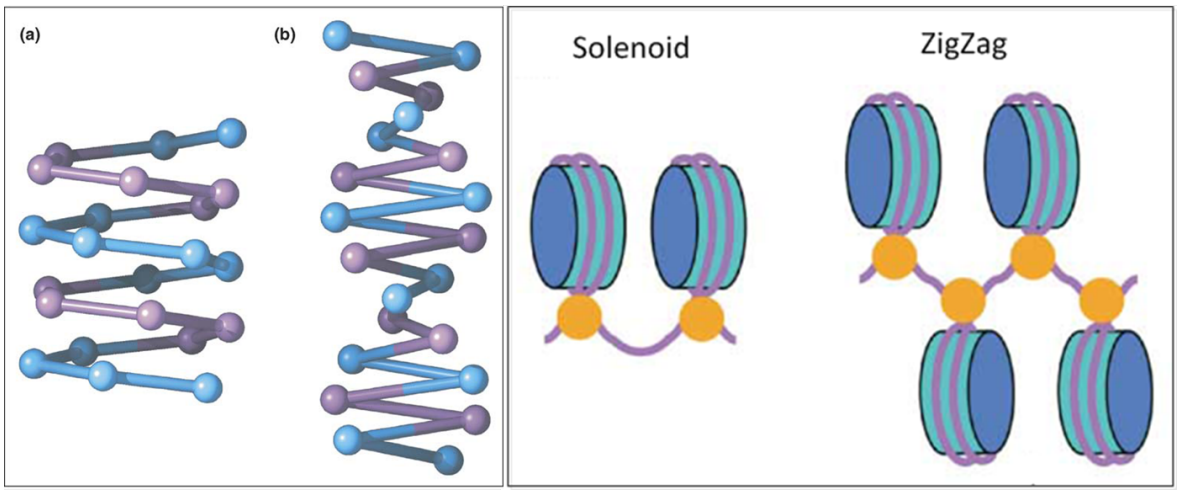


Figure 1.6: Schematic representation of two different topologies for 30-nm chromatin fiber folding. Solenoid (a) VS Zig-zag (b) model. Adapted from [Robinson et al., 2006] and [Khorasanizadeh, 2004].

fiber, which brings about transitions from inactive euchromatin to transcriptionally active state [Cosgrove et al., 2004].

1.3.2 Solenoid VS Zigzag Model

Although it is clear that the hierarchical formation of chromatin higher-order structure to eventually achieve the compacted chromosome, includes both intra- and inter-nucleosome interactions, intra- and inter-chromatin fiber interactions and interactions between chromatin and chromatin architectural proteins, the interplay between these interactions and the detailed arrangement of nucleosomes within chromatin higher-order structures, including 30 nm fibers, still remains largely uncharacterized. None-the-less, over the period of years, scientists have proposed two models to describe how the 30-nm chromatin fiber is organized, i.e. the one-start helix or "solenoid model" and the two-start helix or "Zig-Zag model" [Robinson et al., 2006]; [Khorasanizadeh, 2004], [Figure 1.6]. The solenoids involve 4-6 consecutive nucleosomes arranged in a turn of a helix that can condense into a super coil structure with a pitch of 11 nm. This structure is predicted to be held together by histone-histone interactions. In the solenoid model, a nucleosome array coils up in such a manner that successive nucleosomes are adjacent in the compact structure and connected by variable lengths of interiorly bent linker DNA [Finch and Klug, 1976]. The two-start helix model is based on a zigzag arrangement of nucleosomes, with essentially straight linker DNA connecting nucleosomes on opposite sides of the fiber [Williams et al., 1986]; [Woodcock et al., 1993]. Biochemical and electron microscopy studies favors the zigzag model over solenoid model for the arrangement of nucleosomes in a 30-nm fiber. In the zigzag model, the entry and exit paths of DNA were used to establish the relative positioning of nucleosomes and not protein-protein interaction between nucleosomes. Also, in zigzag model, alternate nucleosomes are physically closer than the adjacent nucleosomes.

1.3.3 Role of Linker Histone

Although condensation has been shown to be an intrinsic property of the nucleosome arrays, the presence of linker histone H1 or its variant H5 helps compaction of the chromatin fiber. Two models have been proposed regarding the compaction of 30-nm fiber, the first one proposed a two-start crossed linker model derived from the crystal structure of a four-nucleosome core array lacking the linker histone [Schalch et al., 2005]. The second model was based on electron microscopy (EM) analysis and proposed an alternative one-start interdigitated solenoid structure [Robinson and Rhodes, 2006]; [Robinson et al., 2006]. These two opposing models of 30-nm chromatin fiber based on the presence or absence of linker histone might reflect the existence of at least two different levels of higher order chromatin structure: one highly compact structure containing the linker histone, which represents the 30 nm fiber, and a looser structure formed in the absence of linker histone. Both structures are likely to be present in the nuclear environment, which promotes nucleosome-nucleosome contacts. Much evidence suggests the two structures represent two functional states, however, it remains to be elucidated how and if the two structural conformations of 30-nm chromatin fiber switches topology.

1.3.4 Chromatin Domains and Chromosome Territories

The 30-nm fiber must compact at least a few hundred-folds to ultimately fit a total of about 2 meters of human DNA within the nucleus [Figure 1.5]. Thus, considerably more levels of higher-order chromatin organization exist leading to interphase and mitotic chromatin states. Organization into larger looped chromatin domains (300-700 nm) occurs through anchoring the chromatin fiber to the nuclear periphery or other nuclear scaffolds via chromatin associated proteins such as nuclear laminins. The association of these chromatin domains to particular regions of the nucleus may lead to the formation of chromosome territories but how these associations are controlled and how they affect the genome regulation needs further attention by the research community. Nonetheless, there is an increasing body of evidence higlighting the correlations of active and silent chromatin configurations with a particular nuclear territory [Cremer and Cremer, 2001]; [Chakalova et al., 2005].

The faithful segregation of chromosomes during mitosis or meiosis depends upon the high degree of condensation of chromatin. This condensation leads to an astonishing reorganization of 2 meters of nuclear DNA into 1.5 μM diamater discrete chromosomes. There is about 10,000-fold compaction that is achieved by the hyperphosphorylation of histone H3 (Serines 10 and 28) and linker histone H1, and the ATP-dependent action of condensin and cohesin complexes, and topoisomerase II.

Apart from overall high degree of condensation of chromatin during metaphase into distinct chromosomes, there also exist specialized chromosome domains, for instance, telomeres and centromeres, which serve dedicated functions in chromosome dynamics and faithful chromosome segregation. Telomeres represent a region of repetitive DNA that mark the ends of chromosome and provide protection and devise a solution for proper replication of the DNA

ends. Centromeres represent a region of DNA in the middle of chromosome and provide an attachment anchor for spindle microtubules during cell division. Both of these regions are strongly heterochromatic and have distinct molecular signatures e.g., hypoacetylated histones and presence of histone H3 variant CenpA in the centromeres, which plays an essential role in chromosome segregation. Intriguingly, its the epigenetic signatures in these domains and not the DNA sequence, which mark these domains.

1.4 Modifications of Chromatin Structure

The structure of chromatin is tightly regulated in the eukaryotic nuclei. Being the building block of the eukaryotic chromatin, the nucleosome plays an essential role in the major nuclear functions (e.g. transcription, replication, recombination and DNA damage response & repair). Packing into chromatin represents a common obstacle for most of the DNA based cellular processes. To overcome this obstacle, eukaryotic cells developed mechanisms to "open" or remodel chromatin to make it accessible to cellular machineries that bring about the above mentioned functions. Three different ways for the cells to act on the nucleosomes have been described so far: post-translational modifications of the histones (mainly their tails), ATP-dependent nucleosome remodeling and incorporation of alternative histone variants. Consistent with the notion that chromatin provides an additional level of control of gene expression in eukaryotes, many regulators of transcription have been shown to have chromatin modifying activity or to be able to recruit various modifiers of chromatin.

1.4.1 Histone Post-translational Modifications

Histones can be subject to a rich variety of post-translational modifications (PTM) [Allfrey et al., 1964] that are central to the regulation of chromatin dynamics, and hence, regulate many biological processes invloving chromatin, such as replication, repair, recombination, transcription, and genome integrity [Ehrenhofer-Murray, 2004]; [Groth et al., 2007]; [Li et al., 2007]; [Kouzarides, 2007a]. Although there are at least nine distinct types of modifications found on histones, namely, acetylation, methylation, phosphorylation, ubiquitination, sumoylation, ADP ribosylation, deimination, proline isomerization, and most recently added to this list, β-N-acetylglucosamine, the most studied among them are the first four [Figure1.7]; [Table1.2], reviewed in [Bannister and Kouzarides, 2011]. Most of these modifications target the protruding amino or carboxy terminal tails of the core histones, however, more recently some modifications have also been described in the globular domain (HFD). A relatively different kind of PTM has also been detected in histones, where there is removal of a part of histone N-terminal tail, a process referred to as "tail clipping" [Santos-Rosa et al., 2009]. There are more than 60 different residues on histones which have found to be modified [Kouzarides, 2007a], however, this number may represent only a tip of the iceberg and with every passing day, with the recent application of mass spectrometry to histone biology, the list is continuously expanding. This vast array of modifications gives enormous potential

for their biological readouts. It is also worth noting that the timing and presence of different modifications on different histones depends upon the signaling conditions within the cell.

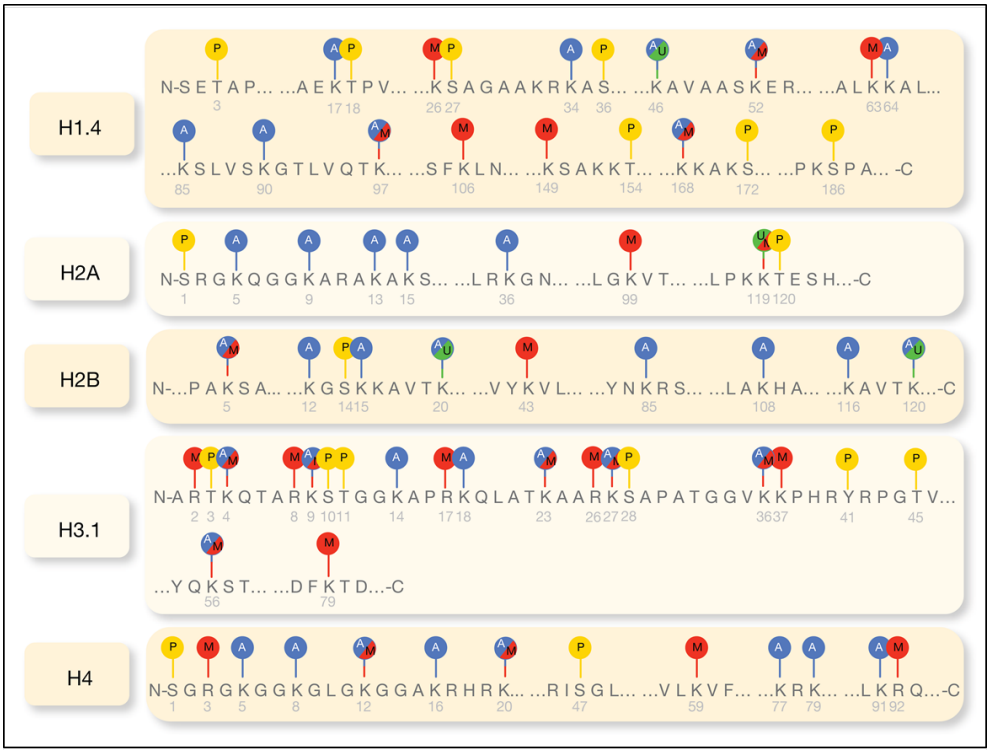


Figure 1.7: Histone modifications chart. All histones are subject to post-translational modifications. The main PTMs are depicted in this figure: acetylation(blue), methylation (red), phosphorylation (yellow) and ubiquitination (green). The number in gray under each amino acid represents its position in the sequence.

Considerable amounts of work has been done in the past decade to elucidate the enzyme systems behind the "writing" and "removal" of these histone modifications. The biochemical and genetic studies have shown a number of chromatin modifying enzymes, being part of large multi-subunit complexes, can catalyze incorporation or removal of covalent modifications on histones. Many of them have remarkable specificity to target residue and the cellular context. The identification of these enzyme groups have lead to the idea of histone modification "writers" and "erasers" along with a third class of proteins the "readers", which recognize a particular modification to bring about its function.

1.4.2 Variant Histones

Studies of chromatin as a carrier of epigenetic information were mostly focused on DNA methylation and histone modifications as epigenetic marks on chromatin. An additional layer of complexity has been added to chromatin dynamics with the renewed interest in alternative variants of the canonical core histones. These paralogues, now known as histone variants, share the same overall structure with the canonical histones, displaying for instance

Table 1.2: Different classes of modifications identified on histones.

Chromatin Modifications	Residues Modified	Functions Regulated
Acetylation	K-ac	Transcription, Repair, Replication, Condensation
Methylation (lysines)	K-me1 K-me2 K-me3	Transcription, Repair
Methylation (arginines)	R-me1 R-me2a R-me2s	Transcription
Phosphorylation	S-ph T-ph	Transcription, Repair, Condensation
Ubiquitylation	K-ub	Transcription, Repair
Sumoylation	K-su	Transcription
ADP ribosylation	E-ar	Transcription
Deimination	R > Cit	Transcription
Proline Isomerization	P-cis > P-trans	Transcription
β-N-acetylglucosamine	S-O-GlcNAc/T-O-GlcNAc	Mitosis

the histone fold, but differ in terms of primary sequence from their canonical relatives [Figure 1.8]. They are found in all eukaryotes and their presence was found to be correlated with particular regions in the genome. Unlike canonical histones which are assembled into bulk chromatin behind the replication fork, variant histones seem to be implicated in replacement processes under certain conditions. Furthermore, at least some of the variant histones appear to be incorporated at specific sites, replacing canonical histones. In addition, this deposition seems to occur outside of S-phase, unlike for the canonical histones. The substitution of a canonical histone by a variant will then create a new nucleoprotein complex, with an alternate structure and stability. Thus, the presence of such variant histones is likely to add another level of complexity, and therefore regulation to the nucleosome arrays. Given very stable association of histones with DNA that can last through all stages of cell cycle, some alternative histone variants also present good candidates for an epigenetic marking of genome.

Histone H3 Variants

The number of histone H3 variants varies among species. Today, we can distinguish five variants in mammals: H3.1 and H3.2 are the replicative histones expressed during the S phase, H3.3 is the replacement histone expressed throughout the cell cycle, CenpA is the H3 paralog found in centromeric regions [Palmer et al., 1991]; [Perpelescu and Fukagawa, 2011], and H3t the testis-specific histone which is not present in somatic cells [Govin et al., 2005]. Human H3.1 and H3.2 are 99% identical; and H3.3 is 96% identical to H3.1, differing at five amino acid positions. Human CenpA is highly divergent and shares only 46% identity with H3.1, whereas H3t differs from H3.1 by just four amino acids (reviewed in [Loyola and Almouzni, 2007]).

How these variants could contribute in establishing epigenetic information and how they could ensure its maintenance is not completely elucidated. However, in the past years many

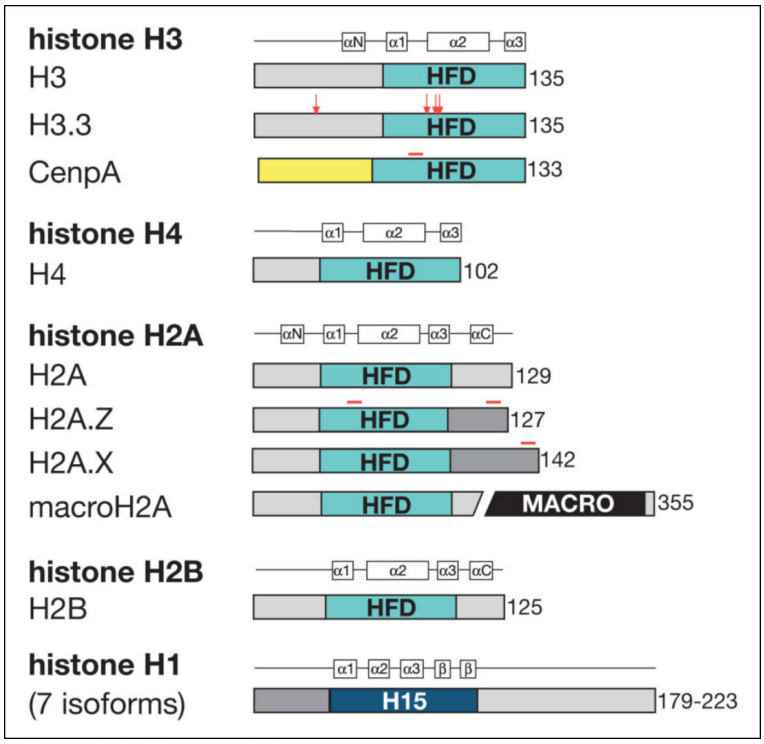


Figure 1.8: Proteins domain structure of canonical histones and histone variants. The regions of proteins that differ in histone variants are shown in red. Adapted from [Allis et al., 2007].

reports have shed light on this issue, proving that these histone variants play an important role in alternative chromatin states' formation and some constitute marks stably transmitted from one cell generation to another.

I **CenpA**

During mitosis, chromosome segregation requires alignment of all the chromosomes at the cell equator. This alignment is due to the attachment of the microtubules of the mitotic spindle to centromeres and this process is controlled by kinetochores. No consensus DNA sequence could be identified in centromeric regions. However, members of the CenH3 family of histone H3 variants including CenpA in human, Cid in Drosophila and Cse4p in yeast have been found in all functional centromeres in eukaryotes. Furthermore, CenpA inactivation has been shown to impair kinetochore formation and chromosome segregation in yeast, Drosophila and mammals [Sullivan, 2002]; [Santaguida and Musacchio, 2009]. In addition, CenpA knock out experiments in mice resulted in mitotic deficiencies and in embryonic lethality, arguing for a key role of this variant in mitosis [Howman et al., 2000]. Despite large divergence in centromere DNA sequences across organisms, CENPs are highly conserved proteins [Malik and Henikoff, 2009]. Human centromeric nucleosomes are thought to be composed of eight histones, and most are homotypic octamers containing two copies of CENPA. More recent work, however, has shown that about 10% of human centromeric nucleo-

somes form a heterotypic octamer containing one CENPA-containing histone and one canonical H3- containing histone [Foltz et al., 2006]. CENPA is a variant of histone H3, and the homology is found mainly at the a-helical carboxy-terminal histone-fold domain [Luger et al., 1997]. The N-terminal tail of CENPA is highly variable between species and is required to recruit kinetochore proteins to the centromere [Van Hooser et al., 2001]; [Henikoff and Dalal, 2005].

Centromeric chromatin is organized as interspersed stretches of CenH3- and H3- containing nucleosomes [Blower et al., 2002]. However, it is still not clear how CenpA assembly is maintained at pre-existing CenpA containing sites. One way to deposit CenH3 in the correspondent sites could be its incorporation following nucleosome loss due to anaphase tension on CenH3 containing chromatin [Ahmad and Henikoff, 2002a]; [Mellone and Allshire, 2003]. The other explanation is that a particular pattern of modification in interspersed and flanking stretches of nucleosomes predispose centromeres for deposition of CenH3 [Sullivan and Karpen, 2004]. More recent reports identify HJURP as a chaperone for soluble CENP-A [Dunleavy et al., 2009]; [Foltz et al., 2009]; [Shuaib et al., 2010]. Once assembled on DNA and flanked by H3 chromatin stretches, CENP-A nucleosomes might impose unusual chromatin arrangements and torsions where they (and implicitly the CATD) are more exposed. It might be that this special distribution provides the apparent difference between the two types of nucleosomes, recognizable by certain CENPs and remodeling factors [Perpelescu and Fukagawa, 2011].

II **H3.3**

The H3.3 variant is nearly identical to canonical histones H3, with only a four amino acid difference in Drosophila and respectively 5 and 4 amino acid difference with H3.2 and H3.1 in mammals. Yet, H3.3 function is different from canonical histone H3.

H3.3 at active sites

Back in 1984, H3S, a H3.3-like histone, was only found in the active macronucleus in ciliates whereas H3F, the canonical H3-like histone was found in the micronucleus where chromatin is usually condensed [Allis and Wiggins, 1984]. More recently, using histones fused with green fluorescent protein GFP, H3.3 was shown to be deposited throughout the cell cycle including S phase in Drosophila cells, whereas canonical histone H3 deposition was seen only during replication [Ahmad and Henikoff, 2002b]. Furthermore, in the same study it was shown that in Drosophila H3.3 deposition localizes to active rDNA arrays and to euchromatin but not heterochromatin. In human cells, H3.3 deposition took place on transgene array when its transcription was activated [Janicki et al., 2004]. The concomitant loss of heterochromatic marks in both Drosophila and human cells reveals that H3.3 takes place following removal of preexisting histones [Loppin et al., 2005]. H3.3 enrichment at promoter regions has been observed not only at active genes but also at inactive genes, possibly accounting for a "poised" state of these genes [Mito et al., 2007]; [Tamura et al., 2009]. Furthermore, H3 replacement also occurs at genic and intergenic regulatory regions in various metazoans [Mito et al., 2007]; [Nakayama et al., 2007]; [Jin et al., 2009].

H3.3 histone was extracted from a Drosophila cell line and mass spectrometry analysis revealed that H3.3 is enriched in marks associated with active chromatin and is deprived of marks of repressed chromatin such as di-methyl lysine 9 [McKittrick et al., 2004]. Consistent with these results, microarrays genome-scale profiling following biotin-H3 and biotin-H3.3 pull down demonstrated that H3.3 deposition occurs mainly on sites enriched in methylated lysine 4 of histone H3, as well as in RNA polymerase II [Mito et al., 2005].

A similar study was conducted on mammalian cells. All three histones, H3.1, H3.2 and H3.3 showed different patterns of modifications, which suggests that they have different biological functions. As in the Drosophila case, the majority of modifications detected on H3.3 were marks of active chromatin (such as methylation of lysine 4 and lysine 79 and acetylation of lysine 14, lysine 18 and lysine 23) whereas H3.2 showed mostly silencing modifications [Hake and Allis, 2006]. Interestingly, H3.1 modification pattern comprised marks of both active and repressed chromatin, though the latter were not the same as those of H3.2. These data suggest that mammalian H3.1 and H3.2 should not be treated as equivalent proteins.

Taken together, these results might suggest that H3.3 plays a causal role in transcriptional activation. However, H3.3 is the dominant H3-subtype in non dividing differentiated cells in vertebrates [Pina and Suau, 1987]; [Urban and Zweidler, 1983], replacing up to 80% of total histones in terminally differentiated cells. It also represents 25% of all histone H3 in Drosophila cells. Considering the fact that only a small percentage of the genome is actively transcribed, these results strongly suggest that H3.3 presence cannot be absolutely correlated with active transcription. Another argument towards the notion that H3.3 deposition is not necessarily coupled with transcription comes from the studies of HIRA-H3.3 role in fertilized oocyte development (see below). These facts suggest an alternative explanation of the correlation between H3.3 and actively transcribed sites of genome. The relative enrichment of H3.3 on active chromatin could the due to two factors: 1. More dynamic state of active chromatin, which renders it more prone to histone exchange, and 2. In non-replicating cells H3.3 is the only histone type that can be involved in the replacement. The former and latter explanations do not necessarily exclude one another, but both possibilities have to be taken into account.

Preexisting modifications before nucleosome assembly are important to evaluate if the enzyme target specific modified substrates. In order to have a better understanding on this issue, H3.1 and H3.3 post translational modifications were characterized before and after chromatin assembly. Pre-deposited and nucleosomal histones showed different patterns. A typical signature of non nucleosomal H3/H4 is the enrichment in acetylated lysine 5 and 12 of the histone H4, regardless of the H3 variant present in the dimer. Moreover, H3.1 and H3.3 are deficient in methylation marks, with the exception of lysine 9 [Loyola et al., 2006]. These results suggest that the establishment of a given modification pattern does not depend only on a specific variant when it is soluble, but is rather affected by the local environment where the histone is incorporated.

Deposition onto chromatin

As described above, H3 and H3.3 containing complexes comprise different protein composition and chromatin assembly factors [Tagami et al., 2004]. CAF-1 copurifies with H3.1 and is responsible for its DNA-synthesis/replication-coupled (RC) dependent deposition. HIRA was found in H3.3 complex and promotes its incorporation into DNA in a replication-independent (RI) manner. Furthermore, the chaperone ASF1 was also shown to be necessary for an efficient nucleosome assembly in vivo [Galvani et al., 2008]. Authors generated H3.1 and H3.3 mutants unable to interact with ASF1A and ASF1B. Although these mutants still bind to their respective chaperones, CAF1 and HIRA , they were not effectively incorporated onto DNA. While HIRA is required for H3.3 deposition at genic regions in mouse embryonic stem (ES) cells, H3.3 enrichment at telomeres and most regulatory elements is HIRA-independent [Goldberg et al., 2010]. Instead, the SWI/SNF-type chromatin remodeler ATRX mediates localization of H3.3 to telomeres [Goldberg et al., 2010]; [Wong et al., 2010].

To test how the ability to participate in RI assembly depends on the aminoacid differences in H3 and H3.3 sequences, Henikoff and coworkers examined the effect of different aminoacid replacement in histones' sequences. Changing any one of the residues in H3.3 to its H3 counterpart did not affect the RI assembly of H3.3. Interestingly, changing any one of the residues in H3 to H3.3 counterpart relieved the block to RI assembly and allowed histone H3 deposition outside of S phase [Ahmad and Henikoff, 2002b]. These findings suggested that RI deposition is the default pathway and that H3 is actively recognized and blocked from the RI deposition. This conclusion is supported by the fact that in the yeast S. cerevisiae, the only H3 histone involved in regular chromatin structure is similar to H3.3, reinforcing the notion that H3.3 assembly is the default H3 deposition pathway. Contrary to what was believed previously [Wolffe, 1992], it was shown that the H3/H4 are deposited as dimers and not as tetramers [Tagami et al., 2004], leading to the proposal of a semi-conservative mechanism of nucleosome duplication. On the other hand, it was also shown that the H3 and H3.3 variants form so called homotypic nucleosomes, having both H3 molecules of the same type in the histone octamer. These findings bear on the possibility that, in addition to post-translation histone modifications and DNA methylation, the 'epigenetic templating' mechanism can also work for chromatin marks of a different kind, e.g., variant histones. Since the new H3 histones are deposited on DNA only in the form of H3/H4 dimers, the deposition machinery must match the types of new and old H3 histones in order to keep the nucleosome homotypic.

H3.3 in development

As mentioned earlier, fertilization study in D. melanogaster revealed an important role of the HIRA-H3.3 deposition pathway as a way to replace protamines to histones in the male pronucleus [Loppin et al., 2005]. Consistent with this result, H3.3 is present in the mice oocytes as a maternal factor and is incorporated into the male pronucleus genome before its activation [Torres-Padilla et al., 2006]. As mentioned before, these data strongly suggest that H3.3 assembly could take place independent from transcrip-

tion. Unlike in mice and flies, H3.3 is incorporated in both male and female pronuclei in C. elegans, suggesting a need to distinguish between paternal and maternal genome in mice and flies but not in worms [Ooi et al., 2006]. Despite potential compensatory mechanisms in flies, HIRA and H3.3 play crucial roles in sexual reproduction in all studied metazoans [Bonnefoy et al., 2007]; [Sakai et al., 2009]; [Hodl and Basler, 2009]. Both male and female H3.3 null flies are sterile. In mice, the hypomorphic gene trap of H3.3A also led to male sub-fertility [Couldrey et al., 1999].

Histone H2A Variants

The H2A family contains so far four subfamilies that differ in terms of sequence and length. In mammals, the members of the first subfamily, H2A.1 and H2A.2, are synthesized mostly during S phase of the cell cycle, in parallel with DNA replication. Unlike the canonical histones, the members of two other subfamilies, H2AZ and H2AX are synthesized at a constant low level throughout the cell cycle. Interestingly, these two proteins are also synthesized in quiescent cells, but at a significantly lower level than in proliferating cells. Finally, the last subfamily members, macroH2A1 and macroH2A2, have approximately three times the size of histone H2A. The first third of the proteins share approximately 65% homology with the sequence of histone H2A. The remaining two thirds of the proteins do not contain histone-fold motives and share 84% homology.

Unlike the genes encoding the canonical histones and histone H1, the variant histones (H2AZ, H2AX and macroH2A) are encoded by single copy genes, containing introns. These genes produce polyadenylated mRNAs with longer 5′ and 3′ untranslated regions than those of the canonical histones mRNAs.

I **H2AX**

H2AX or its homologues are found throughout evolution among animals, plants, fungi and protists, arguing for an important role for this histone variant. In mammals, it is present in all cell types. This H2A variant contains a conserved SQ(E,D)(I,L,F,Y) motif in its C-terminal domain.

As mentioned before, H2AX is encoded by a unique gene, however, it produces two mRNAs. One of these RNAs is polyadenylated mRNA with long 5′- and 3′-untranslated regions which is produced in a replication independent manner, characteristic for an alternative histone variant. The second product is a shorter mRNA produced in replication linked manner [Mannironi et al., 1989]. H2AX mRNA also displays the conserved motives needed to process the replication dependent canonical histones. Thus, H2AX gene expression combines features of both canonical and alternative histones.

H2AX in DNA-damage repair

H2AX has been shown to undergo several post-translational modifications, including acetylation, ADP-ribosylation and phosphorylation. The phosphorylation of serine 139 of the C-terminal tail of H2AX by the PI3K-related ATM kinase (ataxia teleangiectasia

mutated) seems to be important for DNA repair by non-homologous end joining [Rogakou et al., 1998]. After introduction of DNA double strand breaks by ionizing radiations, H2AX is rapidly phosphorylated (H2AX) at the sites of the lesions, along with recruitment to these lesions of proteins involved in DNA-damage repair, such as members of the MRN complex (Mrell/Rad50/Nbs1) , BRCA1 and p53BP1 (Schultz et al., 2000). Furthermore, H2AX phosphorylation is apparently involved in this process, as the recruitment is inhibited when the cells are treated with inhibitors of PI3K kinase family, which includes ATM kinase [Paull et al., 2000]. Also, the kinetic of H2AX phosphorylation is similar to the kinetics of the recruitment of the repair complexes and to the kinetics of the DSB repair. It is apparent within a minute and reaches a peak at approximately ten minutes after irradiation. After thirty minutes, H2AX is dephosphorylated with a half-life of two hours. Thus, H2AX is likely to play a key role in DNA-damage repair initiation and control (Discussed later [Section 1.5.2]).

H2AX in normal cell processes

Interestingly, besides environmentally and metabolically caused DNA double strand breaks, H2AX also seems to play a role in several eukaryotic processes involving generation of double strand breaks, such as recombination. Foci of H2AX have been found around double strand breaks during meiosis in germ cells [Mahadevaiah et al., 2001]. In addition, an accumulation of H2AX has been reported in B and T cell nuclei during immune system development. H2AX is targeted to the locus of the receptor, which is subject to V(D)J recombination to form a genomic library of variable regions [Soulas-Sprauel et al., 2007]. An accumulation of H2AX has also been observed in T lymphocytes at the locus of the T cell receptor (TCR) , also subject to rearrangements. H2AX knock out in mice resulted in an alteration of these recombinations, suggesting an important function of H2AX in this process [Petersen et al., 2001].

Finally, H2AX accumulation has also been observed at the early stages of DNA fragmentation during apoptosis, when high molecular weight DNA fragments appear as a result of caspase-activated DNase. In addition to the replication-independent processes described above, H2AX could take part in the control of DNA replication. When replication is inhibited by hydroxyurea, H2AX is phosphorylated by ATR kinase and H2AX is then accumulated in the arrested replication forks [Ward and Chen, 2001]. As described above, H2AX is recruited to DNA double strand breaks of diverse origins. Because H2AX is a histone, it is tightly bound to DNA when incorporated into nucleosomes, and is likely to function as an anchor for mobile factors when phosphorylated.

II **macroH2A**

The macroH2A variant differs from the other H2A variants by its unusual size, 372 amino acids, which makes it approximately three times larger than canonical histone H2A. The defining feature of this histone variant is the 30 kDa carboxyl-terminal globular domain called a macro domain. MacroH2A is the founding member of a large family of macro domain-containing proteins [Kraus, 2009]. MacroH2A can be divided in two parts; the first third of the protein displays histone fold domains, and shares 64% homology with the sequence of canonical H2A. The remaining two thirds of the protein

are characterized by the absence of histone fold domains. Two sub-types of macroH2A have been identified so far; macroH2A1 and macroH2A2, each one encoded by a single gene containing eight exons. Furthermore, macroH2A1 gene is coding for two isoforms, named macroH2A1.1 and macroH2A1.2, resulting from an alternative splicing. macroH2A1 and macroH2A2 share 84% homology in the histone domain, and 68% in the non-histone domain. In addition, macroH2A1.1 and 1.2 sequences display only a twenty amino acid difference in the non-histone domain. The non-histone region of the macroH2A variants is characterized by the presence of a leucine zipper motif, which is also found in many transcription-related factors, arguing for a potential role of these variants in the control of transcription. Finally, all isoforms are evolutionarily conserved [Pehrson et al., 1997].

MacroH2A1.1 is found in quiescent cells, whereas macroH2A1.2 is generally expressed in proliferating cells [Pehrson et al., 1997], suggesting a different function for these subtypes. Interestingly, macroH2A, which is present both in male and female cell nuclei, seems to be more concentrated in the area of the inactive X chromosome in female cells [Mermoud et al., 1999]. However, the association of macroH2A with the inactive X chromosome seems to occur after the proper extinction mechanism. In addition, a loss of macroH2A at the inactive X chromosome has been reported in several tumor cell lines, indicating a potential role in the maintenance of the inactive state of the X chromosome. Finally, the presence of macroH2A throughout the nucleus in male and female cells seems to indicate a more general role for macroH2A. Recent studies have shown the involvement of macroH2A in repression of transcription. The non-histone part of macroH2A has been shown to repress transcription when targeted near an active promoter [Ouararhni et al., 2006]; [Buschbeck et al., 2009]. In addition, macroH2A, when incorporated into nucleosomes, prevents the recruitment of transcription factors and impairs remodeling by SWI/SNF complex, thus interfering with transcriptional activation [Angelov et al., 2003].

III **H2A.Bbd**

In contrast to the large C-terminal macrodomain of mH2As, the mammal specific H2A Barr body-deficient (H2A.Bbd) has a short C-terminus with a truncation of the docking domain. It is only 48% identical to canonical H2A and lacks the residues that are modified in other H2As. H2A.Bbd seems to be associated with active chromatin, being deficient on Xi chromosomes in fibroblasts and coinciding cytologically with acetylated H4 [Chadwick and Willard, 2001].

In A431 human cell lines, H2A.Bbd tagged with green fluorescent protein (H2A.Bbd GFP) is exchanged more quickly in the nucleosome than H2AGFP [Gautier et al., 2004]. H2A.BbdH2B are preferentially exchanged, in vitro, by nucleosome assembly protein 1-like 1 (NAP1L1) for H2AH2B dimers in reconstituted nucleosomes, and this is more efficient in H3.3 nucleosomes than H3 nucleosomes [Okuwaki et al., 2005]. Reconstituted H2A.Bbd nucleosomes are unstable without DNA, protect only 110130 bp of DNA, have more divergently oriented DNA ends (180°) than canonical nucleosomes and do not undergo remodeling by SWI/SNF properties that are largely attributable to

the H2A.Bbd docking domain [Gonzalez-Romero et al., 2008]. H2A.Bbd has a smaller acidic patch than H2A, which inhibits the folding of reconstituted H2A.Bbd nucleosome arrays into 30-nm fibers and reduces the ability of these arrays to inhibit transcription [Zhou et al., 2007]. Despite the evidence that H2A.Bbd forms accessible chromatin, no specific function has yet been identified for this interesting variant.

The phylogenetic analyses of canonical histones and replacement histone variants revealed that histone H2A.Bbd is a highly variable quickly evolving mammalian replacement histone variant, in striking contrast to all other histones [Malik and Henikoff, 2003]. The incorporation of H2A.Bbd variant into the nucleosomes result in an unfolded highly unstable nucleoprotein complex, which resembles the previously described hyperacetylated nucleosomes associated with transcriptionally active regions of the genome. Nevertheless, the structure of nucleosome core particles reconstituted from H2A.Bbd is not affected by the presence of a hyperacetylated histone complement. This suggests that replacement by H2A.Bbd provides an alternative mechanism to unfold chromatin structure, possibly in euchromatic regions, in a way that is not dependent on acetylation [Eirin-Lopez et al., 2008]

IV **H2AZ**

The histone variant H2AZ has been found in organisms ranging from protists like Tetrahymena to higher eukaryotes, and is highly conserved, with approximately 90% homology in the peptide sequence in these organisms. However, despite a similar structure, H2AZ peptide sequence has only 60% homology with the one of histone H2A. The H2AZ variant is referred to as HV1 in Tetrahymena, HTZ1 or HTA3 in yeast, H2AvD in Drosophila, H2AZ/F in human. For clarity, they will be all referred to as H2AZ.

Structural characteristics of H2AZ

The crystal structure of H2AZ-containing nucleosomes has been resolved and shows a core particle surprisingly similar to the one of H2A-containing nucleosomes, considering the numerous differences in peptide sequence between the two proteins. Actually, no apparent changes in the path of the DNA double helix or in protein-DNA interactions can be observed [Suto et al., 2000]. However, a subset of discrete differences exists between canonical nucleosomes and H2AZ-containing nucleosomes in the tetramer-docking domain, where H2AZ-H2B dimer interfaces with H3-H4 tetramer. In H2AZ containing nucleosomes, the overall shape of the interaction domain is conserved, as well as most of the intermolecular hydrogen bonds, but the substitution of a glutamine (Gln 104 in H2A) by a glycine (Gly 106 in H2AZ) induces the loss of three hydrogen bonds, thus slightly destabilizing the interaction between H2AZ and H3 compared to H2A and H3. In addition, unlike H2A, H2AZ displays two histidine residues, capable of binding a metallic ion in an area of the tetramer-docking domain located at the surface of the nucleosome. The presence of this ion is likely to create an interaction domain with H2AZ-specific interacting proteins. Additional differences exist in the acidic patch formed between amino acids from the docking domain of H2A/H2AZ and H2B at the surface of the nucleosome. This acidic patch seems to be involved

in protein-protein interactions, possibly with histone H4 basic tails from neighboring nucleosomes, making it an important feature of the nucleosome surface [Luger and Richmond, 1998]. In the case of H2AZ, the patch appears to be extended, resulting in a larger acidic surface across the face of the H2AZ-containing histone octamer [Suto et al., 2000].

Interestingly, H2AZ appears to play an essential role in all the tested organisms. Inactivation of the gene coding for H2AZ leads to a decreased growth rate in yeast and to death in Tetrahymena, Drosophila, Mouse and Xenopus leavis [Liu et al., 1996]; [Clarkson et al., 1999]; [Jackson and Gorovsky, 2000]; [Faast et al., 2001]; [Ridgway et al., 2004]. Furthermore, the replacement of the tetramer-interacting region of H2AZ by the one from H2A in Drosophila has been shown to impair fly development, arguing for different roles of H2A and H2AZ in cellular mechanisms, and thus demonstrating the importance of the structural differences between the two proteins. In addition, Drosophila null embryos could be rescued by expressing chimeric constructs including H2AZ tetramer-binding domain. Taken together, these observations indicate a key role for H2AZ specific tetramer-binding domain in mechanisms related to cell proliferation and development.

Recent studies in Tetrahymena showed the importance of acetylation of the amino-terminal tail of H2AZ, which includes six lysines instead of the three normally present in H2A. Complete replacement of the lysine residues by arginines, which prevents acetylation, or by glutamines, mimicking acetylated lysines, was lethal or led to reduced growth rate, indicating the need of H2AZ acetylation to ensure its proper cellular functions [Ren and Gorovsky, 2001]; [Ren and Gorovsky, 2003]. Because of the proximity of the lysines in the amino-terminal tail of H2AZ, it has been hypothesized that this area could behave as a charge patch. The reduction of the positive charge of the histone tail by acetylation could result in a reduced association of H2AZ tail with DNA, making it more accessible for DNA-binding proteins.

In addition, sedimentation and ultra-centrifugation analysis conducted on reconstituted chromatin suggest that H2AZ has a destabilizing effect on intranucleosomal histone-histone interactions [Abbott et al., 2001], and prevents internucleosomal interactions [Fan et al., 2002]. Furthermore, H2AZ has been shown to influence nucleosome positioning in vitro, which may be part of its biological function in vivo.

Finally, the L1 loop domain, which mediates interactions between H2A molecules in the same nucleosome, is altered in the case of H2AZ in such a way that it may favor the interaction with another H2AZ molecule. The pairing of one H2A and one H2AZ molecule in the same nucleosome could lead to sterical clashes, and no stabilizing interaction could be formed between the two molecules [Suto et al., 2000]. Interestingly, the same observation can be made about the L1 loop of macroH2A [Pehrson and Fuji, 1998], suggesting that the L1 loop could ensure the incorporation of only one type of H2A histone into one nucleosome. However, this remains to be demonstrated in vivo.

H2AZ and transcriptional activation

Early observations in Tetrahymena showed that H2AZ was present in the transcrip-

tionally active macronucleus and absent from the micronucleus. In addition, during early conjugation, H2AZ was detected in the micronucleus just before its transcriptional activation [Stargell et al., 1993], thus linking H2AZ to active chromatin. Further studies by immunofluorescence on polytene chromosomes and chromatin immunoprecipitation (ChIP) in diploid chromosomes in D. melanogaster showed that H2AZ was widely but non-randomly distributed across the genome. H2AZ was found to be associated with thousand euchromatic bands, but also with the heterochromatic chromocentre of polytene chromosomes. Furthermore, the ChIP analysis showed that H2AZ was found both in transcribed and non-transcribed genes, as well as in non-coding euchromatic and heterochromatic sequences [Leach et al., 2000]. Similar studies conducted in S. cerevisiae expressing a HA-tagged version of H2AZ showed a non-uniform pattern of distribution of the variant across the PHO5 and GAL1 loci, with a preferential localization in the promoter regions of these genes [Santisteban et al., 2000]. In addition, H2AZ was preferentially found at these locations under repressing conditions, and was decreased after activation, suggesting a role for H2AZ in the establishment of transcription. Furthermore, deletion of H2AZ gene compromised transcriptional activation of these genes. Other experiments showed a defect in the recruitment of RNA polymerase II and the TATA-binding protein on GAL1 promoter, in strains lacking H2AZ gene [Adam et al., 2001], arguing for a role for H2AZ in the recruitment of the basal transcriptional machinery, and thus in the establishment of transcription. Finally, the transcriptional activation defect observed for the deletion of H2AZ gene was further exacerbated when associated with the deletion of GCN5 gene or SNF2 gene, involved in histone acetylation or chromatin remodeling, indicating a partially redundant role of these three chromatin modification pathways in the establishment of transcriptionally active chromatin [Santisteban et al., 2000].

According to these observations, H2AZ could participate, along with ATP-dependent chromatin remodelling complexes and histone modification complexes, in the establishment of higher-order chromatin domains poised for activation. In addition, H2AZ could also serve to recruit the basal transcription complex.

H2AZ and transcriptional repression

Surprisingly, H2AZ was also reported to be involved in transcriptional repression [Svotelis et al., 2009]. Over-expression of H2AZ was able to restore silencing in Sir1-impaired cells. Furthermore, H2AZ was shown to be present at the HMR locus and in telomeres by chromatin immunoprecipitation, and the deletion of H2AZ led to a loss of silencing at the two loci [Dhillon and Kamakaka, 2000]. These results are consistent with the presence of H2AZ in non-transcribed genes and in non-coding heterochromatic regions in Drosophila [Leach et al., 2000] and in pericentric heterochromatin during early mammalian development [Rangasamy et al., 2003], and suggest that H2AZ can be involved both in gene activation and gene silencing. Furthermore, the H2AZ-containing nucleosomes are more stable than canonical ones. According to the authors of this finding, this stabilisation could potentially attenuate the transcription processes by impeding the displacement of a histone dimer [Park et al., 2004].

Interestingly, identification of H2AZ-regulated genes by whole-genome microarray analysis in yeast showed that H2AZ-activated genes are highly enriched near telomeres and often clustered together. Furthermore, H2AZ was found to antagonize Sir-dependent silencing in telomeric locations and in regions flanking the HMR locus. In H2AZ-impaired cells, Sir proteins were shown to spread outside the HMR locus in the neighboring euchromatic regions. In addition, this spreading was concomitant with ectopic transcriptional silencing and with the appearance of histone acetylation and methylation patterns characteristic of heterochromatin formation.

According to all these observations, H2AZ seems to play a critical role in gene expression, maybe by protecting euchromatic genes from silencing. This anti-silencing effect could be due to the establishment of permissive higher-order chromatin structure, resulting from the destabilizing effect of H2AZ on intra- and inter-nucleosomal interactions [Abbott et al., 2001]; [Fan et al., 2002]. However, the exact function of H2AZ in transcriptional activation and silencing remains to be determined.

1.4.3 Histone Modifying Enzymes

Histone modifying enzymes catalyze the addition or removal of a large array of covalent modifications on histone and non-histone proteins. These enzymes can be classified into several distinct classes such as, acetyltransferases, deacetylases, lysine methyltransferases, lysine demethylases, arginine methyltransferases, serine/threonine kinases, ubiquitilases, sumoylases, deiminases, and proline isomerases [Figure 1.9]; [Kouzarides, 2007b]. Histone modifications have found to be dynamic, and enzymes that remove the modification have also been identified. Arginine methylation is an exception as despite of being dynamic, as generally, deiminases can convert methylarginine to citrulline via demethylimination the process that antagonize the arginine methylation [Cuthbert et al., 2004]; [Wang et al., 2004a]. However, Chang et al. reported in vitro and in vivo data that clearly indicated that JMJD6 functions as an arginine demethylase [Chang et al., 2007]. Of all the histone modifying enzymes, the methyltransferases and kinases are the most specific. This might explain the fact why histone methylation is the most characterized modification to date. While considering the specificity of enzymes that modify histones it must be kept in mind that they can be influenced by other factors, for instance, complexes in which enzymes are found may specify a preference for nucleosomal verses free histones [Lee et al., 2005]; proteins that associate with the enzyme may affect its selection of residue to modify [Metzger et al., 2005] or the degree of methylation (mono-, di-, or tri-) at a specific site [Steward et al., 2006]. Here, we will focus our discussion to the enzymes mainly responsible for maintaining the balance between histone acetylation and deacetylation and methylation and demethylation.

Histone Acetylatransferases (HATs) & Histone Deacetylases (HDACs)

Lysine acetylation is a reversible and highly regulated post-translational modification reported on histones in early 1960's [PHILLIPS, 1963]; [Allfrey et al., 1964], while its chem-

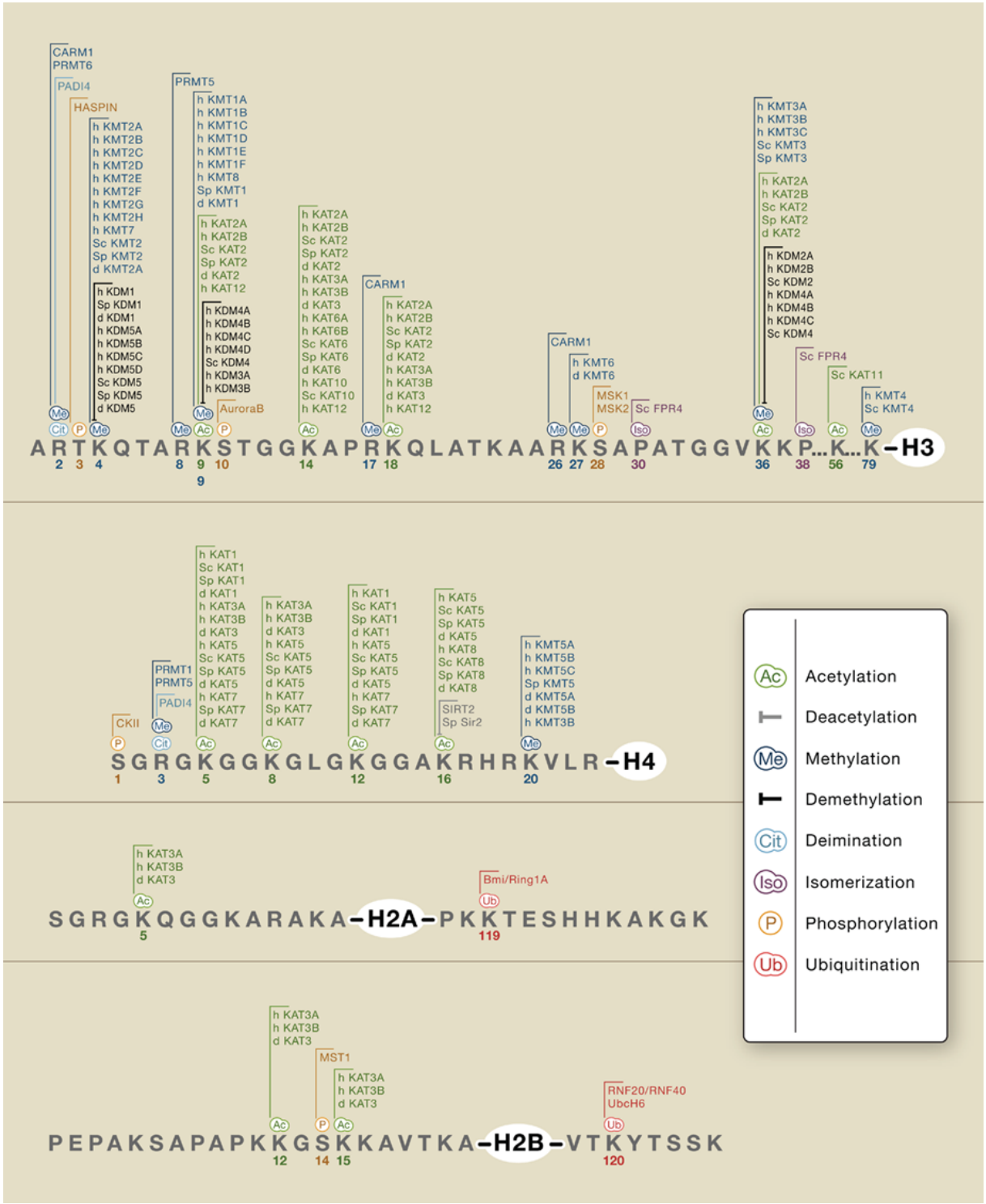


Figure 1.9: Histone modifying enzymes and their sites of modification. Adapted from [Kouzarides, 2007b].

ical nature was defined as N^{ϵ} - acetylation a few years later [Gershey et al., 1968]; [Vidali et al., 1968]. Since its discovery, it took scientists a while to find out the enzymes responsi-

ble for acetyl group addition to or removal from target proteins. These enzymes known as histone acetyltransferases (HATs) and deacetylases (HDACs) , respectively, were identified in mid 1990's and these discoveries lead to the renewed interest in this PTM as they were found to be linked with gene regulation [Brownell and Allis, 1995]; [Kleff et al., 1995]; [Bannister and Kouzarides, 1996]; [Brownell et al., 1996]; [Mizzen et al., 1996]; [Ogryzko et al., 1996]; [Taunton et al., 1996]; [Yang et al., 1996b]; [Yang et al., 1996a]. The general model of transcription is based on the interaction among RNA Pol II, general transcription factors, coactivators, corepressors, and sequence-specific DNA-binding proteins (DBPs), which confer tissue and signal-dependent specificity. Coactivator and corepressor complexes contain a variety of chromatin-modifying enzymes, including HATs and HDACs.

Although all HATs contain an acetyltransferase domain (proven or suspected), but sequence comparison divide them into subfamilies according to additional shared domains. Generally, HATs are classified into two groups, HAT A and HAT B, depending on the mechanism of catalysis and on cellular localization. The HAT B family act predominantly in cytoplasm, acetylating free histones prior to their deposition into chromatin. This class of HATs is highly conserved and all type-B HATs share sequence homology with scHat1, the founding member of this type of HAT. Type-B HATs acetylate newly synthesized histone H4 at K5 and K12 (as well as certain sites within H3), and this pattern of acetylation is important for deposition of the histones, after which these PTM marks are removed [Parthun, 2007]. The members of the HAT A family are more diverse class of enzymes. They are found in the nucleus, where they transfer the acetyl group from Acetyl-CoA to an ϵ-NH2 group of histone N-terminal after the assembly into nucleosomes. The HAT A family can be further divided into three subclasses depending on the amino acid homology with yeast proteins and conformational structure. These classes include GNAT, MYST and CBP/p300 families [Hodawadekar and Marmorstein, 2007]. Each of these enzymes targets multiple sites within the histone N-terminal tails. Indeed, their presence in numerous transcriptional coactivators is perfectly correlated with their ability to neutralize positive charges, thereby disrupting the stabilizing influence of electrostatic interactions between DNA and histones [Yang and Seto, 2007]. It has been shown that acetylation on some histone fold domain residues, specially H3K56, by hGNC5 is also involved in this regulation [Tjeertes et al., 2009]. The H3K56 side chain points towards the DNA major groove, suggesting that acetylation would affect histone/DNA interaction, a situation reminiscent of the proposed effects of acetylating the histone N-terminal tail lysines. Interestingly, knockdown of the p300 HAT has also been shown to be associated with the loss of H3K56ac [Das et al., 2009], suggesting that p300 may also target this site. However, unlike GCN5 knockdown, p300 knockdown increases DNA damage, which may indirectly affect H3K56ac levels [Tjeertes et al., 2009].

HDACs can be grouped into four classes in relation to their phylogenetic conservation [Gregoretti et al., 2004]; [Yang and Seto, 2007]. Class I, class II, and class IV, which are closely related to the yeast Rpd3, Had1, and Hos3 proteins, respectively, encompass the classical family of zinc-dependent HDACs, while class III consists of the NAD^+-dependent yeast Sir2 homologues, which comprise the sirtuin family [de Ruijter et al., 2003]; [Haigis and Guarente, 2006]. HDAC enzymes oppose the effects of HATs and reverse lysine acetyla-

tion, an action that restores the positive charge of the lysine. This potentially stabilizes the local chromatin architecture and is consistent with HDACs being predominantly transcriptional repressors. In general, HDACs have relatively low substrate specificity by themselves, a single enzyme being capable of deacetylating multiple sites within histones. The issue of enzyme recruitment and specificity is further complicated by the fact that the enzymes are typically present in multiple distinct complexes, often with other HDAC family members. For instance, HDAC1 is found together with HDAC2 within the NuRD , Sin3a and Co-REST [Yang and Seto, 2008]. Thus, it is difficult to determine which activity (specific HDAC and/or complex) is responsible for a specific effect. Nevertheless, in certain cases, it is possible to at least determine which enzyme is required for a given process. For example, it has been shown that HDAC1, but not HDAC2, controls embryonic stem cell differentiation [Dovey et al., 2010].

Histone Lysine Methyltransferases (HMTs) & Histone Demethylases (HDMs)

Histone methylation mainly occurs on the side chains of lysines and arginines. Unlike acetylation and phosphorylation, however, histone methylation does not alter the charge of the histone protein but it does affect the hydrophobic and steric properties. Histone methylation increases the affinity of certain proteins, such as transcription factors, toward DNA. Furthermore, there is an added level of complexity to bear in mind when considering this modification, depending on the specific functional properties of the associated methyltransferase; lysines may be mono-, di- or tri-methylated, whereas arginines may be mono-, symmetrically or asymmetrically di-methylated [Ng et al., 2009]; [Bedford and Clarke, 2009]; [Lan and Shi, 2009].

I **Lysine methylation**

Histone methyltransferase (HMTs) can be divided into three classes: SET domain containing lysine methyltransferases, non-SET domain containing lysine methyltransferases, and arginine methyltransferases [Figure 1.10]. All three classes use S-adenosylmethionine (SAM) as a co-enzyme to transfer methyl groups [Smith and Denu, 2009]. Lysine methyltransferases have striking target specificity, and they usually modify one single lysine on a single histone and their output can be either activation or repression of transcription. SUV39H1 that targets H3K9 was the first histone lysine methyltransferase (HKMT) to be identified [Rea et al., 2000]. Numerous HKMTs have since been identified, the vast majority of which methylate lysines within the N-terminal tails. Strikingly, all of the HKMTs that methylate N-terminal lysines contain a so-called SET domain that harbors the enzymatic activity. However, an exception is the Dot1 enzyme that methylates H3K79 within the histone globular core and does not contain a SET domain. Why this enzyme is structurally different than all of the others is not clear, but perhaps this reflects the relative inaccessibility of its substrate H3K79.

The SET domain-containing class of methyltransferases is best characterized and has been associated with metabolic diseases [Lee et al., 2008]; [El-Osta et al., 2008]; [Brasac-

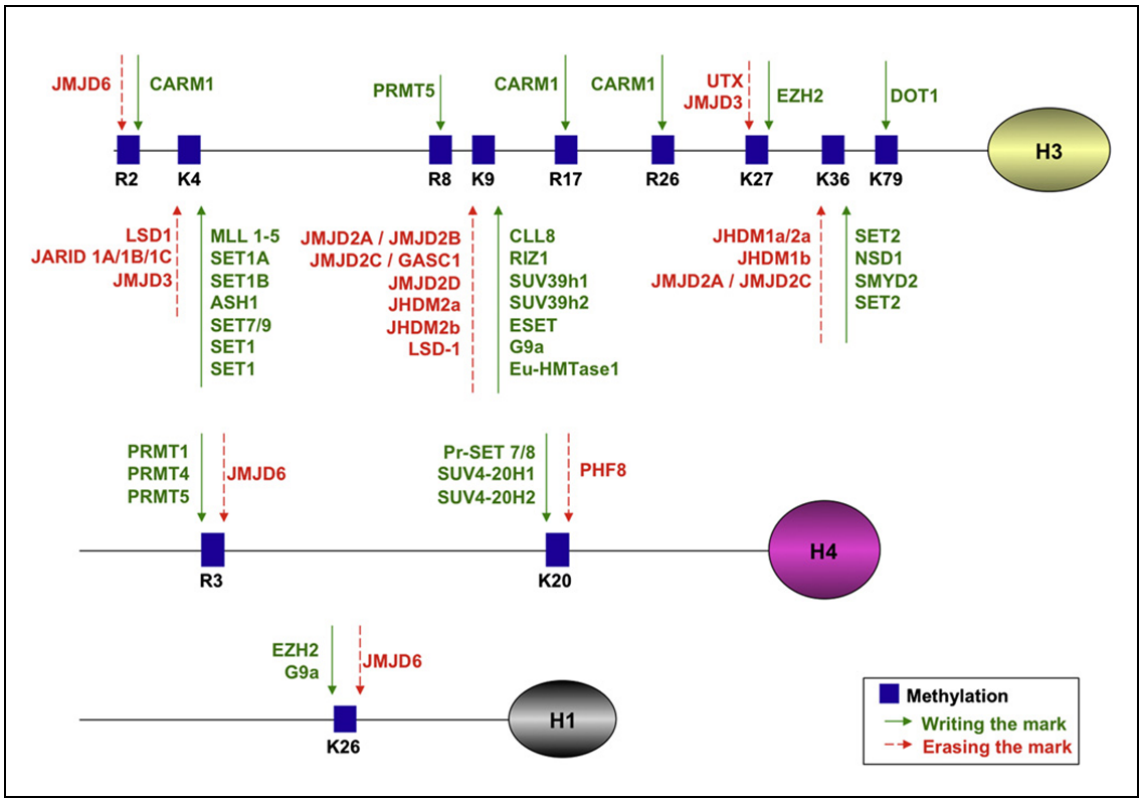


Figure 1.10: Histone lysine methyltransferases and demethylases along with their specificity for particular histone residue.

chio et al., 2009]. The SET domain is an evolutionary conserved domain, initially identified in Drosophila PEV (positive effect variegation) suppressor SU(VAR)39 [Tschiersch et al., 1994], the polycomb group protein Enhancer of zeste [Jones and Gelbart, 1993], and the trithorax group protein Trithorax [Stassen et al., 1995]. Although the major role of these methyltransferases is the modulation of gene activity via histone methylation and alteration of chromatin structure [Rea et al., 2000], they also target several non-histone proteins. In particular, the tumor suppressors p53 [Chuikov et al., 2004]; [Huang et al., 2010]; and pRb [Munro et al., 2010], and the estrogen receptor a [Subramanian et al., 2008] are substrates of SET-domain-containing methyltransferases.

HKMTs tend to be relatively specific enzymes. For instance, Neurospora crassa DIM5 specifically methylates H3K9 whereas SET7/9 targets H3K4. Furthermore, KMT enzymes also modify the appropriate lysine to a specific degree (i.e., mono-, di- and/or tri-methyl state). Maintaining the same examples, DIM5 can trimethylate H3K9 [Tamaru et al., 2003] but SET7/9 can only mono-methylate H3K4 [Xiao et al., 2003]. These specific reaction products can be generated using only the purified enzymes; so the ability to discriminate between different histone lysines and between different methylated states is an intrinsic property of the enzyme. It turns out from X-ray crystallographic studies that there is a key residue within the enzymes catalytic domain that determines

whether the enzyme activity proceeds past the mono-methyl product. In DIM5, there is a phenylalanine (F281) within the enzymes lysine-binding pocket that can accommodate all the methylated forms of the lysine, thereby allowing the enzyme to generate the tri-methylated product [Zhang and Tang, 2003]. In contrast, SET7/9 has a tyrosine (Y305) in the corresponding position such that it can only accommodate the monomethyl product [Xiao et al., 2003]. Elegant mutagenesis studies have shown that mutagenesis of DIM5 F281 to Y converts the enzyme to a monomethyltransferase, whereas the reciprocal mutation in SET7/9 (Y305 to F) creates an enzyme capable of trimethylating its substrate [Zhang and Tang, 2003]. More generally speaking, it seems that the aromatic determinant (Y or F) is a mechanism widely employed by SET domain-containing HKMTs to control the degree of methylation [Collins et al., 2005].

II **Arginine methylation**

There are two classes of arginine methyltransferase, the type-I and type-II enzymes [Figure 1.10]. The type-I enzymes generate Rme1 and Rme2as (asymmetric), whereas the type-II enzymes generate Rme1 and Rme2s (symmetric). Together, the two types of arginine methyltransferases form a relatively large protein family (11 members), the members of which are referred to as PRMTs . All of these enzymes transfer a methyl group from SAM to the ω-guanidino group of arginine within a variety of substrates. With respect to histone arginine methylation, the most relevant enzymes are PRMT1, 4, 5 and 6 [Wolf, 2009]. Methyltransferases, for both arginine and lysine, have a distinctive extended catalytic active site that distinguishes this broad group of methyltransferases from other SAM-dependent enzymes [Copeland et al., 2009]. Interestingly, the SAM-binding pocket is on one face of the enzyme, whereas the peptidyl acceptor channel is on the opposite face. This indicates that a molecule of SAM and the histone substrate come together from opposing sides of the enzyme [Copeland et al., 2009]. Indeed, this way of entering the enzymes active site may provide an opportunity to design selective drugs that are able to distinguish between histone arginine/lysine methyltransferases and other methyltransferases such as DNMTs.

III **Histone demethylases**

For many years, histone methylation was considered a stable, static modification. Nevertheless, in 2002, a number of different reactions/pathways were suggested as potential demethylation mechanisms for both lysine and arginine [Bannister et al., 2002], which were subsequently verified experimentally. Initially, the conversion of arginine to citrulline via a deimination reaction was discovered as a way of reversing arginine methylation [Cuthbert et al., 2004]; [Wang et al., 2004a]. Although this pathway is not a direct reversal of methylation, this mechanism reversed the dogma that methylation was irreversible. More recently, a reaction has been reported which reverses arginine methylation. The jumonji protein JMJD6 was shown to be capable of performing a demethylation reaction on histones H3R2 and H4R3 [Chang et al., 2007]. However, these findings have yet to be recapitulated by other independent researchers.

In 2004, the first lysine demethylase was identified. It was found to utilize FAD as co-factor, and it was termed as lysine-specific demethylase 1 (LSD1) [Shi et al., 2004]. The

demethylation reaction requires a protonated nitrogen and it is therefore only compatible with mono- and dimethylated lysine substrates. In vitro, purified LSD1 catalyses the removal of methyl groups from H3K4me1/2, but it cannot demethylate the same site when presented within a nucleosomal context. However, when LSD1 is complexed with the Co-REST repressor complex, it can demethylate nucleosomal histones. Thus, complex members confer nucleosomal recognition to LSD1. Furthermore, the precise complex association determines which lysine is to be demethylated by LSD1. As already mentioned, LSD1 in the context of Co-REST demethylates H3K4me1/2, but when LSD1 is complexed with the androgen receptor, it demethylates H3K9. This has the effect of switching the activity of LSD1 from a repressor function to that of a coactivator [Klose and Zhang, 2007].

In 2006, another class of lysine demethylase was discovered [ichi Tsukada et al., 2006]. Importantly, certain enzymes in this class were capable of demethylating trimethylated lysines [Whetstine et al., 2006]. They employ a distinct catalytic mechanism from that used by LSD1, using Fe(II) and α-ketoglutarate as co-factors, and a radical attack mechanism. The first enzyme identified as a trimethyl lysine demethylase was JMJD2 that demethylates H3K9me3 and H3K36me3 [Whetstine et al., 2006]. The enzymatic activity of JMJD2 resides within a JmjC jumonji domain. Many histone lysine demethylases are now known and, except for LSD1, they all possess a catalytic jumonji domain [Mosammaparast and Shi, 2010]. As with the lysine methyltransferases, the demethylases possess a high level of substrate specificity with respect to their target lysine. They are also sensitive to the degree of lysine methylation; for instance, some of the enzymes are only capable of demethylating mono- and dimethyl substrates, whereas others can demethylate all three states of the methylated lysine.

1.4.4 Functional Consequences of Histone Modifications

As discussed earlier, histone PTMs, in general, affect genome regulation by altering the nucleosome architecture. Several mechanisms have been described [Allis et al., 2007]; first, histone modifications can directly modify the structure of chromatin by preventing crucial contacts that facilitates certain chromatin conformation or higher order structures. Such effects are termed as *cis*-modifying effects. On the other hand, two other mechanisms can be considered to operate in *trans* i.e. some histone PTMs may disrupt the binding of proteins that associate with chromatin or alternatively, may provide binding platforms to attract various effector proteins. This second *trans*-acting mechanism has been studied and characterized in most detail. Therefore, various modifications or set of modifications on a given histone dictate a set of proteins to bind or occlude from chromatin. These proteins may also carry with them enzymatic activities (e.g., remodeling ATPases) that further modify chromatin. The need to recruit an ordered series of enzymatic activities comes from the fact that the processes regulated by modifications (transcription, replication, repair) have several steps. Each one of these steps may require a distinct type of chromatin-remodeling activity and a different set of modifications to recruit them.

I ***Direct Structural Perturbations or CIS-modifying effects***

Among the modifications that disrupt certain chromatin conformations by reducing the positive charge of histones are acetylations and phosphorylations. These modifications perturb the electrostatic interactions between histones in adjacent nucleosomes or histone DNA interactions. Hence, these modifications may lead to less compact chromatin structure, thereby facilitating DNA access by cellular machineries involved in DNA-based processes, such as transcription, DNA damage response & repair, replication and recombination. Histones are found to be hyper-acetylated on multiple residues (H3K9, H3K14, H3K18, H4K5, H4K8, H4K12, H4K16 [Kouzarides, 2007a]) in particular regions of genome (enhancers, promotors etc. [Wang, 2008]), thereby neutralizing the positive charge, leading to profound effects on the chromatin structure. However, multiple acetylations are not a prerequisite for inducing the structural change in chromatin as for example acetylation on H4K16 can *in vitro* negatively affect the formation of 30-nm chromatin fiber [Shogren-Knaak et al., 2006].

Histone phosphorylation is another example of *cis*-modifying change that alters the chromatin structure by altering the charge on a single site like H4K16Ac. A well known example of this mechanism is H3S10 phosphorylation during mitosis, which occurs genome-wide and associated with highly condensed form of chromatin [Wei et al., 1998]. Although phosphorylation imparts negative charge to the histones and in principle it should decrease its interaction with the negatively charged DNA, it may be that the displacement of heterochromatin protein 1 (HP1) from heterochromatin during metaphase by uniformly high levels of H3S10 phosphorylation [Fischle et al., 2005] is required to promote the detachment of chromosomes from the interphase scaffolding. This would facilitate chromosomal remodeling that is essential for its attachment to the mitotic spindle. Ubiquitination, on the other hand does not impart any charge to the histones but due to its extremely large size, it might affect intra-nucleosomal interactions and/or interactions with other chromatin bound proteins.

II ***Regulating the Binding of Chromatin Proteins or TRANS-modifying effects***

There is an ever-increasing number of chromatin proteins that interact specifically with histones through a variety of distinct domains. These proteins are generally part of large proteins complexes and hence each complex has many distinct domains that allow simultaneous recognition of various modifications, sometimes on the same histone and sometimes on different histones in the same neighborhood. In general, methylations are recognized by chromo-like domains of the Royal family (chromo, tudor, PWWP and MBT domains) and PHD domains [Champagne and Kutateladze, 2009]; [Maurer-Stroh et al., 2003]; [Kim et al., 2006], acetylations are recognized by bromodomains and more recently a PHD domain has also been shown to recognize histone acetylation [Zeng et al., 2010]. Phosphorylation is recognized by a domain within 14-3-3 proteins.

Experimental evidence that histone lysine methylation is recognized by many distinct domain types, underscores the relative importance of this modification. It has been shown that numerous lysine methyl binding domains can recognize the same modi-

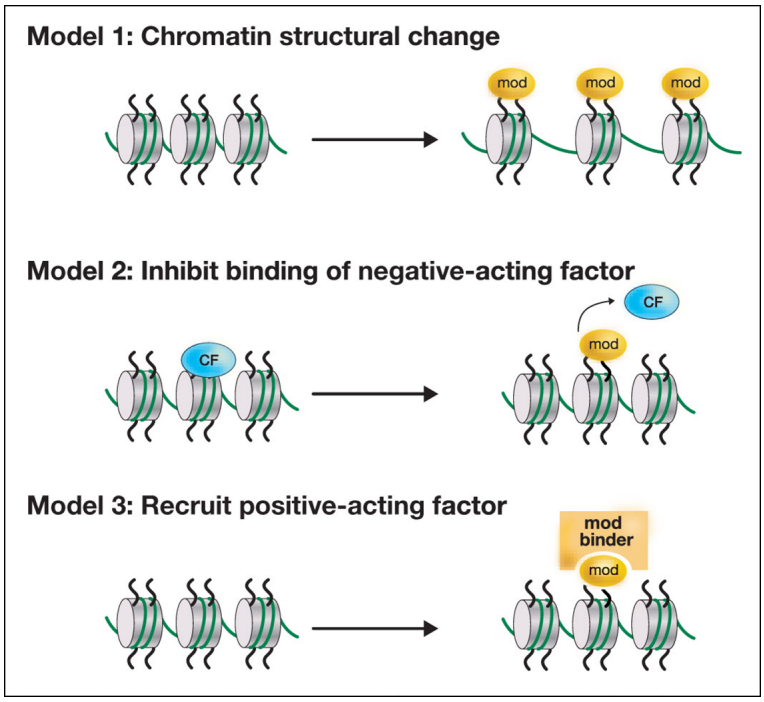


Figure 1.11: Models showing how histone PTMs affect the chromatin template. Model-1 corresponds to *cis*-modifying effects, while Model-2 & -3 corresponds to *trans*-modifying effects. Adapted from [Allis et al., 2007].

fied histone lysine. For example, H3K4me3 - a marker of active transcription, is recognized by the PHD finger domain of ING family of proteins (ING1-5) (The ING protein binding leads to the recruit of additional chromatin modifiers such as HATs and HDACs. For example, ING2 tethers the repressive mSin3a-HDAC1 HDAC complex to active growth-specific genes following DNA damage [Shi et al., 2006]), by tandem chromodomains within CHD1 , an ATP-dependent remodeling complex, capable of repositioning the nucleosomes [Sims et al., 2005], and by tandem tudor domains within JMJD2A, a histone demethylase [Huang et al., 2006]. Another example of specific binding of methylated lysine is the recognition of H3K9me3 by HP1. H3K9me3 is a mark associated with repressive heterochromatin. HP1 binds to this mark via its N-terminal chromodomain [Bannister et al., 2001]; [Lachner et al., 2001] and dimerize via their C-terminal chromoshadow domains to form a bivalent chromatin binder. HP1 has also been shown to bind to methylated H1.4K26 via its chromodomain [Daujat et al., 2005]. All these interactions of HP1 (H3K9me and H1.4K26me) regulate the heterochromatin architecture and plays fundamental role in chromatin compaction.

HATs and several chromatin remodeling complexes often contains bromodomains that specifically recognize acetylated histone lysines [Mujtaba et al., 2007]. For instance, the bromodomain of Swi2/Snf2 targets it to acetylated histones and in turn recruits the SWI/SNF remodeling complex, which "relaxes" the chromatin [Hassan et al., 2002]. More recently, the PHD fingers have been shown to recognize the acetylated histones.

The DPF3b protein is a component of the BAF chromatin remodeling complex and it contains tandem PHD fingers that are responsible for recruiting the BAF complex to acetylated histones [Zeng et al., 2010]. Among the proteins that bind specifically to phosphorylated histones is MDC1, which is involved in the DNA damage recognition and is recruited to the sites of DNA double strand breaks (DSB). MDC1 , through its tandem BRCT domains, binds to γH2AX, the DSB induced phosphorylated H2A variant [Stucki et al., 2005].

Various histone modifications can also function to prevent docking of non-histone proteins onto the chromatin. Examples include H3K4me disrupting the binding of the NuRD complex and H3T3ph preventing the binding of the INHAT complex to the H3 tail. Both of these complexes have transcriptional repressor activity, so their occlusion by modifications associated with active transcription makes sense [Zegerman et al., 2002]; [Schneider et al., 2004b]. H3K4 methylation also disrupts the binding of DNMT3L's PHD finger to the H3 tail [Adams-Cioaba and Min, 2009].

Another way of avoiding the binding of non-histone proteins is to remove respective histone modifications by removing the histone N-terminal tail in which they reside, a process referred to as tail clipping. It was first identified in Tetrahymena [Allis et al., 1980] but much more recently also detected in mouse [Duncan et al., 2008] and yeast [Santos-Rosa et al., 2009]. It has been shown that the endopeptidase activity is sensitive to the modification status of the H3 tail. For instance, the presence of H3K4me3, an active mark, is inhibitory for clipping, whereas the repressive H3R2me2 mark is not. This suggests that the presence of an active mark on the H3 tail impedes binding of the endopetidase. Furthermore, it also suggests that during activation H3 tails methylated at H3K4 may be protected from being clipped, which indicates the selectivity of the tail clipping towards repressive marks, consistent with the idea that clipping may be necessary for the removal of transcription inhibitors at the onset of gene expression. Repressive protein complexes bound to the H3 tail at promoter regions may thus be removed to allow activator protein complexes to take over during the induction process. Thus, H3 tail clipping can be considered as a histone modification that allows "mass clearance" of repressive marks and proteins at promoters, a process that may be a necessary prelude to gene activation.

1.4.5 Histone Modification Cross-talk

To summarize, a decade of extensive research in the field of histone biology has revealed that histones are not merely static architectural proteins whose only function is to wrap DNA, but they are decorated with abundance of various covalent modifications, which adds an extra level of complexity to the dynamic chromatin structure. The presence of these modifications, either on the same histone or on adjacent histones, makes their "cross talk" with each other highly likely, providing grounds for their mechanistic communications during various cellular processes. This communication might occur via several different mechanisms [Bannister and Kouzarides, 2011], which are numerated below with examples, [Figure 1.12];

I There may be competitive antagonism between modifications if distinct type of modifications occur on the same residue(s). This is, in particular true for lysines that can be acetylated, methylated or ubiquitylated. So in this case these modifications are mutually exclusive [Table1.2].

II There are situations where one modification may be dependent upon another or is required for the other modification, either on the same histone or a different histone, to take place. The best studied example of this trans-regulation comes from the work in Saccharomyces cerevisiae, where methylation of H3K4 by scCOMPASS and of H3K79 by scDot1 is completely dependent upon the ubiquitination of H2BK123 by scRad6/Bre1 [Lee et al., 2007]. Intriguingly, this mechanism is conserved in higher eukaryotes too, including humans [Kim et al., 2009]. However, it has also been shown that H3K4 methylation can be interfered by H3R2 methylation [Guccione et al., 2007]; [Kirmizis et al., 2007].

III The docking of a protein to a particular modification can be disrupted or occluded by an adjacent modification. For example, as discussed above, HP1 binds to H3K9me2/3, but during mitosis, the binding is disrupted due to phosphorylation of H3S10 [Fischle et al., 2005]. This eviction of HP1 from H3K9me2/3 can be viewed as a binary 'methyl/phospho switch', which permits the dynamic control of the HP1H3K9me interaction [Fischle et al., 2003]. A similar binary switch has also been discovered in linker histone H1.4, where HP1 binding to H1.4K26me is occluded by phosphorylation of H1.4S27 [Daujat et al., 2005]. For this kind of communication, the modified amino acids do not necessarily have to be directly adjacent to each other. For example, in S. pombe, acetylation of H3K4 inhibits binding of spChp1 to H3K9me2/3 [Xhemalce and Kouzarides, 2010].

IV An enzyme's catalytic activity may be affected due to modification of its substrate recognition site. In yeast, for example, the interconversion of the H3P38 peptide bond is catalysed by scFpr4 proline isomerase and this conformational change affects the ability of the scSet2 methyltransferase to methylate H3K36, which is linked to the effects on gene transcription [Nelson et al., 2006].

V There may be cooperation between modifications, where the presence of certain modifications effectively increases the recognition and binding of a protein to another modification. For example, PHF8 specifically binds to H3K4me3 via its PHD finger, and this interaction is stronger when H3K9 and H3K14 are also acetylated on the same tail of H3 [Vermeulen et al., 2010]. However, this might not be the direct effect of these marks on PHF8, but instead some additional proteins in a complex with PHF8 can ultimately lead to stabilization of PHF8.

VI A cross-talk may exist between histone modifications and DNA methylation. For instance, the UHRF1 protein binds to nucleosomes bearing H3K9me3, but this binding is significantly enhanced when the nucleosomal DNA is CpG methylated [Bartke et al., 2010]. Conversely, DNA methylation can inhibit protein docking to specific histone

modifications. A good example here is KDM2A , which only binds to nucleosomes bearing H3K9me3 when the DNA is not CpG methylated [Bartke et al., 2010].

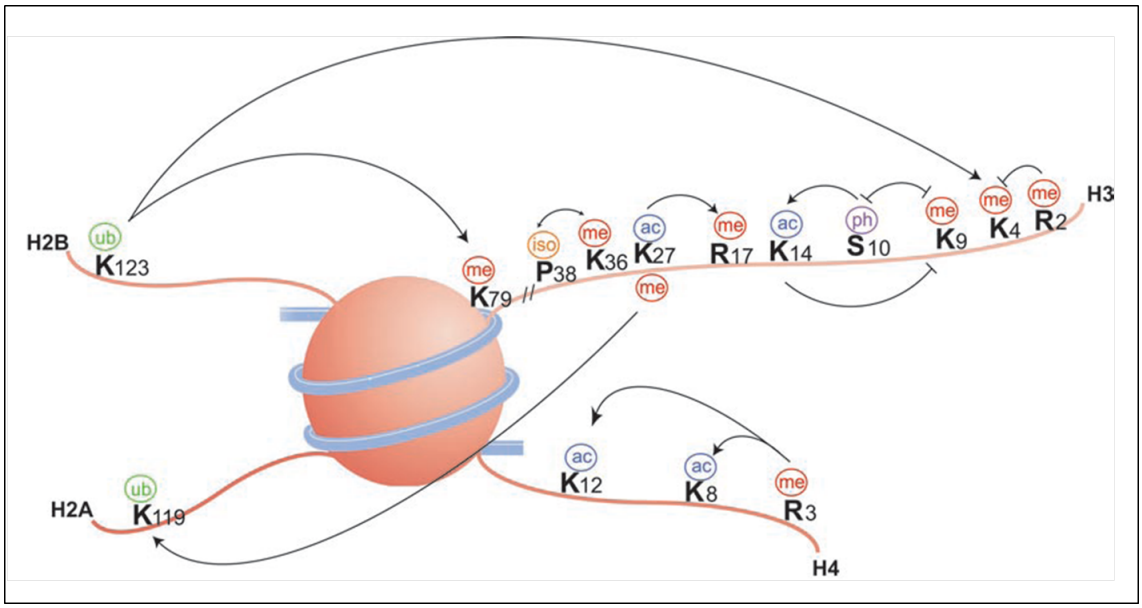


Figure 1.12: Histone modifications cross-talk. Adpated from [Bannister and Kouzarides, 2011].

VII It is known that histone-modifying enzymes are often found in large multi-subunit complexes. In order, for these large complexes, to anchor onto the nucleosome, a modification on the nearby residues can create binding sites for the components of the complex. For example, the PHD finger of Yng1, a subunit of the NuA3 histone acetyltransferase complex, recognizes methylated H3K4 and helps recruit this histone acetyltransferase complex for acetylation of H3K14 [Martin et al., 2006]; [Taverna et al., 2006]. An ING family protein, ING2, through its PHD finger, can also bind methylated H3K4, although it is present in a histone deacetylase complex [Shi et al., 2006]. Therefore, H3K4 methylation can serve as a landing platform for a variety of histone-modifying enzymes with opposing activities.

VIII A crosstalk between two different modifications on histone H3 and H4 tails can regulate transcription elongation through the sequential recruitment of proteins that bind these modifications. For instance, activation of FOSL1 gene requires the binding of a kinase PIM1 to the FOSL1 enhancer. PIM1 phosphorylates histone tail residue H3S10. Phosphorylated H3S10 creates a binding site for 14-3-3, a phosphoserine binding protein, which interacts with histone acetyltransferase MOF to acetylate histone H4 tail residue H4K16. Acetylated H4K16 can in turn provide a binding platform for the bromodomain-containing protein Brd4, a component of P-TEFb , a kinase that phosphorylates the C-terminal domain of RNA Pol II to facilitate transcription elongation. However, the timing, location, and perhaps the identity of the H3 kinase, and not the H3S10 modification alone, determines downstream events, as it has also been shown that at an earlier stage of serum stimulation, an MSK1/2 kinase is recruited to the pro-

moter where it phosphorylates H3S10. 14-3-3 is then recruited to the promoter, but unlike the situation at the enhancer, MOF is not recruited to the promoter [Zippo et al., 2009].

It is obvious from the above examples that various types of histone cross-talk mechanisms, involving numerous histone modifying and nucleosome remodeling complexes, could exist in the cells. This cross-talk leads us towards two functional categories of biological readouts, i.e. the establishment of global chromatin environments and the orchestration of DNA-based biological processes. To establish a global chromatin environment, modifications help to partition the genome into distinct domains such as euchromatin, where DNA is kept "accessible" for transcription, and heterochromatin, where chromatin is "inaccessible" for transcription. Although there is a high degree of overlap between different chromatin regions, there are still regions of demarcation between heterochromatin and euchromatin. These boundary elements have specific histone modification signatures [Barski et al., 2007]; [Ong and Corces, 2009]. On the other hand, in order to facilitate DNA-based functions, modifications orchestrate the unravelling of chromatin to help the execution of a given biological function. This may be a very local function, such as transcription of a gene or the DNA damage response and subsequent repair or it may be a more global genome wide function, such as DNA replication or chromosome condensation. All these biological tasks require the sequentially ordered recruitment of the machinery to unravel DNA, manipulate it and then put it back to the correct chromatin state. In order to correctly undertake these tasks, whether its the establishment of global chromatin domain or execution of a particular cellular function, it requires a definite set of modifications, which act either sequentially or in a combinatorial manner. This idea of combination of various histone modifications for specific biological readouts or alternative chromatin states was the basis of the "Histone Code Hypothesis" [Turner, 2000]; [Strahl and Allis, 2000]; [Jenuwein and Allis, 2001].

1.5 Histone Code Hypothesis

As described previously, chromatin is subject to a variety of post-translational modifications. These modifications are evolutionarily conserved, and have been shown to affect chromatin structure and as a result affect essential processes involving DNA, such as transcription, replication, DNA damage response, recombination and mitotic condensation of chromosomes. Furthermore, some of these changes are inherited from one cell generation to another, thus providing epigenetic information. The extremely high diversity and distinct patterns of these post-translational modifications and their involvement in specific biological processes have lead scientists to believe in existence of so called "Histone Code", encoded onto both N-terminal tails and HFD of histones [Turner, 2000]; [Strahl and Allis, 2000]; [Jenuwein and Allis, 2001]. These histone modifications, which can also be regarded as 'epigenetic marks', are deposited by a large panel of histone modifying enzymes (histone code writers) and subsequently read or recognized by various non-histone proteins (histone code readers). The readers, thus, can act as a translation machinery for the information encoded in the

histone modifications. These epigenetic marks on histones, represent a critical feature of a genome-wide mechanism of information storage and retrieval that is only beginning to be understood. It has becoming more and more evident that "histone code" has the ability to considerably extend and complement the information potential of the genetic (DNA) code.

Thus histone code hypothesis predicts that:

- Distinct modifications on histone tails and HFD generate docking sites for a large number of non-histone chromatin-associated proteins.
- Modifications on the same or different histone tails may be interdependent and generate various combinations or act sequentially on any one nucleosome, thereby creating unique biological outcomes.
- Distinct regions of higher order chromatin, such as euchromatic or heterochromatic domains, are largely dependent on the local concentration and combination of differentially modified nucleosomes and/or the presence of alternative histone variants, and hence, give rise to the idea of alternative chromatin states.

As a further extension of the initial hypothesis of histone code, the idea of "binary switches" and "modification cassettes" has been suggested [Fischle et al., 2003]. Binary switch represents the differential readout of distinct combinations of marks on two neighboring residues, where one modification influences the binding of an effector module onto another modification on an adjacent or nearby residue. A "methyl/phospho" switch on histone H3 can be regarded as the prototype example of this model. Methylation of histone H3K9 is recognized and bound by heterochromatin protein 1 (HP1), however, during mitosis and chromatin condensation, H3S10 is phosphorylated, which reduces HP1 binding to H3K9me. This binding is further reduced when H3 becomes acetylated at K14, which removes HP1 from the mitotic chromosomes [Mateescu et al., 2004]. In addition to the inhibition of HP1 binding to the H3K9me, phosphorylation of H3S10 interferes with the methylation of H3K9 by SUV39H1 [Garcia et al., 2005]; [Rea et al., 2000].

Similarly, the combination of short clusters of modifications, situated at strategic locations in the histones, could be regarded as "modification cassettes". Depending upon their modification state, they can lead to distinct biological readouts. Such short 'cassettes' where at least three out of five sites separated by no more than one residue can be covalently modified are readily spotted in the H3 and H4 tails, e.g. the stretch between H3K9 and H3T11 can carry four different covalent marks. Various combinations of modification on adjacent sites within these short clusters can exist which will influence the recognition and binding of modules to their respective binding mark. Especially since the recognition of short clusters of 510 amino acids is a hallmark of the interaction of most modules with their target interaction marks [Pawson et al., 2001]. Furthermore, modules that directly read complex patterns of marks might exist [Fischle et al., 2003].

1.5.1 Histone Code in Transcriptional Regulation

Scientists have been studying the regulation of gene expression for a very long time and their understanding has grown from identifying transcription factors and their binding sites to include a wide variety of other binding events regulated or affected by the covalent modifications of histones that package the DNA. Experimental evidence suggest that these histone modifications act as signals to attract various proteins, which in turn influence the gene expression. As discussed in [subsection 1.4.5], histone modifications do not occur in isolation but rather in combinatorial fashion, and hence an extensive cross-talk could be imagined among various modifications that can be summed up as the language of histone code. Traditionally, histone code has been studied in the context of transcriptional regulation and various histone marks have been shown to represent alternative forms of gene expression.

It has been demonstrated that reversible acetylation of specific lysines of core histones is correlated with gene activity. Active genes exhibited high levels of histone acetylation, whereas repressed genes usually have low levels of acetylation [Turner, 2000]. Localization in different regions of genes suggests that these modifications have different functions. For example, H3K9ac, H3K14ac and H4K8ac is elevated in promoter regions, consistent with a role in transcriptional initiation [Agalioti et al., 2002]; [Wang, 2008], whereas H4K12ac was detected at high levels in transcribed regions, suggesting a role in transcriptional elongation [Carey et al., 2006].

Active transcriptional enhancers contain relatively high levels of H3K4me1, a reliable predictive feature [Hon et al., 2009]. However, active genes themselves possess a high enrichment of H3K4me3, which marks the transcriptional start site (TSS) [Barski et al., 2007]; [Schneider et al., 2004a]. It has been shown that H3K4me2 exists in both active and repressive genes [Santos-Rosa et al., 2002], which strongly suggests that not only the particular site but also the number of methyl groups in a modification plays an important role in the functional consequences of histone methylation on gene expression. Further on, histone methylations on H3K9, H3K27, and H4K20 are generally associated with a repressed chromatin state. Histone H3K27me3, along with Polycomb repressor complexes, is a marker of facultative heterochromatin, which consists of genomic regions containing genes that are differentially expressed through development and/or differentiation and which then become silenced [Hansen et al., 2008]. However, high levels of H3K9me3, along with HP1 α/β are found in constitutive heterochromatin, which contains permanently silenced genes in genomic regions such as the centromeres and telomeres [Trojer and Reinberg, 2007].

As discussed in 1.4.5, histone modifications do not occur in isolation but rather in combinatorial fashion, and hence an extensive cross-talk could be imagined among various modifications. Evident from the variety and density of the modifications identified so far, it is likely that the histone modifications, in combinatorial way, have an impact on one another, both in a *cis* or *trans*-acting manner. Indeed, H4K20 methylation, which is a mark of silent chromatin, has been shown to prevent acetylation on H4K16, a mechanism involved in gene activation [Nishioka et al., 2002a]. Furthermore, methylation of H3K4 by Set9 has been shown to impair H3K9 methylation by Suv39h1 and to abolish the recruitment of the ATP-dependent

HDAC-containing chromatin remodeling complex NuRD, involved in gene silencing [Nishioka et al., 2002b]. Finally, H3K4 methylation seems to be controlled by the ubiquitination state of H2BK123

A common theme of recent research on histone code interpretation is that the order and mechanism of the addition and removal of histone modifications are important for the transcriptional readout of a gene. The recent examples of histone crosstalk discussed earlier 1.4.5 illustrate this point. In one study, the implementation of H3S10 phosphorylation at two different locations, by two different enzymes, and at two different times after serum stimulation had disparate effects on subsequent histone acetylation at the respective locations [Zippo et al., 2009]. Zippo and colleagues were able to propose a mechanism of gene activation by identifying the histone-modifying enzymes, the histone modifications, and a set of proteins that recognized these modifications on the FOSL1 gene after serum stimulation. In another study [Wang et al., 2009], Wang and colleagues found that artificially recreating histone modifications that correlate with gene expression could result in the recruitment of RNA Pol II, but this was not sufficient for transcription. Thus, simply mapping histone modification patterns without understanding the recruitment, regulation, and interactions of the complexes implementing these marks is not sufficient to understand the mechanisms regulating gene expression. Genome-wide profiling techniques have now become widely adopted, enabling one to map histone modifications, and identify the enzymes implementing these marks, together with the factors that recruit them under different experimental conditions. Hence, the more we look into the histone code, the more it will become obvious that "context" is everything [Figure 1.13].

The idea that certain marks make categoric distinction between active or repressed chromatin may, on some occasions, represent a highly simplistic view as it has been challenged in more recent studies. For example, it has been shown that activation-related H3K4me3 and repression-related H3K27me3 of the same histone can happen in close proximity

These consistent conclusions point out that many modifications occur together, suggesting that combinatorial complexity is limited and that interdependent networks linking different marks could help understanding the modification patterns. However, it is important to understand how the histone-modifying enzymes and ATP-dependent chromatin remodellers are targeted to specific loci in order to regulate gene expression. Evidence suggests that during transcription, the targeting does not seem to occur through direct binding of the effector proteins to the underlying DNA sequence. Instead, interactions between histone-modifying enzymes and sequence-specific transcription factors seem to be necessary. For example, SWI/SNF, an ATP-dependent nucleosome remodeling complex, and HAT complexes are recruited to promoters by direct interaction with transcriptional activators [Ryan et al., 1998]; [Brown et al., 2001]. After a modification pattern has been established, it is "read" by chromatin-remodelling factors (effectors) that increase the efficiency of transcription either by altering the position or mobility of nucleosomes or by promoting the eviction of histones. The large network of histone PTMs appears to represent a mechanism for short and long term regulation of chromatin structure and activity. Fully deciphering this code will certainly impact on our understanding of DNA-templated processes in the cell.

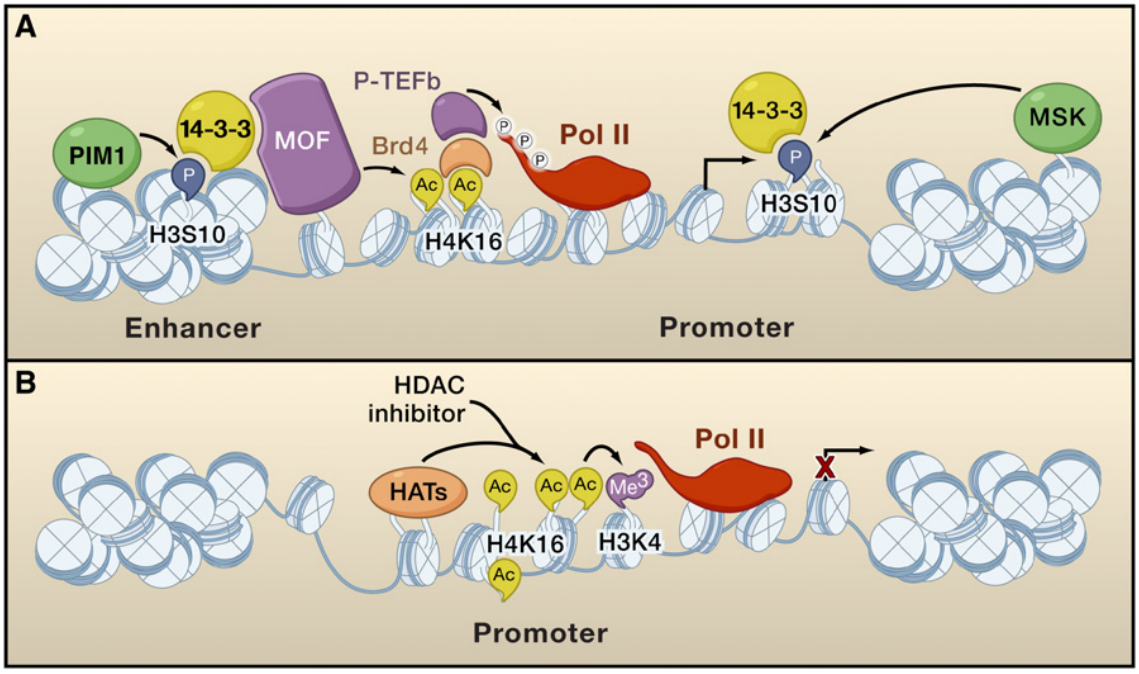


Figure 1.13: Context dependent outcomes of histone code in transcriptional regulation. [**A**]. Activation of FOSL1 requires the binding of PIM1 to the FOSL1 enhancer, which leads to the phosphorylation of H3S10. H3S10ph creates a binding site for a phosphoserine binding protein, 14-3-3, which through its interaction with MOF (HAT) acetylates H4K16. H4K16ac provides a binding plateform for Brd4 through its bromodomain. Brd4 is a component of P-TEFb, which phosphorylates RNA Pol II to facilitate transcription elongation. However, at an earlier stage of serum stimulation, an MSK1/2 kinase is recruited to the promoter where it phosphorylates H3S10. 14-3-3 is then recruited to the promoter, but unlike the situation at the enhancer, MOF is not recruited to the promoter. Thus, the timing, location, and perhaps identity of the H3 kinase, and not the H3S10 modification alone, determines downstream events. [**B**]. Another example of how the order of implementation of histone modifications can affect transcription comes from work from Wang et al. (2009). They report that despite correlations between histone acetylation and H3K4 methylation, artificially increasing acetylation through treatment of cells with deacetylase inhibitors (HDACs) does not lead to productive transcription, despite the presence of H3K4 methylation and Pol II recruitment. Therefore, patterns of histone modifications cannot simply be "read" but instead have distinct effects depending on the cellular context and upstream signaling events. Adapted from [Zippo et al., 2009] & [Wang et al., 2009].

1.5.2 Histone Code in DNA Damage Response

Given that chromatin is the physiological template for all DNA-based cellular processes, its structure fulfills essential functions not only by condensing and protecting DNA but also in terms of maintenance of genome stability and controlling gene repression. However, the complexity and compactness of chromatin structure presents an obstacle in dealing with such structures as DNA lesions, which has to be repaired or eliminated in order to preserve genome stability and to faithfully transmit genetic information across generations. This can only be made possible due to the dynamic nature of chromatin. Recent studies suggest that the chromatin changes not only help to overcome the obstacles to repair machinery, but that chromatin also plays an active role in DNA damage response (DDR) , a complex signal

transduction pathway that starts by sensing DNA damage and transduces this information to trigger cellular responses to DNA damage. Various mechanisms such as DNA methylation [Kulis and Esteller, 2010], incorporation of variant histones [Talbert and Henikoff, 2010], histone PTMs [Bannister and Kouzarides, 2011], and nucleosome repositioning by ATP-dependent remodeling complexes [Clapier and Cairns, 2009] have been suggested to be involved in DDR.

Recent evidences suggest that the mechanisms similar to the histone code in transcription regulation, might exist for DNA damage response [Zhu and Wani, 2010] ; [Mendez-Acuna et al., 2010], where various chromatin modifications function by sensing and marking the lesion, recruiting components of repair machineries, facilitating their action, and finally reassemble the chromatin into its original form [Ikura and Ogryzko, 2003] ; [Escargueil et al., 2008]; [Zhu and Wani, 2010]. Below we describe the functions of various chromatin modifications and their possible combinatorial patterns before, during and after DNA damage.

I **Histone Modifications Before DNA Repair**

While studying the DNA damage response in cellular context, the modifications of chromatin structure have to be considered pivotal, as chromatin is the physiological substrate for the DDR machinery. When we talk about chromatin modifications, it encompass DNA methylation, histone PTMs, incorporation of variant histones and ATP-dependent nucleosome remodeling. However, we will focus our discussion to various histone PTMs that take part in cellular DDR pathways, although ATP-dependent nucleosome remodeling activities are also an integral part of such pathways, making the lesion more accessible to the damage sensing proteins and exposing some constitutive HFD modifications that are usually hidden under normal circumstances [Kruhlak et al., 2006].

- **Histone Phosphorylation**

  H2AX phosphorylation

  Among the histone PTMs in DDR, the earliest, best characterized and still intensively studied hallmark modification is the phosphorylation of H2AX to form γH2AX, which spreads, within minutes, to around 2 Mb of DNA region (in human cells) around the lesion and forms foci that are easily detectable by immunofluorescence microscopy [Rogakou et al., 1998]. In yeast, phosphorylated H2A can be detected as far as 50 kb on either side of double-strand break [Shroff et al., 2004]. H2AX, as described earlier, is one of the H2A variants in humans, that is expressed constitutively, distributed throughout the genome and constitutes about 2-25% of histone H2A depending upon the cell type [Redon et al., 2002]. The phosphorylation of H2AX occurs at the serine 139 of the H2AX in humans, present in a highly conserved SQ(E/D) motif in its extended C-terminal domain (corresponding to yeast S128 of histone H2A [Downs et al., 2000]). H2AX is known to be phosphorylated by members of phosphatidylinositol-3-OH-kinase-like kinases (PIKKs), principally ATM (Tel1 in S. cerevisiae), ATR (Mec1 in S. cerevisiae) and

DNA dependent protein kinase (DNA-PK; also known as PRKDC) in response to genomic insult. ATM and DNA-PK function redundantly to phosphorylate H2AX after ionizing radiation (IR) [Stiff et al., 2004], while ATR phosphorylates H2AX in response to replication stress and UV irradiation [Ward and Chen, 2001]. DNA-PKcs is known to mediate H2AX phosphorylation in hypertonically stressed cells as well as in response to DNA fragmentation during apoptosis [Reitsema et al., 2005]; [Mukherjee et al., 2006].

How do phosphorylation of H2AX influences the state of cellular DDR has been actively investigated. Research has shown that the phosphorylation of H2AX itself does not affect the chromatin organization, but rather has a role in the localization of repair factors at the break. The importance of this modification was highlighted by the knockout studies where H2AX$^{-/-}$ cells or cells with H2AX-S139A mutation showed sensitivity to ionizing radiation or camptothecin [Celeste et al., 2002]. Although the presence of γH2AX is not required for the initial recruitment of signaling and repair factors [Celeste et al., 2003], it is found to be essential for their accumulation and retention at the breaks and for the amplification of DDR signal [Celeste et al., 2002]; [Paull et al., 2000]. It has been shown that H2AX phosphorylation generates a carboxyl-terminal phospho-motif that is then recognized by the tandem BRCA1-carboxyl-terminal (BRCT) domain-containing protein, mediator of DNA damage checkpoint 1 (MDC1) [Stucki et al., 2005]. MDC1 acts as a critical scaffold for mediating downstream events, and along with γH2AX, leads to the recruitment of two downstream protein complexes: MRE11RAD50NBS50 [Chapman and Jackson, 2008]; [Melander et al., 2008]; [Spycher et al., 2008] and an E2/E3 ubiquitin ligase complex, UBC13/RNF8 [Wang and Elledge, 2007]; [Kolas et al., 2007]. It is also to be noted that H2AX is phosphorylated in its C-terminal domain, a region of the nucleosome where the DNA enters and exits. In condensed structures, such as the 30-nm fibre, this region of the nucleosome is obscured at the center of the fiber. In order to efficiently phosphorylate this motif, decondensation of the chromatin might be required through the action of ATP-dependent nucleosome remodeling complexes [Kruhlak et al., 2006]. In this context, it has been shown that in the absence of RSC (Remodels the Structure of Chromatin Complex) or SWI/SNF mediated nucleosome remodeling, although the phosphorylation of H2AX is not lost, it is delayed or reduced relative to that in wild-type cells [Shim et al., 2007]; [Park et al., 2006b]. This suggests that chromatin remodeling might either facilitate the exposure of this motif to the PIKKs or allow the efficient translocation of the PIKKs along the flanking chromatin to phosphorylate the H2AX molecules. Another mechanism by which PIKKs gain entry into the condensed chromatin is through phosphorylation of linker histone H1. It has been noted earlier that the C-terminal phosphorylable motif of H2AX is located inside of the 30-nm chromatin fiber,i.e., very close to where linker histone is bound, accordingly, one can speculate that the phosphorylation of linker histone, known to decondense chromatin [Contreras et al., 2003] ; [Hale et al., 2006], exposes the hidden C-terminal domains of H2AX.

The very large domains of γH2AX around double strand breaks (DSBs) are also thought to be binding platforms for cohesins. These are proteins responsible for tethering sister chromatids together [Unal et al., 2004], which has been shown to support chromosome stability and keep DNA ends in close proximity for the repair process [Unal et al., 2004].

More recently, histone H2AX was found to be phosphorylated at another C- terminal residue Y142, mediated by WSTF (WilliamsBeuren syndrome transcription factor, also known as BAZ1B) , a component of the WICH complex (WSTF-ISWI ATP-dependent chromatin-remodeling complex) [Xiao et al., 2009]. This phosphorylation is critical for the DDR as it may be needed initially at the break to adjust local chromatin structure for the later maintenance of H2AX-S139 phosphorylation. Subsequent dephosphorylation of Y142 could act to enhance MDC1 and ATM recruitment to extend and maintain H2AX phosphorylation after DNA damage. Finally, the absence of Y142 phosphorylation may lead to alteration in the kinetics of phosphorylation/dephosphorylation cycle of γH2AX, such that both the phosphorylation and dephosphorylation of γH2AX on S139 occur much more rapidly, and the foci assemble and disassemble in less than 4 h [Xiao et al., 2009]. At the same time, EYA1 (Eyes Absent) [Krishnan et al., 2009], was shown to be the protein phosphatase targeting H2AX-Y142ph in DNA damage signal dependent pathway. This modification, hence determines the recruitment of either pro-apoptotic or DNA repair complex to γH2AX in response to DNA damage, allowing it to function as an active determinant of repair/survival versus apoptotic responses to DNA damage [Cook et al., 2009].

Histone H1 Phosphorylation

The linker histone H1, which seals the nucleosome at the DNA entry and exit points, is important for the stability of higher order chromatin structure. Incorporation of linker histone into nucleosomes inhibits DNA end-joining by DNA ligase IV/XRCC4 (LX) in vitro. Since LX is important for non-homologous end joining (NHEJ), the linker histone may inhibit NHEJ-mediated DNA repair. However, this inhibition is compromised by the phosphorylation of histone H1 by DNA-PK, which reduces its affinity for DNA and decreases its capacity to inhibit end-joining, thereby facilitating NHEJ-mediated repair [Kysela et al., 2005].

- **Histone Ubiquitination**

  The ubiquitination of specific histone residues has emerged as an important post-translational modification that regulates DNA damage responses. Histone ubiquitination participates in chromatin remodeling and facilitates the accumulation of DNA damage sensing and repair proteins at the sites of DNA damage. As discussed earlier, after induction of γH2AX, MDC1 is recruited, which in turn recruits the UBC13/RNF8 complex to the sites of DSBs. As a consequence, γH2AX becomes ubiquitinated on chromatin surrounding the DSBs. The ubiquitination is initiated by UBC13/RNF8 (E2/E3) ubiquitin ligase complex, and the ubiquitin conjugates are amplified by another E3 ligase RNF168 [Doil et al., 2009]; [Stewart et al., 2009]. Its recruitment depends upon the interaction and docking of FHA

domains of RNF8 onto the phosphorylated TQXF motifs in MDC1. H2AX phosphorylation is a pre-requisite for this ubiquitination of H2AX as its ubiquitination is impaired in S139-mutants [Huen et al., 2007].

Ubiquitination leads to the formation of K63linked polyubiquitin chains on both γH2AX and H2A, leading to subsequent recruitment of Rap80/Abraxas/ Brca1/ Brcc36 complex [Wang and Elledge, 2007]. Not surprisingly, Ubc13deficient cells are compromised for homologous recombination (HR) pathway of DSB repair [Zhao et al., 2007]. It has been shown that it is the ubiquitination of H2A rather than γH2AX, which is a more critical event in subsequent signaling cascade in DDR as mutations of all lysines in H2AX neither abrogate the accumulation of ubiquitin conjugates at DSBs nor affect the recruitment of downstream factors [Doil et al., 2009]. In addition to H2A and γH2AX, histone H2B has also been found as a substrate of RNF8 [Wu et al., 2009]. It is not known whether H2B ubiquitination contributes to DDR in mammalian cells, although absence of histone H2B ubiquitination at K123 in yeast negatively affects the DDR induced by UV damage [Giannattasio et al., 2005]. H2A ubiquitination by UBC13/RNF8 ubiquitin ligase complex also occurs at the sites of UV-induced DNA damage [Marteijn et al., 2009]. Depletion of these enzymes causes UV hypersensitivity, without affecting nucleotide excision repair (NER) , suggesting that UBC13 and RNF8 are only involved in initial UV-induced DDR signaling and not the repair itself. The DDR signaling and recruitment events at UV-induced DNA damage sites are similar to recruitment events at DSBs as MDC1-dependent RNF8 recruitment is observed at the sites of UV damage, but requires ATR as well as NER-generated single stranded repair intermediates. Thus, there may be a conserved common pathway of H2A ubiquitination for both DSBs and UV lesions.

UV irradiation results in a CUL4-DDB-ROC1dependent increase in the levels of H3 and H4 ubiquitination [Wang et al., 2006]. The enzyme CUL4-DDB-ROC1 is also known to ubiquitinate DDB2 and XPC at UV damage sites [El-Mahdy et al., 2006]; [Sugasawa et al., 2005]. A small fraction of histone H3 and H4 (0.3% and 0.1%, respectively) is found ubiquitinated in vivo, and siRNA-mediated knockdown of CUL4A,B and DDB1 decreases the H3 and H4 ubiquitination levels. In addition, the dynamics of CUL4DDBROC1-mediated H3 and H4 ubiquitination is similar to that of XPC. Further evidence showed that CUL4-DDB-ROC1-mediated H3 and H4 ubiquitination results in histone eviction from nucleosomes by weakening DNA-histone interactions that in turn facilitates recruitment of NER repair proteins like XPC to the damaged foci. Hence, these studies point to a role of H3 and H4 ubiquitination in chromatin disassembly at the sites of UV lesions [Wang et al., 2006].

Another DNA damage related mono-ubiquitination has also been observed in histone H4 at histone fold residue K91 [Yan et al., 2009]. BBAP (B-lymphoma and BAL-associated protein) is a RING domain containing E3 ligase that selectively mono-ubiquitinates histone H4K91 and protects cells exposed to DNA-damaging agents. The presence of this modification in the HFD suggests its substantial im-

pact on the nucleosome conformation. As ubiquitin is a large molecule, it alters the histone folding and may expose some buried residues like H4K20me, which are known binding sites for checkpoint mediators like 53BP1 [Botuyan et al., 2006]. Disruption of BBAP-mediated mono-ubiquitination of histone H4K91 is associated with the loss of chromatin-associated H4K20 methylase, mono- and dimethyl H4K20, and a delay in the kinetics of 53BP1 foci formation at sites of DNA damage [Yan et al., 2009].

- **Histone Acetylation & Deacetylation**

  To achieve accurate DNA damage repair, the chromatin needs to be relaxed in order to facilitate access for the repair factors at the site of the DNA lesion. Histone modifiers and ATP-dependent chromatin remodelers are recruited at the break to modulate the chromatin architecture. The destabilization of the nucleosome is thought to require acetylation of histones through the action of HATs. Evidence suggests a link between histone acetylation and DNA repair by analyzing TIP60 histone acetyltransferase (HAT) and the proteins associated with it in the TIP60 complex [Ikura et al., 2000]. Since, a transient acetylation of H3 and H4 in their amino-terminal tails has been found to occur at DSBs in both mammalian and yeast cells [van Attikum and Gasser, 2005], a direct link showing that this acetylation is mediated by Tip60 complex (comprises of Tip60 HAT, SNf2-related p400, TRRAP , and other components) in mammals or its homologue NuA4, SWR1, and INO80 in yeast is provided by Murr and colleagues [Murr et al., 2006]. This large multiprotein complex acetylates several lysine residues (H3 K14, K23, H4 K5, K8, K12, and K16) of core histones in vitro [Kimura and Horikoshi, 1998]. It has been shown that mutations of N-terminal lysine residues in H4 confer sensitivity to ionizing radiation and defective cell cycle checkpoints [Tamburini and Tyler, 2005]; [Murr et al., 2006]. Tip60 is recruited to DSBs, where it binds to and acetylates H4 with cofactor TRRAP [Murr et al., 2006]. Depletion of TRRAP impairs recruitment of DNA repair protein 53BP1 and RAD51 but not MDC1, and affects DSB repair. In addition to H3 and H4 acetylation, the human Tip60 complex also acetylates γH2AX in response to ionizing radiation induced DSBs. Moreover, Tip60 forms a complex with UBC13 to promote the DNA-damage-dependent ubiquitination of acetylated γH2AX [Ikura et al., 2007]. These results suggest that the Tip60-mediated acetylation of γH2AX is most likely a prerequisite for its ubiquitination. Given that UBC13-dependent ubiquitylation of γH2AX is in turn necessary for eviction of this variant, we can assume that TIP60 and ubiquitination facilitate γH2AX dephosphorylation and/or turnover [Keogh et al., 2006]. Hence, recruitment of Tip60 and subsequent acetylation and ubiquitination of γH2AX represent an important cross-talk mechanism during DDR.

  hMOF (males absent on the first) is a Tip60 related histone acetyltransferase (HAT) that has been shown to interact with ATM kinase [Gupta et al., 2005]. This interaction is directly reflected in the activity of ATM and its downstream events in DDR. It has been observed that expression of a dominant negative hMOF mutant blocked IR-induced increases in histone H4K16 acetylation and resulted in

decreased ATM autophosphorylation, ATM kinase activity, and phosphorylation of downstream effectors of ATM and DNA repair, while increasing sensitivity to IR-induced cell killing. In addition, decreased hMOF activity was associated with loss of the cell-cycle checkpoint response to DSBs. Taken together, these results suggest that hMOF functions upstream of ATM and its modification of histone H4 may contribute to DSB sensing.

The histone acetylation is regulated by the concerted action of histone acetyltransferases (HATs) and histone deacetylases (HDACs), which work by adding and removing acetyl groups from lysine residues respectively. The acetylation of histone H3K56 in response to DSB repair in human cells has been a matter of debate [Das et al., 2009]; [Tjeertes et al., 2009]. Das et al. showed that DNA damaging agents, such as IR, UV, hydroxyurea, and MMS induce histone H3K56 acetylation in human cells, and histone acetyltransferase CBP/ p300 is responsible for such an acetylation. Acetylated H3K56 co-localizes with γH2AX immediately following IR, indicating that the acetylated H3K56 is enriched at the sites of DNA damage. Moreover, acetylated H3K56 is assembled into chromatin at the sites of DNA repair by CAF-1 and Asf1. By contrast, Tjeertes et al. observed that the H3K56 acetylation was rapidly and reversibly reduced in response to DNA damage. Histone acetyltransferase GCN5/KAT2A, rather than CBP/p300 acetylates H3K56 in vitro and in vivo. However, a recent work of SP Jackson's group highlighted the role of HDACs in DDR. It has been shown that human histone deacetylases HDAC1 and HDAC2 are the two participants that function in the DDR pathway [Miller et al., 2010]. It was observed that acetylation of histone residues H3K56 and H4K16 was regulated by HDAC1 and HDAC2 and that HDAC1 and HDAC2 were rapidly recruited to DNA-damage sites to promote hypoacetylation of these two residues. Furthermore, HDAC1- and 2-depleted cells were hypersensitive to DNA-damaging agents and showed sustained DNA-damage signaling, phenotypes that reflect defective DSB repair, in particular by non-homologous end-joining (NHEJ) pathway. Collectively, these results show that HDAC1 and HDAC2 function in the DNA-damage response by promoting DSB repair.

- **Histone Methylation**

  Histone H3K79 methylation and H4K20 methylation are important modifications involved in DDR. These modifications represent the constitutive marks, buried in the chromatin structure, not induced by DNA damage. It has been reported that methylation of H4K20 recruits mammalian repair factor 53BP1 (Crb2 in fission yeast/Rad9 in Saccharomyces cerevisiae) to the sites of DSB [Botuyan et al., 2006]. The same work had shown that the tandem tudor domains of 53BP1 recognize and bind specifically to H4K20me2, and this binding may work in conjunction with its BRCT domain-dependent recognition of γH2AX by 53BP1. This methylation is mediated by the histone methyltransferase PR-SET7 [Botuyan et al., 2006], and PR-SET7 appears to be crucial for the maintenance of genome integrity and S-phase progression [Houston et al., 2008];

II **Chromatin Dynamics After DNA Repair**

In order to fully comprehend the distinct roles of histone modifications during the course of DDR, it is imperative to distinguish the events occurring before and after the repair of DNA damage. First, the chromatin structure has to be modulated to allow access of repair factors to the damaged DNA, however, after the damage is removed, the DNA damage signal has to be switched off. It is becoming increasingly apparent that some histone modifications occur as "post-repair" events, as they are linked to the restoration of chromatin after DNA repair. The plethora of histone modifications occurring upon DNA damage induction suggest that the chromatin in the vicinity of damaged sites undergoes substantial processing and rearrangements. Thus, it is intuitive to reason that histone modifications are also critical steps in the execution of post-repair DDR.

- **Histone Phosphorylation & Dephosphorylation**

  H4S1 is phosphorylated by the complex casein kinase II (CK2) in response to DNA damage [Utley et al., 2005]; [Cheung et al., 2005]. CK2 associates with the HDAC complex Sin3Rpd3, which facilitates late events in the repair process. Interestingly, phosphorylation of H4S1 inhibits the ability of NuA4 to acetylate the N-terminal tail of H4, suggesting a mechanism by which the chromatin modifications are re-established. In other words, after the NuA4-dependent acetylation is removed by the HDAC Sin3Rpd3, the associated CK2 phosphorylates H4S1, thus preventing reacetylation by NuA4 and turning off DNA-damage signals.

  Another study showed that at late time points after irradiation, at late time points, the H2B N-terminal tail residue S14 is rapidly phosphorylated at DSB sites [Capetillo et al., 2004]. This induction of H2B-S14ph is independent of γH2AX, however, the irradiation induced foci (IRIF) of H2B-S14ph are dependent on γH2AX. It has also been shown that the kinetics of H2B-S14ph IRIF formation is significantly delayed with respect to γH2AX, which may indicate that H2B-S14ph foci are a late event in response to DSBs [Capetillo et al., 2004]. Furthermore, this modification may lead to greater condensation of chromatin around the break in a manner similar to nuclear condensation during apoptosis, and γH2AX may lead to accumulation of the H2B kinase. This phosphorylation is mediated by sterile 20 kinase (Mst1) and is well known for its role during apoptosis [Cheung et al., 2003], therefore it could play a direct role in regulating chromatin condensation at sites of DSBs.

  The elimination of γH2AX from chromatin surrounding the repaired DSB can be regarded as a signal to the cell that the repair process is achieved. There are two ways to eliminate γH2AX, either by eviction or dephosphorylation. It has been speculated that the presence of SWR1 at the break may allow γH2AX removal from chromatin surrounding the DSB [Altaf et al., 2007]. Alternatively, γH2AX signal can be turned down by dephosphorylation through the action of protein phosphatases PP2A and PP4C [Chowdhury et al., 2005]; [Chowdhury et al., 2008], and more recently added to this list, hPP6 and hWip1 [Douglas et al., 2010]; [Macurek et al., 2010]. It is speculated that γH2AX is evicted from chro-

matin proximal to DSB, whereas it is dephosphorylated in the chromatin farther away. This model is supported by two different dynamics of γH2AX accumulation and/or removal, depending on the position relative to the break [Savic et al., 2009].

- **Histone Ubiquitination**

  Human histone H2A has been found to be mono-ubiquitinated at K119 in the vicinity of UV-induced DNA lesions [Bergink et al., 2006]. The UV-induced ubiquitination requires functional NER, and coincides with the sequestration of PCNA, suggesting that the ubiquitination is linked to DNA repair. Since ubiquitin mutations that eliminate all lysine residues from ubiquitin do not affect the H2A ubiquitination, mono-ubiquitination, instead of poly-ubiquitination, is the most probable feature of this NER-dependent process. Further studies suggest that the observed H2A ubiquitination may be a post-repair process related to chromatin restoration [Zhu et al., 2009]. As determined by the formation of H2A-K119ub foci at local UV damage sites, dynamics of H2A-K119ub and XPC foci formation exhibits a clear temporal difference. For example, XPC foci are seen in almost all the damaged cells within 15 min and completely disappear at 4 h after local UV irradiation. On the other hand, the H2A-K119ub foci persist for 24 h post-irradiation, suggesting that the formation of H2A-K119ub foci is not directly associated with damage recognition. More importantly, the H3H4 chaperone CAF-1 is required for H2A-K119ub foci formation. Depletion of CAF-1 p60 by siRNA impairs recruitment of CAF-1 to DNA damage and eliminates H2A-K119ub foci. Moreover, the H2A ubiquitin ligase Ring2/Ring1b, a component of PRC2, is not detected at damage sites [Zhu et al., 2009]. Therefore, it is very likely that H2A ubiquitination occurs at newly assembled nucleosomes, or the H2A-K119ub is deposited on DNA repair region from an existing pool.

  It must be noted that the UV-induced H2A ubiquitination differs from γH2AX or H2A ubiquitination, observed following IR, in many ways. First, RNF8 and UBC13 are known to form K63-ubiquitin poly-ubiquitin chains [Plans et al., 2006]. Second, the ubiquitination by RNF8 is dependent on γH2AX and MDC1, where BRCT domain of MDC1 binds γH2AX. Deficiency in γH2AX abolished RNF8 recruitment and eliminated IR-induced ubiquitination after IR [Huen et al., 2007]. In contrast, γH2AX deficiency did not affect UV-induced H2A ubiquitination [Bergink et al., 2006]. Therefore, H2AX ubiquitination is integral part of checkpoint signaling in response to DSBs and UV-induced DNA damage, while NER-dependent H2A mono-ubiquitination is related to the chromatin restoration after repair of UV-induced lesions.

- **Histone Acetylation & Deacetylation**

  Recent studies have uncovered a role of histone H3K56 acetylation in chromatin assembly following DNA replication and DSB repair [Chen et al., 2008]; [Li et al., 2008]. It was found in yeast that H3K56 acetylation drives Asf1-dependent reassembly of chromatin after the completion of DSB repair [Chen et al., 2008]. Mutation of K56 to glutamine, mimicking permanent acetylation, partially by-

passes the requirement of Asf1 in resistance to DSB-generating agents. Nevertheless, DSB repair itself was operational in Asf1 mutants. On the other hand, lack of H3K56 acetylation in HAT Rtt109 mutants leads to persistence of DNA damage signals in the form of prolonged phosphorylation of checkpoint protein Rad53. These observations indicated that restoration of chromatin, driven by acetylated H3K56, is a "turn off" signal and points towards the completion of DSB repair. Similarly, CAF-1 is involved in chromatin restoration, driven by acetylated H3K56 [Li et al., 2008]. H3K56 acetylation promotes the association of histone H3 with CAF-1 and Rtt109, and facilitates CAF-1-dependent nucleosome assembly, thus, coordinating the function of H3H4 chaperones in nucleosome assembly.

The action of HDACs is regarded as a late event in the DNA-repair process and functions to reduce the amount of histone acetylation once repair has been completed [Tamburini and Tyler, 2005]. Therefore, HDACs may have a role in the restoration of the original acetylation state, which eventually leads to the recovery of the higher-order chromatin structure, once the DNA repair is finished. Likewise, the reduction of chromatin relaxation upon deacetylation contributes to the termination of the associated checkpoint activity [Murr et al., 2006]. A link between mammalian HDAC3 and the repair of DSBs has recently been discovered [Bhaskara et al., 2008]. Although its precise role in repair is unknown, given that HDAC3 is an essential component of the N-CoR/SMRT repression complex, it can be speculated that it could be involved in reinstating the acetylation status, which would give rise to a repressive chromatin structure.

- **Chromatin Restoration**

  During cell proliferation, histone modifications are tightly associated with DNA replication-coupled nucleosome assembly. As DNA replicates, the parental histones are randomly transferred onto either the leading or the lagging strand [Jackson, 1988]. For chromatin restoration after DNA repair it is logical to assume that histone chaperones, such as chromatin assembly factor-1 (CAF-1), Asf1, and FACT are involved and recruited to the site of DNA damage and hence represent an integral part of DDR, reviewed in [Ransom et al., 2010]. When nucleosomes are disrupted during the repair process, H2A-H2B dimers and one H3-H4 tetramer are separated and later is randomly transfered to one of the nascent DNA strands. These histones then serve as a template for newly synthesized histones to be deposited with the correct histone variant and modifications, which represent the original state of the chromatin. The H3-H4 tetramer deposition involves histone chaperones, and among them, chromatin-assembly factor 1 (CAF-1) plays a pivotal role in chromatin assembly during DNA replication and DNA repair [Green and Almouzni, 2003]; [Polo et al., 2006]. CAF-1 is an evolutionary conserved protein complex, consisting of various subunits (p150, p60, and p48), with a unique ability to preferentially deposit newly synthesized H3 and H4 onto replicating DNA [Shibahara and Stillman, 1999]. It has been shown that histone H3.1/H4 and H3.3/H4 tetramers make complex with two different predeposition complexes: CAF-1 and HIRA, respectively [Tagami et al., 2004], where CAF-

1 is necessary to mediate replication dependent nucleosome assembly. Through its interaction with PCNA, a DNA polymerase processivity factor, CAF-1 is targeted to the sites of DNA synthesis and deposit H3.1-H4 dimers onto nascent DNA strands [Benson et al., 2006]. CAF-1 works alongside with another H3-H4 chaperone Asf1, which synergistically cooperates with CAF-1 in DNA repair and replication-coupled chromatin assembly [Mello et al., 2002]. It has also been shown that CAF-1 is required for Ring1b-dependent mono-ubiquitination of H2A-K119, a mark involved in the chromatin restoration after the UV-induced DNA damage [Zhu et al., 2009].

During chromatin reassembly after DNA repair, it is important to preserve or recover the epigenetic information specific to a particular chromatin domain. Cells have evolved mechanisms to inherit this information by reproducing original histone modification pattern and/or replacement of "correct" histone variant. PCNA molecules, acting as a hub, recruits a number of chromatin-modulating enzymes, for example, DNA methyltransferase DNMT1 and CAF-1, to re-establish epigenetic memory based on DNA methylation and histone modifications. Both DNMT1 and CAF-1 also have the ability to recruit histone modifying enzymes for PTMs. For example, DNMT1 interacts with H3K9 methyltransferase G9a [Esteve et al., 2006] and EZH2 component of PRC2 [Vire et al., 2006], which catalyzes trimethylation of H3K27. PRC2 binds to H3K27me3 and colocalizes with this modification in G1 phase, and with the sites of ongoing DNA replication [Hansen et al., 2008]. The ability of PRC2 to recognize a previously established mark triggers its renewal [Margueron et al., 2009], in accord with the general "epigenetic templating" scheme of reproduction of epigenetic information (see Section1.2. DNMT1 also forms a complex with G9a, which is required for recruitment of G9a to replication sites and for maintenance of H3K9 methylation at epigeneticaly silenced rDNA repeat [Esteve et al., 2006]. CAF-1 also plays a role in maintaining the repressed state of silenced chromatin [Houlard et al., 2006]. Consistent with this, it has also been shown that depletion of CAF-1 p150, In embryonic stem cells, results in the alteration of epigenetic histone methylation marks at the level of pericentric heterochromatin. Additionally, CAF-1 is found to link replication-coupled chromatin assembly to DNA and histone H3K9 methylation [Reese et al., 2003]; [Sarraf and Stancheva, 2004]. CAF-1 forms a replication-dependent complex with H3K9 methyltransferase SETDB1 and MBD1, a protein that binds to methylated CpG dinucleotide. This complex is required for maintenance of H3K9 methylation and stable silencing of certain genes in proliferating cells. During DNA replication, MBD1 recruits SETDB1 to CAF-1 p150 and mediates H3K9 methylation in replication-coupled chromatin assembly. A similar mechanism also operates for H3K9 methylation of newly synthesized H3 deposited by CAF-1 in pericentric heterochromatin [Loyola et al., 2009]. In this case, SETDB1 associates with HP1α-CAF-1 complex and monomethylates H3K9 on non-nucleosomal histone H3, providing H3K9me1 for further trimethylation in pericentric regions. While both HP1 and CAF-1 are reported to localize to the sites of DNA damage [Green

and Almouzni, 2003]; [Luijsterburg et al., 2009], it is not known whether HP1α-CAF-1-SETDB1 complex works in DNA repair-dependent nucleosome assembly.

All this data suggest that histone chaperones along with chromatin modifying enzymes play an essential role in restoring the state of chromatin back to its original pre-DNA damage form. However, further studies are required to reveal the details of molecular interactions among various chromatin modifiers, remodelers, and histone chaperones in repairing DNA and faithfully reinstating the biological information associated with chromatin.

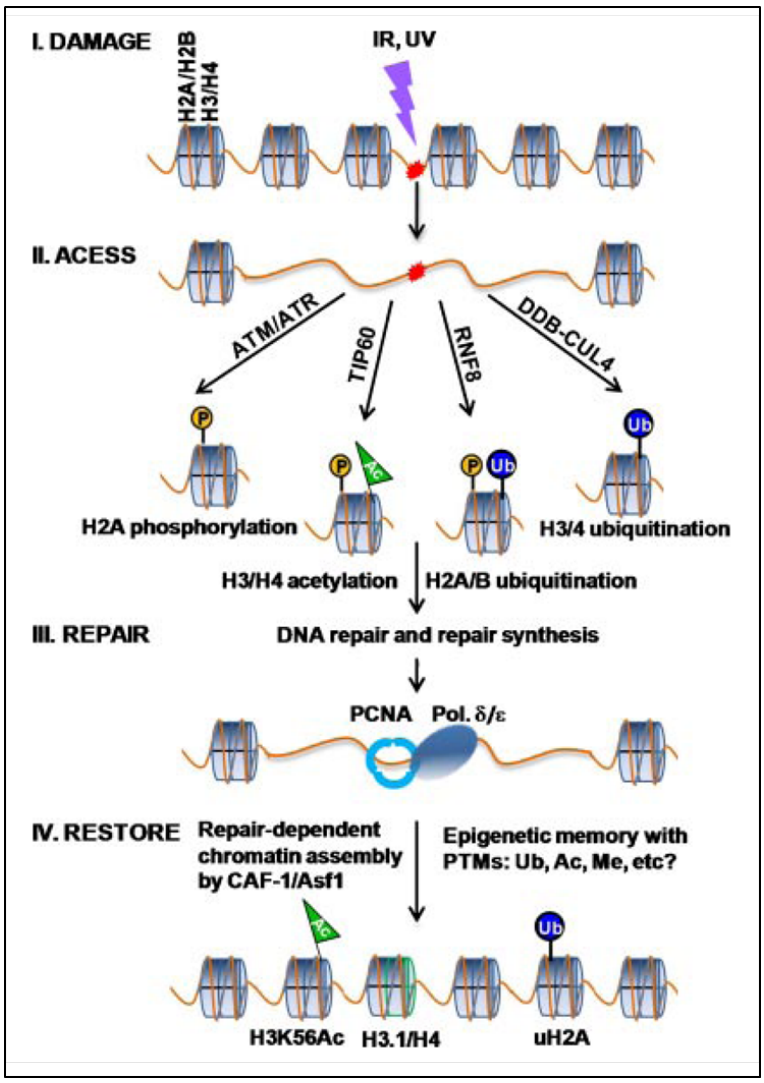


Figure 1.14: Histone modifications in DNA damage response. Adapted from [Zhu and Wani, 2010].

In recent years, a remarkable progress has been made in revealing an orchestra of histone PTMs and their effector proteins in DDR. The ultimate goal for the cells after DNA damage and repair is the restoration of chromatin to its original state. It becomes evident that various histone PTMs organize this response for the prompt execution of checkpoint signaling, DNA

repair and reinstatement of chromatin architecture. Timing and features of modifications appear to be critical factors in determining the physiological effect of a specific modification. For example, H3 acetylation occurs both before and after DSB repair. The former, acetylation at H3K14 and H3K23, leads to γH2AX ubiquitination, checkpoint activation, and DNA repair [Ikura et al., 2007]. On the other hand, the latter, H3 acetylation at K56, leads to restoration of chromatin and termination of checkpoint activation [Chen et al., 2008]. Also, the same concept appears to hold true for histone ubiquitination. The γH2AX and H2A ubiquitination lead to checkpoint signaling, whereas the H2A ubiquitination after NER is associated with chromatin restoration. Similarly, literature indicates that specific combinations of chromatin modifications, read by distinct effector proteins along with ATP-dependent chromatin remodeling allow binding of DNA repair factors and efficient healing of the DNA lesion. For instance, it has been shown that mammalian SWI/SNF remodeler is critical for efficient induction of γH2AX [Park et al., 2006b], because inhibition of its catalytic core subunits, BRG1 and Brm, compromises phosphorylation and subsequent foci formation of H2AX. More recently, Lee and colleagues have shown that SWI/SNF binds to γH2AX containing nucleosomes via an interaction between its bromodomain-containing BRG1 subunit with GCN5-dependent acetylated H3, and this binding is important for DSB repair [Lee et al., 2010].

These scenarios or more precisely different combinations of modifications along with their docking proteins raise an important question whether these modifications are an extension of histone code in the context of DDR? Although there is a degree of overlap in chromatin modifications in DDR to various different DNA damaging agents, evidence shows certain level of specificity of particular modifications occurring in response to particular type of DDR and repair, e.g. poly-ubiquitination of γH2AX in response to DSBs and UV-induced DDR in contrast to NER-dependent H2A mono-ubiquitination. It might be possible that the initial response of the cells to various DNA insults is conserved, however, it is only later the cell decides to follow a certain response trajectory depending upon the DNA lesion and the choice of DNA repair pathway it is going to employ. We can intuitively observe this kind of phenomena by going back to Waddington's epigenetic landscape [Figure 1.1], where dynamic histone modifications and their binding partners, on the crossroads of various possible trajectories, dictate and influence the cellular decision making for DDR and repair. This will again lead us towards our question if a "histone code" really exist for DDR. However, the deciphering of this code requires better understanding of detailed molecular mechanisms by which each modification functions in a specific DDR pathway.

1.6 Methodologies to Study Histone Post-Translational Modifications

Biological research in the post-genomic era has been accelerated by the advent of new methodologies to explore the proteomes, which not only complement the existing information available through the analysis of genomes but also provide opportunities to look beyond the genomes. In this proteomics era, the fundamental challenge of molecular and cell biology is

to unravel the mechanisms and pathways that govern the proper functioning of a cell. Since, all cellular processes inevitably involve proteins, their abundance, post-translational modification (PTM) state, and localization has come into the limelight of proteomics research. The characterization of protein post-translational modifications, in particular, has contributed extensively to our understanding of many key aspects of cellular processes both during normal functioning and diseased states. Different modifications influence each other in an interdependent manner and lead to an intricate network of cellular signaling that by affecting proteinprotein interactions, enzymatic activities and their sub-cellular localizations, regulate the subtle balance of various cellular functions. A wide spectrum of these PTMs exist on proteins ranging from relatively small groups such as methyl, acetyl and phosphoryl groups to the attachment of larger sugar moieties or the addition of small peptides like Ubiquitin and SUMO to a protein of interest [Walsh, 2005].

Histones, small basic proteins that fundamentally function to pack eukaryotic genomes, are among the best characterized proteins in terms of their post-translational modifications, their combinatorial patterns and their pivotal role in various DNA-based cellular processes, such as transcription, replication, DNA damage response etc. ([Section 1.4.4]). As discussed above, histone PTMs function either alone or in a combinatorial fashion to store and translate epigenetic information in the form of a histone code, and thus have an enormous potential for regulating DNA-based processes by selectively and appropriately using this information. One of the biggest challenges for proteomics has been to analyze these codes in a quantitative and reliable way in order to get insight into the complex signaling mechanisms regulated by these codes. Here we review both the historical and recent progress towards routine and rapid histone PTM analysis, mainly focusing on mass spectrometry-based methods for the study of histone PTMs. However, a brief description of non-MS-based methods along with their disadvantages is also presented in the beginning.

1.6.1 Classical Methods

The chromatin in eukaryotic cells is perceived as a dynamic structure, where the interplay of various histone PTMs determines the output of all the DNA-templated cellular processes. Histone PTMs are, therefore, the fundamental regulators of all DNA-based cellular processes and, consequently, their characterization has been the focus of proteomics methodologies during the last decade. Traditionally, targeted approaches were used to measure the abundance of various PTMs of histones, however, due to their inability to measure these changes in a combinatorial way, they only provide a glimpse of a complex system governing the essential DNA-based cellular processes. Classical methods to analyze histone PTMs were mainly based on microsequencing and immunoassays.

- **Microsequencing**

  Classically, histone modifications have been detected using specialized gel systems with or without the incorporation of radioactive precursor molecules, which is followed by a complete protein hydrolysis and the subsequent analysis of resulting amino

acids [Waterborg and Matthews, 1983]; [Waterborg, 1993]. As majority of these modifications were found on N-termini of histones, it was possible to map these modifications using Edman degradation, which is a method of amino acid sequencing [Sobel et al., 1994]. However, although these procedures can be automated, they require large amount of purified material, in addition to multiple time consuming purification steps. Also, using Edman sequencing, only 25-30 residues can be routinely characterized. Due to these reasons, they were not feasible for comprehensive analysis of histone PTMs from small cell numbers and the combinatorial histone codes.

- **Immunoassays**

  The antibody based immunoassays have been developed in late 1980's when researchers, in the quest of new and more sensitive methods, that allowed for detailed study of histone PTMs, developed sets of specific antibodies that could recognize specific histone PTMs with distinction of modification sites [Hebbes et al., 1989]; [Turner et al., 1992]; [Turner, 2002]. These antibodies, on one hand, could be used in western blots to study changes in specific modifications, while on the other hand were also used in immunofluorescence experiments, which allowed for the localization of specifically modified histones within particular regions of genome [Turner et al., 1992]; [Jeppesen and Turner, 1993]. Moreover, chromatin immunoprecipitation using antibodies against specific histone PTMs and subsequent analysis of DNA sequences could also provide information regarding its location and association with particular DNA sequences.

  Although the sensitivity of such methods was quite high they suffer from certain drawbacks. The similarity between different modification sites, on the same or an adjacent histone molecule, leads to significant cross-reactivity of antibodies, rendering them less specific [Cheung, 2004]. Secondly, the combinatorial nature of histone PTMs may leads to strong interference in epitope binding. In this case, a neighboring modification may occlude the epitope that the antibody recognizes despite the presence of this PTM mark. This shortcoming can be resolved by making antibodies specific for combinations of modifications, however, considering the huge number of combinations even in a short stretch of amino acids, it is virtually impossible to generate antibodies for each set of modifications. Thirdly, different antibodies have varying affinities toward their epitopes and therefore the signal obtained by one antibody cannot be compared on absolute scale with that of another antibody. Finally, it is impossible to detect new modifications using immunoassays as they require a priori knowledge of the site of PTM under question.

It is clear from the above discussion that these classical methods had quite a few serious pitfalls that needed to be addressed in order to comprehensively analyze histone PTMs. Scientists, in the search of newer and more direct methods to study combinatorial histone PTM networks, found their heaven in mass spectrometry (MS) based proteomics, which has enormous potential to study histone code at a high throughput level.

1.6.2 Mass Spectrometry

The complete structural and functional characterization of all proteins in the cell has always been the goal of proteomics. Proteins can be described by many features including but not limited to expression, localization, interaction, domain structure, post-translational modification, and activity. If the proteome of an organism is considered together with all these aspects of the proteome, this represents a considerably complex body of information in comparison to the genome of an organism, which generally does not vary from cell to cell. Traditionally, the proteomics techniques have focused on understanding the structure and function of individual proteins, which, no doubt, improves the general understanding of the working of organisms, however, do not allow systems level understanding of complete proteome.

For the comprehensive and large scale high throughput proteome analysis, mass spectrometry (MS) has gained much popularity in the last decade owing to its ability to handle the complexities associated with the proteome. Other techniques such as two-dimensional gel electrophoresis (2DE), two-hybrid analysis, and protein microarrays fail to achieve the depth of informative proteome analysis seen with MS. Mass spectrometry has been widely used to analyze biological samples and has evolved into an indispensable tool for proteomics research.

Mass spectrometry is an analytical technique that measures the mass to charge ratio of charged particles. By definition, a mass spectrometer consists of an ion source, a mass analyzer that measures the mass-to-charge ratio (m/z) of the ionized analytes, and a detector that registers the number of ions at each m/z value. Electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) are the two techniques most commonly used to volatize and ionize the proteins or peptides for mass spectrometric analysis [Karas and Hillenkamp, 1988]; [Fenn et al., 1989]. Mass spectrometry based proteomics has matured into a useful tool for the global analysis of protein composition, post-translational modifications, and dynamics of proteins in various physiological and pathological cellular states [Walther and Mann, 2010]. Mass spectrometry based proteomics has seen enormous growth during the last decade owing to the great many advances in the development and innovation in the range of mass spectrometers and analysis techniques. In the field of chromatin biology, mass spectrometry has emerged as an important platform for characterizing the specific locations of various types of histone post-translational modifications. Moreover, the quantitative analysis of combinatorial histone codes has been made possible due to the recent advances in high resolution mass spectrometry.

[Figure 1.15] shows the work flow of a typical MS-based histone PTM analysis.

Chromatography

MS-based proteomics is tightly linked and highly dependent on separation technologies that simplify incredibly complex biological samples before subjecting them to MS analysis. Because proteins are identified by the mass-to-charge ratios of their peptides and fragments, sufficient separation is required for unambiguous identifications. Therefore, both accuracy

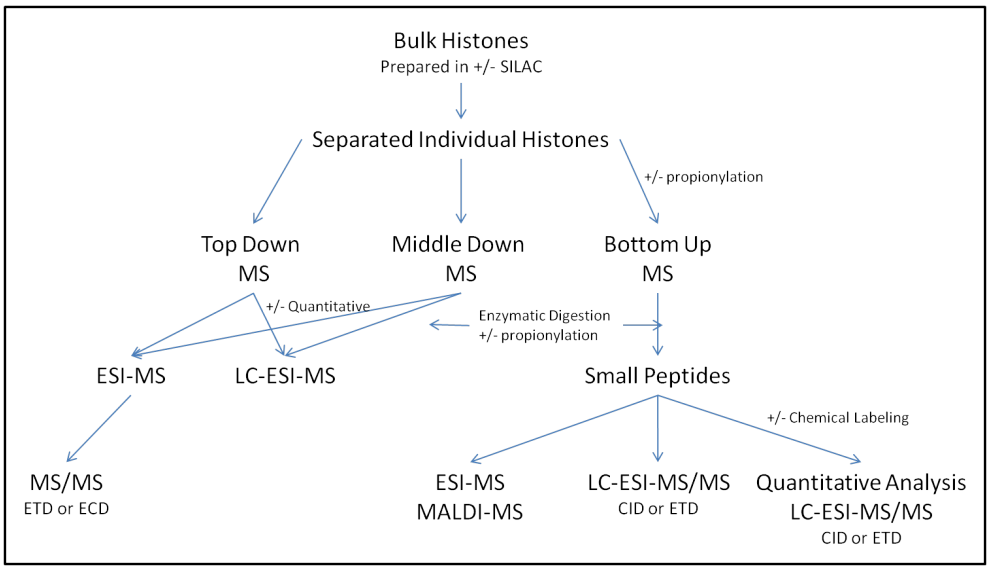


Figure 1.15: Work flow of a typical MS-based histone PTM analysis.

and sensitivity of a mass spectrometric experiment rely on the efficient separation. Selection of appropriate separation methods is often the first step in designing the proteomic application. Two major approaches to separation widely used in proteomics are gel based and gel free.

Gel based methods generally rely on 2D-PAGE, where proteins are separated in two dimensions, first on the basis of charge and second according to mass. With the advent of latest MS instruments coupled with online high performance liquid chromatography (HPLC), the 2D-PAGE based separation of proteins is largely out of practice in large scale proteomics studies. Also, to separate proteins or peptides, HPLC is advantageous because liquid handling is well-suited to automation, an important consideration in high throughput analysis. The following types of HPLC chromatographic techniques are most commonly used in MS-based proteomics: reverse phase (RP), ion exchange (IEX) , hydrophilic interaction chromatography (HILIC), affinity chromatography, such as IMAC, and hybrid techniques.

Among the hybrid techniques, gel-enhanced liquid chromatography mass spectrometry (GeLCMS) is worth mentioning as it combines pre-separation of proteins by gel electrophoresis followed by analysis of the peptides resulting from an in-gel digest step by online high performance liquid chromatography-tandem mass spectrometry (HPLC-MS/MS) [Shevchenko et al., 1996]; [Lundby and Olsen, 2011]. It is beneficial to separate complex protein mixtures into several fractions to reduce sample complexity and increase the dynamic range of peptide identification. Splitting the proteins to be analyzed in several gel slices, which are individually processed, increases the chances of identification of proteins and their PTMs. A considerable disadvantage of the in-gel protein digestion procedure compared to that of in-solution digest is the risk of contaminating the samples with, for instance, skin and wool keratins, when handling or excising protein bands from the gel.

Separation of intact histones, histone variants and peptides derived from the enzymatic digest of histone is an important task in order to get a clean, high purity sample before MS analysis. Gel electrophoresis methods have been used to separate histones for over half a century [NEELIN and CONNELL, 1959]; [NEELIN and NEELIN, 1960]. These electrophoretic techniques can also be used in a 2D gel separation for even greater resolution [Bonner et al., 1980]. Although these gel based approaches have proved to be excellent ways to separate complex samples, such methods are difficult to incorporate into high throughput methods as they are both labor intensive and not directly compatible with mass spectrometry.

Capillary electrophoresis (HPCE) [Gurley et al., 1991]; [Lindner et al., 1992] is a powerful separation technique with many distinct advantages including low sample requirement, better resolution of histones, ease of separation and high efficiency. One of its major limitation is related to the fact that while analyzing histones, their highly basic tails interact strongly with the surface of the fused silica tubing used in HPCE. A lot of effort has been done to minimize this problem. One such example is the work of Aguilar et al. [Aguilar et al., 2001], in which they used hydropropylmethylcellulose (HPMC)-coated capillaries and an MS-compatible solvent system to perform capillary electrophoresis-electrospray ionization-mass spectrometry (CE-ESI-MS) on calf thymus histones. Such an approach definitely continues to have potential in histone PTM analysis however, it received scant attention. Another extension of HPCE is shown by the work of Mizzen [Mizzen and McLachlan, 2000], where they introduced some degree of cation exchange behavior to the capillary surface. Overall, HPCE provides excellent histone separations and has the capacity to be used on-line with mass spectrometry, making it potentially high-throughput; however, recent work in histone PTMs has primarily focused on liquid chromatography approaches.

In the vast majority of the chromatin studies, where histone PTMs and variant histones are analyzed by mass spectrometry, reverse-phase HPLC was used to chromatographically separate the digested histone peptides. While satisfactory for the separation of histone families and some variants, other chromatography techniques can be adopted for reducing spectrum complexity (i.e., enhance chromatographic resolution) or for enriching lower level modification states. For instance, cation exchange HPLC has been demonstrated as an effective tool which can separate hyperacetylated histone forms primary by their degree of acetylation. Zhang et al. used this technique to analyze an Arg-C digestion of histone H3 from HeLa cells and observed a progressive link between H3 acetylation and H3K4 and H3K79 methylation [Zhang and Freitas, 2004].

An important problem in histone PTM analysis is examining the roles of phosphorylation, which are known to be involved in the transition from interphase to mitotic (i.e., condensed) chromatin, but whose exact mechanistic roles in this process have not yet been determined. Immobilized metal affinity chromatography (IMAC) has emerged as a very useful separation protocol for enriching phosphopeptides. For instance, IMAC was used in a MS analysis of the five major isoforms of H1 (i.e., H1.1, H1.2, H1.3, H1.4, and H1.5) and H1.X from asynchronous HeLa cells [Garcia et al., 2004]. In this study, 19 phosphorylation sites were identified, where 14 fit the CDK consensus motif, and the H1.4 peptide exhibited serine phosphorylation adjacent to lysine methylation, suggesting a possible "binary switch" hypothesis for

linker histones which had been previously suggested for core histones [Fischle et al., 2003]. A similar IMAC-MS/MS experiment was conducted for protease digests of histone H3 from mitotically arrested HeLa cells [Garcia et al., 2005]. The study identified other potential binary switches, including H3T3ph/K4me1, H3K9me1-3/S10ph, and H3K27me1-3/S28ph. It was notably observed that phosphopeptides containing lysine methylation were found in much higher abundances than phosphopeptides containing lysine acetylation.

The resolution capabilities of cation-exchange chromatography and IMAC were combined for the MS interrogation of H1 in Tetrahymena Thermophila cells [Garcia et al., 2006]. Five phorphorylation sites (T34, S42, S44, T46, T53) in H1 were confirmed and two novel sites of phosporylation were identified (S4 and S5), in addition to the discovery of two novel minor sites of acetylation (K77 or K78 and K154). The cation-exchange chromatography also revealed a hierarchical pattern of phosphorylation, where T34, T46, and T53 are phosphorylated first, followed by S4 and S5 next, and then finally S42 and S44. Lastly, it was shown by Freitas and co-workers that offline hydroxyapatite (HA) chromatography using salt step or salt gradient elution can improve the dynamic range of subsequent online reversed-phased liquid chromatography for histone analysis [Su et al., 2009]. The hydoxyapatite separates the histones based on their relative binding affinity to DNA and resulted in the detection of lower abundance histone variants and their PTMs when compared to the traditional acid extraction of histones.

The compatibility of RP-HPLC with ESI-MS makes it an ideal choice for analysis of histones. However, the similarity of highly basic histone variants limits the effectiveness of this technique. Hydrophilic interaction LC (HILIC) is a powerful technique that has proved to be well suited for histone isoform separation. Linder and coworkers [Lindner et al., 1996] optimized this technique to attain excellent resolution of core and linker histones. Later Sard and colleagues used HILIC and achieved separation and identification of K20 methylation states of acetylated H4 isoforms [Sarg et al., 2002]; [Sarg et al., 2004]. Their results revealed that H4 hyperacetylation blocked tri-methylation of H4K20. Furthermore, H4K20 methylation in mammals was related to aging. HILIC separations, often at the peptide level, in conjunction with mass spectrometry approaches have been essential to many methods that thoroughly and effectively characterize combinatorial histone codes [Sarg et al., 2006]; [Pesavento et al., 2006]; [Taverna et al., 2007]; [Garcia et al., 2007]. HILIC methods, however, rely on a non-volatile mobile phase additive, such as a chaotropic agent to improve resolution, and/or salt gradients for elution of the histone forms. This makes the most effective separation not directly compatible with mass spectrometry, which is the most effective means to sequence and quantitate the separated combinatorial histone codes. Recently, Garcia's group has addressed this long-standing problem in histone analysis by creating a mass spectrometry-compatible weak cation exchange-HILIC method (WCX-HILIC) [Young et al., 2009]. They removed all non-volatile components from the mobile phase and used a pH gradient to neutralize the cation exchange character of the stationary phase. This was done in conjunction with a decreasing organic solvent gradient for elution based on purely HILIC mechanisms. In this way, a highly effective WCX-HILIC-based separation is achieved that is also directly compatible with mass spectrometry.

Bottom-Up MS Analysis

Currently, mass spectrometric analysis of proteins relies upon two complementary techniques, i.e., bottom up and top down approaches, which represent two different trends in MS-based proteomics [Chait, 2006]. The bottom-up approach it still considered as the stronghold of proteomics research, although much more recently the focus is drawn toward the top down analysis of proteins. Each of these two approaches has its own advantages and drawbacks and relative fields of application. The bottom-up approach is mainly used for identifying proteins and determining details of their sequence and post-translational modifications. Classically, in bottom up approach, proteins are digested with proteases (usually trypsin) and the corresponding peptides are analyzed by electrospray ionization (ESI) or matrix assisted laser desorption/ionization (MALDI). In this approach, the ESI or MALDI based mass spectrometric analysis take place in two steps. First, the masses of the intact tryptic peptides are determined; next, these peptide are fragmented to produce information on their sequence and modifications.

Bottom up proteomics has many undeniable advantages: (i) It is relatively easy to perform, because trypsin, due to its versatility, reproducibility and reliability, works well in a variety of different conditions (denaturing agents, organic solvents, relatively extreme pH values). (ii) Small peptides (500 2000 Da) can be easily analyzed with every MALDI and ESI instruments, with high sensitivity (reaching the femtomoles range in ordinary instrumental setups). (iii) Good MS/MS spectra of one or two peptides can be sufficient to confidently assign a protein hit. (iv) Extremely large and complex proteomes, such as serum or cell lysates, can be confidently analyzed with the multidimensional protein identification technology or "shotgun" approach [Delahunty and Yates, 2007]; [Kislinger et al., 2005] and a list of hundreds of trustworthy protein hits can be obtained in a single automated experiment. (v) Several protocols for absolute and relative quantitative proteomics [Barnidge et al., 2004]; [Ramus et al., 2006]; [Mann, 2006]; [Perkel, 2009] have been implemented over the last 10 years, yielding reliable data of protein expression levels often ranging over several orders of magnitude in very complex matrices.

However, the bottom up approach is suboptimal for determining protein PTMs and alternative splice variants or alternately modified variants of proteins, as usually only a small fraction of tryptic peptides are detected and only a fraction of these yield useful fragmentation ladders. The peptides, after tryptic digestion, represent pieces of a jigsaw puzzle where many of the pieces are missing. Therefore, we can identify the protein on the basis of a limited number of peptides, but we have no clue to the effective complete sequence of the intact protein and a comprehensive full set of its PTMs. Even if we manage to detect all the pieces of the puzzle, the picture would still be incomplete as we cannot reconstitute the intact protein with its full set of modification pattern. Given the combinatorial nature of various PTMs and their correlations, there exist many different forms of the same protein containing different sets of modifications. However, using the bottom up approach, by identification of individual peptides, we certainly lose information regarding the combinatorial pattern of PTMs.

Regardless of these limitations, bottom up approaches have been extensively used for the identification and characterization of histone PTMs in various cell lines and cellular states. It has been shown more than a decade ago that mass spectrometry, employing bottom up approaches, can be used for discovering novel histone PTMs [Strahl et al., 2001]; [Ng et al., 2002]; [van Leeuwen et al., 2002]. Burlingame and co-workers were the first to publish comprehensive mass spectrometric investigations of histone H3 and H4 in chicken erythrocytes [Zhang et al., 2002a] and found abundant methylations and acetylations at known residues. They have also found a "zip" model of H4 acetylation at N-terminal residues K5, K8, K12 and K16 in HeLa cells, where acetylation progressed from K16 down to K5 and deacetylation followed in opposite direction [Zhang et al., 2002b]. Recent mass spectrometry based studies have also identified or confirmed a variety of modifications such as novel methylation and acetylation sites [Ye et al., 2005]; [Morris et al., 2007]; [Garcia et al., 2007a]; [Drogaris et al., 2008]; [Das et al., 2009], as well as new or unique modifications including biotinylation, propionylation, butyrylation and formylation on many histone proteins [Camporeale et al., 2004]; [Chen et al., 2007]; [Zhang et al., 2009]; [Jiang et al., 2007b], however, the significance of many of these PTMs during nuclear processes still remain unclear.

As have been already mentioned, one major limitation of bottom-up mass spectrometry, which involves extensive digestion of the histones, is that it can provide very limited combinatorial information. However, some early examples of the analysis of combinatorial pattern of the histone code were in fact shown by such methods. For example, Fischle et al. blocked lysine residues of histone H3 by propionylation and trypsin digested to achieve reproducible ArgC-like peptides. They showed that HP1, which normally binds the silencing mark H3K9me3, is released from chromatin during mitosis by phosphorylation of the adjacent S10 residue [Fischle et al., 2005]. Later Garcia et al. used ion metal affinity chromatography (IMAC) and performed an analysis of phosphorylations on histone H1 from Tetrahymena thermophila. They found as many as three phosphorylations on a single small peptide [Garcia et al., 2006]. Although these examples showed some utility of bottom-up MS in the histone code combinatorial analysis, they fail to document the full complement of combinatorial information embedded on histone proteins, which can sufficiently be preserved by top-down and middle-down approaches (discussed later).

Chemical Derivatization

A fundamental step in the Bottom Up MS approach is the proteolytic digestion of the proteins into smaller peptide sequences, which are more amenable to the standard LC-MS protocol. Typically, the enzyme of choice is trypsin since it produces peptides that have a basic C-terminal arginine or lysine residue, which usually results in more complete ion series coverage during collision induced dissociation (CID) since a positive charge is always retained with the C-terminal moiety. However, the N-terminal tails of histone proteins are extremely basic and contain an unusually large number of lysines and arginines (e.g., the 1-50 N-terminal tail of H3 contains 15 and the 1-23 N-terminal tail of H4 contains 9). A standard tryptic digest of a purified mixture of H3 or H4 modified histone forms generates a large number of extremely short peptide sequences, resulting in several analytical problems due to poor fragmentation, poor retention on RP-HPLC columns, and lack of reproducibility.

One should note that the use of the Arg-C protease does result in cleavage only after arginine residues, but is known to be less efficient and less specific than trypsin. Alternatively, proteolysis with enzymes that cleave after acidic residues (e.g., Glu-C) generates larger multiply charged peptides whose MS/MS spectra are difficult if not impossible to interpret. Therefore, several groups have utilized chemical derivatization of histones peptides to facilitate downstream MS experiments. The derivatization strategies reported in the literature exploit the unique amino acid composition of the histone proteins and the PTMs known to be present predominantly on the lysine residues.

One of the earliest chemical derivatization approaches used deuterated acetic anhydride to analyze H3 changes in Drosophila embryos [Bonaldi et al., 2004] and the 4-17 amino acids of H4 in mutant cells [Smith et al., 2002]; [Smith et al., 2003]. Acetic anhydride is known to readily acetylate alcohols and amines, and thus reacts with the unmodified lysine residues and the N-terminus of histones. This subsequently blocks tryptic digestion on the carboxyl side of lysine residues, leaving arginine as the sole site of enzymatic cleavage. A drawback to this approach is that the introduction of a deuterated acetyl group on the lysine residues is very close in mass with endogenous lysine acetylation or trimethylation. To get around this issue, alternative derivatization methods were developed.

In particular, several research groups have pursued the use of propionic anhydride for the chemical derivatization of endogenously unmodified or monomethylated lysines. This reaction is very similar to that for acetic anhydride, with the exception that an additional CH2 group is added to the side-chain modification, thereby inducing an observed shift of 56 Daltons per derivatized lysine. One of the earliest applications of this approach enabled the investigation of the quantitative effects of H3 methylation after knocking down specific methyltransferases [Peters et al., 2003]. Since then, it has guided the discovery of new histone modifications, enabled the investigation of PTM changes under distinct cellular conditions and elucidated modification differences between the lower and higher eukaryote species [Johnson et al., 2004]; [Syka et al., 2004b]; [Garcia et al., 2007a]; [Drogaris et al., 2008].

Propionic anhydride derivatization serves two practical purposes: 1. By blocking all lysine residues by the addition of a chemical group (propionyl) or by endogenous modifications, trypsin digestion results in proteolysis only C-terminal to arginine residues. This reaction mimicks an Arg-C digestion, but with the efficiency and reproducibility of trypsin. 2. The derivatization step neutralizes the charge at the N-termini, and unmodified and monomethylated lysine residues thereby making the resulting histone peptides less hydrophilic. They can now be easily resolved by standard RP-HPLC; in addition, the peptides produce doubly and triply charged ions in electrospray ionization MS, and so are pseudo-tryptic-like fragments whose MS/MS spectra are facile to interpret. This is an advantage over using the Arg-C protease, which will cleave at Arg residues, but does not remove the charge from lysine residues [Garcia et al., 2007b].

Quantification

The quantification of differences among different physiological states of a biological system is not only the most important but also the most challenging technical tasks in proteomics.

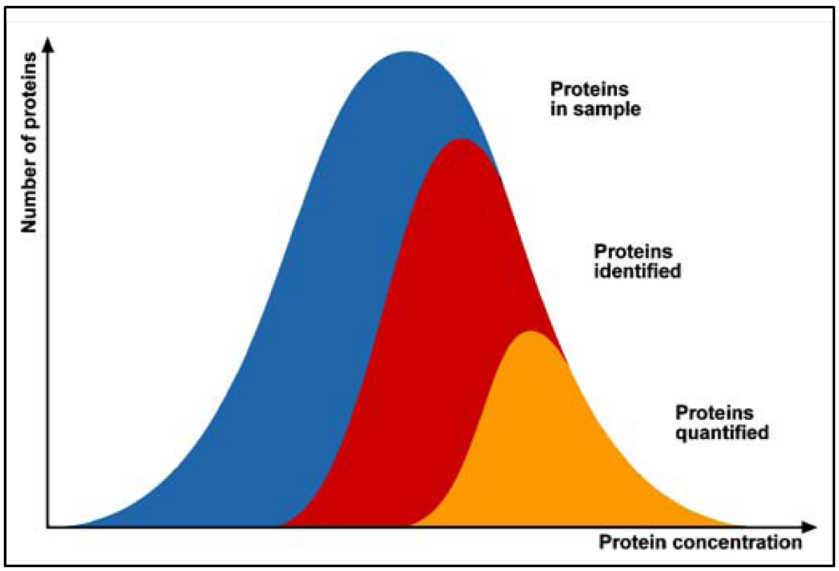


Figure 1.16: Schematic representation of the fraction of a proteome that can by identified or quantified by MS-based proteomics. Adapted from [Bantscheff et al., 2007].

The ability to accurately measure how the levels specific modifications change in response to different conditions, such as cell cycle changes, disease states or treatment with therapeutic agents, is essential for fully understanding the mechanisms by which histone modifications act [Knapp et al., 2007]. The primary function of the mass spectrometer in proteomics is the identification of peptides and proteins, however, it is not limited to mere identification of proteins in a given sample. The second important function of mass spectrometer is to provide quantitative information relating to relative or absolute peptide abundance. Mass spectrometry is, however, not inherently quantitative because proteolytic peptides exhibit a wide range of physico-chemical properties such as size, charge, hydrophobicity, etc. which lead to large differences in mass spectrometric response. For accurate quantification, it is therefore generally required to compare each individual peptide between experiments. During the last decade, several methods that add a quantitative dimension to mass spectrometric measurements have been developed [Ong and Mann, 2005]; [Bantscheff et al., 2007]; [Vermeulen and Selbach, 2009].

Despite the phenomenal impact of mass spectrometry and peptide separation techniques on proteomics, the identification and quantification of all of the proteins in a biological system is still an unmet technical challenge. While for unicellular organisms proteomic coverage of the genome has been occasionally achieved beyond 50%, coverage for higher organisms rarely exceeds 10%. For protein quantification, these figures are significantly smaller due to the fact that the data quality, in terms of information content, required for quantification by far exceeds that for protein identification [Figure 1.16] [Bantscheff et al., 2007].

In most proteomics workflows, these quantification strategies can be divided into three groups: chemical labeling, stable isotope labeling and label free approaches.

I **Chemical Labeling**

The simplest form of chemical labeling involves ^{18}O method, where isotope label is introduced into peptides by digesting one sample using trypsin or Glu-C in the presence of H_2 ^{18}O while the other sample is digested in ordinary aqueous solution to differentially label at the carboxyl terminus [Yao et al., 2001]; [Reynolds et al., 2002]; [Ye et al., 2009]. This has originally been employed to aid de novo sequencing of peptides by mass spectrometry [Rose et al., 1983] but has recently also been applied to quantitative proteomic applications [Miyagi and Rao, 2007]; [Ye et al., 2009]. A practical disadvantage is that full labeling is rarely achieved and that different peptides incorporate the label at different rates which complicates data analysis [Ramos-Fernandez et al., 2007].

The chemical derivatization procedure can be labeled by esterification using deuterated alcohols. This is accomplished by converting the carboxylic acid functional groups on the histones to their corresponding esters via either ethyl or methyl esterficiation. For instance, after propionylation and trypic digesion of the two samples, the peptides in the control sample are derivatized into d0-ethyl or d0-methyl esters using d0-ethanol/HCl or d0-methanol/HCl, and the peptides in the experimental sample are converted to their corresponding d5-ethyl or d3-methyl esters using d6-ethanol/HCl or d4-methanol/HCl, respectively. The two samples can then be combined in equal amounts and analyzed in a single LC-MS/MS experiment, where the same modified histone peptide from the two samples will be simultaneously detected with a +5 Da shift per label for ethyl esterfication and a +3 Da shift per label for methyl esterfication.

The stable-isotope labeling via ethyl esterfication has allowed for the quantification of the PTMs between physiologically different cell types [Syka et al., 2004b], differential expression analysis of histones from wild-type and methyltransferase mutant Arabidopsis Thaliana [Johnson et al., 2004], and quantification of PTM differences between a mouse model of lupus and control mice, where it was observed that site-specific hypoacetylation states on H3 and H4 were corrected with treatment by trichostatin A (TSA, a histone deacetylase inhibitor) and resulted in an improvement in disease phenotype [Garcia et al., 2005]. The latter study also reported novel histone PTMs corresponding to H3K18 methylation and K4K31 methylation and acetylation. Likewise, methyl esterification derivatization has been used to quantify the levels of modifications between all of the histone H3 variants [Hake and Allis, 2006].

Although an extremely useful quantitative tool, the esterification reaction is known to be particularly sensitive to moisture [Garcia et al., 2007b]. This requires samples to be completely dried prior to the reactions, which can potentially lead to sample loss. Furthermore, retained water and fluctuating reaction temperatures can result in reduced product yield or undesired side products. Motivated by these issues, a one pot approach was recently introduced which can identify and quantify nearly all histone variants and their corresponding PTMs in a single LC-MS/MS experiment without the need for prior HPLC or SDS-PAGE purification of the individual histone families. A second derivatization step using d0- and d10-propionic anhydride for the control sample and sample of interest, respectively, incorporates a stable isotope-label mass

shift on the new peptide N-termini created from the tryptic digest. Since d10-propionic anhydride incorporates a stable d5-propionyl amide label on the N-terminus of the peptides in the infected sample, the modified peptides appearing in both the control and infected samples are simultaneously detected with a +5 Dalton shift between them [Plazas-Mayorca et al., 2009]. While the incorporation of stable isotopes provides an important means for identifying and quantitating peptides containing post-translational modifications, the accuracy of these methods critically depends on the reproducibility of the derivatization reactions.

Post-biosynthetic labeling of proteins and peptides can also be achieved by incorporation of an isotope-coded affinity tag by chemical derivatization. This method known as the isotope-coded affinity tag (ICAT) approach uses cysteine residues specifically derivatized with a reagent containing either zero or eight deuterium atoms as well as a biotin group for affinity purification of cystein-derivatized peptides and subsequent MS analysis [Gygi et al., 1999]; [Shiio and Aebersold, 2006]. The rarity of cysteine makes ICAT and related methods advantageous by significantly reducing the complexity of the peptide mixture when highly complex samples are analyzed. However, ICAT is obviously not suitable for quantifying the significant number of proteins that do not contain any (or a few) cysteine residues and is of limited use for analysis of post-translational modifications and splice isoforms.

Another group of labeling reagents targets the peptide N-terminus and the epsilon-amino group of lysine residues. This is realized via the very specific N- hydroxysuccinimide (NHS) chemistry or other active esters and acid anhydrides as in, e.g., the isotope coded protein label (ICPL) [Schmidt et al., 2005], isotope tags for relative and absolute quantification (iTRAQ) [Ross et al., 2004], tandem mass tags (TMT) [Thompson et al., 2003]. The isobaric mass tagging in iTRAQ and TMT introduces tags that initially produce isobaric labeled peptides which precisely comigrate in liquid chromatography separations. Only upon peptide fragmentation are the different tags distinguished by the mass spectrometer and the relative quantification is performed in the MS/MS spectra, in which each tag generates specific fragment ions. This permits the simultaneous determination of both identity and relative abundance of peptide pairs in tandem-mass spectra. The commercially available iTRAQ reagent [Ross et al., 2004] provides a further refinement of this approach, allowing multiplexed quantitation of up to eight samples. This has turned out to be particularly useful for following biological systems over multiple time points or, more generally, for comparing multiple treatments in the same experiment.

Chemical labeling is applicable to all sample types. The primary disadvantage of the method is the inevitable presence of chemical side products, which can interfere with the analysis of rare PTMs and adversely affects quantification results.

II **Metabolic Labeling**

The metabolic labeling approaches introduce a stable isotope signature (light or heavy) into proteins through the growth medium or food during cell growth and division. This allows for the samples to be mixed immediately after the experiment at the level

of intact cells before any protein extraction, thus, minimizing experimental errors introduced by biochemical and mass spectrometric preparation procedures as these will affect both protein populations the same way.

Metabolic labeling was initially described for global labeling by replacing all nitrogen atoms in the medium by ^{15}N [Oda et al., 1999], although this technique is generally restricted to specialized applications, such as bacterial and plant species, because it leads to broad isotope distributions in the heavy forms. Furthermore, the cost and time required to routinely apply this technology is neither possible nor practical. More targeted approach of metabolic labeling is SILAC (Stable Isotope Labeling by Amino acids in Cell culture), where incorporation is achieved by replacing essential amino acids with their heavy counterparts.

SILAC, first described in the literature almost a decade ago [Ong et al., 2002]; [Blagoev et al., 2003], is conceptually a straight forward experimental setup: it involves growing two populations of cells, one in a medium supplemented with "light" form of an amino acid and the other contains a "heavy" amino acid. The heavy amino acid usually contains either ^{13}C or ^{15}N or both. Arginine and lysine labeling are most commonly used. Cells are cultured in heavy medium for at least five cell doublings to allow the incorporation of the heavy amino acid. After five cell doublings, 97%of the proteins should be present in the heavy state. In practice, the proteomes reach full incorporation of the heavy amino acid more rapidly through a combined process of cell division, new protein synthesis and degradation. Once, full labeling is achieved, cells are mixed in equal numbers, their proteomes extracted and analysed by LC-MS [Figure 1.17-A]. Since, trypsin cleaves at the C-terminal of arginine and lysine residues, it ensures that every tryptic peptide contains at least one labeled amino acid. For every peptide, two isotope clusters can be observed, one from each sample, a so-called SILAC pair [Figure 1.17-B]. The mass difference between the SILAC pairs is exactly the mass difference between the light and heavy amino acids. Generally Arg10 and Lys8 are used, generating shifts of 10.0083 Da and 8.0142 Da, respectively. For every identified peptide, a ratio can be directly assigned , and this ratio represents the relative expression level of that protein in the two experimental samples. A large variety of different established cell lines, including mouse and human ES cells, have already been labeled and the proteome is quantified [Graumann et al., 2008]; [Prokhorova et al., 2009]; [Tian et al., 2011]; [Rigbolt et al., 2011].

Analysis of PTMs is one of the daunting task in MS-based proteomics [Mann and Jensen, 2003]. Understanding PTMs and their regulation demands knowledge of their change in abundance in distinct cellular conditions. This regulation could occur at the level of protein abundance through gene expression, cellular localization and by site-specific PTM modulation. Any change in the levels of modification often leads to important biological consequences and therefore there is intense interest in the quantitative study of modifications. Stable isotope labeling in various forms is now being used to quantify protein levels on a large scale. A well-designed quantitative proteomics experiment could provide such key pieces of information in a single experiment. Two strategies for quantification of PTMs are available: either labeling of the peptides them-

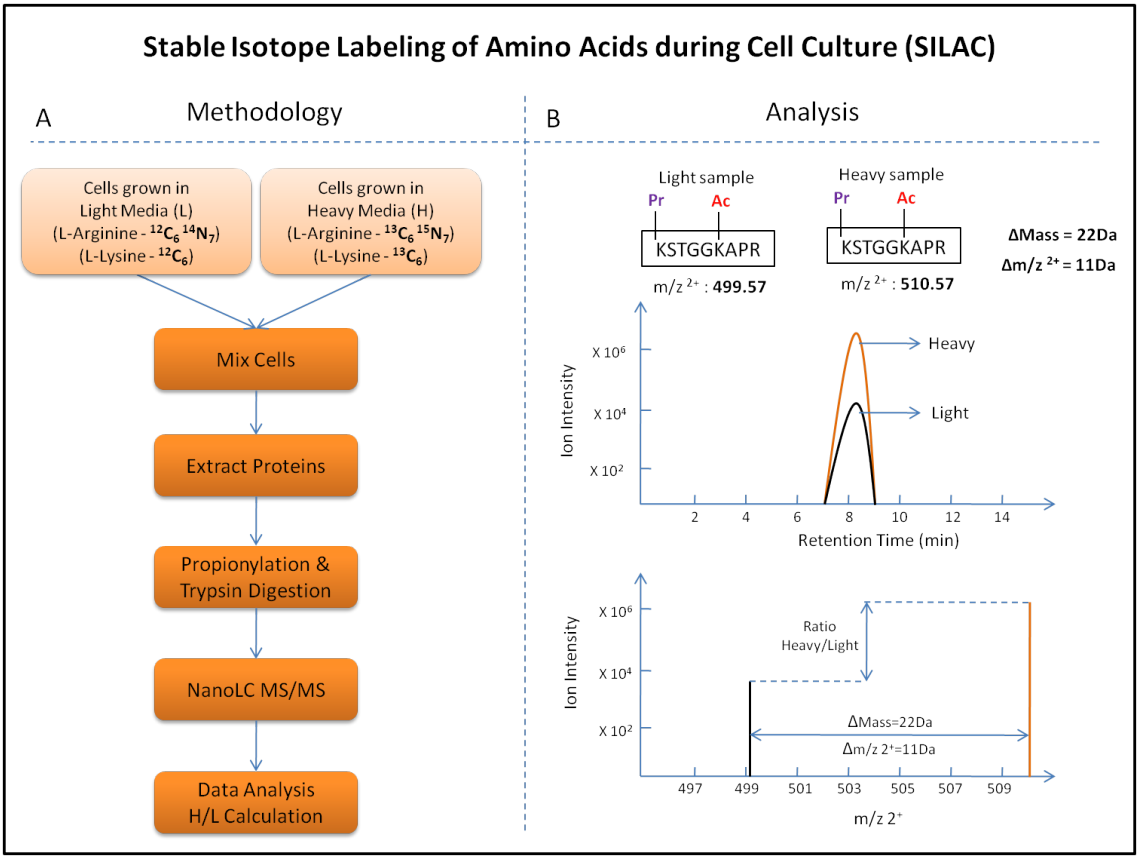


Figure 1.17: SILAC principle. A, on mixing of samples from two SILAC states, histones can be extracted and processed further for nanoLC-MS/MS analysis. B, SILAC pairs can be observed for every peptide. The chromatogram shows the co-eluting heavy and light peptide (upper graph), while the mass spectrum is used to calculate the relative abundances of light and heavy peptides (lower graph).

selves by a stable isotope method as described above, or introduction of the label into the modification. The first approach can be used for any modification and is mainly applied to achieve differential quantification of a modification of interest in two different samples. The latter approach has so far only been described for methylation, a SILAC strategy called "heavy methyl SILAC" [Ong et al., 2004].

Histone PTMs holds a fundamental place in chromatin biology and MS has proved to be the most reliable and unbiased method to identify and quantify these PTMs [Eberl et al., 2011]. Freitas and co-workers used SILAC to quantify changes in histone H4 with respect to HDAC inhibitor treatment, such as depsipeptide (DDP). The histones were first labeled with D4-lysine, since this residue is most modified, and then separated using acid-urea polyacrylamide gel electrophoresis (AU-PAGE). After in-gel digestion with trypsin, the histones are analyzed using MADLI-TOF MS. This approach was able to detect time-dependent changes in histone H4 during apoptosis after DDP exposure [Su et al., 2007a]. An Arg-C digestion of H4 followed by LC-MS/MS was also used to uncover patterns regarding the ordering of acetylations, where K16 was found to be acetylated first, K12 subsequently, and then lastly K8 or K5 [Zhang et al., 2007]. SILAC

has also been used for characterizing the modification patterns of the core histones in HeLa cells across the cell cycle (i.e., for cells arrested in G1, S1, and G2/M) [Bonenfant et al., 2007]. It was observed that the modifications were essentially identical in the G1 and S phases, but certain phosphorylation events were observed to occur exclusively in the G2/M phase for H3 and H4 (but not for H2A) in combination with a loss of acetylation across all histones. Furthermore, Jensen and coworkers used a knock-down of Suz12, a Prc2 component in mouse ES cells and found a methyl/acetyl switch in a SILAC based quantitative mass spectrometry analysis of histones H3.2 and H3.3. This knock-down led to reduction of K27 di- and trimethylation but also to an increase in K27 acetylation [Jung et al., 2010]. A SILAC approach was likewise applied to compare PTMs on histones H3 and H4 between breast cancer cell lines and normal epithelial breast cells, which demonstrate the possibility of defining cancer-specific methylation signature [Cuomo et al., 2011].

PTM enrichment strategies can be applied to enrich peptides containing a particular modification. Lysine-acetylated peptides, for example, can be enriched by appropriate antibodies [Zhao et al., 2010]; [Wang et al., 2010]. Choudhary et al. applied this strategy to identified as well as used SILAC to quantify 3600 lysine acetylation sites (majority were novel sites) in human cells upon treatment with the deacetylase inhibitors suberoylanilide hydroxamic acid and MS-275 [Choudhary et al., 2009]. These studies revealed a great diversity acetyl-regulated cellular processes. Another PTM that has frequently been the target of large scale proteomics studies is phosphorylation. Macek et al. described the major techniques available for enrichment for phosphopeptides [Macek et al., 2009]. These include 1) enrichment of peptides containing phosphorylated tyrosines by antibodies specific for this modification, 2) immobilized metal affinity chromatography (IMAC) utilizing coordination of phosphopeptides by metal ions, 3) titanium-dioxide-based enrichment, typically with a specificity enhancing reagent, and 4) enrichment in the flowthrough and early fractions in strong cation exchange chromatography (SCX), because of the presence of additional charge on phosphate at acidic pH, which leads to reduced retention times of phosphorylated peptides. Furthermore, Matsuoka et al. performed a large-scale SILAC-based proteomic analysis of proteins phosphorylated in response to DNA damage on consensus sites recognized by ATM and ATR and identified more than 900 regulated phosphorylation sites encompassing over 700 proteins. Functional analysis of a subset of this data set indicated that this list is highly enriched for proteins involved in the DDR [Matsuoka et al., 2007]. Recently a titanium dioxide enrichment strategy was used to identify and quantify, using SILAC, more than 10,000 phosphorylation sites throughout the HeLa cell cycle by Mann and coworkers [Olsen et al., 2010]. These findings suggest that mining of the data from large-scale, high-accuracy screens could provide useful insights and potential leads for further studies in elaborating the role of chromatin in essential cellular processes.

Apart from cell lines, whole organisms can also be metabolically labeled [Heck and Krijgsveld, 2004]. Thus far, S. cerevisiae (lysine auxotroph strains growing in minimal medium with heavy lysine) [Gruhler et al., 2005], Drosophila melanogaster (feeding

on heavy yeast) [Sury et al., 2010] and mice (by means of a special lysine-free diet supplemented with heavy lysine) [Kruger et al., 2008] have been labeled and used for functional in vivo proteomics experiments.

Although SILAC is a powerful tool for histone quantification, it is important to note that not all samples are compatible with this labeling technique, including certain tissues and body fluids. However, as an alternative to the classical SILAC approach, a different approachthe use of SILAC as an internal or "spike-in" standard [Geiger et al., 2010]; [Geiger et al., 2011], wherein SILAC is only used to produce heavy labeled reference proteins or proteomes. These are added to the proteomes under investigation after cell lysis and before protein digestion. The actual experiment is therefore completely decoupled from the labeling procedure. Spike-in SILAC is very economical, robust and in principle applicable to all cell- or tissue-based proteomic analyses. Spike-in SILAC is especially advantageous when analyzing the proteomes of whole tissues or organisms.

While many cell lines can be labeled quite readily, some do require special attention. For example, some cell lines require careful titration of the amount of arginine in the medium in order to prevent metabolic conversion of excess arginine into proline which in turn complicates data analysis [Ong et al., 2003]. Cell lines that are sensitive to changes in media composition or are otherwise difficult to grow or maintain in culture may not be amenable to metabolic labeling at all. A further limitation of metabolic labeling is the restricted number of available labels. For SILAC, a limited number of labels can be used, meaning that a limited number of conditions can be compared in one experimetnal setting, however, the use of spike-in SILAC allows for the comparison of more conditions than previously possible. Because of the early combination of samples, metabolic labeling and SILAC, in particular, is probably the most accurate quantitative MS method in terms of overall experimental process. This makes it particularly suitable for assessing relatively small changes in protein levels or those of post-translational modifications [Blagoev et al., 2004]; [Park et al., 2006a]; [Olsen et al., 2006]; [Choudhary et al., 2009]; [Olsen et al., 2010]; [Monetti et al., 2011].

III **Label Free Quantification**

Currently, three widely used label-free quantification strategies can be distinguished: (a) Relative Quantification by Peak Intensity of LC-MS, (b) counting and comparing the number of fragment spectra identifying peptides of a given protein [Liu et al., 2004], (c) exponentially modified protein abundance index (emPAI) [Ishihama et al., 2005]. These approaches rely on the fact that the signal intensity from electrospray ionization (ESI) correlates with ion concentration. Hence, more abundant the protein is in a given sample, the more MS/MS fragmentation spectra it will generate as compared to low-abundance proteins. An advantage of these methods is their applicability to any qualitative dataset, including low-resolution data. Spectral counting and emPAI can estimate the absolute amounts of proteins in a sample, as well as the relative amounts under two or more different sets of conditions. However, a significant drawback is that the correlation between the number of times that peptides are sequenced and protein amount is only approximate. In particular, proteins with low peptide counts show

high quantification variability. Also, a general disadvantage in label free approaches is that samples have to be processed and measured separately and that the accuracy is therefore compromised by experimental variability between different runs. A general workflow of label free proteomics is given in [Figure 1.18].

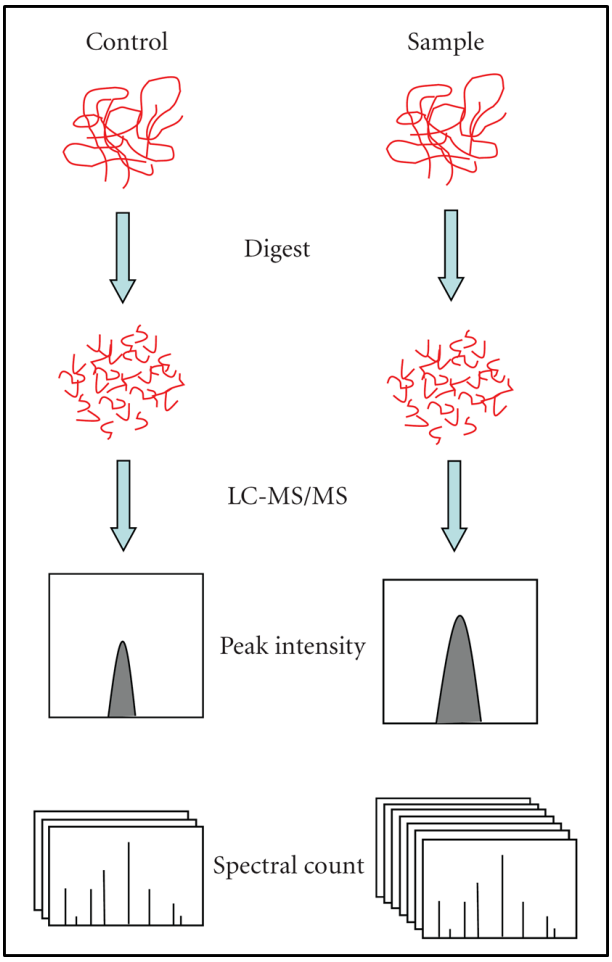


Figure 1.18: Label-free quantitative proteomics. Control and sample are subject to individual LC-MS/MS analysis. Quantification is based on the comparison of peak intensity of the same peptide or the spectral count of the same protein.

Label free approaches have been used in chromatin biology to quantify changes in histone PTMs as an alternate to stable isotope-based strategies. Using label free quantification, it was shown for an Arg-C digestion of H3 extracted from Drosophila cell lines that H3.3 accounts for 25% of total histone H3 in bulk chromatin, enough to essentially pack all actively transcribed genes. Modifications were observed that are consistent with the formation of euchromatin, or transcriptionally active genes. It had been previously believed that H3.3 was interchangeable with H3, however this study demonstrated the presence of H3.3 at active loci [McKittrick et al., 2004]. Thelabel-free approach has also been successfully applied in a number of cancer studies. Fraga et al. utilized high-performance capillary electrophoresis (HPCE) to resolve histone H4

forms by their degree of acetylation and methylation and LC-MS/MS to analyze 25 human cancer cell lines [Fraga et al., 2005]. It was determined that a hallmark common to all cancers is the loss of monoacetylation at H4K16 and trimethylation at H4K20, predominantly located at DNA repetitive sequences. In a study of small cell lung cancer [Beck et al., 2006], SDS-PAGE followed by in-gel tryptic digestion and reverse phase LC-MS/MS revealed 32 modifications on all core histones, including several novel modification sites for H2B, H2A, and H4. Triplicate runs for 6 different concentrations of the HDAC inhibitor PXD101 were shown to significantly change the abundances of many modified peptides. Alignment and calibration of the data with respect to peptide retention times and m/z values were used to enhance the quantification accuracy for the core histones. One important issue with the label-free approach for quantification that should be considered is the requirement that reproducible peptide sequences can be routinely generated. The ETD fragmentation technique has been applied to large histone fragments, enabling quantification of 74 unique H4 modification combinations in differentiating human ES cells by a labelfree approach [Phanstiel et al., 2008].

Top-Down MS Analysis

In the previous sections, we have described Bottom Up mass spectrometry as an invaluable and sensitive tool for the discovery and characterization of the post-translational modifications present on histone proteins. However, the necessary enzymatic digestion of the histone proteins into smaller peptide sequences loses important information regarding the connectivity of modifications on the molecular level. The ability to simultaneously analyze the modifications across an intact sequence is of paramount importance, as there is well-established cross-talk between sites of modification in histones and quantitative changes within the individual sites may not accurately reflect their collective changes. Recent evidence also supports the hypothesis that specific combinations of histone modifications are the driving force behind important cellular and nuclear events. For instance, it has been shown that differences in global histone modifications in tissues can be used to predict the clinical outcome of cancer patients [Seligson et al., 2005]; [Barlesi et al., 2007]; [Seligson et al., 2009], supporting the theory that the Histone Code works on a combinatorial basis.

In recent years, top down proteomics has emerged as another effective MS platform for PTM identification and quantification, where the digestion step is omitted and electron transfer dissociation (ETD) or electron capture dissociation (ECD) is utilized to fragment the intact protein sequence (described in the next section). The ETD and ECD fragmentation techniques have advantages over CID in that (a) their fragmentation is not inhibited by basic amino acids (e.g., arginine, lysine, and histidine) and (b) the PTM is retained on the residue side chain, whereas oftentimes in CID the PTM bond is the energetically favorable cleavage site.

Top Down MS usually involves initial separation of proteins from a complex mixture down to a single protein or few proteins. The protein sample is then either directly injected into the mass spectrometer through ESI, or further fractionated by on-line RP-HPLC followed by ESI. After injection, the intact protein precursor ion mass is normally measured at high mass

accuracy, usually on high-resolution instruments such as on Orbitrap or a Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR-MS). Tandem mass spectrometry is then performed in a targeted or data-dependent mode and data is then analyzed, often manually, or with specialized software. Additionally, entire protein sequences are deduced by top down MS, this results in 100% protein sequence determination, which is near impossible to generate with bottom up approaches as well.

It has been demonstrated that ion abundances from top down and middle down fragmentation (described below) in association with the ion abundances from the parent ions can be used for quantification of isomeric histone forms in the same mixture [Pesavento et al., 2006]. A review of recent studies is presented below, which have applied ECD or ETD for the top down MS analyses of histone variants and their PTMs.

The first report regarding the application of top down MS for the analysis of histone post-translational modifications was published by Kelleher research group. FT-ICR ECD data for histone H4 extracted from asynchronously grown HeLa cells was searched against a database derived from known histone modifications using ProSight Retriever. Several acetylated and methylated forms were observed and sodium butyrate was used to enhance the signal of hyperacetylated forms [Pesavento et al., 2004]. Protocol enhancements for the analysis of histone variants were achieved by the use of mild performic acid for oxidizing methionine into methionine sulfone and cysteine into cysteic acid [Pesavento et al., 2007]. This alleviated issues associated with partial oxidation, since the masses of oxidization (16 Da) and methylation (14 Da) are similar and can result in overlapping isotopic distributions.

Mizzen and co-workers combined the quantitative labeling approach SILAC with Top Down MS to detect quantitative changes in H4K20me throughout the cell cycle for asynchronous HeLa cells [Pesavento et al., 2008]. It was reported that newly synthesized H4 is gradually methylated during the G2, M, and G1 phases of the cell cycle and within two to three cell cycles is almost entirely dimethylated. Interestingly, new H4 is methylated regardless of the prior acetylation state and acetylation is observed to occur on H4 that is dimethylated, which was in contradiction to the former hypothesis that H4K20 methylation antagonizes H4 acetylation. Top Down ECD tandem MS confirmed the previously observed ordering of H4 acetylation, where K16 is acetylated first, then K12 subsequently, and finally K8 or K5.

Reverse phase HPLC was used in the majority of the aforementioned top down MS studies. This approach fails to chromatographically resolve isobaric modifications and results in the analysis of bulk histones. As a result, there is a decrease in both the specificity and sensitivity due to competing ionization efficiencies among the modified forms. However, the ability to detect and analyze the combinatorial modifications across single histone forms is of great interest to chromatin biologists. Previous studies have notably demonstrated the utility of acid-urea cation exchange chromatography [Thomas et al., 2006] and acid-urea gel electrophoresis [Jiang et al., 2007a] for separating modified histone forms based upon their degree of acetylation. The later study also showed a top-down analysis of intact yeast histones H2A, H2B, H4 and H3 and characterized over 50 distinct total histone forms. Subsequent experiments using methyltransferase knockout strains revealed the cointerdepen-

dence of histone H3 acetylation with H3K4 methylation as well.

Successful off-line techniques (in which the eluate is not directly coupled to a tandem mass spectrometer) for the separation of highly-modified histone forms have been achieved using cation exchange hydrophilic interaction chromatography (HILIC). Two-dimensional liquid chromatography (i.e., RP-HPLC and HILIC) coupled with top down ECD MS was used to identify and quantify 42 modified H4 forms from asynchronous and cycling HeLa cells [Pesavento et al., 2008]. Two novel modifications, namely H4K20me2 and H4N-acR3me2K20me2, were identified by this approach. A major disadvantage of these HILIC approaches is that the off-line nature of the experimental protocol is extremely time consuming (on the order of months) and thus prohibits the ability to conduct multiple runs for high-throughput studies and statistical validation.

Top down analysis of intact histone H2B in Tetrahymena using ECD fragmentation found 13 versions of this protein (variants and their modifications) [Medzihradszky et al., 2004], while the human version revealed the presence of seven unmodified H2B variants in addition to one monomethylated isoform [Siuti et al., 2006]. A similar analysis on H2A found at least 14 forms of this protein expressed in human cells. These experiments also detected the presence of an H2Aubiquitin- conjugated protein [Boyne et al., 2006]. Multiple approaches, including a top-down analysis, were used in one study to reveal the modifications of histone H4 [Zhang et al., 2004]. Lastly, top down analysis lead to the characterization of human histone H3.1, H3.2 and H3.3 modification site occupancy. Histone H3 was found to be mostly and pervasively post-translationally modified as less than 3% of the unmodified protein was found for all three H3 variants [Thomas et al., 2006]. Again, none of these methods can be considered high throughput, as they involve both extensive off-line prefractionation and the subsequent analysis of many fractions. Progress has been made towards top-down analysis on a chromatographic time scale, including some characterization of histone H4 combinatorial forms in the process [Parks et al., 2007]. Such an approach is very appealing, but there remain technical challenges to overcome before on-line top-down LC-MS will be used routinely in the proteomics labs as the most effective approach.

Middle-Down MS Analysis

Middle down mass spectrometry, an alternate version of top down approach (analysis of peptides with mass greater than 3000 Da) has also been recently popularized. Middle down mass spectrometry requires the proteolysis of proteins with enzymes that target less abundant amino acid residues than trypsin, such as GluC or AspN, before a similar path as Top Down MS is taken. A particular advantage of top down and middle down approaches is that if a protein is multiply modified at residues that are not very close in sequence, this combinatorial modification pattern can be deduced by Top Down MS or partially by Middle Down MS, but would mostly likely not be deciphered by bottom up mass spectrometry methods. Middle down approach gives rise to shorter but still highly charged peptides that are ideal for thorough analysis by ETD or ECD analysis. However, the physical separation of these smaller, but still substantial, peptides is a significant problem that has to be overcome. In essence, the middle down approach is currently considered as a thorough and reasonably

high-throughput method.

Middle down approach, until recently, have utilized primarily off-line separations followed by analysis of fractions, as is the case with top-down methods. These off-line methods usually require large amounts of sample and are quite long for the analysis of a single polypeptide. With such off-line separations, the analysis of all fractions of a single sample representing one biological state could take more than 100 h of analysis time. Yet these still relatively new approaches have led the way toward more high-throughput approaches.

Pesavento et al. performed middle-down analyses on histones, first without up-front separation [Pesavento et al., 2004], and later fractionating with HILIC chromatography [Pesavento et al., 2006], and were able to achieve reasonable results on the relatively simple H4 system. Reversed phase and PolyCAT A-based HILIC separations have been used to fractionate out the H3-(150) peptide from Tetrahymena into distinct acetylation states before ETD analysis on a modified Orbitrap [Taverna et al., 2007]. This identified over 40 modified forms of histone H3 from this lower eukaryote. Around the same time, modified histone H3.2 forms extracted from asynchronously grown and butyrate treated HeLa cells were fractionated into 30 fractions and over 150 unique modified forms were identified using middle down ECD, revealing partial ordering patterns for H3K14, K18, and K23 acetylation and H3K4 methylation correlations with acetylation [Garcia et al., 2007]. Here, the separation occurs first by separating isoforms by acetylation status (weak cation exchange) and then by a second mode of separation based on the methylation (HILIC). HILIC fractions were statically infused into an FT-ICR mass spectrometer, and fragmentation was induced using ECD [Garcia et al., 2007]. In another study, middle down approach was used and HeLa cells were treated with 5-aza-2'-deoxycytidine, a DNA methylating inhibitor, which resulted in a significant reduction in global methylation states for H3.1 and a simultaneous rise in relative acetylation levels [Siuti and Kelleher, 2010].

A reasonable online separation with an ETD capable of fragmentation on chromatographic timescale has only recently been achieved when Phanstiel et al. perfomed reverse phase LC-MS/MS using a custom-built ETD-capable quadrupole ion trap mass spectrometer. 74 discrete combinatorial codes on the tail of histone H4 from human embryonic stem (ES) cells were detected in this study [Phanstiel et al., 2008]. Due to the reverse phase LC, most of the combinatorial forms, however, co-eluted with others and had to be deconvoluted by mathematical analysis. Similarly, Nicklay et al. analyzed both H3 and H4 from Xenopus laevis, but only the H4 was analyzed by middle-down methods. They also used on-line reversed phase LC-MS methods in conjunction with ETD on the H4 123 peptide tail, where they observed 66 discreet isoforms [Nicklay et al., 2009].

Young et al. recently developed a method to identify and quantitate combinatorial histone codes, utilizing a pH gradient to elute histone codes from a weak cation exchange and hydrophilic interaction (WCX-HILIC) nano flow liquid chromatography column directly into a mass spectrometer using electron transfer dissociation [Young et al., 2009]. This was demonstrated on the histone H3 and histone H4 N-terminal peptide (first 50 and 23 amino acids, respectively). With this approach, they have shown an effective chromatographic separation

of the structural isomers of histones H3 and H4 and they have characterize over 200 modified forms of histone H3.2 in one LC-MS/MS experiment and over 70 modified forms from histone H4. This was a dramatic improvement over previous approaches that required weeks worth of work to produce similar data. There remains room for improvement in this current state of high throughput histone code analysis, yet it is truly a high-throughput method, and the quality and confidence in the data are greater than previous lower throughput methods. Cation exchange HILIC chromatography had been used previously, but with non-volatile buffers, and was based on an ionic strength gradient [Garcia et al., 2007], however, the critical step here was to convert this chromatography into a high-throughput on-line LC-MS method. This was achieved by the use of a pH gradient to take advantage of the weak cation exchange properties of the PolyCAT A stationary phase. In this manner, the cation exchange retention mechanism could be disrupted while being compatible with directly introducing the eluent into the mass spectrometer.

Dissociation Methods for MS/MS Analysis

I **Collision Induced Dissociation (CID)**

Tandem mass spectrometry (MS/MS) is a key technique for protein or peptide sequencing and PTM analysis. Collision-induced dissociation (CID) [Shukla and Futrell, 2000] has been the most widely used MS/MS technique in proteomics research. In this method, gas-phase peptide/protein cations are internally heated by multiple collisions with rare gas atoms. This leads to peptide backbone fragmentation of the CN bond resulting in a series of b-fragment and y-fragment ions. Because of the slow-heating, energetic feature associated with this method, the internal fragmentation and neutral-losses of H_2O, NH_3, and labile PTMs are common. This also results in limited sequence information for large peptides (>15 amino acids) and intact proteins.

II **Electron Capture Dissociation (ECD)**

A new fragmentation technique, electron-capture dissociation (ECD) was introduced by the McLafferty laboratory in [Zubarev et al., 2000]; [Zubarev, 2006], by which the capture of a thermal electron by a multiply protonated peptide/protein cation induces backbone fragmentation at the NCα bond to produce c-type and z-type fragment ions. ECD provides more extensive fragmentation resulting in richer MS/MS spectra and better sequence coverage, and the nonergodic feature of ECD preserves labile PTMs. Therefore, it has become a powerful tool for top-down analysis of intact proteins [Medzihradszky et al., 2004]; [Thomas et al., 2006]; [Zabrouskov and Whitelegge, 2007]. However, ECD is most often constrained to the expensive, highly sophisticated FTICR instruments.

III **Electron Transfer Dissociation (ETD)**

An analogous technique, electron-transfer dissociation (ETD) was developed by the Hunt laboratory in 2004 and extends electron-capture-like fragmentation to more common bench top mass spectrometers [Syka et al., 2004a]. In this process, the electrons transfer from radical anions with low electron affinity to multiply protonated peptide

cations initiating backbone fragmentation to produce c-ion and z-ion series. Because the ion/ion reaction is highly efficient and fast, ETD can easily be performed with femtomole quantities of peptides on a chromatographic timescale. ETD MS/MS provides superior sequence coverage for small-sized to medium-sized peptides and is highly complementary to conventional CID for proteome identification applications [Good et al., 2007]. ETD can be utilized to analyze very large peptides as well as intact proteins with a sequential ion/ion reaction, proton transfer/charge reduction (PTR) by which the ETD produced multiply charged fragments are deprotonated with evenelectron anions resulting in singly and doubly charged ions that are readily measured by the bench-top low resolution ion trap instrument [Coon et al., 2005]; [Chi et al., 2007b]. This allows for the sequence analysis of 1540 amino acids at both N-terminus and C-terminus of the protein. ETD has also shown great promise in labile PTM analyses such as phosphorylation and glycosylation [Chi et al., 2007a]; [Navajas et al., 2011]; [Wang et al., 2011].

1.7 DNA Damage

Damage to our genetic material is an ongoing threat to both our ability to faithfully transmit genetic information to our offspring as well as our own survival. To respond to these threats, living organisms have evolved the DNA damage response (DDR). The DDR is a complex signal transduction pathway that has the ability to sense DNA damage and transduce this information to the cell to influence cellular responses to DNA damage. To maintain genomic integrity, DNA must be protected from damage induced by environmental agents or generated spontaneously during DNA metabolism.

Environmental DNA damage can be produced by physical or chemical sources. Examples of physical genotoxic agents are ionizing radiation (IR) and ultraviolet (UV) light from sunlight,which can also induce up to 105 DNA lesions (pyrimidine dimers and 64 photoproducts) per cell per day [Hoeijmakers, 2009]. Ionizing radiation (from, e.g., cosmic radiation and medical treatments employing X-rays or radiotherapy) can induce oxidation of DNA bases and generate single-strand and double-strand DNA breaks (SSBs and DSBs, respectively). Chemical agents used in cancer chemotherapy can cause a variety of DNA lesions: alkylating agents such as methyl methanesulfonate (MMS) and temozolomide attach alkyl groups to DNA bases, while crosslinking agents such as mitomycin C (MMC), cisplatin, psoralen, and nitrogen mustard introduce covalent links between bases of the same DNA strand (intrastrand crosslinks) or of different DNA strands (interstrand crosslinks or ICLs). Other chemical agents, such as the topoisomerase inhibitors camptothecin (CPT) and etoposide, which inhibit topoisomerase I or II respectively, induce the formation of SSBs or DSBs by trapping topoisomerase-DNA covalent complexes. Cigarette smoking, one of the most common mechanisms of self-inflicted DNA damage, causes a wide variety of adducts and oxidative damage in lung and other tissues. Hydrolytic reactions and non-enzymatic methylations also generate thousands of DNA-base lesions. DNA damage can also be generated

by reactive species that attack DNA leading to base loss, DNA single-strand breaks (SSBs) or adducts that impair base pairing and/or block DNA replication and transcription.

To counteract DNA damage, repair mechanisms specific for many types of lesion have evolved. Different DNA-repair pathways exist and perform major roles at both cellular and organismic levels. These pathways include (1) mismatch repair (MMR) pathway [Jiricny, 2006], where mispaired DNA bases are replaced with correct bases, (2) SSBs are repaired by single-strand break repair (SSBR) [Caldecott, 2008], (3) the nucleotide excision repair (NER) pathway, which repairs more complex lesions, such as pyrimidine dimers and intrastrand crosslinks, through the removal of an oligonucleotide of approximately 30 bp containing the damaged bases, while ICLs are excised by ICL repair employing Fanconi anemia pathway [Moldovan and D'Andrea, 2009], (4) the base excision repair (BER) pathway, where small chemical alterations of DNA bases are repaired through excision of the damaged base and finally, (4) the homologous recombination (HR) pathway, and (5) the non-homologous end joining (NHEJ) pathway, which deals with the DSBs. While NHEJ promotes the potentially inaccurate religation of DSBs, HR precisely restores the genomic sequence of the broken DNA ends by utilizing sister chromatids as template for repair.

I **Mismatch Repair (MMR)**

The MMR pathway plays an important role in both prokaryotes and eukaryotes in repairing mismatches, which are small insertions and deletions that take place during DNA replication [Jiricny, 2006]. Failure of MMR commonly results in microsatellite instability (MSI). Several homologues of the bacterial MMR genes MutS and MutL have been identified in yeast and mammals. The importance of the MMR pathway became evident upon identification of mutations in certain human MMR genes in hereditary non-polyposis colorectal cancer (HNPCC), a highly penetrant autosomal dominant cancer syndrome [Vasen et al., 2007]. Whereas most HNPCC human individuals carry heterozygous germline mutations of MMR genes, which predisposes them to cancer, mice heterozygous for MMR mutations do not appear to have an increased risk for developing cancer. This difference is not specific for MMR mutations, as heterozygous mutations in certain genes involved in other DNA-damage repair pathways are also able to predispose humans, but not mice, to cancer.

II **Base Excision Repair (BER)**

The BER pathway deals with base damage, the most common insult to cellular DNA [Wilson and Bohr, 2007]. Two sub-pathways, short-patch BER and long-patch BER, are involved in BER. The short-patch BER sub-pathway typically replaces a single nucleotide, whereas the long-patch sub-pathway results in the incorporation of 213 nucleotides. recent studies have demonstrated the existence of a human disorder linked to defective BER. This autosomal recessive disorder, referred to as MUTYH-associated polyposis (MAP), is associated with biallelic germline mutations of the human MUTYH, and is characterized by multiple colorectal adenomas and carcinomas [Cheadle and Sampson, 2007]. Despite the presumed important role for the BER pathway in

maintaining genomic integrity, mutations in this pathway have not significantly predisposed mutant mice for cancer. This is in contrast to mutations of other excision repair pathways, such as NER and MMR.

III **Nucleotide Excision Repair (NER)**

The NER pathway is a multistep process that serves to repair a variety of DNA damage, including DNA lesions caused by UV radiation, mutagenic chemicals, or chemotherapeutic drugs that form bulky DNA adducts. Over 30 different proteins are involved in the mammalian NER, whereas only three proteins (UvrA, UvrB, and UvrC) are required by prokaryotes [Truglio et al., 2006]. Two NER sub-pathways that have been identified are as follows: the global genome NER (GG-NER) that detects and removes lesions throughout the genome, and the transcription-coupled NER (TC-NER), which repairs actively transcribed genes. NER begins with the recognition of the DNA lesion, followed by incisions at sites flanking the DNA lesion, and culminates in the removal of the oligonucleotide containing the DNA lesion. NER has attracted a great deal of attention due to its role in three rare human syndromes characterized by increased cancer frequencies, neurodegeneration and ageing. These syndromes are xeroderma pigmentosum (XP), Cockayne syndrome (CS), and trichothiodystrophy (TTD) [Thoms et al., 2007].

IV **Homologous Recombination (HR)**

DSB repair can be mediated by two major repair pathways depending on the context of the DNA damage, HR or NHEJ repair pathways [Kanaar et al., 2008]. In bacteria and yeast, DSBs are preferentially repaired by HR, whereas more than 90% of DSB in mammalian cells are repaired by NHEJ. Both pathways are well defined and their impairment is associated with defects and pathologies, including increased cell death, cell-cycle arrest, telomere defects, genomic instability, meiotic defects, immunodeficiency, and cancer [Sung and Klein, 2006]. HR is a multistep process that requires several proteins and operates at the S or G2 phase of the cell cycle. Although it accounts only for the repair of 10% of DSBs in mammalian cells, HR defects can have severe consequences, as demonstrated by the human syndromes AT-like disorder (ATLD) and the NBS [Thompson and Schild, 2002]. The predisposition to either syndrome has been linked to mutations in the MRN complex, which is important for the resection of DSBs [Thoms et al., 2007]. However, it is important to note MRN functions are not restricted to HR, as is also involved in NHEJ, checkpoint activation, and telomere maintenance [Niida and Nakanishi, 2006]. BRCA1 and BRCA2, the early-onset breast cancer-susceptibility genes, have also been demonstrated to partake in HR-mediated DSB repair. Thus the human syndromes and early onset of breast cancer, and the developmental defects, increased genomic instability and tumorigenesis associated with impaired HR in mouse models, all demonstrate the in vivo requirement for HR-mediated DNA-damage repair.

V **Non-Homologous End Joining (NHEJ)**

NHEJ is the predominant pathway of DSBs repair in mammalian cells [Kanaar et al.,

2008]. This repair pathway is active especially at the G1, but is error prone. NHEJ is also essential for T-cell receptor α & β and Ig V(D)J recombination, and thus this repair pathway is required for the development of the T and B-cell repertoires. The core protein components of the mammalian NHEJ include the Ku subunits (Ku70 and Ku80), DNAPKcs, XRCC4, DNA ligase IV (LigIV), Artemis, and the recently identified CernunnosXLF (also known as NHEJ1). The radiosensitivity, genomic instability, immunodeficiency, growth retardation, embryonic development, and cancer predisposition associated with defective NHEJ all demonstrate the major role this DNA-damage repair pathway plays in vivo.

The coordination of DNA repair processes plays a critical role in allowing the proper development and survival of organisms. They are responsible for preventing numerous human diseases and conditions, including cancer and aging. Therefore, a better and more in depth understanding of various DNA damage responses along with other associated cellular pathways is one of the essential goals in developing therapeutics approaches for human diseases

1.8 Fanconi Anemia - A Model to Study DNA Damage Response

Fanconi anemia is a rare, recessive (autosomal or X-linked) inherited disease, characterized by severe bone marrow failure, a multitude of congenital defects and elevated risk of developing malignancies including acute myeloid leukemia, squamous cell carcinomas of head and neck, and hepatic tumors [D'Andrea and Grompe, 2003]; [Mace et al., 2007]; [Wang et al., 2007]; [Moldovan and D'Andrea, 2009]. FA is a highly heterogeneous genetic disorder: it consists of at least 13 complementation groups (A, B, C, D1, D2, E, F, G, I, J, L, M and N), each of which is associated with a distinct gene disease. A possible 14th FA gene, RAD51C, has recently been identified [Vaz et al., 2010]. Eight FANC proteins are assembled to form a large E3 ubiquitin ligase nuclear complex known as "FA-core complex" [Kennedy and D'Andrea, 2005], which mono-ubiquitinates the FANCI-FANCD2 complex after DNA damage [Smogorzewska et al., 2007]; [Alpi et al., 2008]. Ubiquitinated FANCI-FANCD2 is recruited to the chromatin where it colocalizes with DNA repair factors [Taniguchi et al., 2002]; [Wang et al., 2004b]. The FANC proteins, all working together in concert, constitute a common pathway in the DNA damage repair signaling. The pathway, termed as "Fanconi Anemia Pathway", is a major signaling cascade in response to stalled replication forks and DNA interstrand crosslinks (ICL) in addition to other types of DNA damage including ionizing radiation (IR) induced DNA double strand breaks and UV-induced damage [Moldovan and D'Andrea, 2009].

Interstrand crosslinks covalently link both strands of a DNA, thereby blocking its unwinding by DNA helicases and stopping the progression of the replication and transcription machineries. These kinds of lesions pose an extreme danger to the cell's fate if not repaired or bypassed [Scharer, 2005]. The FA-pathway is primarily activated during S-phase [Taniguchi et al., 2002]; [Raschle et al., 2008]; [Thompson and Hinz, 2009], and it has been shown that

DNA ICLs are removed almost exclusively during replication [Akkari et al., 2000], although FA-pathway might also play a minor role in ICL repair during G1-phase of the cell cycle [Ben-Yehoyada et al., 2009]. These results suggest that ubiquitinated FANCI-FANCD2 controls ICL repair in S phase, but the underlying mechanism how the 'Fanconi Anemia Pathway' functions is unknown. It is, however, hypothesized to regulate molecular processes that stabilize and/or resume the arrested fork by influencing nucleotide excision repair, homologous recombination and translesion DNA synthesis [D'Andrea and Grompe, 2003]; [Rosselli et al., 2003]; [Wang et al., 2007]; [Moldovan and D'Andrea, 2009].

At molecular level, the cells derived from patients of FA show spontaneous chromosome instability and striking sensitivity to DNA interstrand cross-linking (ICL) agents like mitomycin C (MMC), methoxypsoralen (8-MOP), or Cisplatin. These two features are thought to be responsible for the cancer susceptibility aspect of the disease [Niedernhofer et al., 2005]. The hypersensitivity of FA cells to DNA-ICL agents and increased chromosomal breakage with clinically strong predisposition to cancer implies a profoundly impaired ability to recognize or repair DNA ICL [Mathew, 2006].

The fact that Fanconi anemia pathway is activated not only by DNA-ICLs inducing chemicals, but also by other DNA damaging agents such as ultraviolet radiation (UV), ionizing radiation (IR), hydroxyurea, and even spontaneously during replication [Dunn et al., 2006]; [Howlett et al., 2002]; [Taniguchi et al., 2002], suggests the FA pathway is likely to be involved in the replication-dependent repair of many types of lesions. But unlike ICLs, most of these lesions are also repairable by other pathways. Thus, loss of FA proteins only mildly or not at all affects survival rates in response to IR or UV [Bridge et al., 2005]; [Dunn et al., 2006]; [Niedzwiedz et al., 2004]. However, the FA pathway is the principal mechanism that can remove crosslinks efficiently, accounting for the hypersensitivity of FA patient cells to ICLs.

1.9 Project Presentation

Cells are constantly exposed to endogenous (metabolic) and environmental agents (e.g. ionizing radiation) that induce potentially harmful DNA lesions. Therefore, preventing genomic instability is essential for maintaining integrity of the organisms. Accordingly, cells have developed diverse systems that integrate DNA damage detection and checkpoint mechanisms to coordinate repair and cell cycle progression [Hoeijmakers, 2007]; [Lobrich and Jeggo, 2007]. Loss of these mechanisms invariably results in accumulation of genetic changes and often leads to cancer predisposition [Lobrich and Jeggo, 2007]. However, in eukaryotes, the packaging of DNA into chromatin constitutes a physical barrier to enzymes and regulatory factors, generally preventing their access to DNA necessary for its proper functioning, such as replication, transcription, recombination and repair. Recent work has also revealed the importance of histone PTMs in other aspects of genome functioning, both as a mechanism for factor recruitment and for signal propagation [van Attikum and Gasser, 2009]; [Bannister and Kouzarides, 2011]. These observations underscore the importance of chromatin, and not

naked DNA, as the natural substrate of most genome based transactions.

The role of covalent histone modifications in transcriptional regulation and replication have been studied extensively, however, their implication in DNA damage response has started gaining interest more recently. The recent research suggests that the language of covalent histone PTMs termed as Histone Code [Strahl and Allis, 2000]; [Jenuwein and Allis, 2001] plays an essential role in maintaining the genome integrity and multi-step process of carcinogenesis [van Attikum and Gasser, 2005]; [Hake et al., 2007]; [Escargueil et al., 2008]. This code is believed to regulate chromatin accessibility either by disrupting histone-DNA contacts or by recruiting non-histone proteins to chromatin [Section 1.4.5]. Phosphorylation of histone H2AX (γH2AX), one of the most well-characterized modifications, caused by DNA DSBs (as well as by other kinds of DNA damage), on the chromatin surrounding the DNA lesion, is required for the recruitment and stabilization of DNA repair proteins to the damaged chromatin. It has also been shown to be required for the recruitment of FANCD2 protein to the stalled replication forks [Bogliolo et al., 2007]. However, it is also not clear whether induction of histone modifications is dependent upon functional FA-pathway. Accordingly, the role of histone code "writers" and "erasers" needs to be elaborated in detail in order to get an insight into the epigenetic regulation of DNA damage response.

The first goal of this thesis was to study histone code in the context of Fanconi anemia, in order to elucidate the role of chromatin in DNA-ICL damage response. AIn this regard, we have quantitatively analyzed the dynamics of histone PTMs in FA mutated and corrected cell lines after inducing DNA-ICL damage, using high throughput mass spectrometry approach. For quantitative comparisons, we have used an established quantitative approach, SILAC, which offers efficient metabolic labeling of cells. FA mutated cells provided a good model to study DDR to DNA-ICL, as these cells (derived from FA patients) are unable to sense and repair the DNA-ICLs. Accordingly, the FA corrected cells (FA mutated cells corrected by introduction of wild type gene) provided an adequate control for our experiments.

Out of the various DNA-ICL inducing chemicals, we used 8-methoxypsoralen (8-MOP) (10 μM). 8-MOP belong to a class of the group of furocoumarins, present in plants and cosmetics [Scott et al., 1976]. 8-MOP can be activated by UVA light to induce thymine mono-adducts and ICLs between thymines at d(TpA) sequences in DNA. The ratio of ICL to mono-adduct can be influenced by the wavelength and dose of UVA. In this way, up to 40% of the adducts can be converted to ICLs at a suitable UVA dose [Brendel and Ruhland, 1984]; [Dronkert and Kanaar, 2001].

Histones are covalently modified by chromatin modifying proteins. Our histone PTM analysis revealed interesting differences in various modifications after induction of DNA-ICL damage, and we were next intrigued to understand what histone modifying enzymes could be responsible for the observed PTM patterns. As a second step, using tools of functional genomics, we have analyzed the gene expression level of various histone acetyltransferaes (HATs), histone deacetylases (HDACs), histone methyltransferases (HMTs), and lysine demethylases (KDMs), as our histone code analysis was focused on histone lysine acetylations and lysine methylations.

DNA interstrand cross-links (ICLs) create a serious threat to cell survival by posing a complete blockade to DNA replication and transcription. To improve our general understanding of DDR to DNA-ICL at transcriptome level, we next decided to study the global transcriptional response to DNA damage in FA mutated cells after inducing two different kinds of DNA damage i.e., ICL and IR damage. We have chosen to study IR induced DDR in FA screen, in addition to DNA-ICL, in order to understand the role of functional FA pathway in different kinds of DNA damage and also to find out if the FA-screen based transcriptional profile shows differences in the two DDRs. This would ultimately help us to narrow down our analysis of genes, which are affected, after DNA-ICL induction, by the presence or absence of functional FA pathway.

In the "histone code analysis" part of this thesis, we have performed a global analysis of histone PTM changes in response to DNA damage. However, DNA repair occurs in distinct and localized regions of the nucleus, as evident from the appearance of γH2AX foci after induction of DNA DSBs. It is therefore important to study histone PTM profile at a more local level than on global scale. The third and final part of this thesis comprised of a developing a new proteomics methodology to study local changes in histone modification profile, in contrast to global analysis that we have performed earlier. For developing this technique, we have used a different model system based upon an E3 ubiquitin ligase Rad18, which forms detectable foci in the nucleus. Due to technical limitations we could not use FA model. We, however, believe that this technique has a much larger scope, and can be eventually used to study histone PTM changes related to DDR in Fanconi anemia.

2

Results - I

2.1 Mass Spectrometry-Based Histone Modification Profiling of Cellular Response to DNA-ICL

A rapidly emerging evidence suggests that histones, the architectural proteins that package DNA into nucleosomal arrays, are major carriers of epigenetic information. This information, mostly in the form of histone post-translational modifications, is likely to have an impact on many essential cellular processes including DNA damage response. In this part of work, we sought to examine the histone PTM profile in response to DNA damage of a particular kind, namely DNA interstrand cross links (DNA-ICL). We have chosen Fanconi Anemia (FA) as a model for this study since the cells derived from patients of FA, show striking sensitivity to DNA-ICL inducing agents, such as mitomycin C (MMC), 8-methoxypsoralen/UVA (8-MOP), or Cisplatin. In order to comprehensively analyze the various histone PTM marks in both FA mutant and control cells, after induction of DNA-ICL damage, we decided to use the mass spectrometry (MS)-based high throughput approach. For quantitative analysis of histone PTM, a stable isotope labeling technique, SILAC (Stable Isotope Labeling of Amino acids during cell Culture) was used, which is based on incorporation of stable isotope-coded lysines and arginines into newly synthesized proteins through normal metabolic processes. Before performing the actual experiment, we found it necessary to optimize the detection of various histone PTMs on our mass spectrometer and devise ways to quantify the detected peptides. Accordingly, first we describe how we achieved this preparatory goal, and then we will describe what kinds of changes in histone PTMs we have observed after subjecting cells to DNA-ICL damage, in both FA mutant and control cell lines.

2.1.1 MS-Based Identification of Histone PTMs

For the analysis of histone PTM profiles, automated nano-flow liquid chromatography coupled with tandem mass spectrometry (Agilent's nano-HPLC ESI-Ion Trap) was used. As an experimental model we selected FA-C cells, which corresponds to fibroblasts obtained from a FA patient having mutation in FANCC gene. The protein product of FANCC gene participates in the formation of "FA core complex", which later leads to the mono-ubiquitination of "FA ID complex", one of the critical events in the cellular response to DNA-ICL damage. FA-C corrected fibroblasts (corrected for FANCC mutation by introducing the wild type FANC gene [Zanier et al., 2004] were also taken along with the mutated cells to serve as a control.

For induction of DNA-ICL, we used a DNA intercalating agent 8-methoxypsoralen (8-MOP) (10 μM). Cells were kept at 37°C for 20 minutes and then irradiated with a suitable dose of UVA (10 KJ/m^2) (8-MOP can be activated by UVA light to induce thymine mono-adducts and ICLs between thymines at d(TpA) sequences in DNA). Out of several known agents inducing DNA-ICL, 8-MOP was selected because it can induce up to 30-40% of DNA ICLs depending upon the UVA dose [Brendel and Ruhland, 1984]; [Dronkert and Kanaar, 2001]. Cells were then harvested at different time intervals. Histones were subsequently extracted, by acid extraction method [Chapter 6], and histones from different time points were mixed together in order to increase the chances of detection of modified peptides. Samples were then quantified on Nano Drop and about 5 μg was loaded on 15% SDS-PAGE gel electrophoresis. Gel bands corresponding to different core histones were excised, treated with propionic anhydride and in-gel trypsin digested and run on nanoLC-MS/MS as described in Materials and Methods [Chapter 6]. Although the mass spectrometer can only detect ions, for our convenience we will use the term "peptide" instead of ions throughout this thesis. The purpose of treatment with propionic anhydride was to block trypsin digestion of unmodified and mono-methylated C-terminal lysines [Garcia et al., 2007b], which significantly simplifies the comparison between modified and unmodified peptides.

In preparation for this analysis, a list of histone peptides (modified and unmodified) along with their mass to charge (m/z) ratios was compiled using previously published data for core histones H3 and H4 [Zhang et al., 2002b]; [Zhang et al., 2003]; [Beck et al., 2006]; [Su et al., 2007b]; [Garcia et al., 2007a]. In addition, as we were hoping to identify new, previously unobserved modifications, the predicted m/z ratios for several potential candidate peptides were also included in the list (Table2.1 & Table2.2). The samples were then loaded onto the LC-MS/MS instrument and analyzed in multiple reaction monitoring mode (MRM) to search for ions with specific m/z ratios corresponding to the peptides present from the list ([Table2.1]); ([Table2.2]). A total of 47 peptides of H3 (including 16 modified and 8 unmodified) and H4 (including 15 modified and 8 unmodified) were searched, and 28 were detected, among them 15 were of H3, 13 were of H4. Several new modifications, which have not yet been described in the literature, have also been observed. These include H4K77 mono-methylation, and H3H56 mono-methylation. For histones H2A and H2B, MS analysis was difficult to perform due to the low ion intensities of modified peptides and the presence of numerous histone aminoacid sequence variants. Also, in case of H2A and H2B, the peptides

obtained after in-gel propionylation and trypsin digestion were either too small or too large to be identified on our mass spectrometer, hence, making it difficult for us to include them in our analysis.

Table 2.1: List of Histone H3 Peptides.

Histone H3 Peptide Sequence	Modification(s)	m/z^{+2}
Residues 3-8, TKQTAR	K4 Unmodified	380,43
Residues 9-17, KSTGGKAPR	K9,K14 Unmodified	506,58
Residues 18-26, KQLATKAAR	K18,23 Unmodified	549,65
Residues 27-40, KSAPATGGVKKPHR	K27,36,27 Unmodified	801.50
Residues 54-63, YQKSTELLIR	K56 Unmodified	653.76
Residues 64-69, KLPFQR	K64 Unmodified	422.51
Residues 73-83, EIAQDFKTDLR	K79 Unmodified	695.77
Residues 117-128, VTIMPKDIQLAR	K122 Unmodified	720,00
Residues 3-8, TKQTAR	K4 ac	372,92
Residues 3-8, TKQTAR	K4 me1	387,21
Residues 64-69, KLPFQR	K64 me1	429,52
Residues 18-26, KQLATKAAR	K18 ac and K23 ac	536,65
Residues 18-26, KQLATKAAR	K18 ac or K23 ac	542,70
Residues 27-40, KSAPATGGVKKPHR	K27 me1 and 36 me1	815,46
Residues 54-63, YQKSTELLIR	K56 me1	660,77
Residues 54-63, YQKSTELLIR	K56 ac	646,76
Residues 73-83, EIAQDFKTDLR	K79 me1	703,28
Residues 73-83, EIAQDFKTDLR	K79 ac	689,50
Residues 117-128, VTIMPKDIQLAR	M121 Ox	728,00
Residues 117-128, VTIMPKDIQLAR	K122 me1	727,00
Residues 117-128, VTIMPKDIQLAR	K122 me1 and M121 Ox	735,00
Residues 9-17, KSTGGKAPR	K9 ac or K14 ac	499,57
Residues 9-17, KSTGGKAPR	K9 me1 or K14 me1	521,09
Residues 9-17, KSTGGKAPR	K9 me1 and K14 me1	514,09

In MRM mode the samples were separated using nano LC gradient, and the ion trap was set to isolate, fragment and generate MS/MS (MS2) scans of six parental peptides having predetermined m/z ratios. Collision induced dissociation (CID) was used for internal fragmentation of the peptides, which creates "daughter" peptides for every given "parent" peptide. We have used reverse-phase (RP) chromatography, which separates peptides on the basis of their hydrophobicity. This was reflected on the retention times (RT) of the peptides, as the more hydrophobic the peptide was, the higher was the corresponding RT value. The relative quantity of each peptide in the different fractions was estimated by comparison between the peak areas in the Extracted Ion Chromatograms (EIC) for each fragment or daughter peptide obtained from Total Ion Chromatogram (TIC) in MRM analysis of these fractions. The

Table 2.2: List of Histone H4 Peptides.

Histone H4 Peptide Sequence	Modification(s)	m/z^{+2}
Residues 4145, GGVKR	K44 Unmodified	286,34
Residues 4655, ISGLIYEETR	Unmodified	590,66
Residues 20-23, KVLR	K20 Unmodified	285,86
Residues 68-78, DAVTYTEHAKR	K77 Unmodified	673,73
Residues 24-35, DNIQGITKPAIR	K31 Unmodified	691,29
Residues 56-67, GVLKVFLENVIR	K59 Unmodified	721,88
Residues 4-17, GKGGKGLGKGGAKR	K5,8,12,16 Unmodified	747,38
Residues 79-92, KTVTAMDVVYALKR	K79,91 Unmodified	854,03
Residues 68-78, DAVTYTEHAKR	K77 me1	680,74
Residues 24-35, DNIQGITKPAIR	K31 me1	698,31
Residues 20-23, KVLR	K20 me1	292,88
Residues 20-23, KVLR	K20 me3/ac	278,88
Residues 4145, GGVKR	K44 me1	293,34
Residues 4-17, GKGGKGLGKGGAKR	K5,8,12,16 ac(tetra)	719,33
Residues 4-17, GKGGKGLGKGGAKR	K5,8,12,16 ac(tri)	726,34
Residues 4-17, GKGGKGLGKGGAKR	K5,8,12,16 ac(di)	733,35
Residues 4-17, GKGGKGLGKGGAKR	K5,8,12,16 ac(mono)	740,37
Residues 4-17, GKGGKGLGKGGAKR	K5,8,12,16 me(tetra)	754,38
Residues 4-17, GKGGKGLGKGGAKR	K5,8,12,16 me(tri)	761,38
Residues 4-17, GKGGKGLGKGGAKR	K5,8,12,16 me(di)	768,38
Residues 4-17, GKGGKGLGKGGAKR	K5,8,12,16 me(mono)	775,38
Residues 56-67, GVLKVFLENVIR	K59 ac	728,88
Residues 79-92, KTVTAMDVVYALKR	K79 me1 or K91 me1	868,03

RT's of various peptides (with or without modifications) were highly reproducible between different samples. The RT of the parent peptide along with the alignment of EICs of a given peptide at the same RT helped us to identify the correct peptide [Figure 2.1]. For a given peptide, the sum of the intensities of its daughter peptides, observed in the MS2 spectrum was taken to calculate the intensity of the peptide in a given sample [Chapter 7]. Agilent's "Data Analysis" software was used to visualize the chromatograms from each MRM and MS2 spectrum at each RT point of the corresponding chromatogram.

2.1.2 Quantitative Analysis of Histone PTMs in Fanconi Anemia

After learning how to detect various histone modifications with our LC-MS/MS instrument, we proceeded to analyze the dynamics of these modifications in response to DNA-ICL damage. For quantitative comparison of various histone PTMs, we used an established approach in quantitative proteomics, termed SILAC. This method involves growing two populations

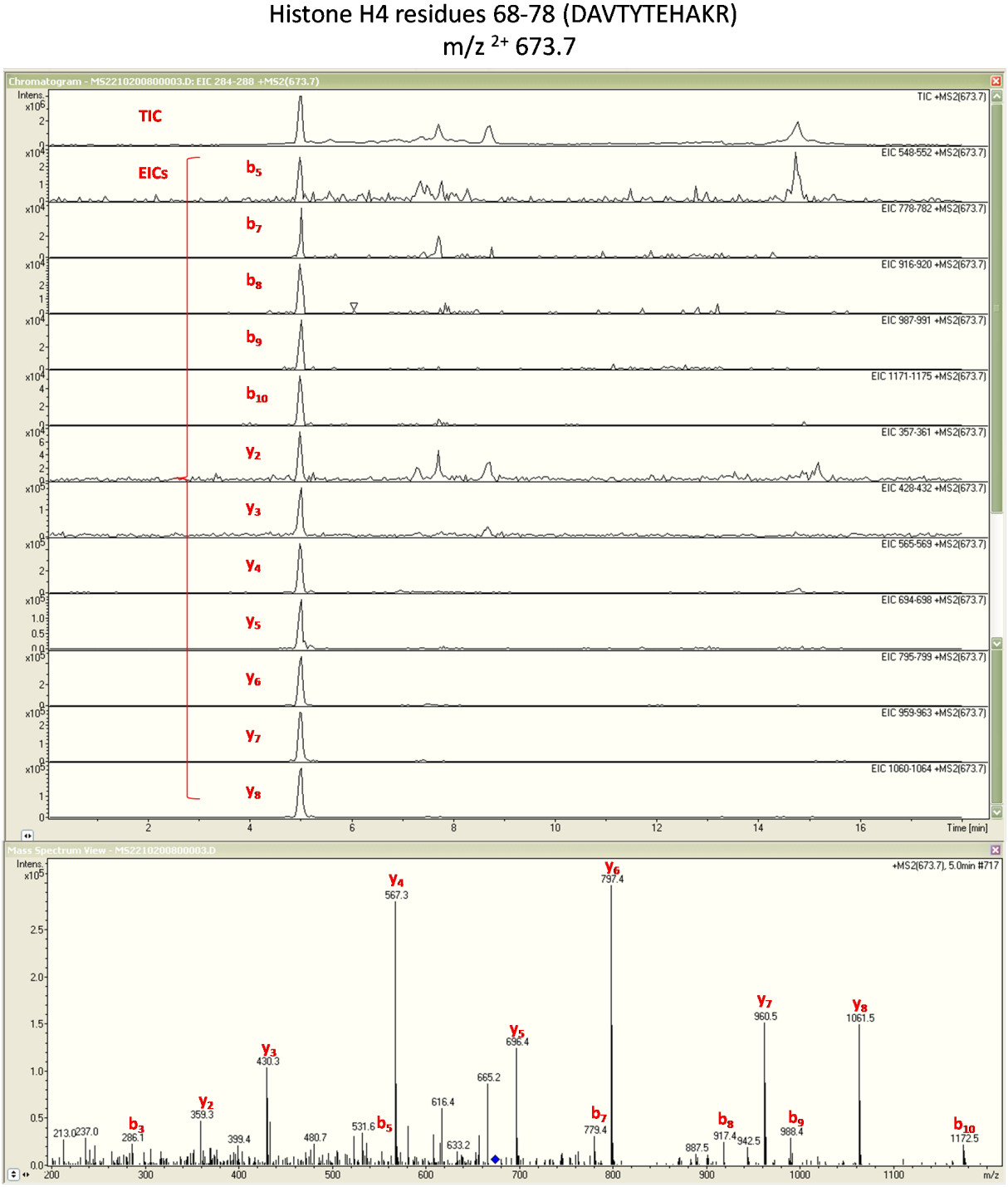


Figure 2.1: Representative TIC of Histone H4 residues 68-78 (K77UM) along with EICs of daughter peptides and MS/MS spectrum.

of cells, one in a medium that contains "light" (normal) amino acids and the other in a medium that contains "heavy" (stable isotope labeled) version of the same amino acids. The heavy amino acids contain C^{13} instead of C^{12}, and/or N^{15} instead of N^{14}. Incorporation of the heavy amino acid into a protein leads to a detectable mass shift in its tryptic peptides as compared to the peptides with the light version of the same amino acid (for example, 6 Da in the case of ${}_6C^{13}$-Lys), but no substantial physico-chemical changes are expected to occur. Usually, Lysine and Arginine are chosen to be labeled by stable isotopes, because in most of the cases tryptic peptides contain only one of these amino acids located at the C-terminus, which significantly facilitates the calculation of mass shifts. Cells are cultured in SILAC light and heavy medium for at least five cell doublings to allow the incorporation of these amino acid. After five cell doublings, 97% of the proteins were expected to be present in the light and heavy state in the corresponding cell population [Ong and Mann, 2006]. For many proteins, such level of incorporation is achieved more rapidly because of contribution of protein turnover.

The cells from light and heavy samples were mixed in equal ratio (1:1) and from this time on were processed in the same manner during the proteome fractionation steps and preparation for MS and mass spectrometry analysis. Because of being treated in the same mixture, the presence of heavy and light versions of peptides originating from the two samples are not subject to variations between runs. This gave us more confidence in interpreting the differences in their abundance, measured by LC-MS/MS, as reflecting the true differences between the samples compared. Thus, SILAC is generally expected to be more accurate in the quantitative comparison of two proteomes due to the presence of internal standard. H/L ratios are then calculated either manually by comparing the intensities of MS2 fragments of the two versions of the same peptide [Figure 2.2] or by using specialized softwares (e.g. Spectrum Mill), which automatically gives the H/L comparison of various proteins identified in the two conditions. The representative MS/MS spectra of various modified and unmodified histone peptides along with the lists of theoretical masses of fragment peptides are presented in the Annex - I at the end [Chapter 7].

In our preliminary experiments, the sufficient levels of incorporation efficiency after five cell doublings was confirmed when we detected less than 5% of "light" version of a histone peptide in histones prepared from "heavy" sample. Also, comparison of the cells grown in either light or heavy SILAC medium did not show any substantial differences in growth rate, viability and overall cell morphology.

[Figure 2.3] shows the scheme of SILAC experiment that we have performed. We used Fanconi anemia cell lines FA-C (pd331 + empty vector) and FA-C Corr (pd331 + FANCC). FA-C cells are the fibroblasts obtained from a Fanconi anemia patient having a mutation in "FANCC" gene, stably transfected with an empty vector, whereas the control cells, FA-C corr, are the same cells stably transfected with functional FANCC gene. Both of these cell lines were separately grown, for at least five cell doublings, in SILAC medium containing the heavy isotope versions of lysine (${}_6C^{13}$) and arginine (${}_6C^{13}$ and ${}_4N^{15}$). Cells were treated with 8-MOP+UVA, harvested after 1h and 6h and mixed in equal ratios (1:1) with the control cells (untreated FA-C & FA-C corr), which were grown in SILAC medium with a light

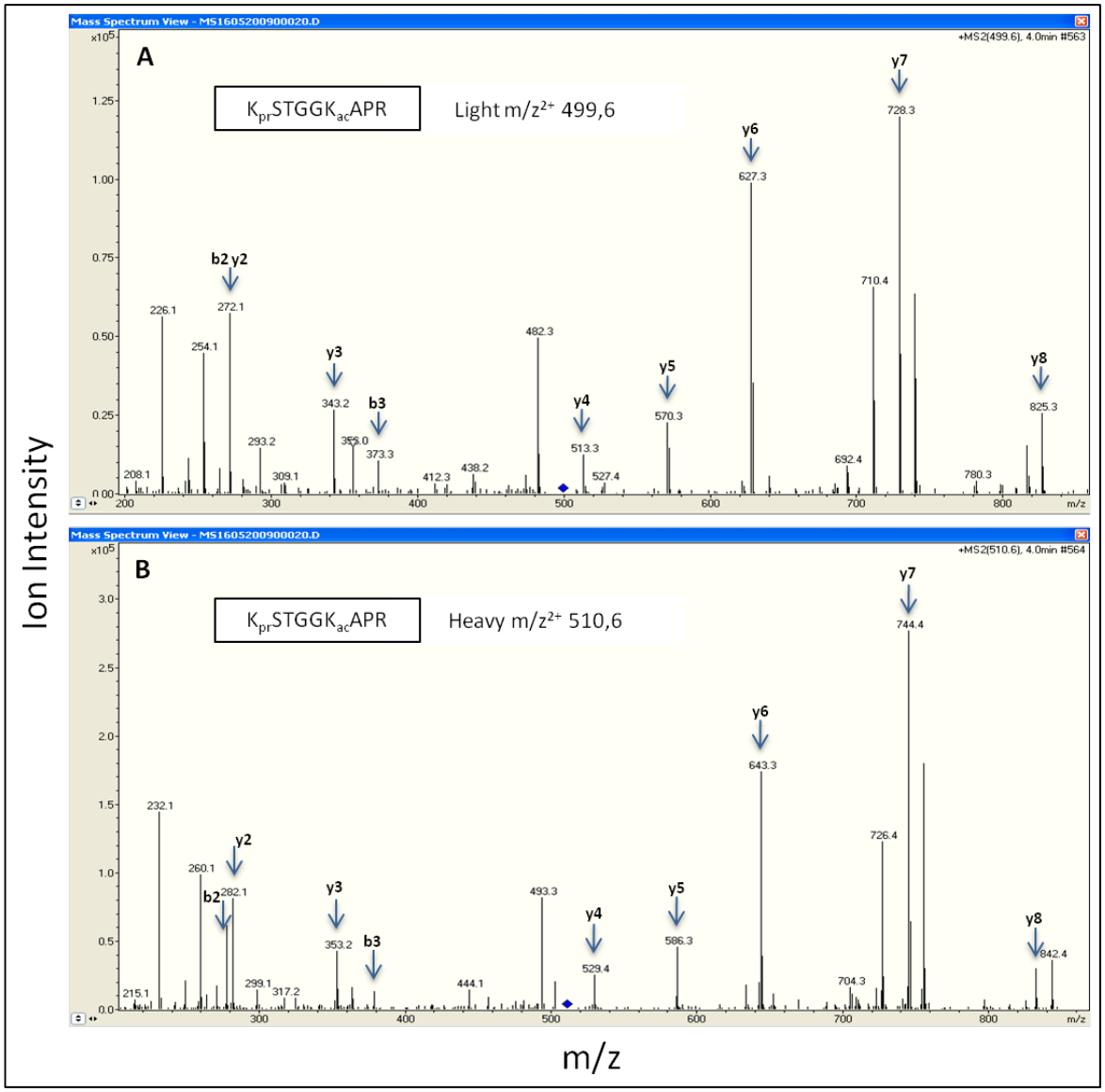


Figure 2.2: Representative "Light" (A) and "Heavy" (B) MS2 Spectrum of Histone H3 peptide 9-17 (KSTGGKAPR).

isotope version of these amino acids ($_6C^{12}$ for Lys and $_6C^{12}$, $_4N^{14}$ for Arg).

The histones were extracted from these samples with the acid extraction protocol. The purified histones were quantified and loaded onto 15% SDS-PAGE gel electrophoresis as shown in [Figure 2.4]. As the four core histones were well separated by SDS-PAGE, the bands corresponding to histones H3 and H4 (each band now containing light and heavy versions of the same protein) were excised and propionylated using 75% propionic anhydride in methanol. After several washings, the bands were subjected to overnight in-gel trypsin digestion. Later, the digested peptides were analyzed by nano-LC-MS/MS [Chapter6].

As explained previously, we limited our histone PTM analysis to only H3 and H4 histones.

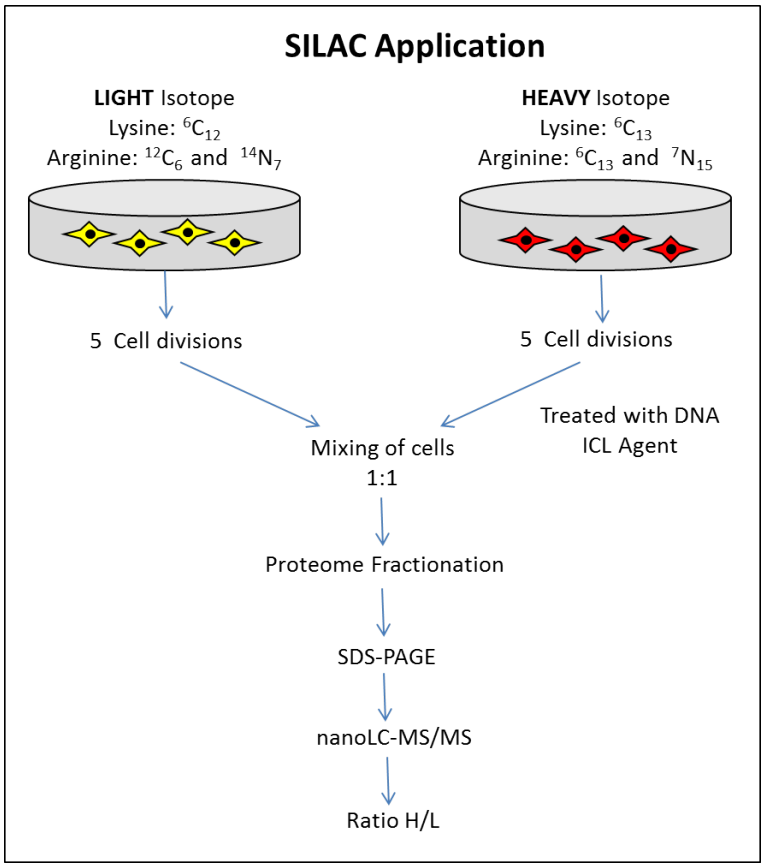


Figure 2.3: SILAC schematics and workflow diagram.

A total of 94 different peptides (both light and heavy versions) from core histones H3 and H4 were analyzed by mass spectrometer in MRM mode. The heavy m/z ratios for the heavy version of every peptide were calculated individually by taking into account the number of lysine and arginine residues in that peptide. A mass of 6 Da for each lysine and 10 Da for each arginine was added into the "heavy" version of the peptide. The H/L ratios of relative intensities of each peptide were calculated by taking the sum of the intensities of corresponding light and heavy daughter fragments. For normalization, the H/L ratios calculated for each modified peptide was divided by the H/L ratios calculated for the unmodified version of the same peptide from the same sample. These values (normalized H/L ratios) are shown as graphs in the next section. In these graphs, light (L) represents the untreated cells and heavy (H) represents cells harvested either 1h or 6h after DNA-ICL induction. We term the change in modification status after 1h as "early response" and 6h as "late response" to DNA-ICL induction. We show here only those modifications that manifest an interesting dynamics in our analysis.

I **Histone H3 PTM Profile** In case of histone H3, we have identified mono-methylations on three different lysine residues (H3K9, H3K56 and H3K79), showing different responses to DNA-ICL damage in Fanconi anemia corrected and mutated cells. Among

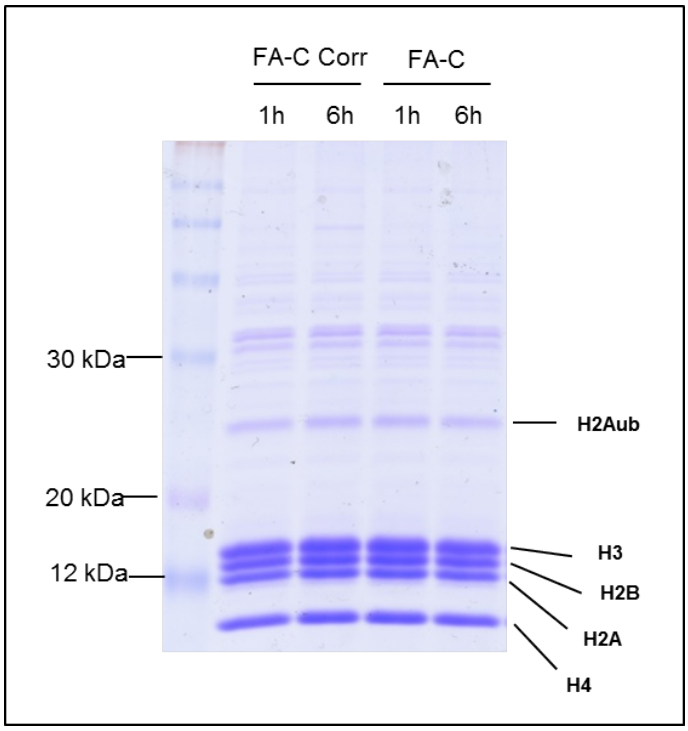


Figure 2.4: 15% SDS-PAGE of FA-C corr and FA-C SILAC Light and Heavy samples, where FA-C corr stands for corrected cells and FA-C stands for mutated cells

acetylations, we were able to detect only H3K14ac in all the samples. Figure 2.5 shows these responses.

- **H3K79me**: Among various histone H3 PTMs identified the strongest response to DNA-ICL damage was shown by mono-methylation on H3K79. In FA-C corr cells, a substantial increase was observed in H3K79me after 1h of damage induction which subsided at 6h time point. In sharp contrast, no such increase was observed in FA-C mutated cells. The absence of response of H3K79me to DNA-ICL damage in mutated cell line suggests that this modification is dependent on functional FA pathway [Figure 2.5-A].
- **H3K9me**: In case of H3K9me, a similar response profile was observed in both corrected and mutated cells. The most prominent increase is seen after 6h, whereas it is barely detectable after 1h of DNA damage induction, thus the response is qualified as 'late'. However, the intensity of the change is detectably lower in the mutant cells, indicating that the functional FA-pathway is involved in the appearance of this modification after DNA-ICL damage [Figure 2.5-B].
- **H3K56me**: H3K56me has also been detected in our screen. This modification is notably decreased in FA-C mutated cells at 1h after induction of DNA-ICL damage, whereas no decrease was observed in the case of control FA-C corr cells. After 6h of treatment, the same profile was observed for both cell lines, which qualifies this modification as an 'early responder' to DNA-ICL damage. Importantly,

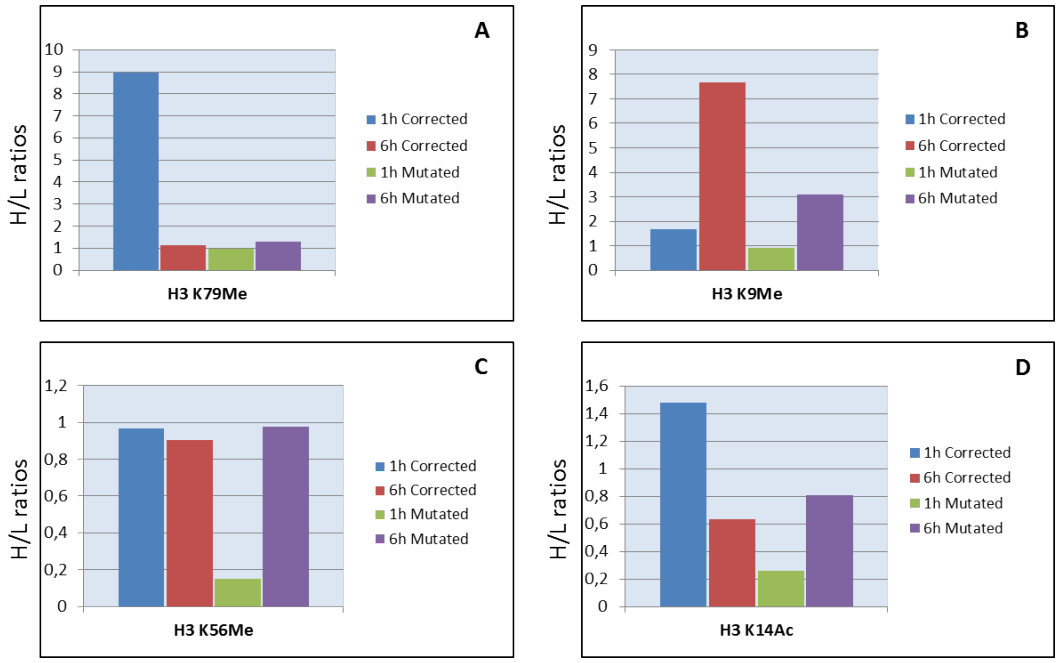


Figure 2.5: Histone H3 PTM profile after DNA-ICL Induction. The H/L ratios represent single experiment where L represents the untreated cells and H represents cells harvested either 1h or 6h after DNA-ICL induction.

however, this response happens in the absence of the functional FA-pathway [Figure 2.5-C].

- **H3K14ac**: We also observed differential acetylation of lysine 14 of histone H3 in Fanconi anemia corrected and mutated cells in response to DNA damage. While comparing the FA-C corr cells treated and untreated, an initial increase after 1h of treatment was observed, which subsequently decreases after 6 hours of treatment. A reverse kinetic was observed in case of FA-C mutated cells, where an initial decrease in H3K14ac after 1h of DNA-ICL induction marked the early response, which became less prominent after 6 hours [Figure 2.5-D].

II Histone H4 PTM Profile

- **H4K(5,8,12,16)ac**: The analysis of core histone H4 revealed an increased acetylation response on the N-terminal tail residues K5, 8, 12, and 16 after 1h of treatment in case of corrected cells. Importantly, no such response was observed in case of mutated cells, suggesting that functional FA pathway is required for this modification to occur in response to ICL damage. Evidently, a slight decrease in H4 N-terminal acetylation was observed in mutant cells after 1h of DNA-ICL damage. The response after 6h treatment in both cases was more similar as a significant decrease in H4 acetylation status was observed in both cell lines, suggesting that the initially increased acetylation response was transient in nature [Figure 2.6-A].

- **H4K77me**: Similarly, a slight increase in mono-methylation was observed on H4K77 1h after damage induction in corrected cells, while no such response was observed in mutated cells [Figure 2.6-B]. Interestingly, an increase was also observed in FA-C mutated cells, however after 6h, instead of 1h. This profile suggests that in the absence of functional FA pathway, the response is delayed.
- **H4K31me**: Other noteworthy modification that was detected in histone H4 included H4K31 mono-methylation. No difference was observed in mutated and corrected cells after 1h suggesting that it was not involved in the early cellular response to DNA-ICL damage. However, a slight decrease in this modification was observed in FA-C mutated cells after 6h when compared to the response in FA-C corr cells. This may indicate that mono-methylation of H4K31 is required in late response to DNA-ICL damage, however, due to the absence of functional FA pathway in mutated cells, this modification is decreased, which qualifies this modification as an example of a histone PTM that is affected by the absence of functional FA proteins [Figure 2.6-C].

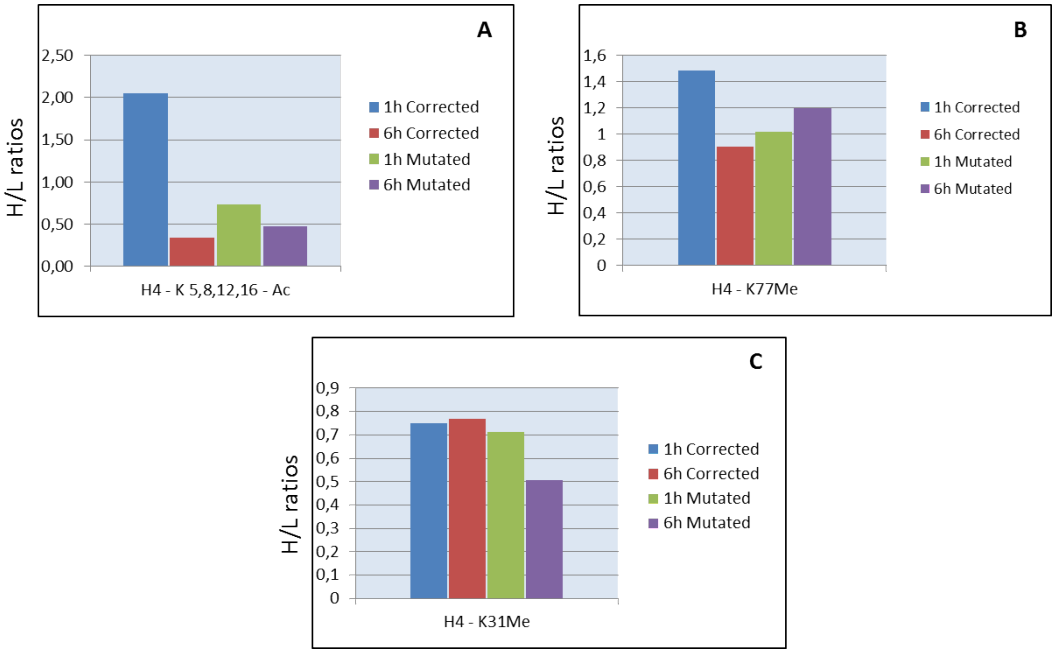


Figure 2.6: Histone H4 PTM profile after DNA-ICL Induction. The H/L ratios represent single experiment where L represents the untreated cells and H represents cells harvested either 1h or 6h after DNA-ICL induction.

2.2 Discussion

In this part of our work we performed analysis of "histone code" part of cellular response to DNA-ICL damage. We have chosen Fanconi anemia as a model, as the cells derived from

FA patients are hypersensitive to DNA-ICL damage, which manifests their inability to adequately respond to such kind of DNA lesions. Overall, we have identified several acetylations and mono-methylations on both histone H3 and H4, which showed dependency on functional FA pathway.

We have observed an FA-dependent early increase in acetylation of the N-tails of histones H3 and H4 ((K14 in H3 & K5, 8, 12, 16 in H4). As shown earlier, these residues play an important role in chromatin dynamics during DDR especially in the case of DSBs, as mutations in H4 N-terminal tail confers sensitivity to ionizing radiation [Tamburini and Tyler, 2005]; [Murr et al., 2006]. Moreover, a transient acetylation was observed at these residues at the sites of DSBs and it has also been shown that this acetylation is mediated by Tip60 histone acetyltransferase complex [Murr et al., 2006]. Our results appear consistent with these studies, as we also observe an early increase in acetylation of H4 N-terminal tail. However, in our model, this acetylation depends on functional FA pathway, as no such increase was observed in case of FA mutant cells. If we assume that these modifications are similarly due to Tip60 complex, the most natural next assumption is that FA proteins function upstream to Tip60 complex in this response pathway. The mutation in FANCC, which we used, does not allow assembly of the FA core complex. The established function of core complex is to mono-ubiquitinate the FA-ID complex, through the action of FANCL, an E3 ubiquitin ligase. However, it is possible that the FA core complex has other substrates too. We can speculate that it mono-ubiquitinates Tip60, leading to its activation or physical recruitment to the DNA-ICL lesions. On the other hand, since the levels of histone acetylation are influenced by both histone acetyltransferases (HATs) and histone deacetylases (HDACs), we can expect that the increase in acetylation might also be due to decreased activity of HDACs, which may be directly or indirectly influenced by functional FA proteins.

In addition to Tip60 HAT, other HATs might also be responsible for this pattern of H4 acetylations as well. These include MOF (MYST1/KAT8), HBO1 (MYST2/KAT7), CBP (KAT3A), and EP300 (KAT3B) [Kouzarides, 2007b]. It must be worth looking for their possible interactions with FA-proteins as they might be regulated or affected by the mutation in FA proteins or alternatively, they may play a role in DDR to DNA-ICL damage.

In any case, it appears that that the acetylation of N-tails of H4 and H3 is a generic response to many kinds of DNA damage. Although we have observed it in response to DNA-ICL damage, other studies have detected these modifications in the case of DSBs, a different kind of damage. One simple explanation of why this acetylation is important is the well established fact that acetylation of histone H4 N-terminal tails leads to an open conformation of chromatin [Kouzarides, 2007a], which might be a prerequisite for any kind of repair activity. Apart from making chromatin more accessible for the repair machinery, acetylation on these residues can also participate in sensing DNA damage. For example, histone acetyltransferase hMOF was shown to function upstream of ATM in sensing DSB, via modification of histone H4K16 [Gupta et al., 2005].

Among methylations, histone H3K79 methylation and H4K20 methylation have shown interesting dynamics in DDR to DNA-ICL. Although H3K79 methylation represents a con-

stitutive mark, and so far has not been shown to be induced by DNA damage, our results surprisingly showed a substantial increase in this modification after induction of DNA-ICL. Moreover, this increase was only observed in the wild type cells and early after the damage, strongly suggesting that this modification requires functional FA proteins or at least a functional FA-core complex formation. Study of the mechanistic role of this modification in DDR has shown that it provides a docking site for 53BP1, which recognizes this modification through its tandem tudor domains [Huyen et al., 2004]; [Wysocki et al., 2005]. 53BP1 interacts with p53 and is an important checkpoint protein that is recruited to the DNA DSBs in minutes after exposure to ionizing radiation and colocalizes with γH2AX. The lack of consistency of our results with the current state of knowledge regarding this modification could reflect a differential DDR pathway in case of DNA-ICL, as H3K79Me has been so far studied only in the context of DNA DSBs, which represent a different kind of lesion. This, H3K79Me might be one of the modifications that distinguishes the cellular response to DNA-ICL from that to DSBs and potentially other kind of lesions.

A relatively uncharacterized modifications H3K56me and H4K77me were also identified in our screen as showing differential response to DNA-ICL damage in wild type and mutant cell lines. These two residues, H3K56 and H4K77, are located in the core domains of H3 and H4, respectively, and participate in DNA-histone contacts. It is possible that any modification of these residues would affect the nucleosomal structure either positively or negatively. The striking early decrease in H3K56 mono-methylation in mutant cells [Figure 2.5-C] could reflect either an active demethylase activity or eviction of H3K56me carrying histone H3, however, the later is less probable due to the early nature of the response. The absence of such response in wild type cells suggests that the demethylase activity is controlled either directly or indirectly by FA-pathway. In case of H4K77me, the detectable increase in wild type cells after 1h of DNA-ICL induction and its delayed response in mutants suggests that it could provide a docking site for a downstream effector protein, which works in DNA damage signaling in the manner similar to H3K79me-53BP1. Therefore, it would be intriguing to search for its binding partners in DNA damage response.

Mono-methylation on histone H3K9 is a mark associated with euchromatin. Our results showed a remarkable increase in this modification in wild type cells but only after 6h of DNA-ICL induction. An attenuated response was also seen in FA mutant cells after 6h. This late responding modification might signal the end of DNA repair in wild type cells and due to either delayed repair or persistent damage in FA mutant cells, this late response was not as significant as in wild type cells. Another mono-methylation, on H4K31, also showed a delayed decrease in FA mutant cells while no change was observed in wild type cells at the two time points. This decrease in methylation, as discussed earlier may be due to the activity of demethylases or due to the removal of histone H4 carrying this modification.

Overall, our results showed a number of histone PTMs that responded dynamically to the DNA-ICL damage, often in FA-dependent manner. We also were able to classify these modifications as early responders and late responders. This dynamic behavior of histone PTMs lead us to propose the following model of 'histone code' part in DDR to DNA ICL. We hypothesize that immediately upon induction of DNA damage, chromatin is locally destabi-

lized to facilitate the access to DNA damage sensing and repair machinery. The PTMs most likely responsible for this part are acetylations of histones H3 and H4 N-terminal tails, as these modifications are believed to loosen up histone-DNA interactions and hence create a relaxed nucleosome. This relaxation leads to additional modifications, now on the histone core domain residues, such as H3K79 and H4K77, which are expected to become more accessible, both to the modifying enzymes and to the downstream effector proteins. The latter specifically recognize these marks and bring about important changes in recruitment of other proteins and checkpoint activation. H3K79me-53BP1, for instance, is an example of such binding event. The late response modifications, on the other hand, most likely signal the end of repair. In our system, the H3K9me is increased late after the DNA damage induction and may signal an increase in transcriptional activity. The proposed interpretation of the histone PTM changes is consistent with the general Access-Repair-Restore model for DNA damage and repair activities [Section1.5.2], supporting the notion that, similar to other kinds of DNA damage, the efficient execution of cellular response to DNA ICL requires changes in chromatin structure.

3

Results - II

In the previous chapter we have analyzed the dynamics of various histone post-translational modifications in response to induction of DNA-ICL lesions and observed that some of these PTMs interestingly depend on the functional Fanconi Anemia pathway. There are many different enzymatic activities in the cell responsible for histone modifications, thus in order to obtain further insight into the role of histone PTMs in Fanconi Anemia, we resorted to functional genomics approach. We analyzed the transcriptome of normal and FA-defective cells after induction of DNA-ICL damage. One goal of such an approach was to monitor the changes in expression of known histone modifying enzymes, in order to find genes that best correlate in their expression profiles with the changes in histone PTMs that we have observed. Given that our analysis was focused on histone acetylations and methylations, this part of analysis was mostly focused on enzymes responsible for these modifications or for their removal.

However, our functional genomics analysis had a second goal, extending it beyond the monitoring of histone modifying enzymes. Regardless of the implications for histone code, we also decided to study the general differences between FA mutant and corrected cell lines at transcriptional level and the transcriptional response to DNA-ICL damage in these cells. One of the goals was to understand what part of the cellular response to DNA-ICL is dependent on functional FA pathway. The second goal was to understand what else can happen to the cells after DNA-ICL damage if the FA pathway is not working. For this purpose, the functional genomics experiments were designed in a way that allowed to apply the so called interaction analysis to the data obtained [Section 3.3.3]

In this global transcriptome analysis, in addition to the transcriptional response to DNA-ICLs, we also decided to include a different kind of DNA damage, i.e., ionizing radiation

induced damage, which generates mostly double DNA strand breaks (DSB). As FA mutant cells are highly sensitive to DNA-ICL damage, the goal of this comparative analysis was to understand what part of the transcriptional response is specific for DNA-ICL damage, and what part of the same response could reflect a more general cellular response to any kind of DNA damage.

We also have to admit two caveats in our analysis. As described earlier, the DNA damage response to DNA-ICLs depends upon the functional FA pathway, which is based upon the concerted activities of at least 13 different FA proteins. A mutation in any one of these proteins leads to inactivation of the whole FA pathway and thus affects the DDR to DNA-ICL damage. Our histone PTM study [Chapter 2] was performed using FA-C mutated cells, whereas our transcriptome analysis was done with FA-D2 mutation. We would like to point out that although FANCC, a core complex protein, functions upstream of FANCD2 protein, and is essential for the activation of FANCD2, the mutation in either of them leads to the similar phenotypic changes in general, i.e. inactivation of FA-pathway. We acknowledge, however, that the results of our transcription data may not be reflected directly on the histone code changes, as the experiments were performed in two different cell lines, representing two different stages of FA-pathway.

The second caveat in our analysis is due to a different ways of comparison between the samples that we have used in our SILAC-based analysis of histone PTMs and in our transcriptome analysis. Because there was no direct comparison between the reference and every sample in our histone PTM analysis, this somewhat limited the possibility to make direct comparisons, especially for the early response in the FA mutant cell lines. Nevertheless, even with these limitations in mind, we still hope that our analysis allowed us to rule out some explanations of the observed changes in histone PTM patterns and thus to narrow down the number of potential mechanisms of the cellular response to DNA-ICL.

3.1 Experimental Setup

In order to setup our experiments, FA-D2 mutated and corrected cells were incubated with 8-MOP at 37°C for 20 minutes and then irradiated with a suitable dose of UVA (10 KJ/m^2). In order to study the kinetics of DNA damage response (DDR) in these cells after DNA-ICL induction, two time points were chosen, i.e. 1h and 6h. For ionizing radiation, which induces double stand breaks on DNA (DSB), we used Cesium based irradiator. Cells were irradiated at 10 Gy and later incubated for 1h and 6h as for DNA-ICL induction. These two time points allowed us to quantify and classify changes in gene expression in FA-D2 corrected and mutated cells as early and late responses to DNA damage induction. [Figure 3.1] describes the schematics of microarray experiment.

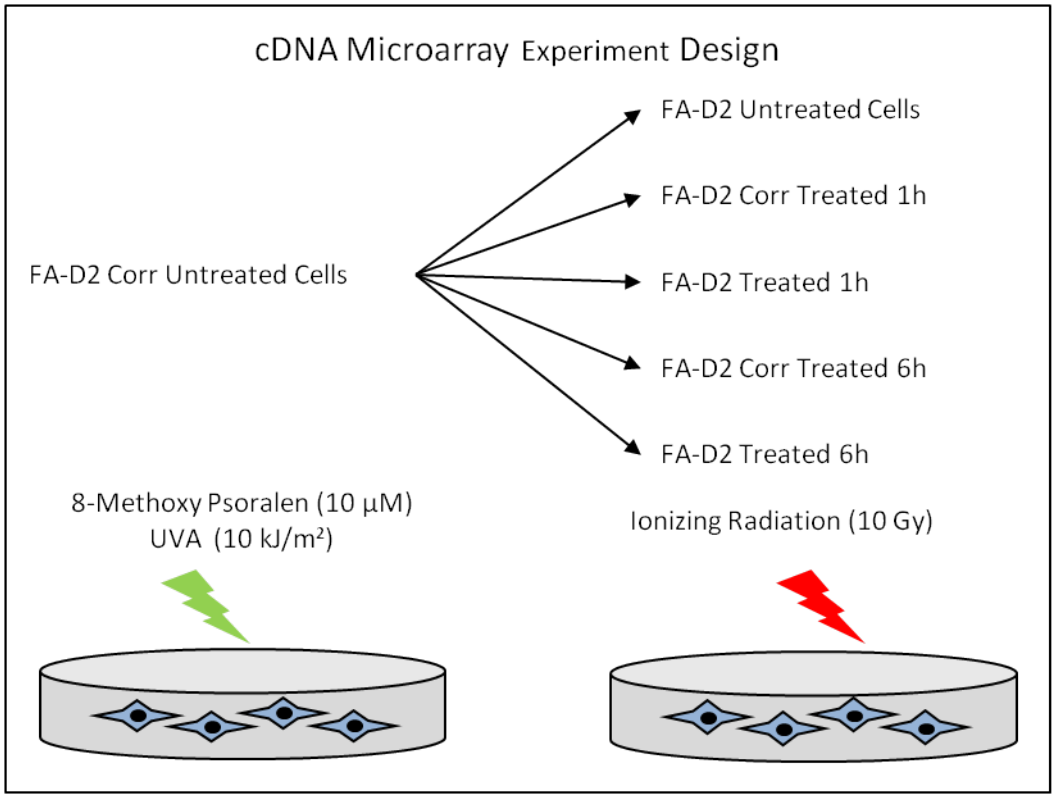


Figure 3.1: cDNA microarray experiment design.

In microarray experiments FA-D2 corrected untreated cells were taken as reference for all the samples. RNAs were prepared from different samples using Trizol reagent and RNeasy columns as described in [Chapter 6].

3.2 Transcriptome Analysis of Histone Modifying Enzymes in Fanconi Anemia Model

The first goal of the functional genomics analysis was to obtain clues on what histone modifying enzymes could be responsible for the patterns of modifications that we have observed in our PTM analysis. Accordingly, we specifically searched for the histone modifying enzymes, aiming to identify those genes that show changes in their levels that correlate and thus could explain the changes in particular histone modification. Here, we focused our analysis towards histone acetyltransferases (HATs), histone deacetylases (HDACs), histone methyltransferases (HKMTs) and histone demethylases (HKDMs), as we detected and analyzed mainly histone acetylations and methylations in proteome analysis part. In general, the gene expression profile analysis of these enzymes showed notable differences between Fanconi anemia mutated and corrected cell lines, which indicates that the expression of some of these enzymes is directly or indirectly affected by the mutation in Fanconi Anemia genes.

We again have to admit a design flaw in our experiment, as the samples were not compared in exactly the same way as in the previous section, which somewhat limited our possibility to make direct comparisons, especially for the early response in the FA mutant cell lines. The second interpretational caveat, common for all functional genomics data, is that the protein activity can be modulated in different ways independent on total protein/RNA levels. Nevertheless, we still hope that this analysis allowed us to rule out some simple mechanisms and thus narrow down the search.

3.2.1 Histone Acetyltransferases (HATs)

We analyzed the expression profile of 10 histone acetyltransferases, including the GNAT-family (GCN5 & PCAF), MYST-family (Tip60, HBO1, MOF, MOZ, MORF) and others including p300, CBP and HAT1.

We have noticed that, among these HATs, the changes in expression of PCAF and MORF acetyltransferases appear to match most closely the changes in acetylation of the H4 terminal tail [Figure 2.6-A]. Indeed, both of these HATs were up-regulated 1h after MOP/UVA treatment, however, the levels go down at 6h time point. In the mutant cell lines, the transcriptional response is also less pronounced 6h after treatment as compared to the 1h time point [Figure 3.2]. Note that we cannot comment directly on the levels of 1h response in mutant cells, as we did not compare mutant and wild type cells in our SILAC experiments. With this caveat in mind, both of these HATs could be considered as plausible candidates responsible for the observed dynamics of H4 acetylation. However, taking into account that PCAF does not acetylate H4 histone but is responsible for H3 acetylation, the simplest interpretation of the observed changes in H4 acetylation is that they are due to the changes in the levels of MORF histone acetyltransferase (see however, next section about a potential role of histone deacetylases).

Unlike H4 acetylation, the profile of H3 acetylation [Figure 2.5-D] was not possible to explain by the changes in expression of one HAT only. Given their specificity towards H3K14 residue, GCN5 and PCAF histone acetyltransferases appear as the most plausible candidates for the H3 acetylation. Moreover, the levels of PCAF mirror the dynamics of the H3K14ac 1h and 6h after treatment. However, in the mutant cells, the levels of H3K14ac are higher 6h after damage as compared to 1h, whereas the levels of PCAF go down from 1h to 6h after treatment. Thus, additional factors should be considered in explanation of this pattern (see the HDAC section).

3.2.2 Histone Deacetylases (HDACs)

As discussed earlier, the function of HDACs is opposite to that of HATs in that they are responsible for the removal of acetyl groups from the lysines of target proteins. Therefore, their dynamics can also contribute to the explanation of observed acetylation patterns, if it

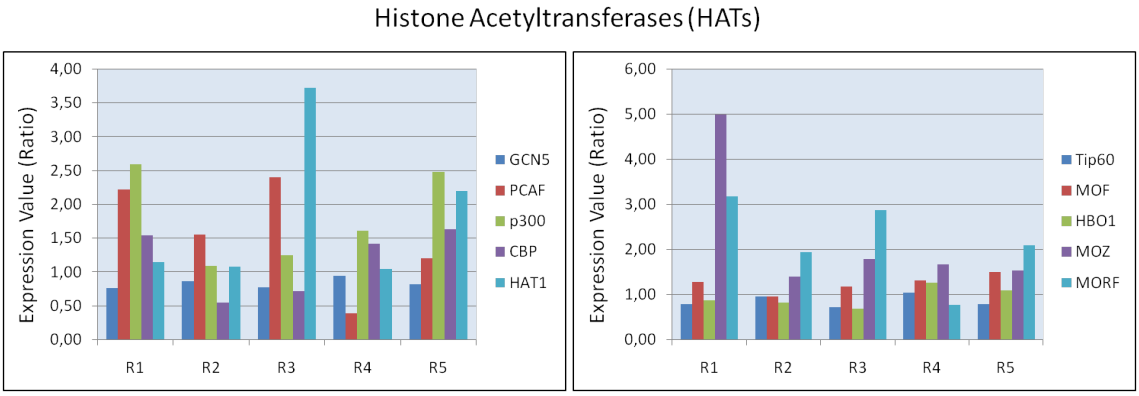


Figure 3.2: Gene expression profile of HATs in MOP/UVA treated FA-D2 corrected and mutant cells, where, R1 is an array Mutated Untreated/Corrected Untreated, R2 is Corrected 1h Treated/Corrected Untreated, R3 is Mutated 1h Treated/Corrected Untreated, R4 is Corrected 6h Treated/Corrected Untreated, R5 is Mutated 6h Treated/Corrected Untreated. The bars represent ratios of expression values of corresponding genes from single microarray.

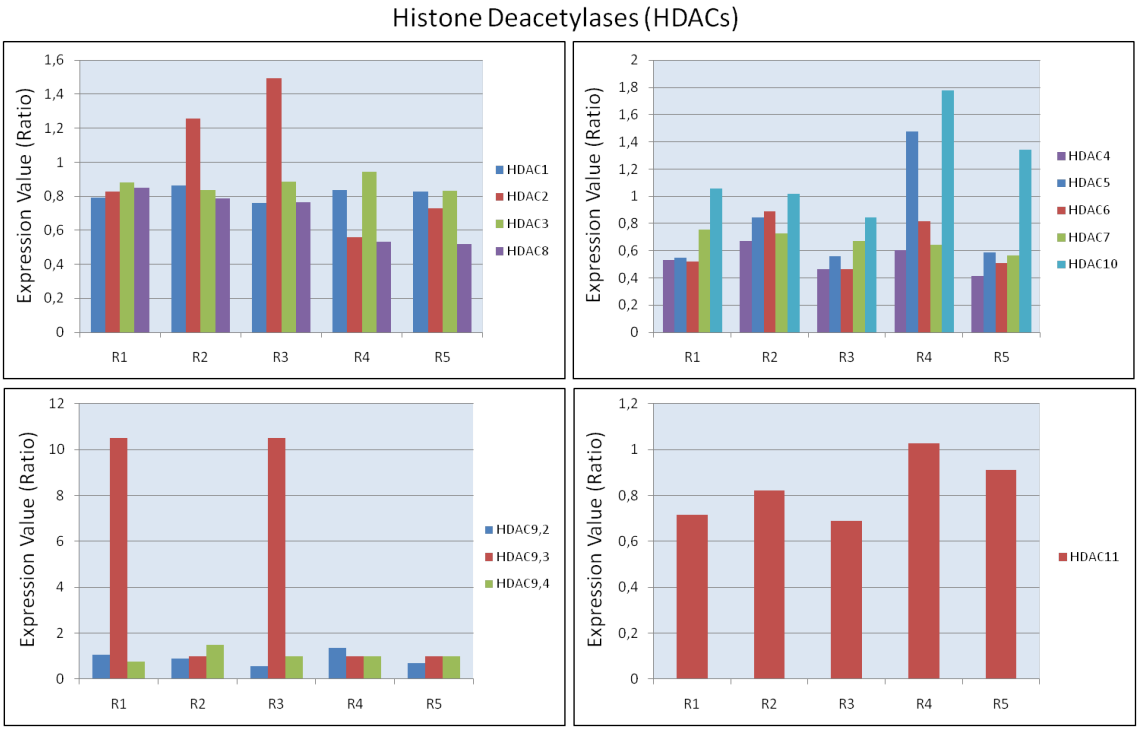


Figure 3.3: Gene expression profile of HDACs in MOP/UVA treated FA-D2 corrected and mutant cells. The bars represent ratios of expression values of corresponding genes from single microarray.

anti-correlates with the acetylation levels. Accordingly, we analyzed expression levels of 11 HDACs in our experiment.

Concerning H4 acetylation, the pattern of expression of HDAC10 and HDAC11 could be the best candidate to explain the dynamics of acetylation of H4 histone, as their levels of expression are higher 6h after treatment than 1h after treatment. However, HDAC10 could not be responsible for the early increase in H4 acetylation, because its levels are not changed 1h after treatment [Figure 3.3]. On the other hand, the levels of HDAC11 do not change as dramatically as the acetylation levels. Therefore, we still believe that the simplest explanation of the observed changes in H4 acetylation is due to the changes in the levels of MORF HAT.

Concerning H3 acetylation, no HDAC gene appears to be solely responsible for the observed dynamics of this modification. However, the HDAC2 and HDAC6 show a significant decrease in expression in the mutant cells 6h after treatment as compared to 1h time point [Figure 3.3]. Assuming that PCAF is responsible for the H3 acetylation pattern, the behavior of these deacetylases could be responsible for the increase in H3 acetylation 6h after treatment in mutant cells, which could not be explained solely via the changes in PCAF levels.

Overall, we have noticed that HDAC genes manifested less varied response to DNA damage induction and FA mutation, as compared to HAT genes. One strong exception was HDAC9 (transcriptional variant 3), that showed a dramatic 10 fold increase in expression in the FA defective cell line [Figure 3.3]. Also notably, the expression levels of this gene went down 6h after MOP/UVA treatment. Therefore, the changes in the levels of this protein could also explain the observed increase in H3 acetylation in mutant cells 6h after treatment, if taken together with PCAF dynamics. One have to keep in mind, however, that this particular variant of HDAC9 is catalytically inactive [Petrie et al., 2003].

3.2.3 Histone Methylation

We next applied similar logic to the analysis of 23 histone methyltransferases and 15 histone demethylases. Histone lysine methyltransferases are much more specific than histone acetyltransferases, as for example, histone methyltransferases SUV39H1/2, SETDB1/2, EHMT1/2 and PRDM2 are specific for methylation on histone H3K9. Likewise, MLL, SET1A/B, SMYD1-/3, SET7/9 are specific for methylation on histone H3K4. Most histone KMTs have characteristic SET (Su(Var)3-9, Enhancer of Zest, Trithorax) domain that methylates specific residues on histones and non-histone proteins. The only exception to this rule is H3K79 methyltransferase, DOT1L, which does not have SET domain [Feng et al., 2002]; [Martin and Zhang, 2005]. We grouped and analyzed the gene expression profile of various histone KMTs according to their specificity of residues on histones [Figure 1.10].

Unlike in the case of histone acetylation and despite the greater variety of analyzed genes and a larger number of modified residues with different profiles, it was not easy to find a single methyl-transferase and/or demethylase gene that could explain the observed profile for any of the H3 or H4 methylations. The closest match was the expression of PRDM2 and EHMT2 and the profile of H3K9 methylation [Figure 3.4]. Indeed, both the levels of these

enzymes and the histone modification levels increase late after DNA damage in the wild type cells, and they are also higher 6h than 1h after damage in the FA mutants. The second case was the profile of H4K31 methylation and the expression of KDM1B demethylase that was sharply up-regulated late after treatment in mutant cell lines, which could explain the decrease in the histone modification in the same sample [Figure 3.6].

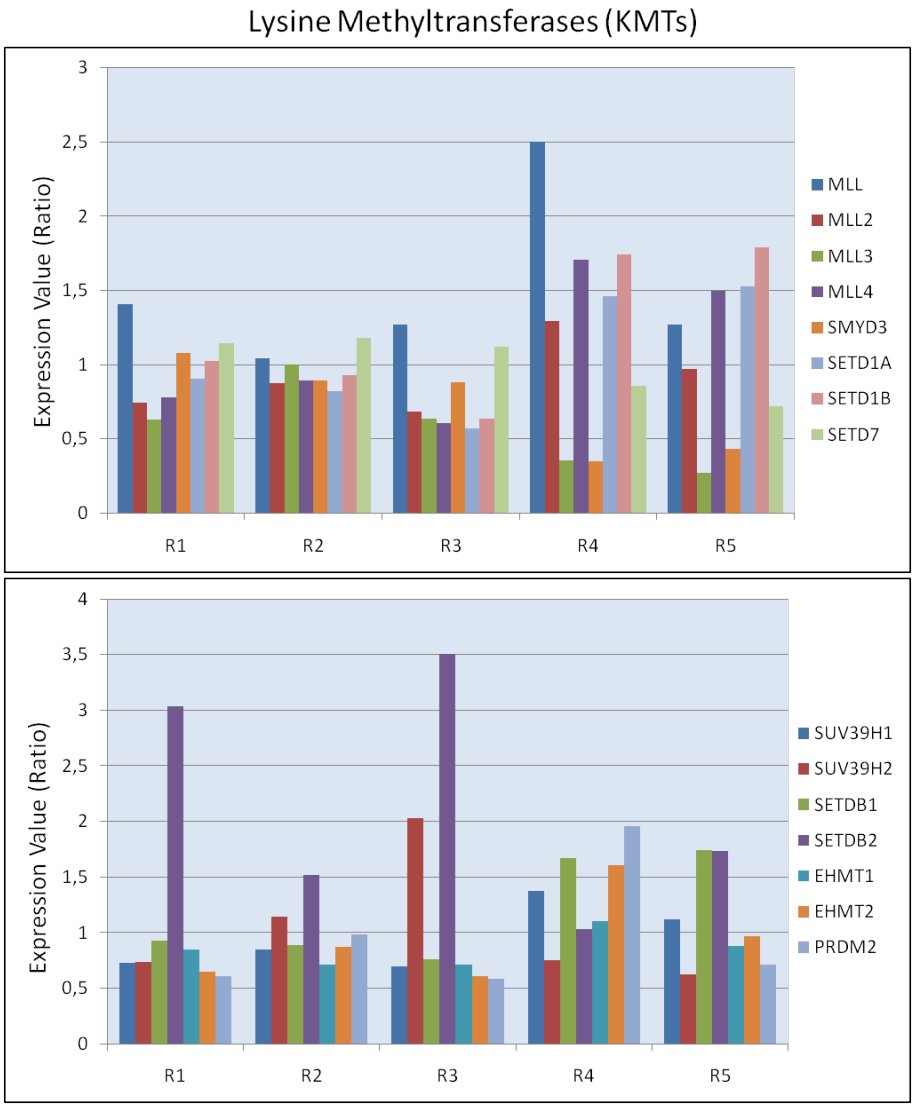


Figure 3.4: Gene expression profile of HMTs in MOP/UVA treated FA-D2 corrected and mutant cells. The bars represent ratios of expression values of corresponding genes from single microarray.

Concerning the H3K79 methylation, which was increased 10 fold early after damage (only in the wild type cells), only EZH1 and SetDB2 were up-regulated at this time point [Figure 3.5]. However, the former gene had even higher expression levels later after treatment, whereas the later gene manifested significant differences in expression in the mutant cell lines between the 1h and 6h sample. Thus, no single gene expression could account for the pattern of H3K79 methylation. However, a number of demethylases, such as KDM2B,

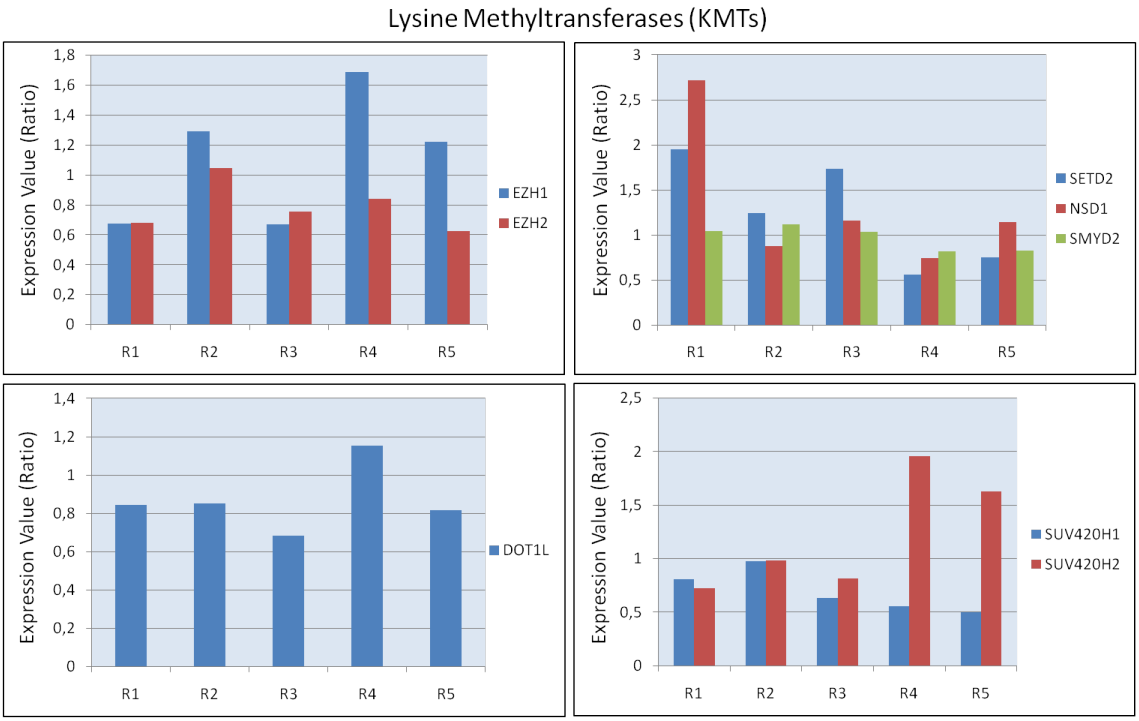


Figure 3.5: Gene expression profile of HMTs in MOP/UVA treated FA-D2 corrected and mutant cells. The bars represent ratios of expression values of corresponding genes from single microarray.

KDM4A and KDM6B had an increased expression levels late after treatment, suggesting that the concerted action of these two classes of enzymes could be the simplest explanation of the observed pattern of H3K79 methylation, based on our data [Figure 3.6].

The simplest explanation of the dramatic decrease in H3K56 methylation in the mutant cells 1h after treatment would be an increased level of a histone demethylase. However, we could not find any gene that showed a desired expression pattern. The KDM3A demethylase showed a little increase, but it was not significant to explain the observed effects. Also, we could not find a methyltransferase that could, in combination with the KDM3A, account for the observed pattern of this histone modification [Figure 3.6].

To summarize, by performing functional genomics analysis, we were able to find enzymes that could be responsible for some of the changes in the histone modifications that we have observed after induction of DNA-ICL damage in the wild type and FA-defective cells. In some cases, such as H4 acetylation, we found a single gene with the expression levels that were matching the changes in this modification in the samples studied. In most of the other cases, it was possible with two enzymes having opposite activities (such as HATs vs HDACs).

One have to keep in mind, however, that the changes in gene expression level do not necessarily translate into increased protein levels, let alone to the increased levels of enzyme activity. Various post-translational regulation steps could also be involved in the observed

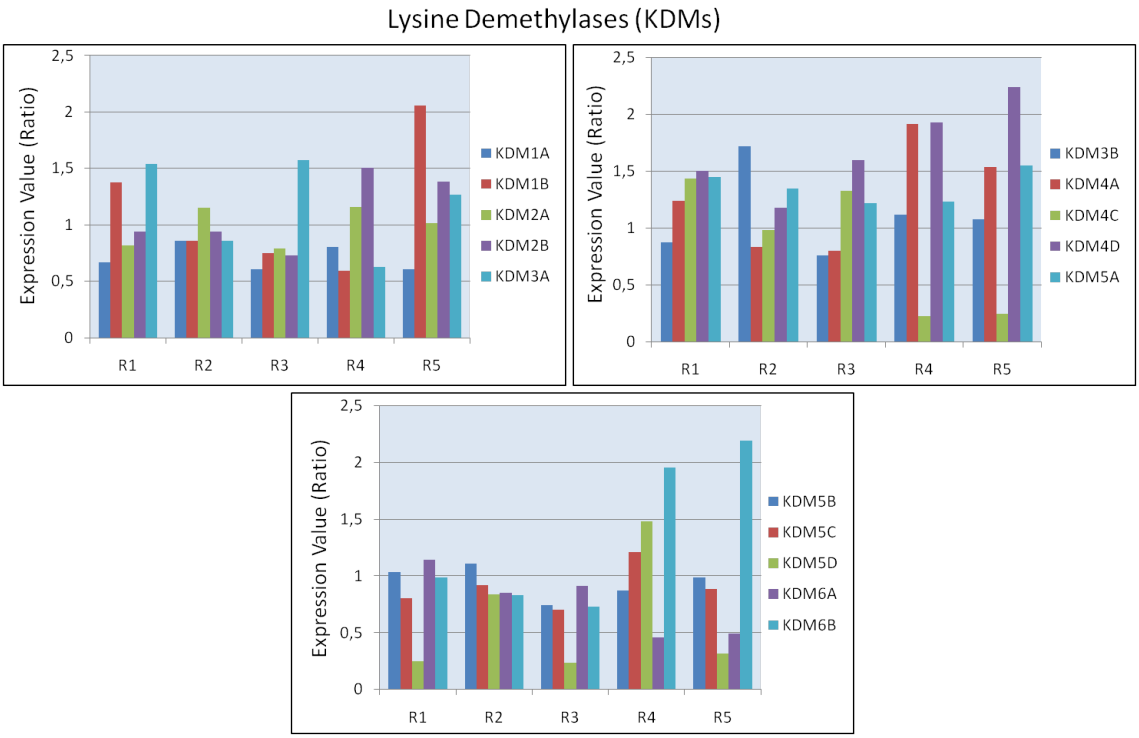


Figure 3.6: Gene expression profile of KDMs in MOP/UVA treated FA-D2 corrected and mutant cells. The bars represent ratios of expression values of corresponding genes from single microarray.

patterns of histone modifications, including activation or repression of the enzymes responsible, as well as their recruitment to chromatin. Such analysis, however, was beyond the scope of our study. In any case, the results of transcriptional analysis of histone modifying enzymes showed intriguing differences between wild type and mutated cells both at the basal level and also after induction of DNA-ICL damage. These differences may reflect the dependence of these enzymes on the presence of functional FA-pathway. Given that many histone modifying enzymes have, in addition to histones, other protein targets, these data might prove useful for better understanding of cellular response to DNA-ICL damage.

3.3 Global Transcriptional Analysis

The second goal of transcriptional analysis was to analyze the total transcriptome after damage in mutant or wild type cells. We also included in this analysis, in addition to DNA-ICL, ionizing radiation induced DNA damage, as we wanted to dissociate more generic response to DNA damage from a more specific changes that are associated with the role of FA pathway to the DNA ICLs.

3.3.1 Transcriptional responses to DSB and DNA ICL are different

At first, we analyzed the generic transcriptional response to two kinds of DNA damage in FA-D2 corrected cells, which are considered as "wild type" or normal cells. For this, we compared the transcriptomes from untreated cells with the same cells treated with MOP/UVA or IR at 1h and 6h time points. [Figure3.7-A & -B] shows the numbers of differentially expressed genes in cells where either DNA-ICL and DSB were induced, at the early (1h) and late (6h) times after treatment. These signature sequences were selected by filtering the data at fold change ± 1.5 and P<=1.0E-5. Curiously, we have observed that the early response was significantly more prominent in the case of MOP/UV treatment, as compared to the IR. Moreover, the overwhelming majority of the genes affected were up-regulated, whereas the IR-treatment affected genes showed opposite pattern.

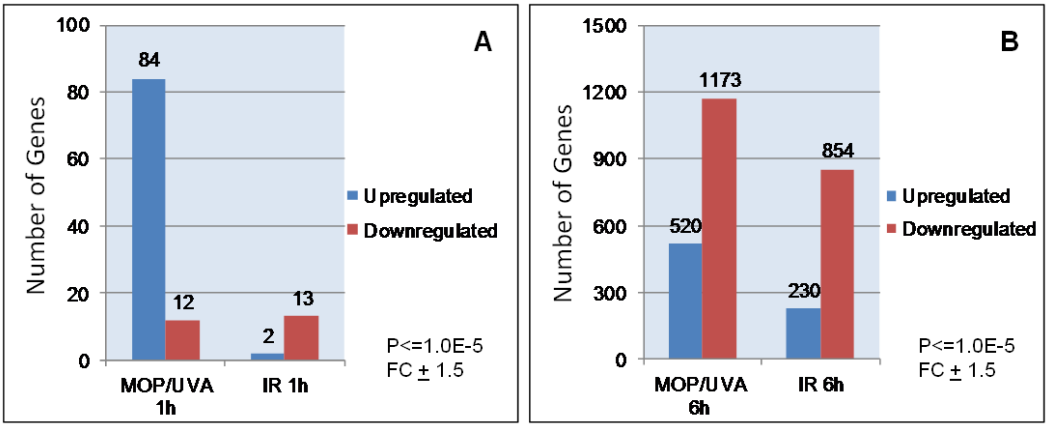


Figure 3.7: The overall transcriptional response to two types of DNA damage. The bars represent the number of genes. The filter was set at fold change ± 1.5 and P<=1.0E-5.

We performed the functional annotation analysis of affected genes in case of MOP/UVA treated cells after 1h. For convenience, we divided this analysis into two categories, i.e. the functions enriched in up-regulated and down-regulated genes. As seen in [Table 3.1], the analysis of up-regulated genes in 1h-MOP/UVA treated cells showed that mainly the sequences corresponding to transcription related activity and kinase based signaling pathways were enriched. As expected from the much smaller number of down-regulated genes, the enrichment in a particular GO class was less specific, although transcriptional regulation and nucleosome assembly related functions were enriched with low p-values [Table 3.2].

Table 3.1: Gene Ontology (GO) analysis of up-regulated sequences in FA-D2 corrected (1h-MOP/UVA) cells VS FA-D2 corrected untreated cells.

Primary GO Term Name	P-value
Transcription factor activity	3,64E-06
Transcription corepressor activity	0,00007
Negative regulation of transcription factor activity	0,0014
Regulation of transcription, DNA-dependent	0,00255
Positive regulation of transcription from RNA polymerase II promoter	0,00451
Transcription factor binding	0,00855
Mitogen-activated protein kinase kinase binding	4,13E-06
Negative regulation of lipopolysaccharide-mediated signaling pathway	0,00001
Regulation of MAP kinase activity	0,00006
MAP kinase tyrosine/serine/threonine phosphatase activity	0,00055
Inactivation of MAPK activity	0,00263
JNK cascade	0,00291
Protein kinase inhibitor activity	0,00306
Negative regulation of protein kinase activity	0,00456
Mitogen-activated protein kinase binding	0,00612
Protein kinase activity	0,00634

Table 3.2: Gene Ontology (GO) analysis of down-regulated sequences in FA-D2 corrected (1h-MOP/UVA) cells VS FA-D2 untreated cells.

Primary GO Term Name	P-value
nucleosome	0,0376
nucleosome assembly	0,04322
positive regulation of transcription, DNA-dependent	0,04463
transcription from RNA polymerase II promoter	0,05715
chromosome	0,06487

[Table 3.3] & [Table 3.4] show similar analysis of IR treated cells. As seen from the [Table 3.3], among the up-regulated genes, apart from transcription related activity, there is a notable enrichment of sequences related to histone acetylation, which is consistent with the established fact that histone acetylation is increased early after DNA damage. Moreover, unlike in the case of MOP/UVA treatment, cell cycle arrest function is also enriched. Among the down-regulated genes [Table 3.4], cell cycle progression, especially mitosis and spindle checkpoint related functions are enriched, also consistent with a faster cell cycle response in the case of IR induced DNA damage, as compared to the DNA-ICLs.

Table 3.3: Gene Ontology (GO) analysis of up-regulated sequences in FA-D2 corrected (1h-IR) cells VS FA-D2 untreated cells.

Primary GO Term Name	P-value
Histone acetyltransferase complex	0,00029
N-terminal peptidyl-lysine acetylation	0,00034
Positive regulation of transcription factor activity	0,00039
Acetyltransferase activity	0,00112
Histone acetylation	0,00117
Histone acetyltransferase activity	0,00175
Transcription factor complex	0,00777
Cell cycle arrest	0,00797
Transcription coactivator activity	0,01213
Positive regulation of transcription from RNA polymerase II promoter	0,01576

Table 3.4: Gene Ontology (GO) analysis of down-regulated sequences in FA-D2 corrected (1h-IR) cells VS FA-D2 untreated cells.

Primary GO Term Name	P-value
Positive regulation of mitosis	0,00003
Mitotic cell cycle spindle assembly checkpoint	0,00316
Positive regulation of translation	0,00442
Mitotic spindle organization	0,00568
Positive regulation of anti-apoptosis	0,00694
Single-stranded DNA binding	0,02316
Protein tyrosine kinase activity	0,03272
Positive regulation of transcription, DNA-dependent	0,04825
Transcription factor complex	0,04946
Transcription from RNA polymerase II promoter	0,06177

Next we performed the GO analysis of the late transcriptional response (i.e., after 6h of DNA damage) of the same cells treated with MOP/UVA or IR. Again we divided our analysis into functional annotation of up-regulated sequences and down-regulated sequences. The MOP/UVA-6h experiment [Table 3.5] again showed enrichment of transcription factor activity with high p-values among the up-regulated genes. Also, whereas at the early response, we could not detect effect on cells cycle after induction of interstrand cross-links, at the later time, such effects could be readily seen, especially, in down-regulated genes [Table 3.6]. In addition, ubiquitin-protein ligase activity was also detected, which, on one hand, pointed toward increased protein turnover through proteosomal degradation (poly-ubiquitination), while on the other hand, could be related to intra-cellular signal transduction (mono-ubiquitination) [Table 3.6]. Out of 96 genes affected early upon MOP/UV treatment, 20 remained affected 6h after damage, represented, among others, by transcrip-

tion corepressor activity, SMAD binding and DNA methylation [Figure 3.8-A & Table 3.7]. Similarly, the IR-6h experiment showed mainly the transcription related activity among the up-regulated genes. Interestingly, some chromatin related functions were also enriched including histone demethylase, acetyltransferase, and ubiquitination activity [Table 3.8]. The down-regulated sequences again showed cell cycle related functions and response to stress and radiation, as an expected late response [Table 3.9]. Among the chromatin related activities, HDAC binding was observed. In case of IR treated 1h and 6h samples [Figure 3.8-B & Table 3.10], 9 genes consistently showed altered gene regulation after 6h of damage, representing stress response (mostly related to heat shock) and histone acetyltransferase activities among the most prominent functions indicating that histone acetylation is an integral part of IR induced DNA damage response at both early and late time points.

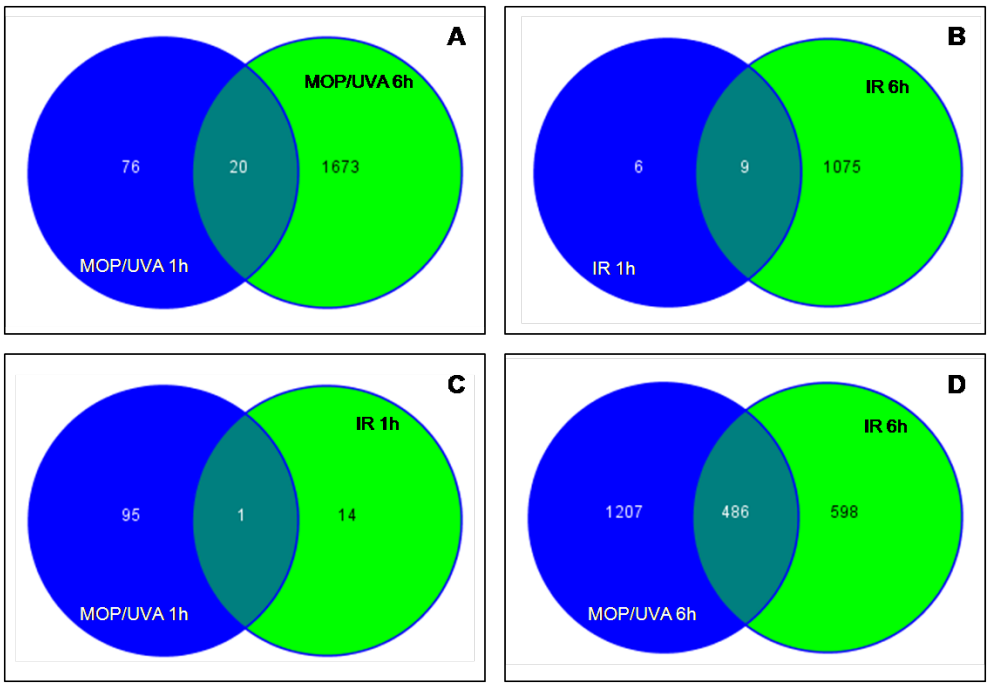


Figure 3.8: Venn dagrams showing comparisons of MOP/UVA 1h and IR 1h (A), MOP/UVA 6h and IR 6h (B), MOP/UVA 1h and 6h (C), and IR 1h and 6h (D).

Next, we analyzed whether there is anything common between the transcriptional responses to MOP/UVA and IR treatments. As shown in [Figure 3.8-C] only one gene was moderately upregulated 1h after both treatments, which corresponds to PKD2 (poly-cystic kidney disease 2), a membrane protein involved in calcium transport. Intriguingly, the literature search showed no link of this gene to DNA damage response. However, a significant overlap was observed in the late responding genes, as shown in [Figure 3.8-D]. As one might expect, gene ontology analysis [Table 3.11] revealed the most similarity in cell proliferation related genes, as well as in the induction of apoptosis.

To summarize, the cellular responses of the wild type cells to the two kinds of DNA damage appear to be quite different, especially at the early times. In the case of DNA-ICLs, transcrip-

Table 3.5: Gene Ontology (GO) analysis of up-regulated sequences in FA-D2 corrected (6h-MOP/UVA) cells VS FA-D2 untreated cells.

Primary GO Term Name	P-value
Sequence-specific DNA binding	9,79E-13
Regulation of transcription, DNA-dependent	2,58E-12
Transcription factor activity	2,01E-10
Nucleus	4,67E-06
Zinc ion binding	5,23E-06
DNA binding	0,00001
Negative regulation of cell proliferation	0,00064
Cyclin-dependent protein kinase inhibitor activity	0,00082
Cell growth	0,00091

Table 3.6: Gene Ontology (GO) analysis of down-regulated sequences in FA-D2 corrected (6h-MOP/UVA) cells VS FA-D2 untreated cells.

Primary GO Term Name	P-value
Cell cycle	5,02E-13
Cell division	5,41E-11
Ubiquitin-protein ligase activity	1,65E-10
Mitosis	1,99E-10
Protein amino acid phosphorylation	1,09E-08
Negative regulation of transcription	9,05E-06
Cell cycle arrest	0,00002
Cellular response to UV	0,00002
ATP-dependent helicase activity	0,00005
G1/S transition of mitotic cell cycle	0,00006
Positive regulation of transcription from RNA polymerase II promoter	0,00007
Negative regulation of DNA recombination	0,00009
Protein serine/threonine kinase activity	0,00012
Positive regulation of phosphorylation	0,00019
Intra-S DNA damage checkpoint	0,00022
Negative regulation of cell growth	0,00023
Response to DNA damage stimulus	0,00025
Single-stranded DNA binding	0,00026
Protein ubiquitination	0,00032
Negative regulation of mitotic cell cycle	0,00038
Double-stranded DNA binding	0,00063
Ubiquitin ligase complex	0,00093

tion regulation appears to be more significantly affected than in the case of double-strand breaks. On the other hand, histone acetylation appears to play a more significant role in the

Table 3.7: Gene Ontology (GO) analysis of common sequences in 1h and 6h MOP/UVA treated FA-D2 corrected cells VS FA-D2 untreated cells.

Primary GO Term Name	P-value
Transcription corepressor activity	0,00063
SMAD binding	0,00177
DNA methylation	0,01131
SMAD protein signal transduction	0,01131

Table 3.8: Gene Ontology (GO) analysis of Up-regulated sequences after 6h IR treated FA-D2 corrected cells VS FA-D2 untreated cells.

Primary GO Term Name	P-value
Regulation of transcription, DNA-dependent	2,10E-08
Nucleus	2,24E-07
Zinc ion binding	4,93E-06
Induction of apoptosis	0,00003
Transcription factor activity	0,00003
Negative regulation of cyclin-dependent protein kinase activity	0,00009
Cyclin binding	0,00046
Cell cycle arrest	0,00238
Chromatin organization	0,00264
Response to DNA damage stimulus	0,00276
Negative regulation of apoptosis	0,00325
Negative regulation of phosphorylation	0,00401
Histone demethylation	0,0167
Histone demethylase activity (H3-K36 specific)	0,0222
Histone acetyltransferase complex	0,03312
Nucleosome positioning	0,03312
Chromatin silencing complex	0,03853
Histone ubiquitination	0,04392

response to DSB as compared to ICL. Also, consistent with a more rapid response to DSB, the cell cycle regulation is prominently affected early after IR treatment, and only 6h after the MOP/UVA treatment. By comparing transcriptional DDR between DNA-ICL and DSB, we were also able to detect the genes that most probably reflect a generic response to a DNA damage. Not surprisingly, cell proliferation and apoptosis-related genes were most prominent among this category.

Table 3.9: Gene Ontology (GO) analysis of Down-regulated sequences after 6h IR treated FA-D2 corrected cells VS FA-D2 untreated cells.

Primary GO Term Name	P-value
Nucleus	7,98E-32
Cell cycle	4,12E-23
Cell division	1,45E-20
Transcription factor activity	1,47E-12
Positive regulation of transcription from RNA polymerase II promoter	6,72E-11
Cell proliferation	4,13E-09
Protein serine/threonine kinase activity	5,64E-09
Transcription repressor activity	2,38E-08
Apoptosis	1,48E-07
Negative regulation of cell growth	0,00001
DNA damage response	0,00008
Negative regulation of transcription factor activity	0,00021
Histone deacetylase binding	0,00045
Single-stranded RNA binding	0,00056
SMAD binding	0,00058
Response to stress	0,00058
Response to radiation	0,00069
Mitotic cell cycle spindle assembly checkpoint	0,00096
Ubiquitin protein ligase binding	0,00123
Chromatin binding	0,01298
G1/S transition checkpoint	0,02095
Nucleosome binding	0,04114
DNA replication	0,05995

Table 3.10: Gene Ontology (GO) analysis of common sequences in 1h and 6h IR treated FA-D2 corrected cells VS FA-D2 untreated cells.

Primary GO Term Name	P-value
Response to heat	0,00006
Heat shock protein binding	0,00036
Unfolded protein binding	0,00089
N-terminal peptidyl-lysine acetylation	0,00826
Acetyltransferase activity	0,00826
Histone acetylation	0,00826
Histone acetyltransferase complex	0,00826

3.3.2 Transcriptome is dramatically affected by defect in FA pathway

Before analyzing how the cellular response to DNA damage is affected in the cells defective in the FA pathway, we wanted to see the differences in transcriptional profile between FA-D2

Table 3.11: Gene Ontology (GO) analysis of common sequences in 6h MOP/UVA and IR treated FA-D2 corrected cells VS FA-D2 untreated cells.

Primary GO Term Name	P-value
Mitosis	0,00177
Induction of apoptosis	0,00703
Cell division	0,00711

corrected and mutated cells without induction of any DNA damage. Approximately equal numbers of genes were up- and down-regulated in the mutant cells [Figure 3.9]. Quite surprisingly, the GO analysis of up-regulated genes showed many chromatin related functions to be affected by FA-D2 mutation, most prominently including histone acetyltransferase activities and ubiquitination. Most notable genes in this category are presented in [Table 3.12]. In this respect, the cellular response to FA-D2 mutation was more reminiscent of the wild type cell response to the DNA double strand breaks than to the DNA-ICLs. This was somewhat surprising, however, might be reflecting the fact that defect in FA pathway leads to permanent elevation of DSB levels in the cells (e.g., resulting from stalled replication forks).

The analysis of down-regulated genes in FA-D2 mutant vs FA-D2 corrected (both untreated) does not show any significant enrichment of pathways related to chromatin modifying activities. However, protein demethylation and dephosphorylation activities were enriched along with cell cycle arrest and anti-apoptosis activities [Table 3.13].

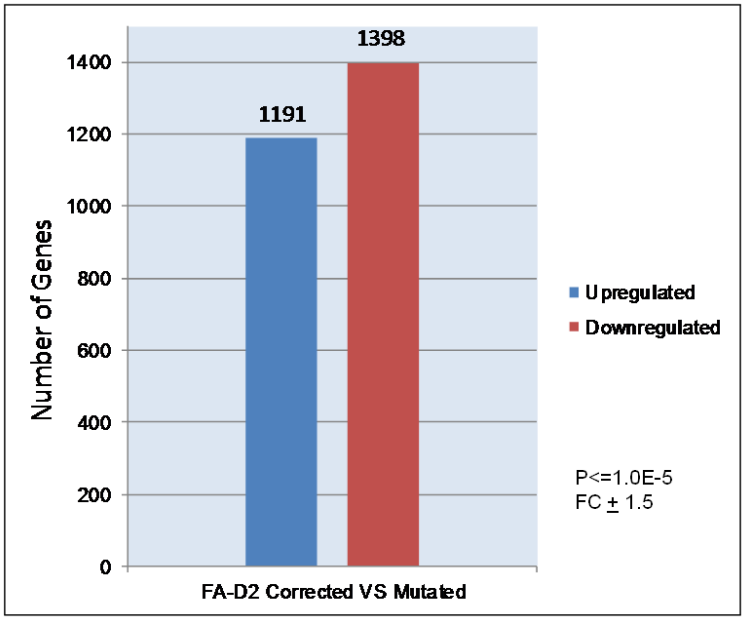


Figure 3.9: Numbers of genes up and down-regulated in FA-D2 mutated cells, as compared to the wild type (FA-D2 corrected)without any DNA damage induction.

Table 3.12: Up-regulated chromatin-related genes in untreated FA-D2 mutant VS wild type cells.

Gene Name & Symbol	Fold Change	P-value
Cullin 9 (CUL9)	5,49321	8,49163E-16
MYST histone acetyltransferase 3 (MYST3)	4,9904	5,04243E-13
MYST histone acetyltransferase 4 (MYST4)	3,17932	2,43832E-09
Chromodomain helicase DNA binding protein 7 (CHD7)	3,16401	1,69855E-10
E1A binding protein p300 (EP300)	2,59095	3,93592E-17
High-mobility group nucleosome binding domain 5 (HMGN5)	2,5607	9,49204E-17
Chromodomain helicase DNA binding protein 7 (CHD7)	2,45127	7,80561E-08
Cullin 4A (CUL4A), transcript variant 1	2,36822	1,02539E-07
K(lysine) acetyltransferase 2B (KAT2B)	2,2195	7,98776E-06
M-phase phosphoprotein 8 (MPHOSPH8)	2,21176	6,40402E-07
K(lysine) acetyltransferase 2B (KAT2B)	2,07674	3,91118E-06
N(alpha)-acetyltransferase 16 (NAA16), transcript variant 1	1,66772	0,00062
CREB binding protein (CREBBP), transcript variant 1	1,54194	0,0032

Table 3.13: Gene Ontology (GO) analysis of genes down-regulated in untreated FA-D2 mutant cells as compared to the wild type (FA-D2 corrected).

Primary GO Term Name	P-value
Protein amino acid demethylation	0,00002
Positive regulation of cell cycle	0,00056
Anti-apoptosis	0,00219
Cell proliferation	0,00502
Protein amino acid dephosphorylation	0,0052
Response to stress	0,00751
Cell cycle arrest	0,00857
Mismatch repair	0,00864

3.3.3 Role of Fanconi Anemia pathway in the Transcriptional Response to DNA Damage

The next step of our analysis was to compare the responses to the two kinds of DNA damage between the wild type and FA-defective cells. Given that FA-negative cells are hypersensitive to the ICL type of DNA damage, this analysis could help us to understand what part of the cellular response to ICL that we have observed in the wild type cells requires functional FA pathway. These FA-dependent changes most likely constitute the bona-fide cellular response to ICLs. Equally importantly, by finding genes that are affected by DNA damage only in the mutant cells, the same experiments should also help us to find out what unwanted consequences of DNA damage the FA pathway protects the cells from.

At first, we compared the gene expression profiles between wild type and mutant cells for either MOP/UVA-(1h and 6h samples), [Figure 3.10] or IR-(1h & -6h treated samples), [Figure 3.11] induced damage. For this purpose, we used cluster analysis to visualize general similarities between different samples. Clustering in two dimensional space is often used to give ordering by identifying groups of genes that show the most similarity in their behavior and then arranging these groups so that the closest groups of genes are adjacent to each other. In addition, the samples are also ordered in clusters according to their similarities among each other. The numbers of samples are significantly smaller that the numbers of genes, and thus they are easier to directly compare on the heatmaps. Accordingly, at first we compared two different DNA damage treatments in terms of similarities between the samples. We have noticed an intriguing difference between MOP/UVA treatment and IR treatment in this respect. In both types of DNA damage, all samples that correspond to mutant cells cluster closest to each other, with the untreated and 1h samples being most close, and the 6h treatment being slightly more apart. The clustering of mutant samples most likely reflects the abundant differences between mutant and wild type cells, which keeps them more similar to each other regardless of what DNA damage treatment was used. However, we observed a difference between MOP/UVA and IR treatment in how it affected the wild type cells. The 1h and 6h IR-treated wild type samples were closer to each other than to any of the mutant IR-treated samples. On the other hand, the 1h and 6h MOP/UVA-treated wild type samples did not cluster together, but the 1h sample was closer to the 'mutant cluster' than the 6h sample. This observation suggests that the dramatic long term response to ICL damage (6h) that is observed in the case of wild type cells is significantly compromised in the FA-negative cells. This is consistent with the known role of FA pathway in response to ICL damage.

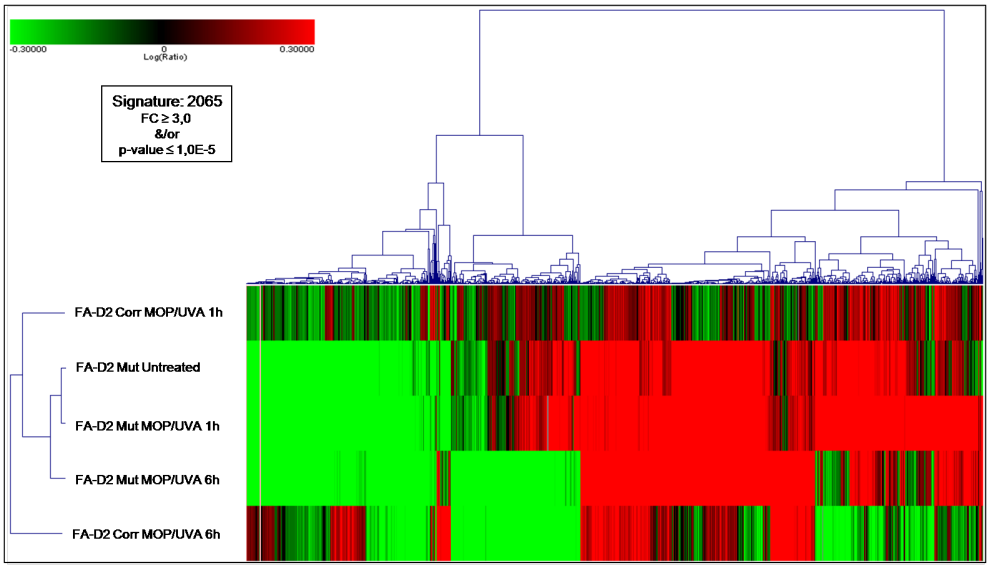


Figure 3.10: 2-Dimensional cluster of MOP/UVA treated samples. Sequences having FC ≥ 3.0 and/or p-value $\leq 1.0E-5$ in any one of the five experiments were included in the cluster analysis.

After comparing between different samples in the cluster analysis, we set out to find genes

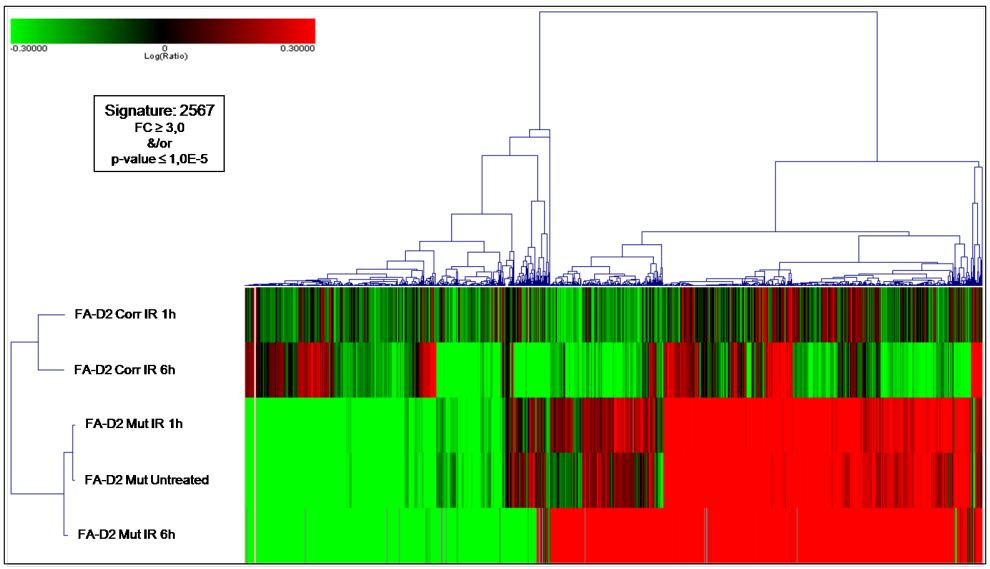


Figure 3.11: 2-Dimensional cluster of IR treated samples. Sequences having FC $\geq$ 3.0 and/or p-value $\leq$ $1.0E-5$ in any one of the five experiments were included in the cluster analysis.

that are affected differently by DNA damage in the wild type and FA-defective cells. Such analysis should help us to address two different questions:

I What part of the transcriptional response to DNA damage requires functional FA-pathway?

II What happens to cells after DNA damage if FA pathway is inactive?

In order to extract this information from our functional genomics data, we applied interaction analysis. We can define the notion of interaction or non-additivity of effects as "The joint effect of two factors differs from what might be predicted on the basis of their independent effects alone. It is the measurement of the extent to which the presence or absence of one factor affects the differential expression of a gene in response to the presence or absence of the other factor" [Yang and Speed, 2002].

First we will use a toy example to explain the notion of interaction between two factors in a given functional genomics experiment [Figure 3.12].

We have "factor A" and "factor B", which could affect the expression of a given gene. These factors could be a particular experimental condition (DNA damage) or a particular mutation, etc. Now we consider an experiment where cells affected by factor A or by factor B are compared to a control C. We also have an experiment where we combine these two factors in an experiment AB and the expression of gene is measured in A, B, and AB. Now, if the factors A and B act on the gene independent from each other, then the fold change in the level of expression of the gene in AB, as compared to control C (we designate it Rab), should be the product of the fold changes in its expression in A and B experiments : Rab = Ra*Rb.

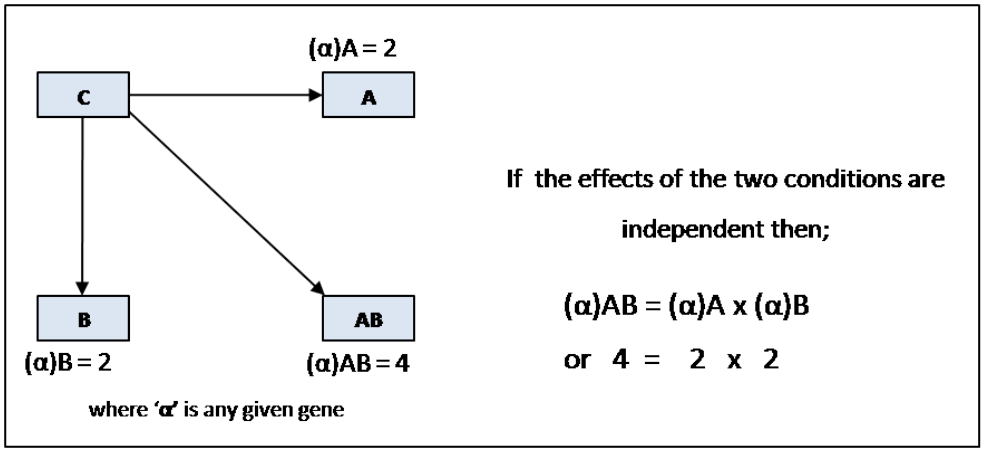


Figure 3.12: 2x2 Factorial experiment and the idea of interaction. In this factorial experiment, we have factor "A" and "B" which could affect the expression of a given gene. We consider an experiment where cells affected by factor A or by factor B are compared to a control C. We also have an experiment where we combine these two factors in an experiment AB and the expression of gene is measured in A, B, and AB. If the factors A and B act on the gene independent from each other, then the expression of the gene in AB, as compared to control C (we designate it Rab), should be the product of the fold changes in its expression in A and B experiments : Rab = Ra*Rb. This situations means no interaction between factor A and B in their effects on the gene expression. Any deviation from this value will point toward the existence of interaction.

In this case, we can say that there is no interaction between factor A and B in their effects on the gene expression. Any deviation from this value will point toward the existence of interaction.

Applying this analysis to our system, we have two factors, the mutation in FA-D2 gene and particular DNA damage, represented by R1 and R2 respectively for 1h time point experiments and R1 and R4 for 6h experiments [Figure 3.13]. Their combined effect is represented by R3 and R5 in 1h and 6h experiments, respectively. If we denote interaction by 'ϵ' it can be calculated for 1h and 6h experiments by the following equations;

$$\epsilon = \frac{R3}{R1 \times R2} \tag{3.1}$$

$$\epsilon = \frac{R5}{R1 \times R4} \tag{3.2}$$

The interaction analysis corresponds to search of genes for which there is a significant deviation of epsilon = 1. Accordingly, we calculated the "ϵ" values for each 1h and 6h according to equations 1 and 2 respectively.

As shown on the [Figure 3.14], most of the genes do not show interaction. In order to identify the genes which manifest most significant levels of interaction, we have applied another filter on the "ϵ" values such that only the sequences having values ≥ 3.0 or ≤ 0.3 were selected for both 1h and 6h. [Figure 3.15] shows the number of sequences selected after applying this

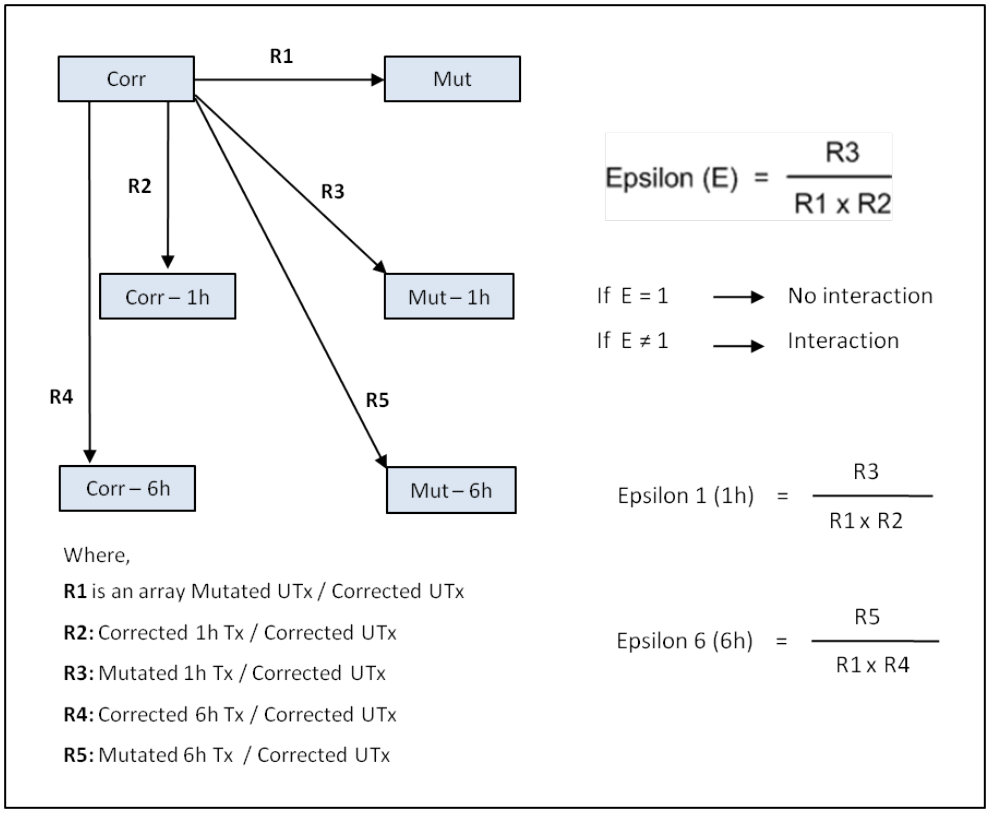


Figure 3.13: Interaction analysis in FA-D2 corrected and mutant cells after induction of DNA damage (MOP/UVA & IR). We applied this factorial approach to our microarray data where we tried to find out the effect of DNA damage (factor A) and mutation (factor B) on gene expression.

filter. To our big surprise, we found that significantly more genes show interaction in the case of IR, than in the case of MOP/UVA treatment at the later stage of response, indicating that late IR induced DDR highly depends on the functional FA pathway.

The simple interaction analysis shown above is not sufficient, because it does not allow one to differentiate between different explanations for the observed interactions between two factors for every individual gene. In two simplest cases, the epsilon could deviate from 1 (unity) for the following reasons : a) The defect in FA pathway suppresses the response to DNA damage and b) The FA defect 'activates' the response to DNA damage for a gene that was otherwise not sensitive to it in the wild type cells. The first class of genes corresponds to the part of the transcriptional response that requires functional FA pathway, whereas the second class of genes reflects what happens with the cells after DNA damage when the FA pathway is not functioning properly. Thus, this step in analysis is necessary in order to address the two questions posed at the beginning of this section [Figure 3.16].

To find the genes that demonstrate a FA-dependent suppression of the effect of DNA damage, we filtered the data on the genes that show interaction, according to the following constraint :

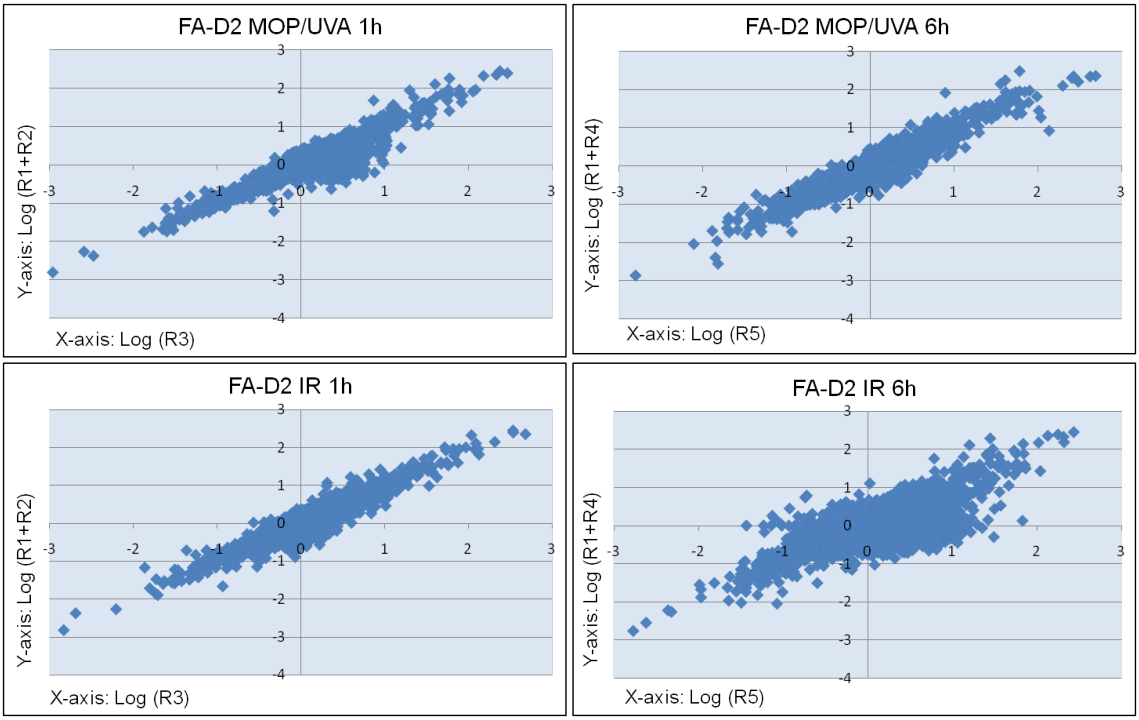


Figure 3.14: Scatter Plots representing Log(R1+R2)-Y-axis VS Log(R3)-X-axis in case of MOP/UVA 1h while Log(R1+R4)-Y-axis VS Log(R5)-X-axis in case of 6h treatment revealed that majority of genes show no interaction.

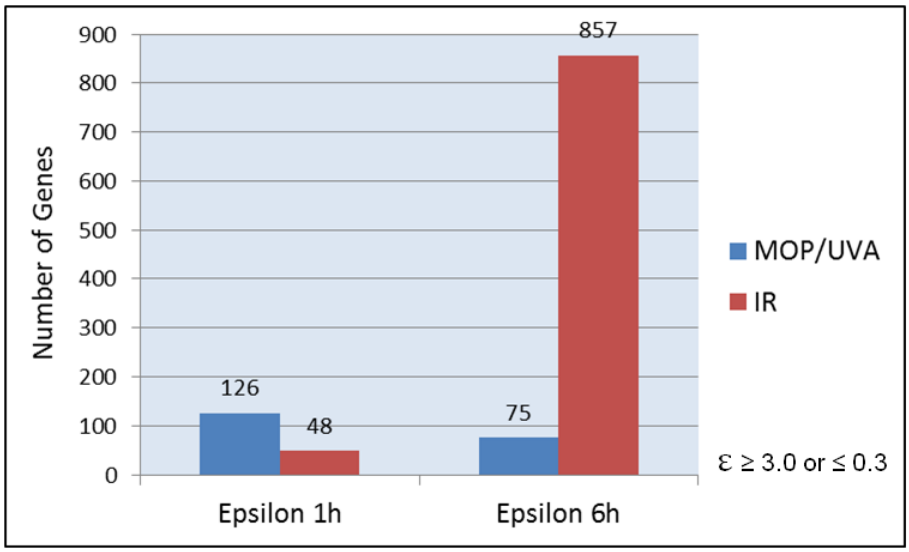


Figure 3.15: Graph showing number of sequences having ϵ values $\geq$ 3.0 or $\leq$ 0.3 in MOP/UVA and IR samples.

$$[\frac{R3}{R1} \sim 1] and [\epsilon \neq 1] \tag{3.3}$$

In other words, we were looking for the genes that showed a response in the case of wild type (R2 or R4), but no response in the case of the FA mutant (R3, R5 = 1) [Figure 3.16].

Next, we uploaded the list of sequences corresponding to each of these categories for functional annotation analysis to DAVID Bioinformatics Resources web page [[Huang et al., 2009a]; [Huang et al., 2009b]] in order to look for the functional correlation of the epsilon values. Given that wild type cells responded to MOP/UVA treatment by largely changing expression of transcription regulation factors, it is not surprising that at least some of DNA binding activities and regulation of transcription was affected by the FA mutation, indicating that many of these genes require functional FA pathway in order to respond to DNA-ICL damage. The similar analysis for 1h-IR treated set of genes revealed different functional enrichment, including functions related to phosphoproteins, cell cycle, metabolic processes, zinc finger proteins, DNA damage and proteins acetylation related pathways.

Next, we looked at the genes showing dependence on FA pathway after 6h of DNA damage. In case of MOP/UVA we observed mostly the chromosome condensation and mitosis related functions as dependent on functional FA pathway, while in case of 6h-IR we observed functions related to cytoskeleton, intra-cellular signaling, and metabolic processes. Again the cellular response to these two kinds of DNA damage at later stages of damage induction show striking differences, which indicates activation of different pathways in response to different kinds of DNA damage.

To find the genes that normally do not respond to DNA damage, but start to show a response when the FA pathway is inactivated, we applied a different constraint to the same set of genes that demonstrated interaction [Figure 3.16], given by the following equation;

$$[R2 \sim 1] and [\epsilon \neq 1] \tag{3.4}$$

As discussed before, these genes should reflect 'what could happen to cells after DNA damage that does not happen in the presence of functional FA-pathway', i.e., most likely, the unwanted consequences of not having FA pathway in the case of DNA damage. The functional annotation analysis in case of 1h MOP/UVA showed enrichment of mostly the cell cycle and mitosis related functions along with phosphorylations and chromatin modifications and chromatin remodeling activities. As discussed before, we have not seen any cell cycle and mitosis effects that early after the DNA-ICL induction, and can conclude that inactivation of FA-pathway sensitizes some cell-cycle regulatory genes that otherwise would not be responding to MOP/UVA treatment. However, the similar analysis of 1h-IR treated samples showed enrichment of various zinc finger proteins related functions, which might be related to mono- or poly-ubiquitinations, and also, DNA-binding, transcriptional regulation and DNA damage response functions. The difference between functions enriched in these two genes sets might reflect that differences in the early cellular response to these two kinds of DNA damage inductions.

The similar analysis for 6h post-MOP/UVA treated experiment revealed functions related to cytoskeleton, proteins post-translational modifications including phosphorylations and

acetylations and proteins turnover, and again cell cycle checkpoint. The analysis of 6h post-IR treated sample showed enrichment of zinc finger proteins related function with highly significant p-values and also transcriptional regulation, DNA binding and phosphorylations. In general, we have observed disproportionally many genes that showed sensitization to FA-inactivation at the later stages of response to DSB damage, as compared to DNA-ICL damage.

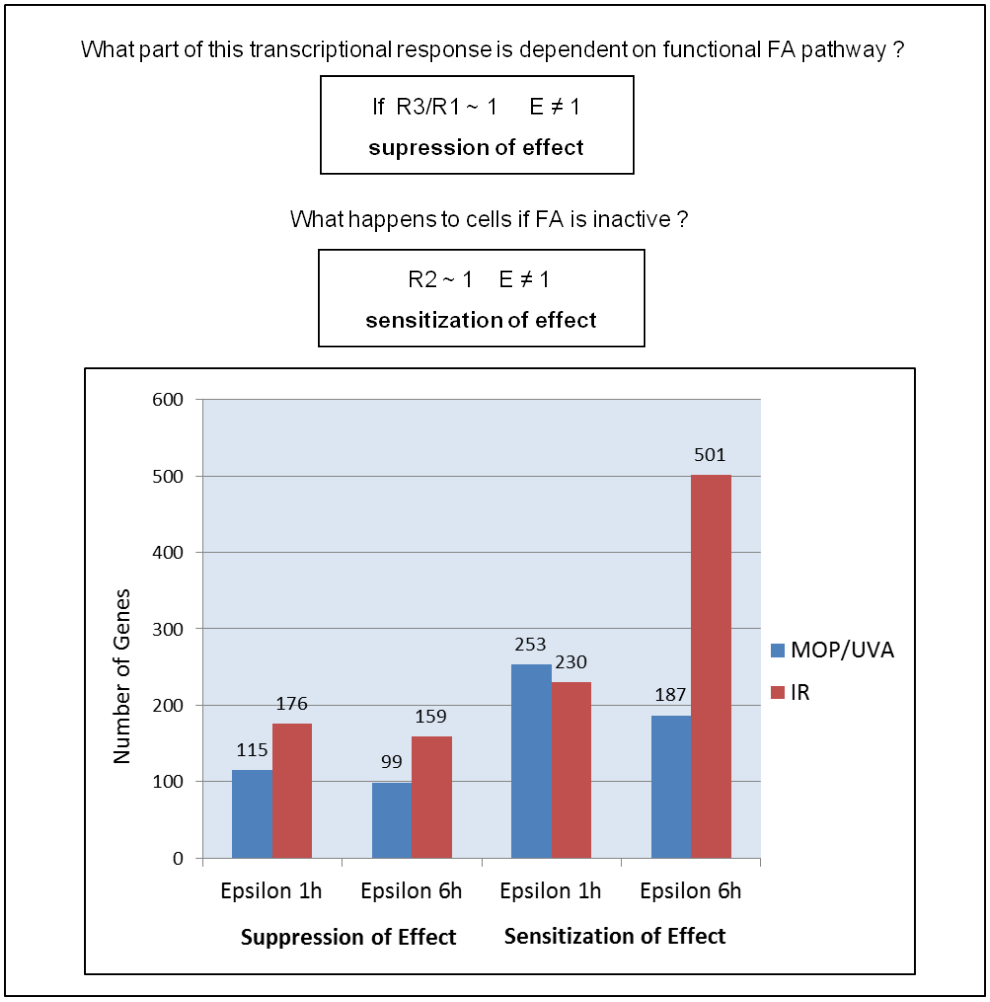


Figure 3.16: Graph showing number of sequences in various sub-categories based on ϵ values in MOP/UVA and IR samples.

To summarize this part of our analysis, we were able to detect many differences in the transcriptional response to DSB and DNA-ICL. Early after damage, significantly more genes are up-regulated in the case of DNA-ICL, mostly comprising transcriptional regulation and signal transduction. Interestingly, in the case of DSB damage, we detected an enrichment in histone acetylation activity, already early after treatment. Also, consistent with the faster response to IR, cell cycle arrest-related signature was detectable early after DSB induction, whereas DNA-ICL was causing such response only at later times.

Intriguingly, practically no common genes were detected among early responders to DNA-ICL and IR. Later, however, both treatments were inducing cell proliferation arrest and apoptosis activation, which most likely constitute a generic cellular response to any kind of DNA damage.

Comparison of transcriptomes of FA mutant and control cells revealed dramatic differences between these cells. Intriguingly, many chromatin-related activities were affected by the FA mutation, which was somewhat reminiscent of the cellular response to DSB. This surprising observation might reflect the fact that defect in FA pathway leads to permanent elevation of DSB levels in the cells (e.g., resulting from stalled replication forks).

Finally, by performing an interaction analysis, we were able to find the part of transcriptional response to DNA damage that was requiring functional FA pathway, as well as find the genes that were sensitized to DNA damage by the inactivation of FA pathway. We found that the number of transcriptional regulators are indeed dependent on FA pathway in their response to DNA-ICL. Surprisingly, we observed that inactivation of FA-pathway sensitizes some cell-cycle regulatory genes to DNA-ICL.

3.4 Discussion

Understanding the molecular details of complex signaling cascade of DNA damage response inevitably requires the analysis of transcriptional changes in the various players involved in DDR. Chromatin modifying enzymes are among one of the important classes of effector proteins that take part in DDR signaling as chromatin the physiological substrate of DNA damage repair. Although chromatin acts a physical barrier to the detection and repair of DNA lesions, the chromatin modifying enzymes confer a degree of flexibility to the normal compacted state of chromatin by making it more dynamic and responsive to the needs of a damaged cell. Histone modifying enzymes, among the various chromatin modifiers were the focus of our attention owing to our interest in the role of histone PTMs in DDR. Since our focus remained on the acetylations and methylations of histones in response to DDR, in this part of the project we looked at the transcriptional profile of enzymes related to said modifications, which include HATs, HDACs, HMTs, and KDMs.

Although a kind of an unorthodox approach, we tried here to correlate the histone PTM data with the transcriptional profile of enzymes responsible for these modifications. This analysis proved to be fruitful as we have found a few correlations among the histone acetylations and methylations after inducing DNA-ICL damage and the expression levels of possible genes, whose protein products could be responsible for creating these marks. To our knowledge, no study has been reported in the literature that attempts to correlate histone modifications with the gene expression profile of the enzymes responsible for them. This kind of analysis, a reverse functional genomics analysis, can be considered as a new and potentially interesting way to study the chromatin response to DNA damage at transcriptional level by reverse correlating the increase or decrease in particular protein PTMs with the gene expression profile of the enzymes responsible for such modifications.

In general, a few studies report the transcriptional profile of histone modifying genes but mostly in the context of various cancers [Ozdag et al., 2006]; [Lucio-Eterovic et al., 2008]. In these studies, the expression patterns of HATs, HDACs, and HMTs suggest that these genes have important implications in neoplastic transformation and have characteristic patterns of expression depending on tissue of origin. Another study reported the increased expression of SETDB1, an H3K9 methyltransferase, and suggested that this HMT acts as an oncogene significantly accelerating melanoma formation in zebrafish [Ceol et al., 2011]. This study also speculated that SETDB1 over-expression can increase the activity of the H3K9 methyltransferase complex, leading to alterations in its target specificity. Inactivating mutations in histone methyltransferases and histone demethylases were recently described in renal cell carcinoma [Dalgliesh et al., 2010]; [Staller, 2010]. Moreover, SETDB1 is focally amplified in a broad range of malignancies, suggesting that alterations in histone methyltransferase activity could define a biologically related subset of cancers. Another study correlated the various transcripts encoding histone methyltransferases and demethylases and altered proteasome inhibition using drugs &/or hormones. As proteasome inhibitors are currently in clinical trials as therapy for multiple myeloma, HIV/AIDs and leukemia, the possibility that some of the target molecules are regulated by chromatin modifying enzymes is intriguing in this era of epigenetic therapy.

Our results showed an increased expression of certain HATs (MOZ and MORF) and SETDB2, an HMT, in mutant FA-D2 cells compared to the corrected FA-D2 cells without any DNA damage induction. Generally, increased histone acetylation, due to increased expression of HATs or decreased expression of HDACs, leads to increased gene transcription. If this altered acetylation pattern is observed at any oncogene promotor, it inevitably leads to increased expression of that gene and can contribute to neoplastic changes. Similarly, the recent implication of SETDB1 over-expression in melanoma lead us to suspect another H3K9 methyltransferase, which is over-expressed in FA-D2 mutant cells, as a potential target for further investigation towards it plausible role in malignant transformation. Fanconi anemia patients have increased prevalence of leukemias and head and neck squamous cell carcinnomas, therefore, a more comprehensive approach towards identifying the potential genes involved in malignancies in FA patients is worth an effort.

DNA-ICLs induced by 8-MOP plus UVA (8-MOP/UVA) are critical genotoxic lesions and have strong anti-proliferative effects [Zheng et al., 2006] by initiating a cascade of events leading to cytotoxic, mutagenic and carcinogenic responses. Transcriptional activation plays an important part in these responses. Therefore, keeping in mind the specificity and the complexity of the damage induced by 8-MOP/UVA, it was of particular interest to analyze the global transcriptional response after induction of 8-MOP/UVA damage in human cells. Not much work is done on elucidating the transcriptional response elicited by this kind of DNA damaging drug, however, a comprehensive genome wide study was performed by Dardalhon and colleagues in Saccharomyces cerevisiae [Dardalhon et al., 2007], where they used DNA microarrays to examine genome-wide transcriptional changes produced after induction of 8-MOP/UVA photo-lesions. In order to gain a functional overview of the genes involved in the transcriptional response, they have performed a Gene Ontology (GO) anal-

ysis. The GO assignments show that 8-MOP/UVA response genes are implicated in a wide variety of biological processes that concern three main activities: (1) DNA metabolism and cell cycle-related functions; (2) cellular functions and maintenance of cellular structures; and (3) general metabolism. They have found that the up-regulated genes belong to DNA damage response and cell cycle related. Our results, on the other hand, showed the transcription related and kinase based signaling functions to be enriched in the early (1h) up-regulated genes, however, in late response (6h), our results also show cell cycle related functions to be enriched but mostly in the down-regulated gene set. Among the repressed genes, our results showed some enrichment of nucleosome assembly related functions, which is consistent with the findings of Dardalhon et al. as they have also found some functions enriched in repressed genes involved in structural modification of chromatin after DNA damage. These differences in transcriptional response may well be explained by the nature of experimental system used as human cells are much more complex and dynamic than lower eukaryotes, such as Saccharomyces cerevisiae.

Most experimental investigations of the DDR are carried out in commonly used cell lines, most of them derived from human tumors. Such cell lines are expected to exhibit a wide spread variation in their response to DNA damage, most notably to IR. This variation stems primarily from genetic variation among the donors, which is known to markedly affect their radiation response [Chistiakov et al., 2008]; [Pugh et al., 2009]; [Popanda et al., 2009]; [Niu et al., 2010]. Variations in response to IR also come from the multiple genetic changes that take place in the course of prolonged culturing, particularly in tumor cell lines. In a recent study where, Rashi-Elkeles and coworkers [Rashi-Elkeles et al., 2011] performed a large-scale expression analysis to examine the transcriptional response induced by 5 Gy of IR in five malignant and non-malignant cell lines. They compared three different time points 0h, 3h and 6h after irradiation. The majority of the transcriptional response of these cell lines to IR was cell line-specific, with substantial differences from one line to another. There was, however, a core of induced or repressed genes common to the different cell lines, with the induced ones consisting almost exclusively of validated p53 targets. Interestingly, hierarchical clustering demonstrated that cell line was the principal determinant of the expression profiles, as samples from the same cell line (treated and untreated) were clustered closest to each other. Our results are consistent with this finding. As shown in [Figure 3.11], samples from FA-D2 corrected cells were clustered close to one another while those from FA-D2 mutant cells, treated or untreated, were clustered closer to each other than to the corrected cells.

The IR-induced DNA DSB response involves a marked modulation of the cellular transcriptome [Begley and Samson, 2004]; [Elkon et al., 2005], as numerous large-scale expression studies collectively profiled cellular responses to a multitude of DNA-damaging agents [Heinloth et al., 2003a]; [Heinloth et al., 2003b]. They demonstrated that the transcriptional response to genotoxic stress goes beyond the core DDR pathways of DNA repair and cell cycle regulation to affect most aspects of cell physiology, including modulation of cell death pathways, energy metabolism, cell-cell communication, and RNA processing. While looking at the early response (1h) after IR-induced DSBs, our data, on one hand, is in line with the previous observations that IR-induced DNA damage lead to the cell cycle modulation, while

on the other hand, reveals enrichment of histone acetyltranferase activity related genes in the up-regulated category. This finding, although not reported before, is consistent with the observation that histone acetylation is induced early after DNA DSBs [Tamburini and Tyler, 2005]; [Murr et al., 2006]. The late response (6h) in FA-D2 corrected cells again showed histone acetyltransferase activity among up-regulated sequences, which clearly indicates this function as an integral part of IR-induced DNA damage response. At six hours, other chromatin related activities were also observed among the up-regulated genes including histone demethylase activity, histone ubiquitination and chromatin organization. Among the down-regulated sequences, cell cycle related functions and response to stress and radiation were enriched, which is consistent with an expected late response. Moreover, HDAC activity was also observed, which probably indicates the end of repair and restoration of chromatin to its original state by histone deacetylation. In general, the transcriptional response to IR-induced DNA damage in FA-D2 corrected fibroblasts clearly demonstrates the increased activity of chromatin related functions, which highlights the importance of chromatin signaling cascades in DDR.

In order to improve our understanding of Fanconi anemia syndrome, bearing in mind the genome instability and chemical sensitivity of FA cells to DNA-ICL inducing agents, it is imperative to study the deficiencies of transcriptional response in FA mutant cells as compared to their corrected "wild type" counterparts. A few studies previously reported the transcriptional profile of FA cells [Zanier et al., 2004]; [Martinez et al., 2008]. However, firstly, they did not compare the transcriptional response of FA cells before and after inducing DNA-ICL, secondly, they mostly focused on p53 dependent transcriptional changes in FA cells, and thirdly, they worked with FA cells mutated in core complex proteins, i.e., FANCC and FANCA respectively. We, on the other hand, performed a large scale analysis, where we not only compared the wild type versus mutant cells, but also analyzed the transcriptional response in both cells lines after inducing DNA-ICL damage. Also, we used FA-D2 gene mutated cell line (and its corrected cell version), which represents a different level of FA-pathway, down stream of core complex. At basal level our results indicate an increased HAT activity in mutant cells as compared to wild type cells, which resembles more the transcriptional response to IR in wild type cells. This finding suggests that FA-D2 mutant cells behave like IR damaged normal cells as if the defect in FA-pathway leads to accumulation of DSBs in these cells.

Next, we sought to find out the genes that are differently affected in FA mutated and wild type cells. Here, the analysis was not a straight forward direct comparison of genes affected in the two cell lines after induction of DNA-ICL damage. Rather, the special nature of experimental design lead us to apply interaction analysis. Our experimental design belongs to a more complex class of functional genomics experiments, where there is involvement of more than one factor, and thus a factorial design approach is required. Factorial experiments are used to study differences that result from the joint effect of two or more factors [Yang and Speed, 2002]; [Glonek and Solomon, 2004]. Our design corresponds to 2x2 factorial experiments, where we analyzed the transcriptional response to DNA damage in FA mutated and corrected cells. Here, we have two factors that might affect this transcriptional response to a

varying degree, these include, DNA damage and FA mutation. For some genes the response may be interactive (i.e., the effect of one factor depends on the other factor), while for other genes the response may be additive (i.e., the effect of one factor does not depend on the other factor). Therefore, in this part we tried to measure the interactive response in our model to answer two basic questions, i.e., 1- What part of the transcriptional response to DNA damage requires functional FA-pathway? and, 2- What happens to cells after DNA damage if FA pathway is inactive? We applied a simple epsilon analysis to our microarray data in an attempt to answer these questions according to equations 3.1 and 3.2, where a value of "1" was considered "no interaction".

In a recent study, Wong and colleagues [Zhou et al., 2010] employed the idea of interaction, where they used factorial analysis to analyze the impact of age on the temporal gene response to burn injury in a large-scale clinical study and revealed that 21% of the genes responsive to burn are age-specific. Their analysis is based upon the classical statistical method of analysis of variance (ANOVA), which they used to model the dependency of gene expression on the experimental factors. Our approach of factorial analysis, although different from this study, showed interesting differences in the transcriptional response to DNA-ICL and IR in FA mutant and corrected cells. In spite of the fact that most of the genes do not show any interaction, we have found much more genes showing interaction in IR treated cells than MOP/UVA treated cells at the later stages of response (6h), which either indicates that late response in IR damaged cells is highly dependent on functional FA-pathway or this may be due to the secondary and indirect effects on gene transcription. The interaction analysis not only helped us to identify the part of transriptional response to DDR that requires functional FA-pathway, but also revealed the genes that were sensitized to DNA damage due to the inactivation of FA-pathway. This kind of analysis, on one hand, certainly improves the current understanding of the transcriptional response to DDR in FA mutated and corrected cells, while on the other hand, provides a simple way of analyzing multi-factorial interaction data based on two or more factors.

4

Results - III

In the previous two sections, we have studied global changes in histone modification profile after inducing DNA damage, as well as the global transcriptional response to this damage, both in the context of their dependence on functional Fanconi anemia pathway. Some of our results are consistent with the existing literature, however we have also presented some previously uncharacterized modifications that manifest an interesting dynamics in response to DNA-ICL damage.

However, one big shortcoming of such global analysis resides in the fact that the cell is a structured system, highly organized in space. For example, nucleus is not a homogeneous object, but it contains a number of structures and domains, each with its own function. Accordingly, it might be significantly more informative, instead of studying the global changes in histone PTMs, to focus on the PTMs that are specific to a particular nuclear domain that is most relevant for the cellular response to a DNA damage. A canonical example here is γH2AX foci, which mark the regions of active DNA damage and repair very early after DSB induction. Accordingly, it would be important to study the histone PTMs in a more localized regions of the chromatin that bear most relevance to DNA repair (such as repair foci), which could be significantly more informative for understanding the role of histone code in DDR. However, so far, there was no methodology that would allow one to accomplish such a task in hight throughput manner.

In our next section we describe a new approach to study local changes in histone PTM profile in the vicinity of a particular protein of interest. Due to technical limitations (mainly, large size of FANCD2 protein), we used an experimental model, unrelated to Fanconi anemia, to develop and validate this methodology. We have chosen an E3 ubiquitin ligase, Rad18, which is convenient, as it forms detectable foci in the nucleus. These foci represent the repli-

cation factories, although Rad18 foci increase significantly after UVC irradiation induced DNA damage. Using our novel methodology, we were able to purify histones in the proximity of this protein and to analyze their PTMs with methods of quantitative proteomics. Although this approach has been validated on Rad18/UVC model, we believe that it has a much larger scope, and can be eventually used to study histone PTM changes related to DDR in Fanconi anemia.

4.1 Design and features of the system

PUB-MS (Proximity Utilizing Biotinylation) has been developed in our laboratory, originally as a new method to study protein-protein proximity (PPP-networks) in vivo using mass spectrometry (the vectors are described in [Figure 4.1-B]). Similar method, although not designed for mass-spectrometry analysis, was also developed in the laboratory of Dr. A Y Ting [Fernandez-Suarez et al., 2008].

The general principle of the method is presented in [Figure 4.1-A]. Two proteins to be tested for their proximity in vivo are co-expressed in the cells of interest, the first protein being expressed as a BirA fusion, and the second one as a fusion to the Biotin Acceptor Peptide (BAP). More efficient biotinylation of the BAP is expected when the two proteins are in proximity to each other, e.g., when interaction occurs [Figure 4.1-A(bottom)], as compared to the background biotinylation arising from random collisions between noninteracting proteins [Figure 4.1-A(top)]. The biotinylation status of the BAP fusion can be further monitored by Western blot or mass spectrometry.

To implement this principle, we constructed two types of vectors: for expression of the BirA fusion and for expression of the BAP fusion, correspondingly [Figure 4.1-B(top)]. We constructed each vector in two forms, with the protein expression regulated either by a strong CMV or a weaker MoMuLV enhancer. In this work, we used the CMV enhancer for the expression of both BirA fusions and BAP fusions. Different amounts of BirA fusion and BAP fusion vectors are cotransfected, which allowed us to achieve an excess of the BAP fusion (biotinylation target) over the BirA fusion (biotinylating enzyme), which is essential for the further quantitative analysis.

In addition, we significantly modified the sequence of the BAP, by introducing two flanking arginines that yield a peptide of an 'MS/MS-friendly' size after trypsin digestion of the BAP sequence. We also included a 7XHis tag in the BAP-domain, to have the option of purifying or/and monitoring the amounts of the target protein regardless of its biotinylation status [Figure 4.1-B(bottom)].

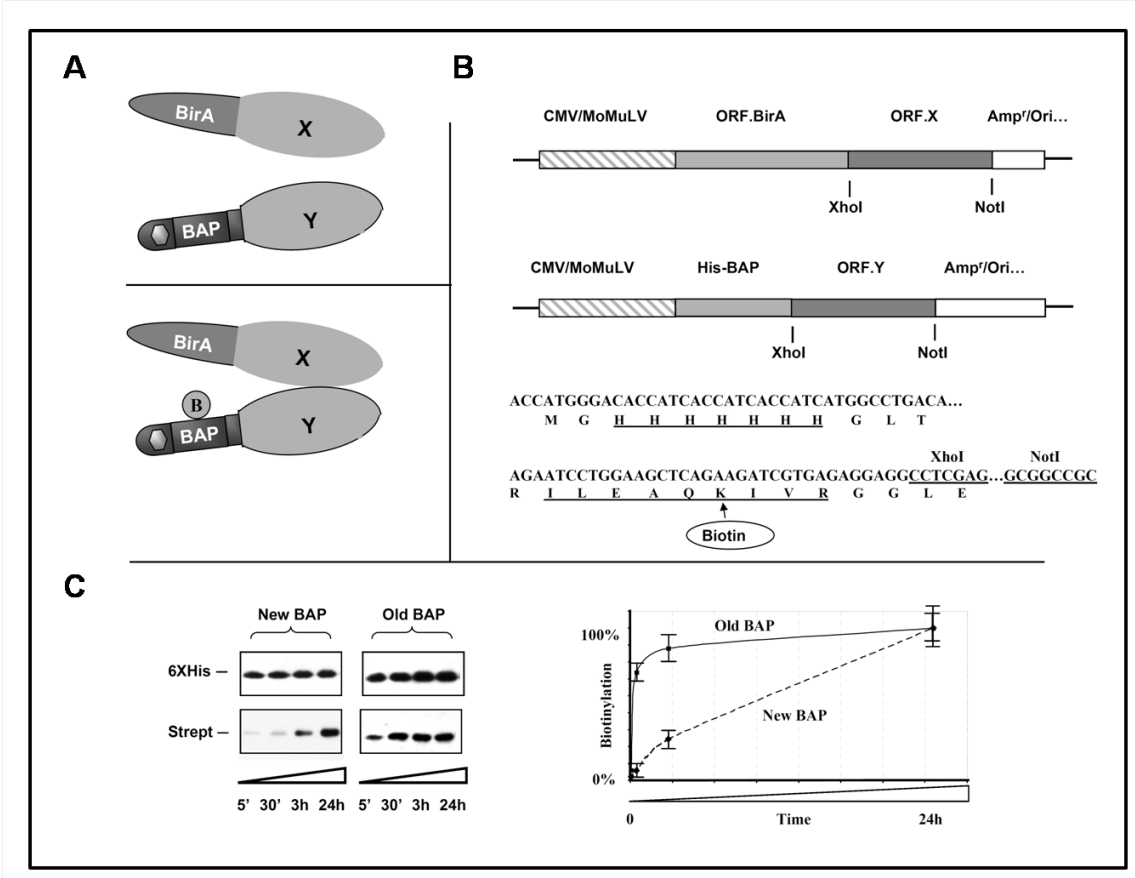


Figure 4.1: Design and features of the system. [**A**] The principle. Two candidate interaction proteins X and Y are co-expressed in the cells of interest, the first protein being expressed as a BirA fusion, and the second one as a fusion to the Biotin Acceptor Peptide (BAP). The cells are pulse-labeled with biotin. Where there is interaction (top), the BAP-X fusion should be more efficiently biotinylated. [**B**] Vector design. Top - the positions of the CMV/MoMuLV enhancers, and of the BAP and BirA sequences, relative to the cloned ORF, are indicated. Bottom - nucleotide and aminoacid sequences encoding the newly designed His-BAP are shown. [**C**] New BAP is biotinylated with slower kinetics as compared to the old BAP. Left - Western analysis of biotinylation of the new BAP-GFP (ILEAQKIVR, left) versus previously used efficient old BAP-GFP (MAGLNDIFEAQKIEWHE, right) in the presence of BirA-GFP fusion. (top) -α-6XHis-HRP signal, (bottom) - streptavidin-HRP signal. Right - quantification of the kinetics of biotinylation of the two BAPs. The streptavidin signal was first normalized by taking a ratio with the respective -6XHis signal. For every individual experiment, the normalized value of the biotinylation level for each time point was divided by the respective value for the 24h time point from the same experiment, which was taken as 100%. Plotted are averaged results from three independent experiments.

To test if the newly designed BAP could be biotinylated by BirA, we co-transfected 293T cells with BirA-GFP and BAP-GFP. Regardless of the fact that GFP can form dimers at high concentrations [Chalfie and Kain, 2005], the main aim of this experiment was not to detect the GFP-GFP interaction, but rather to test the BAP biotinylation, as even background biotinylation due to random collisions between non-interacting proteins could be sufficient to give detectable signal. As seen in [Figure 4.1-C], the biotinylation was indeed observed; moreover, as expected, the level of biotinylated BAP-GFP linearly increased with the time of incubation

of the cells with biotin. Notably, in parallel co-transfections, the GFP carrying a different BAP sequence (MAGLNDIFEAQKIEWHE), used previously [Viens et al., 2004], was biotinylated with significantly faster kinetics (the time of half-saturation was 15 min, contrasting with 680 min for the newly designed BAP). Because saturation in the BAP biotinylation levels obscures differences in biotinylation efficiency, the slower kinetics of biotinylation of the new BAP was welcome given our ultimate goal to correlate BAP biotinylation with protein-protein proximity.

4.2 Analysis of post-translational modifications of a specific protein fraction using PUB

The same protein can exist in several complexes in the same cell, and its properties can vary depending on the proximity partners (PP). Given that PUB allows labeling a PP-dependent fraction of the BAP-fusion of the protein of interest, one can expand the scope of the PUB approach beyond the mere study of protein-protein proximity. Namely, one can purify a particular PP-dependent subfraction of a BAP-fused protein biochemically, and then study its properties, e.g., its post-translational modifications (PTMs).

To test this possibility experimentally, we have analyzed the ubiquitination status of the subset of PCNA protein that is located in proximity to the translesion polymerase PolH after irradiation of cells with UV. Ubiquitination is a convenient model to study PTMs, because it is easy to detect as a shift in protein mobility on SDS-PAGE. UV irradiation induces ubiquitination of PCNA, which then recruits PolH to the sites of DNA damage [Kannouche and Lehmann, 2004]; [Prakash et al., 2005]. The PolH has a PCNA-binding and ubiquitin binding domains (PIP and UBZ, respectively), both believed to be involved in the interaction.

To analyze the ubiquitination status of PCNA located in proximity to PolH in vivo, the cells were transfected with the plasmids expressing BAP-PCNA and BirA-PolH. Two days later, the UV damage was induced and 6h afterwards, the cell were pulse-labeled with biotin. The cells were harvested, and the biotinylated BAP-PCNA was separated from the non-biotinylated BAP-PCNA using streptavidin-sepharose pulldown. The ubiquitination status of the two fractions was analyzed by Western with PCNA antibodies. As one can see from the [Figure 4.2-B, lane 5], a significant part of the biotinylated BAP-PCNA is present in ubiquitinated form, whereas such form is absent in the non-biotinylated BAP-PCNA fraction remaining in the flowthrough [Figure 4.2-B, lane 3]. Control experiment, performed in parallel with BirA-GFP fusion, showed no such enrichment in the biotinylated BAP-PCNA fraction [Figure 4.2-B, lane 2 & 4].

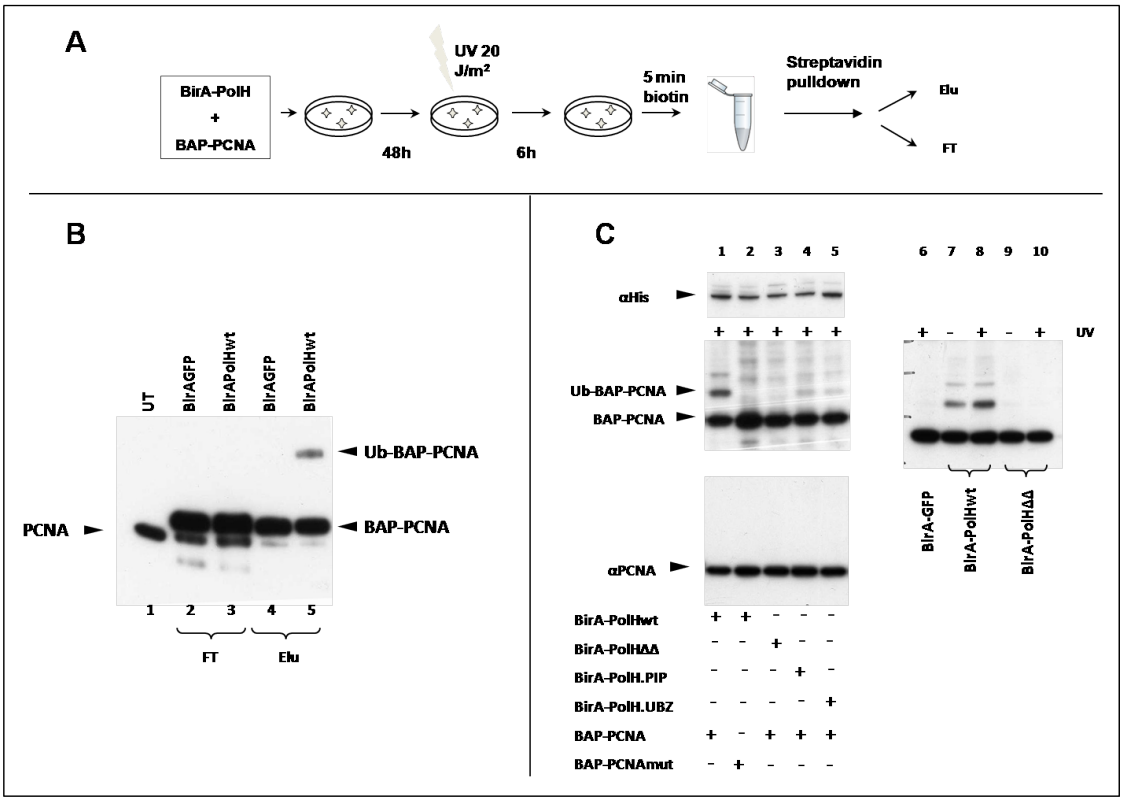


Figure 4.2: Analysis of post-translational modifications of a specific protein fraction using "PUB". [**A**] Experimental scheme. 293T cells are transfected with the plasmids expressing BAP-PCNA and BirA-PolH. Two days later, the cells are treated with UVC (20 J/m^2) and 6h afterwards, the cell are pulse-labeled (5′) with biotin. The cells are harvested, and the biotinylated BAP-PCNA is separated from the non-biotinylated BAP-PCNA using streptavidin-sepharose pulldown. [**B**] Ubiquitination status of PolH-proximal PCNA. Western with α-PCNA antibodies. 1 - untrasfected cells. 2,4 - BAP-PCNA cotransfected with BirA-GFP fusion. 3,5 - BAP-PCNA cotransfected with Bir-PolH fusion. 2,3 - Flowthrough fraction, 4,5 - Eluate. Note that the BAP-PCNA from the flowthrough fraction was further purified via Ni-NTA chromatography, in order to decrease the signal from the endogenous PCNA. The endogenous PCNA, the BAP-PCNA fusion and the ubiquitinated BAP-PCNA are indicated by arrows. [**C**] Ub-BAP-PCNA is preferentially biotinylated due to its specific interaction with PolH. Western analysis of nuclear extracts from transfected 293T cells. Middle - strepatividin-HRP, detecting two forms of biotinylated BAP-PCNA, in the presence of different Bir-PolH fusions, treated or untreated with UV. Top left: Expression of different forms of BirA-PolH, detected by α-His Western. Bottom left: PCNA in the transfected 293 cells, showing that BAP-PCNA constitutes undetectable fraction of total PCNA in the transfected cells.

To confirm that the ubiquitinated fraction of BAP-PCNA is preferentially biotinylated due to its specific interaction with PolH, we used mutations in PolH lacking the PIP or/and UBZ domains in our next experiments. This time, the ubiquitination status of the biotinylated BAP-PCNA was analyzed directly in nuclear lysates by Western with streptavidin-HRP. As one can see on the [Figure 4.2-C middle], the higher molecular form of biotinylated PCNA (lane 1) was decreased after we used the BirA-PolH PIP and BirA-PolH UBZ, and practically disappeared when the double mu- tant BirA-PolH was used (lanes 4,5 and 3, respectively). Consistent with the notion that the higher molecular fraction of BAP-PCNA corresponds

to ubiquitinated BAP-PCNA, a BAP-PCNA carrying a mutation in the ubiquitination site showed no such fraction [Figure 4.2-C lane 2], when cotransfected with BirA-PolH in a parallel experiment.

We conclude from these experiments, that the purification of biotinylated fraction of a BAP-fusion of a protein of interest allowed us to study PTMs specific for the particular fraction of this protein that is located in proximity to another protein of interest, fused to BirA. Whereas in the case of ubiquitination, the PTM analysis could be done by Western of cellular extracts with streptavidin-HRP, in most other cases it will require biochemical purification of the biotinylated fraction of the protein.

4.3 PUB-NChIP reveals a specific pattern of H4 acetylation in the Rad18-proximal chromatin

Next, we applied the same principle for the analysis of chromatin in proximity to a protein of interest fused to BirA. For this purpose, we fused the BAP domain to the core histone proteins H3.1 and H2A. After incorporation to chromatin, their biotinylation by a protein of interest fused to BirA is expected to lead to specific labeling of chromatin located in the proximity of the BirA-fusion. Purification and analysis of such biotinylated chromatin is the principle beyond the PUB-NChIP approach.

To verify that only the BAP-histones that are located in proximity to a BirA fusion are biotinylated in such conditions, we used repair-related Rad18 protein as a convenient model. This protein is known to form characteristic foci in the nuclei, clearly seen after UV treatment of cells (these foci are believed to correspond to specialized compartments in the nuclei, containing repair and replication proteins, including PCNA). HEK 293T cells were co-transfected with the BirA-Rad18 fusion together with BAP-H2A. Two days after transfection, the cells were irradiated with UV and pulse-labeled with biotin 6h later, i.e., at the time when Rad18 forms distinct foci in the nuclei. The cells were fixed immediately after the biotin pulse. The staining of the pulse-labeled cells with streptavidin-Cy3 revealed foci of biotinylation signal, which colocalized with Rad18 [Figure 4.3-A top], detected by specific antibodies. Importantly, staining with H2A antibodies that detect H2A regardless of its biotinylation status, gave very different pattern, consistent with known homogeneous distribution of H2A histone over the nucleus [Figure 4.3-A bottom]. Western analysis [Figure 4.3-B bottom left] shows that only the BAP-H2A histone is biotinylated in cells under these conditions, thus the Cy3 signal that we observe in microscopy should reflect the localization of biotinylated BAP-H2A. On the other hand, pulse labeling for different time lengths shows that during the 15 minutes of exposure to biotin, only a small part of BAP-H2A is biotinylated [Figure 4.3-B bottom right]. These data clearly demonstrate that only the part of BAP-H2A containing chromatin located in proximity to Rad18 was biotinylated in our experiments.

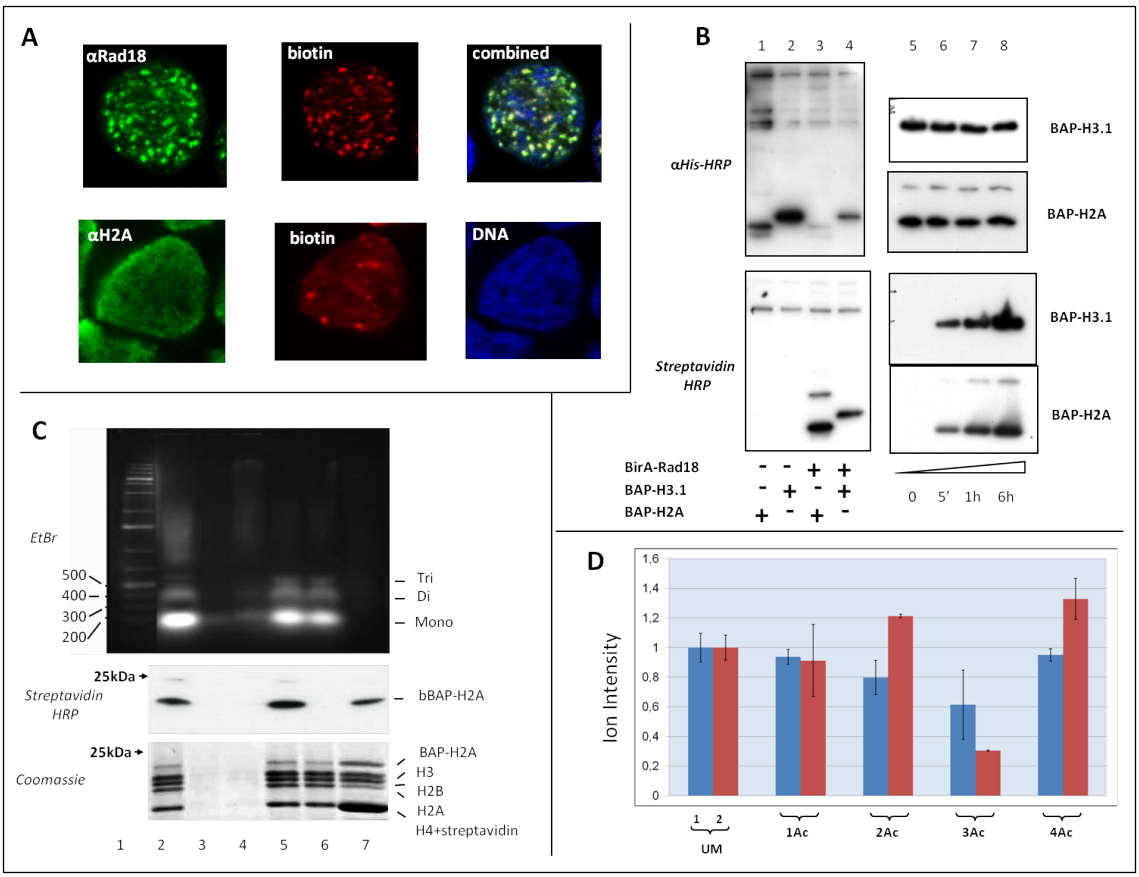


Figure 4.3: PUB-NChIP reveals a specific pattern of H4 acetylation in the Rad18-proximal chromatin. **[A]** Chromatin is biotinylated at the sites of RAD18 foci. Confocal microscopy analysis of colocalization of (top) Rad18 and biotinylated chromatin or (bottom) H2A histone and biotinylated chromatin. **[B]** Biotinylation signal is due to BAP-H2A in proximity to Rad18-BirA. Left: Western blot analysis with streptavidin-HRP, showing that a specific signal appears only after BirA-Rad18 and BAP-histones are cotransfected. 1 - BAP-H2A, 2 - BAP-H3, 3 - BirA-Rad18 + BAP-H2A, 4 - BirA-Rad18 + BAP-H3. Top - α-His antibodies, bottom - streptavidin-HRP. Right: Western blot analysis of biotinylation status of BAP-H2A or BAP-H3 after different times of biotin-labeling. Top -α-His antibodies, bottom - streptavidin-HRP. **[C]** Purification of biotinylated chromatin. Bottom: SDS-PAGE gel stained with coommassie blue, with the identities of histones indicated. The H4 band in elution also contains streptavidin, and thus is thicker than the other histone bands. Middle: Western analysis of the same samples with streptavidin-HRP, showing that all biotinylated BAP-H2A was pulled down with immobilized streptavidin. Top: Agarose gel electrophoresis and EtBr staining for the same samples, showing that mostly the mononucleosome fractions were purified. **[D]** Relative acetylation status of H4 N-terminal tail (aa 4-17). The relative presence of a particular form of H4 tail (UM - unmodified, 1Ac, 2Ac, 3Ac, 4Ac - mono, di- tri- and tetra- acetylated forms, respectively), as a ratio to the unmodified form. 1 - control BirA-GFP cotransfected with BAP-H2A, 2 - BirA-Rad18 cotransfected with BAP-H2A.

The biotinylated chromatin was purified on streptavidin sepharose without prior crosslinking. As one can see from the [Figure 4.3-C bottom, lane 7], the composition of the purified histone octamers is consistent with the presence of one BAP-tagged H2A/H2B protein. Whereas in both input and flowthrough samples [Figure 4.3-C bottom lane 5 & 6, respectively], all core histones are present in stoichiometric amounts, there is approximately twice as less of H2A in the eluate. Judging by the Western with streptavidin-HRP, all biotinylated

BAP-H2A is purified [Figure 4.3-C middle, lane 7]. However, confirming that not every BAP-H2A molecule is biotinylated in our experiment, the flowthrough also contains BAP-H2A band stained by Coomassie [Figure 4.3-C bottom, lane 6]. Note that after elution from streptavidin beads by boiling with Laemmli buffer, significant amounts of streptavidin appear in the sample, overlapping with the H4 histone band [Figure 4.3-C bottom, lane 7]. As expected, the comparison of DNA content of the input, flow-through and eluate samples indicates that only a small fraction of total chromatin has been purified in this experiment [Figure 4.3-C top, lane 5, 6 & 7, respectively].

Next we set out to analyze the post-translational modifications, and compare them with control BirA-GFP biotinylated chromatin. We used SILAC (Stable Isotope Labeling with Aminoacids in Culture) in order to quantitatively compare the presence of specific modifications in the chromatin in proximity to Rad18, as opposed to the GFP-BirA fusion. The cells were grown in medium, containing either light version of lysine and arginine aminoacids or their heavy counterparts (labeled with C^{13} and N^{15}). The heavy labeled cells were cotransfected with BAP-H2A and either BirA-RAD18 or BirA-GFP. The control, light labeled cells were cotransfected with BAP-H2A and BirA-GFP. As in the previous experiment, two days after transfection, the cells were UV irradiated, and 6h afterwards they were pulse-labeled with biotin for 15 min. Then the cells were harvested, mixed, and biotinylated chromatin was purified as described above. The bands corresponding to histones were excised, treated with trypsin and then MRM analysis was performed to quantitatively detect the presence of different histone PTMs and compare their presence between the experimental and control samples (Rad18 and GFP, respectively).

The [Figure 4.3-D] shows that the H4 histone in proximity to BirA-Rad18 has an increased acetylation of its N-terminal tail (aa 4-17). Notable, the di- and tetra- acetylated forms are over-represented, as compared to H4 from the control chromatin biotinylated by GFP-BirA-fusion. Intriguingly, we have also observed under-representation of the tri-acetylated form of H4 in the Rad18 proximal chromatin.

4.4 The Rad18-specific pattern changes after the proximity with Rad18 has been diminished

The distinctive feature of the PUB-NChIP method is that chromatin in proximity to the BirA-fusion is left with a permanent molecular mark (i.e. biotin), which can persist after the proximity between BAP- and BirA-fusions have been lost. We decided to take advantage of this property to ask what happens with the H4 acetylation pattern of the Rad18-labeled chromatin at a determined time after the biotinylation.

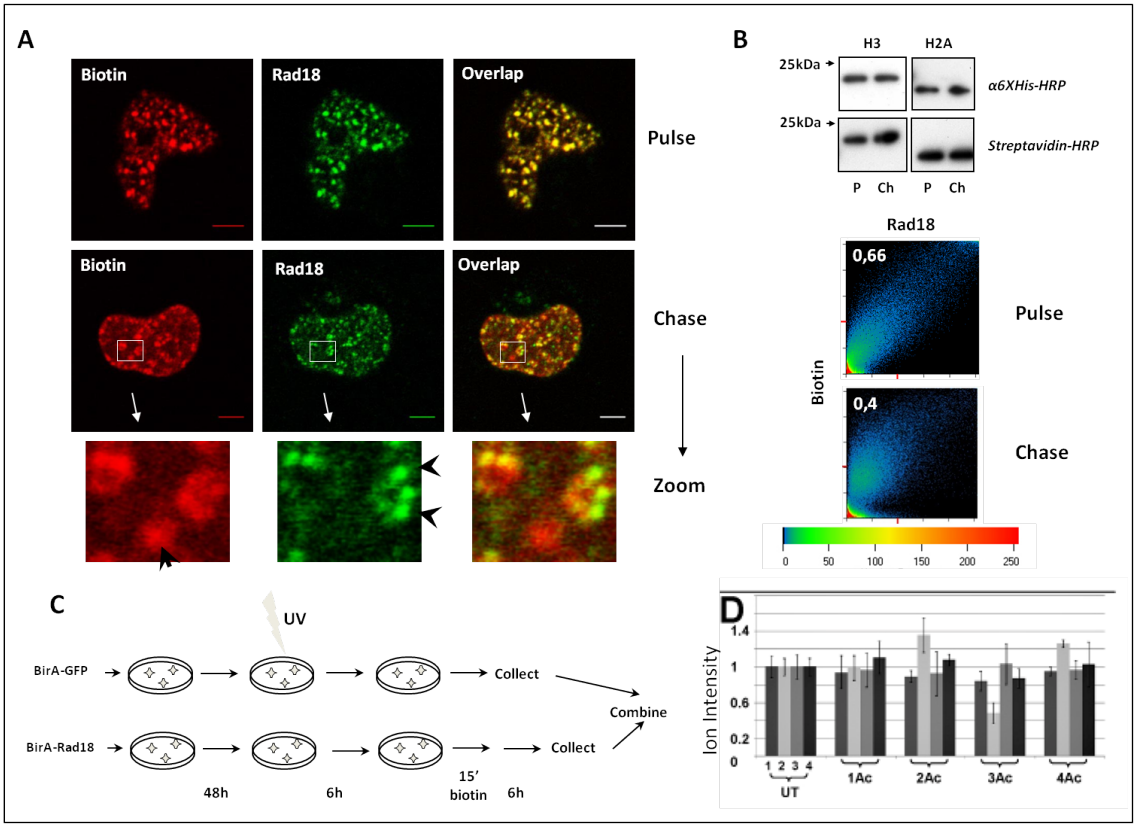


Figure 4.4: The Rad18-specific pattern changes after the proximity with Rad18 has been diminished. **[A]** Decrease in colocalisation between Rad18 and biotinylated chromatin after 6h chase. Top - Confocal microscopy showing strong colocalization between Rad18 and biotinylated chromatin in the nucleus of 293T cell immediately after pulse labeling. Middle - Confocal microscopy showing different localization of Rad18 and biotinylated chromatin 6h after pulse labeling. Bottom - zoomed area from the chase sample showing an example of biotinylated foci that do not colocalize with the Rad18 foci and vice versa (indicated by arrows). Right - scatterplot with the coefficient of correlation (top left corner) between the intensities of the red and green signals for every pixel (above a background threshold), showing stronger colocalization of the biotin and Rad18 signals in the pulse sample. **[B]** No increase in biotinylation signal after chase. Western analysis showing that the level of biotinylation of BAP-H3 and BAP-H2A is not increased 6h after the cells have been washed from biotin. Top - α-His antibody, bottom - strepavidin-HRP. P- pulse sample, Ch - chase sample. **[C]** The scheme of pulse-chase experiment. Cells grown in light SILAC medium are transfected by BirA-GFP and BAP-H2A (reference sample). Cells grown in heavy SILAC medium are transfected with BirA-Rad18 (or BirA-GFP) and BAP-H2A. The heavy labeled cells are either harvested immediately after biotin labeling or washed from biotin and left for 6h before harvest. The heavy and light labeled cells are mixed and chromatin is prepared as in [Figure 4.3]. **[D]** Acetylation status of H4 N-terminal tail. Relative presence of a particular form of the H4 tail (marked as in the [Figure 4.3]). Heavy labeled samples were mixed with the light labeled 'BirA-GFP + BAP-H2A pulse' sample. The heavy samples are: 1 - BirA-GFP pulse, 2 - BirA-Rad18 pulse; 3 - BirA-GFP chase; 4 - BirA-Rad18 chase.

First, we used immunofluorescent microscopy to monitor the distribution of biotinylated chromatin 2h after its labeling by BirA-Rad18. Until the biotin labeling, the cells were treated as in previous experiment. Then they were either fixed immediately after the biotin pulse (pulse sample), or else intensively washed to eliminate biotin and then left in fresh medium

for 2h before fixation (chase sample). Western analysis shows that the biotinylation levels are not increasing after the chase [Figure 4.4-B bottom]. Consistent with the previous experiments, streptavidin-Cy3 staining of the pulse-labeled cells revealed foci of labeled BAP-H2A colocalizing with Rad18 [Figure 4.4-A top]. However, the chase sample manifested a different pattern [Figure 4.4-A middle]. Whereas Rad18 formed foci, the biotin signal appeared more diffuse, and we generally observed a smaller number of foci per cell, whereas the colocalization of the remaining biotinylated foci with the PCNA or Rad18 foci was considerably less pronounced [Figure 4.4-A bottom for an example].

The observations are consistent with the notion that after BAP-H2A was biotinylated in the proximity of BirA-Rad18 at the moment of the biotin pulse labeling, the Rad18 started to associate with different parts of chromatin, as could be seen by the loss of colocalization of the Rad18 (and PCNA) foci with the biotinylated histones. This observation demonstrates that one can monitor the state of proximally-biotinylated chromatin at later times after labeling - e.g., after the proximity with Rad18 could have been lost.

At the next step, we purified the PUB-labeled chromatin after 6h chase and analyzed the pattern of H4 acetylation. The SILAC labeling was again used to quantitatively compare the modifications between different samples. All experimental samples were grown in heavy medium, and as a common reference sample (grown in the light medium), the pulse-labeled "BirA-GFP + BAP-H2A" transfected cells were used. Judging by the more modest H/L ratios in the case of the chased sample, the characteristic under-representation of the tri-acetylated form, which was clearly seen in the RAD18- proximal chromatin at the moment of biotinylation pulse (i.e., 6h after UV irradiation), became less pronounced and more resembling the modification pattern of the control chromatin, i.e., biotinylated by GFP-BirA.

4.5 The pattern of H4 acetylation near Rad18 is different in the case of H2AZ containing chromatin

Another advantage of PUB-NChIP method resides in the fact that, instead of the canonical core histones (e.g., H3.1, H2A), one can use their alternative forms, which have special role in chromatin functioning and have been shown to preferentially associate with different functional states of chromatin.

Immunofluorescent microscopy together with Western analysis shows that, similar to the canonical histone H2A, co-transfection of BirA-Rad18 with BAP-H2AZ, BAP-macroH2A and BAP-H2A.BBD leads to appearance of biotinylated foci in the nuclei that are due to the BAP-histone fusion [Figure 4.5-A & B]. As expected, BAP-macroH2A foci are mostly coincident with DNA-dense areas of nuclei, whereas BAP-BBD foci associate with less densely stained areas. Given a more complicated distribution of H2AZ in the nucleus [Viens et al., 2006], no correlation with DNA density was detected for this protein. Overall, the results are consistent with the notion that using alternative histones in PUB-NChIP should lead to labeling of functional chromatin states associated with these histone variants.

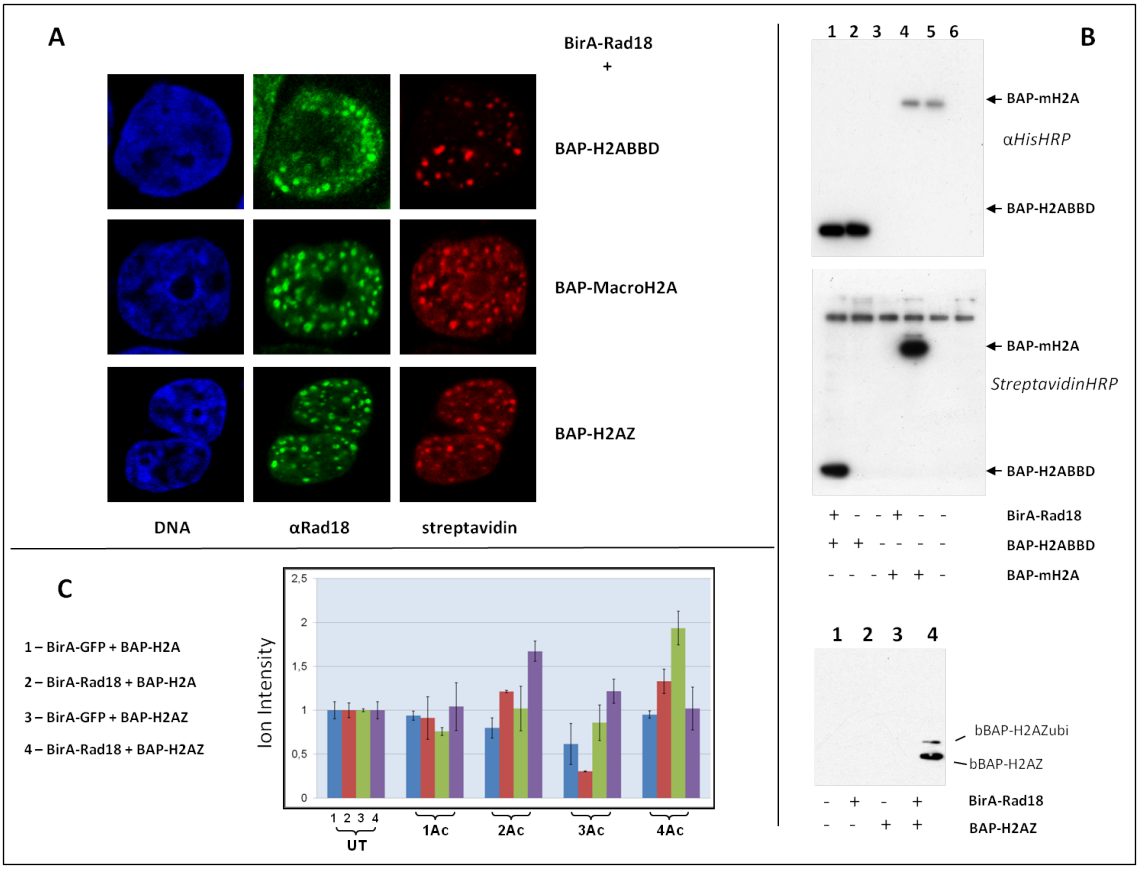


Figure 4.5: Use of alternative histone variants. **[A]** Alternative histones are biotinylated at the locations of Rad18. Confocal microscopy of 293T cells cotransfected by BirA-Rad18 and BAP-H2ABBD (Top), BAP-macroH2A (Middle) and BAP-H2AZ (Bottom). Shown are staining with Topro-3 (left), Rad18 (middle) and streptavidin (right). **[B]** BAP-histone fusions. Top, Middle: Western analysis, showing that a specific biotinylation signal appears only after BirA-Rad18 and BAP-macroH2A or BAP-H2ABBD fusions are co-transfected. 1,4 - BAP-histone + BirA-Rad18; 2,5 - BAP-histone; 3,6 - control untransfected samples, Top - α-His antibodies, Middle - streptavidin. Bottom: The Western analysis for the BAP-H2AZ fusion. The two forms of H2AZ (non-ubiquitinated and ubiquitinated), both biotinylated in this experiment, are indicated by arrows. **[C]** Acetylation status of H4 N-terminal tail. Relative presence of a particular form of the H4 tail (marked as in the [Figure 4.4]). Heavy labeled samples were mixed with the light labeled 'BirA-GFP + BAP-H2A pulse' sample. The heavy samples are: 1 - BirA-GFP + BAP-H2A, 2 - BirA-Rad18 + BAP-H2A; 3 - BirA-GFP + BAP-H2AZ; 4 - BirA-Rad18 + BAP-H2AZ.

To test the possibility to study histone PTMs both in the vicinity of a protein of interest and in context of a particular alternative histone, we performed SILAC experiment with alternative histone H2AZ, which is implicated in many cellular functions, including cellular response to DNA damage [Ahmed et al., 2007]; [Kalocsay et al., 2009]; [Morillo-Huesca et al., 2010].

As before, all experimental samples were labeled in heavy medium, and as a reference (light) sample, the pulse-labeled "BirA-GFP + BAP-H2A" transfected cells were used. The relative abundance of heavy versus light peptides was analyzed by MRM. As the [Figure 4.5-C] shows, when BAP-H2A is replaced by BAP-H2AZ, the levels of di-acetylated forms of H4 clearly increase in the Rad18-proximal chromatin. The same modification also appears to

increase in the control GFP-BirA labeled chromatin, containing BAP-H2AZ. The most parsimonious interpretation of this observation is that the two factors (Rad18-proximity vs GFP, and presence of H2AZ vs H2A) affect the di-acetylated form of H4 independently, resulting in the additive effect in the case of 'Rad18 + H2AZ' combination.

Strikingly, however, we have observed a non-additive response to the two factors in the case of tri- and tetra- acetylated forms of H4. Unlike in the case of 'Rad18 + H2A', there was no decrease in the tri-acetylated form in the case of the 'Rad18 + H2AZ' combination. On the other hand, the four-acetylated form was under-represented in the case of 'Rad18 + H2AZ', as compared to 'Rad18 + H2A'.

The results show that the patterns of post-translational histone modifications in proximity to Rad18 protein depends on the histone composition, and moreover, the histone replacement can affect the modification pattern in a non-trivial way.

4.6 Discussion

We describe here a new methodology to study chromatin in proximity of the protein of interest in vivo. The well established approach for such analysis is chromatin immunoprecipitation (ChIP). However, it suffers from the following limitation. Due to the cross-linking of DNA to histones, only the DNA part of chromatin purified after ChIP is useful, and the protein part is discarded, being irreversibly damaged by the cross-linking procedure. On the other hand, the analysis of histone modifications and replacement histone variants could provide valuable information about the state of chromatin around the particular protein of interest. So far, the correlations between a protein and histone PTMs could be inferred only indirectly, after comparison between the results of parallel ChIP experiments that use antibodies against the protein and antibodies against particular PTMs. Here, a more direct method to access the chromatin state is proposed.

The principle of PUB-NChIP is somewhat similar to the DamID approach, an alternative to ChIP way to map DNA binding sites of nuclear proteins [van Steensel and Henikoff, 2000]; [Greil et al., 2006]. This methodology has been used in several high profile studies and is based on expression of a fusion of bacterial Dam-methylase to a protein of interest that leads to labeling the protein's genomic binding site by adenine methylation. The label on DNA can be detected by various methods. Our approach is conceptually similar, but instead of Dam-methylase, we fuse biotin-ligase BirA to the protein of interest and, instead of DNA, we label the histone part of chromatin. The scopes of these approaches are also different, as PUB-NChIP is mostly aimed to analyze the histone PTMs. However, we cannot exclude that PUB-NChIP can be also used for DNA analysis as well. Although we have not explored such possibility in the present work, one can point to several potential advantages of PUB-NChIP in this respect. There is a wide variety of commercially available resources for detection and/or purification of biotinylated macromolecules, and the NChIP with biotinylated chromatin has been shown to work [Ooi et al., 2010]. In addition, by substituting

the canonical core histone by its replacement variants, the analysis can be made dependent on the context of particular functional state of chromatin.

As far as the histone PTM analysis is concerned, we present here a proof of principle that it can be done with the PUB-NChIP purified chromatin, focusing mostly on the acetylation status of H4 N-termini. Our approach to analyze histone PTM is based on SILAC, which allowed us to quantitatively analyze the histone modifications in the Rad18-proximal chromatin, their dynamics after labeling and their dependence on histone variant. This particular approach is well suited for the type of mass-spectrometer (ion-trap) used. However, because it requires trypsin digestion, the information about correlations between different histone modifications on individual histone molecules is not accessible. On the other hand, the more powerful mass-spectrometry instruments (such as FTICR and Orbitrap) can be used in the so called Top-down approach that targets histones without tryptic digestion [Armirotti and Damonte, 2010]; [Macht, 2009]. Using these instruments for downstream analysis of the PUB-NChIP prepared chromatin, one will be able to detect correlations between PTMs on individual histone molecule. Notably, such analysis is impossible with the traditional ChIP technique.

Another type of additional information that could be provided by PUB-NChIP is based on the fact that biotinylation leaves the BAP-histone fusion with a permanent modification, which persist after the proximity to the BirA-fusion could have been lost. We presented a way to take advantage of this feature that adds kinetic dimension to the studies of the histone PTMs in proximity to a protein of interest. Using BAP-fused histone H2Aand BirA-Rad18 as a model, we demonstrated that one can use a pulse-chase labeling with biotin to monitor the fate of histone PTMs in the chromatin that was in proximity to a given BirA-fused protein at the moment of labeling. Thus, in addition to helping to analyze histone PTMs in chromatin located close to a protein of interest, our approach has a potential to add temporal dimension to such studies.

In respect to the biological significance of our findings, the simplest interpretation is that the chromatin is vicinity to Rad18 is locally destabilized to facilitate the access to DNA damage repair machinery. This is consistent with the notion that acetylation of N-terminal tails of H4 loosens up histone-DNA interactions and hence creates a relaxed nucleosome. After the repair is completed, the chromatin returns to its original state.

The acetylation pattern of the H2AZ containing chromatin in proximity with Rad18 is more difficult to interpret. It is possible that the chromatin containing H2AZ histone in the proximity to Rad18 corresponds to a specific stage of DNA lesion processing, which could be reflected in the different histone PTM signature. Regardless of a specific mechanistic interpretation of these observations, our data strongly demonstrate that one can analyze how changes in specific histone modifications in the proximity of the nuclear protein of interest depend on the particular chromatin type, represented by a specific histone variant.

5

Conclusions and Future Lines of Research

5.1 High Throughput Analysis of Histone PTMs in Response to DNA-ICL

Cells have evolved complex signal transduction pathways to detect and repair different kinds of DNA lesions. These pathways constitute cellular response to DNA damage (DDR - DNA Damage Response), which in eukaryotes occurs primarily in the context of chromatin. The chromatin, a DNA-protein complex, that exists as arrays of nucleosomes containing two copies of each H2A, H2B, H3, and H4, wrapped with 146 bp of DNA, is a dynamic structure that regulates DNA accessibility during essential nuclear events, such as replication, transcription, recombination, and DNA damage repair. The chromatin structure has several levels of compactness and needs to be modulated in the event of DNA damage. This modulation is regulated by various mechanisms, the most important among them being post-translational histone modifications mainly occurring on histone tails protruding from the nucleosome.

Histone PTMs, in the form of "histone code" have largely been studied and most well characterized in the context of transcription regulation, however, their precise role in DDR remained in the shadows. Apart from the most well known example of γH2AX and its role in DDR, the roles of other histone modifications that accompany this response remain imprecise and mostly unknown. Given an increasing evidence that histone PTMs are critical in modulating many process related to chromatin, it is of utmost importance to characterize their role in cellular response to various kinds of DNA damages.

Accordingly, in this thesis we have quantitatively analyzed various core histone PTMs in re-

sponse to DNA-ICL damage, a particular type of DNA damage where the two DNA strands are covalently linked and form a formidable barrier to replication fork progression. The cells respond to this kind of DNA damage in a sophisticated way by activating a specialized pathway named the Fanconi anemia pathway, which is dependent upon the concerted activities of at least 13 FA proteins. The cells mutated in any one of these FA proteins can serve as a model to study cellular response to DNA-ICL damage.

In the first part of this thesis, we have presented data on histone PTM changes in cellular response to DNA-ICL damage, where we were able to identify several histone PTM modifications that show interesting dynamics in response to DNA-ICL, often in the manner dependent on the functional FA pathway.

Some of these modifications, for example, acetylation of the N-terminal tails of histone H3 and H4, have a relatively straightforward interpretation, whereas the role of various monomethylated lysines requires further investigation.

In any case, the variable behavior of histone PTMs in the presence and absence of functional FA lead us to propose a model of "histone code" role in response to DNA-ICL damage. According to this model, DNA-ICL induction leads to local destabilization of chromatin around the DNA damage to facilitate access to not only DNA damage sensing and repair machinery but also to additional chromatin modifying and remodeling factors. The local relaxation of chromatin leads to surfacing of core domain residues, which may become more efficiently modified and in turn serve as docking sites for repair factors and checkpoint proteins. The late acting modifications leads to the restoration of chromatin to its original state. The dependence of some modifications on functional FA pathway indicates that FA proteins and hence the FA pathway may be the prerequisite for such modifications to occur, and this could be partly how the FA complex coordinates the response to DNA-ICL damage.

Concerning further lines of research of the histone PTMs in DNA-ICL damage response, it would be interesting to take into account other modifications, such as phosphorylation and ubiquitination. Moreover, given the fact that other core histones, such as H2A and H2B, are equally prone to PTMs as histone H3 and H4, they should also be included in the histone code analysis in the context of DNA damage response. Our limitation, however, was the choice of mass spectrometer, where we could only perform "bottom up" mass spectrometry, using trypsin digested samples. The comprehensive analysis of global histone modifications on all the core histones using a high resolution and more accurate mass spectrometer, such as Orbitrap or FT-ICR employing the "top down" or "middle down" approach, would be much more informative.

Whereas we had done some search for the potential histone code 'writers' and 'erasers' that could be responsible for the observed histone PTM pattern, the downstream effects of these modifications in response to DNA-ICL remains elusive. One perspective line of research is to search for the histone code 'readers' that recognize these modifications. This could be accomplished by immobilization of the histone peptides carrying the modifications of interest on a solid support and then performing affinity pulldown experiments followed by mass-spectrometry, with the aim of identifying proteins that specifically recognize these PTMs.

5.2 Transcriptional Profile in DNA-ICL Damage Response

The histone PTMs are result of coordinated activities of various histone modifying enzymes, i.e., histone code "writers" and "erasers". In order to gain insight into their plausible role in DNA-ICL damage response in FA, we employed functional genomics approach. One goal of such an approach was to monitor the changes in expression of known histone modifying enzymes, mainly related to lysine acetylations and methylations, in order to find genes that best correlate in their expression profiles with the changes in histone PTMs that we have observed.

We were able to find candidate enzymes that could be responsible for some of the changes in the histone modifications that we have observed after induction of DNA-ICL damage in the wild type and FA-defective cells. In some cases, such as H4 acetylation, we found a single gene with the expression levels that were matching the changes in this modification through the samples studied. In most of the other cases, it was possible with two enzymes having opposite activities (such as HATs vs HDACs).

However, we again have to address several technical and interpretational caveats, discussed previously. Two of these caveats are related to the inconsistency between our histone code and functional genomics analysis. First, the more correct way to compare the histone code data and functional genomics data would be to perform the experiments on the same cell lines. Although the two FA-C and FA-D2 cell lines give the same phenotype, they still represent different steps of FA-pathway. Therefore, the direct comparison of proteomics data with genomics data would only be possible by performing the experiments on the same cell line. Second, the sample comparison in our proteomics analysis should have been done in the same way as in the functional genomics analysis. Namely, control untreated cells should have been taken as a reference sample, and labeled with one isotope, whereas all other samples (including the control untreated as well) labeled with another isotope and mixed with the reference sample. In fact, such an analysis that takes these two caveats into account, is currently underway in the lab.

Finally, one has to keep in mind that the changes in gene expression level do not necessarily translate to the changes in protein levels, let alone to the increased levels of enzyme activity. Various post-translational regulation steps could also affect enzymatic activity, without changes in the protein amounts. Also, their recruitment to chromatin could be affected via changes in interaction with other proteins. Such analysis, however, was beyond the scope of our study. In any case, the results of transcriptional analysis of histone modifying enzymes showed intriguing differences between wild type and mutated cells both at the basal level and also after induction of DNA-ICL damage. These differences may reflect the dependence of these enzymes on the presence of functional FA-pathway. Given that many histone modifying enzymes have, in addition to histones, other protein targets, these data might prove useful for better understanding of cellular response to DNA-ICL damage.

The second goal of functional genomics analysis was to analyze the total transcriptome after damage and in mutant and wild type cells. We also included in this analysis IR irradiation,

as we wanted to dissociate more generic response to DNA damage from a more specific changes that are associated with the role of FA pathway to the DNA ICLs.

We were able to detect many differences in the transcriptional response to DSB and DNA-ICL. Early after damage, significantly more genes were up-regulated in the case of DNA-ICL, mostly comprising transcriptional regulation and signal transduction. Interestingly, in the case of DSB damage, we detected an enrichment in histone acetylation activity, already early after treatment. Also, consistent with the faster response to IR, cell cycle arrest-related signature was detectable early after DSB induction, whereas DNA-ICL was causing such response only at later times.

Intriguingly, practically no common genes were detected among early responders to DNA-ICL and IR. Later, however, both treatments were inducing cell proliferation arrest and apoptosis activation, which most likely constitute a generic cellular response to any kind of DNA damage.

Comparison of transcriptomes of FA mutant and control cells revealed dramatic differences between these cells. Intriguingly, many chromatin-related activities were affected by the FA mutation, which was somewhat reminiscent of the cellular response to DSB. This surprising observation might reflect the fact that defect in FA pathway leads to permanent elevation of DSB levels in the cells (e.g., resulting from stalled replication forks).

Finally, by performing an interaction analysis, we were able to find the part of transcriptional response to DNA damage that was requiring functional FA pathway, as well as find the genes that were sensitized to DNA damage by the inactivation of FA pathway. We found that a number of transcriptional regulators are indeed dependent on FA pathway in their response to DNA-ICL. Surprisingly, we observed that inactivation of FA-pathway sensitizes some cell-cycle regulatory genes to DNA-ICL.

As far as the global cellular response to DNA-ICL is concerned, it would be beneficial to complement the transcriptome data by proteomics data. In addition to histone PTM analysis, the SILAC technique also allows to perform the comparison of various sub-fractions of proteins, which will complement both our analysis of histone PTMs as well as transcriptome comparisons. Again, with the ion trap mass spectrometer, the proteomic analysis requires so called gel-LC/MS protocol, a tedious and time consuming procedure, which allows to detect maximum of 500 proteins. However, with more powerful instruments, such as Orbitrap or FT-ICR one can perform in-solution trypsin digestion and directly analyze the resulting peptide mixture on LC-MS/MS, reliably detecting up to 2000 different proteins. In addition, one can reduce the dynamic range of the samples by using commercially available "ProteoMiner" columns [Boschetti and Righetti, 2008a]; [Boschetti and Righetti, 2008b], which when combined with SILAC not only allows to reduce the dynamic range of proteins in a given sample, but also permits to keep the quantitative information about compared proteomes.

5.3 PUB-NChIP

In the last part of this thesis, we presented a new methodological approach to study histone PTMs in a more localized context instead of global histone PTM analysis. The technique we have developed, called PUB-NChIP (Proximity Utilizing Biotinylation and Native Chromatin ImmunoPrecipitation) is based on co-expression of a) the protein of interest, fused with the bacterial biotin ligase BirA together with b) a histone fused with BAP (biotin acceptor peptide), which is specifically biotinylated by BirA. With the incorporation of BAP-histones into chromatin, the chromatin located in proximity of the BirA-fusion of protein of interest is preferentially biotinylated. The further steps of the procedure take advantage of the strong interaction between core histones and DNA, that allow one to avoid cross-linking for the purification of the biotinylated chromatin (so called native ChIP). In the absence of the crosslinking-induced modifications, the analysis of the protein component of the chromatin (e.g., histone post-translational modifications) becomes feasible.

We have chosen an E3 ubiquitin ligase, Rad18, as an experimental model to develop and validate PUB-NChIP. However, we believe that this approach has a much larger scope, and can be eventually used to study histone PTM changes related to DDR in Fanconi anemia.

In this part of the thesis we also demonstrated additional advantages of the PUB-NChIP approach.

First, the distinctive feature of this method is that chromatin in proximity to the BirA-fusion is left with a permanent molecular mark (i.e. biotin), which can persists after the proximity between BAP- and BirA-fusions have been lost. We took advantage of this property by showing that we can follow the dynamics of histone PTMs over a given period of time in a quantitative way. Namely, by using biotin pulse-labeling, we could see what happens with the histone modifications in the chromatin that was in proximity to a protein of interest at a determined moment of time before cell harvesting.

The second advantage of our technique is the possibility to use, instead of the canonical histones, their alternative variants expressed as BAP-fusions. Given that different functional states of chromatin are enriched in different alternative histone variants, one can use these histones to narrow down the PTM analysis to chromatin corresponding to a specific functional state.

We have to note that, because of the scarcity of material obtained in PUB-NChIP, we were able to analyze only the histone H4 modifications in proximity to Rad18, because they are relatively easy to detect. Further research will focus on other histone PTMs.

Concerning further perspectives of PUB-NChIP, our analysis of PTMs was based on SILAC, which allowed us to quantitatively analyze the histone modifications in the Rad18-proximal chromatin, their dynamics after labeling and their dependence on histone variant. However, due to technical limitations of ion-trap mass-spectrometer, we could not study the combinatorics of histone PTMs on the individual histone molecules. The more powerful mass-spectrometry instruments (such as FT-ICR and Orbitrap) can be used in "Top-down"

approach that targets histones without tryptic digestion [Armirotti and Damonte, 2010]; [Macht, 2009]. Using these instruments for downstream analysis of the PUB-NChIP prepared chromatin, one will be able to detect correlations between PTMs on individual histone molecule. Importantly, such analysis is impossible with the traditional ChIP technique.

In addition to histones, many other proteins sufficiently strongly associate with DNA and thus are expected to be purified in PUB-NChIP. Thus, instead of the histone component, one can also use PUB-NChIP for analysis of the non-histone fraction of the chromatin in proximity to the nuclear protein of interest. This task is complicated in the case of regular ChIP methodology because of the use of chemical crosslinking that irreversibly damages the protein part of chromatin, making it difficult to analyze by mass-spectrometry.

In addition to the analysis of the protein component of chromatin in proximity to the protein of interest, the scope of our PUB-NChIP approach can be broadened by using it for the analysis of its DNA fraction as well. Although we have not explored such possibility in the present work, one can point to several potential advantages of PUB-NChIP, as an alternative to the traditional ChIP. There is a wide variety of commercially available resources for detection and/or purification of biotinylated macromolecules, and the NChIP with biotinylated chromatin has been shown to work [Ooi et al., 2010]. In addition, we can also perform this analysis in the context of a particular functional state of chromatin by using replacement histone variants, which is impossible in the case of regular ChIP methodology.

To summarize, the PUB-NChIP methodology that we have developed in this part of the work, can significantly expand the repertoire of techniques available for epigenetic research. In particular, it will allow one to analyze the role of histone PTMs in cellular response to DNA damage from a new, and hopefully, more informative perspective.

5.4 General Conclusions

Differential gene expression between two different samples is commonly determined by mRNA profiling, with the aid of microarrays and recently also by high throughput sequencing technology. Although these methods achieve impressive coverage of global mRNA expression, they are based on the fundamental assumption that mRNA levels are accurately reflecting protein abundance. This assumption is not always true, as protein post-transcriptional regulation, for example by miRNAs or by targeted protein degradation, escapes detection in mRNA profiling experiments. Therefore, as an aid to large scale transcriptional profiling, MS-based proteomics is particularly attractive as it can be employed to sequence proteomes as comprehensively as possibly.

In this thesis we have addressed the role of chromatin in DNA damage response using various methodologies encompassing functional genomics and proteomics. First, we analyzed histone code in the context of a specific kind of DNA lesions in Fanconi anemia screen using quantitative proteomics methodology, SILAC, which enabled us to quantify changes in histone PTMs in the presence and absence of DNA-ICL damage. We have successfully identified

a number of histone PTM marks in histone H3 and H4, acetylations and methylations, which have shown dependence upon functional FA-pathway. Since, chromatin is the physiological template for all DNA-templated cellular process, histone PTMs most invariably represent an important signaling mechanism that take part in DDR cross talk and coordination. Therefore, identification of various histone PTM marks, which takes part in FA-pathway of DDR signaling is a first step towards enlightening the role of chromatin in DDR.

The histone PTM analysis represents a "global" form of histone code investigation, which although quite informative yet suffers from intrinsic limitations. Some limitations were of technical nature, while others were related more to the methodology. The technical limitations were related to the choice of mass spectrometer and the use of bottom up proteomics approach. We analyzed our samples using ion trap mass spectrometer, which due to its low to medium resolution, is not an instrument of choice for such kind of analysis. Secondly, we performed MRM analysis using this machine, which again is possible, but has its own limitations. Ion trap has no separate collision cell for MS/MS fragmentation, however, it still has the capability to preselect ions by resetting the RF (Radiofrequency) voltage of the three electrodes corresponding to the predetermined m/z ratios and letting all other ions out of the machine. We have tried to perform a comprehensive histone code combinatorial analysis, which ideally requires a middle down or at least top down approach. We were unable to use any of these approaches, instead we used bottom up approach, as our mass spectrometer is unable to detect larger peptides or proteins, which usually is the case with middle down and top down approaches. Despite of all these difficulties, we managed to perform a relatively large scale quantitative analysis of histone code and we confidently detected the alternately modified and unmodified versions of many peptides, which is verified by the accurate MS/MS fragmentation spectra that we have observed.

As a next step, we applied a functional genomics approach to study DDR in FA cells. In this analysis we first looked at the expression profile of histone modifying enzymes related to histone acetylations and methylations. We used this approach based upon the assumption that the histone PTM profile reflects the levels of various enzymes that establish them. Opposite to the approach usually found in literature, where first at the mRNA levels of histone modifying enzymes are analyzed and then they are correlated with the changes in protein modifications, we first analyzed the protein PTMs and then targeted our functional analysis to the mRNA levels of the enzymes responsible for them. Interestingly, we have found some correlations in the histone PTM profile and the gene expression levels of some histone modifying enzymes, although conclusive evidence warrants further investigations. Nonetheless, this kind of approach has not being reported before and we believe that it could be a very useful form of analysis, as it has the advantage of focusing only on those genes whose protein products have found to be implicated in important cellular signaling events.

For identifying the specific transcriptional response we have applied a statistical methodology to our complex data. We have designed our experiments in 2x2 factorial way and applied interaction analysis in order to decipher more specific response from the generic DDR response. This kind of methodology is very useful when two or more than two factors are being considered in one experiment, as we can dissociate the additive from non-additive

response. The next step after finding out the which genes are specifically involved in a particular type of DNA damage response, is to verify them using q-PCRs and also at protein level by western blots. Once, specific target proteins are identified, siRNAs can be used to block the protein expression and observe the effects on DDR. Identification of proteins, which are specifically involved in a particular kind of DNA damage response bears important significance in painting the overall picture of how a cell reacts and responds to various genomic insults.

In the final part of the thesis, we attempted to solve one of the major limitations that is usually encountered in the histone code analysis. Ideally, any investigation regarding histone code in DNA damage response should involve isolation of histones from DNA damage regions, i.e., from the chromatin surrounding the damage site. However, the current proteomics methodologies are unable to purify histones from a particular region of the chromatin, instead they rely on whole cell lysis and mostly acid extraction of all the histones. This "global" analysis does not reflect the exact changes in histone PTM levels as a result of DNA damage, as histones from surrounding the DNA lesion (most likely accounting for less than 1% of total chromatin), are diluted by the histones from the rest of the genome. The current approaches to purify histones from particular loci involves classical chromatin immunoprecipitation, which due to involvement of formaldehyde crosslinking renders the protein part mostly unavailable for MS-based proteomics. We have proposed a novel methodology, which is based upon the biotin tagging of histones surrounding a protein of interest and hence, does not involve any crosslinking, enabling us to purify histones from specific loci, and subject them to large scale MS-based histone code investigation. A time dimension can also be added to our approach as we can follow the modification status of particular set of histones once they get biotinylated. Another advantage is the use of alternate variant histones and study their modification profile in different conditions, as they represent different functional states of chromatin. This methodology certainly has an edge on current techniques to study histone codes surrounding a particular protein of interest or in particular cellular states.

In general, in this thesis we have applied various functional genomics, biostatistics, and proteomics approaches to study the role of chromatin in DNA damage response. These methodologies have allowed us to understand the chromatin signaling pathways in response to various kinds of DNA lesions. One of the future developments essential to understanding the biology of combinatorial histone codes is to connect them back to specific genomic locations in a manner similar to how ChIP-seq technologies are able to with single modifications. We believe that our PUB-NChIP methodology has the potential to perform such kind of analysis where both DNA and protein from the same sample are equally good for subsequent high throughput analysis. This kind of approach will enable chromatin researchers to answer many relevant questions in chromatin biology - and also generate MORE questions.

6

Materials and Methods

6.1 Recombinant DNA

For the pCMV enhancer vectors, the pCDNA3.1 (Invitrogen) backbone was used. The BirA ORF was PCR amplified from the vector pBBHN [Viens et al., 2004]. The BAP domain fragments were prepared by designing oligonucleotides and sequential PCR, followed by restriction digest and insertion into the pCDNA3.1 vector. The insert sequences were confirmed by sequencing. The vectors generated were used to clone the ORFs of Rad18, GFP, H2AZ, H2ABBD, macroH2A, H3.1, and H3.3.

6.2 Cell Culture and Cell Lines

FA-D2, FA-C (corrected and mutated), and HEK293T cells were grown in Dulbecco's modified Eagle's medium with high glucose (PAA) and 10% fetal bovine serum (FCS, PAA). For transient transfection and Western analysis, a standard calcium phosphate precipitation method was used (described later), and the cells were analyzed one or two days after transfection, as indicated. For the biotin labeling in vivo, cells were grown for several days before transfection in the DMEM supplied with dialyzed FCS, and for the specified time of labeling, biotin (Sigma) was added to a final concentration of 5 μg/ml, while the pH was stabilized by addition of 50 mM HEPES (pH 7.35) to the medium. For SILAC experiments, the cells were grown in either DMEM with lysine ${}_6C^{12}$ and arginine ${}_6C^{12}$, ${}_4N^{14}$ (Light Medium) or lysine ${}_6C^{13}$ and arginine ${}_6C^{13}$, ${}_4N^{15}$ (Heavy Medium) (Thermo Scientific, cat 89983) for at least 5 divisions before transfection, and kept in the same medium until harvest. For the Western

analysis, 3*10^5 of cells (corresponding to one well in a 6-well plate) were used for transfection. For the mass-spectrometry analysis, 3*10^6 of cells were transfected per data point. DNA damage was induced either with 8-MOP/UVA (10 μM of 8-MOP + 10 KJ/m^2 of UVA), UVC (20 J/m^2), or Ionizing radiation (10 Gy). The time of incubation after various treatments is indicated in the results section.

6.3 Biochemistry and Western Blot Analysis

Except where indicated otherwise, cell nuclei were used for analysis. They were prepared by cell disruption in CSK buffer (100 mM NaCl, 300 mM Sucrose, 10 mM Tris pH 7.5, 3 mM $MgCl_2$, 1 mM EGTA, 1.2 mM PMSF, 0.5% Triton X-100), and centrifugation for 5 min at 4000 rpm. For the analysis of chromatin-associated histones (chromatin fraction), the nuclei were first incubated in CSK buffer containing 450 mM NaCl by 30 min rotation at 4°C, then spun at 4000 rpm, and the supernatant containing soluble histones was discarded. For Western analysis, 1x NuPAGE LDS Sample buffer (Invitrogen, NP0007) with DTT (10 mM) was added, the nuclei were sonicated, boiled for 5 min at 96° C and loaded on 4-12% gradient Novex Tris-Glycine precast gels (Invitrogen, NP0315). After separation, the proteins were transferred to PVDF membranes and probed with HRP-conjugated streptavidin (Sigma, S5512) or HRP conjugated anti-PentaHis antibody (QiaGen, 34460) according to the manufacturers protocol, except that for the detection of biotinylated proteins by the HRP-conjugated streptavidin, 500 mM NaCl was added to he washing buffer (PBS + 0.1% Tween). The blocking of membranes as well as preparation of streptavidin-HRP or anti-PentaHis-HRP dilutions were made in 5% ECL blocking agent (GE Healthcare, RPN418V) as milk powder contains molecules that interact with streptavidin, resulting in background HRP activity. For protein visualization, the gels were stained with PageBlue (Fermentas, R0579). For the densitometric analysis of Western blots, the program ImageJ 1.42q (freely available online) was used. To compare the biotinylation levels between different samples, the value of the streptavidin-HRP signal was normalized by taking a ratio with the value of PentaHis signal, which reflects the amount of the transfected protein regardless of its biotinylation status. Every Western analysis was performed three times, with a representative figure shown.

6.4 Mass Spectrometry Analysis

The protein bands were excised from the gel and processed as in [Shevchenko et al., 1996]. The gel slices were dehydrated with 300 μL of 50% acetonitrile followed by 300 μL of 100% acetonitrile, then re-hydrated with 300 μL of 50 mM ammonium bicarbonate. A final dehydration was performed with 2 washes of 300 μL of 50% acetonitrile, followed by 2 washes of 300 μL of 100% acetonitrile. Each wash was carried out for 10 min at 25°C with shaking at 1400 rpm. The gel slices were dried in a SpeedVac at 35° C for 10 min. For trypsin digestion, the gel slices were pre-incubated with 7 ml of 15 ng/mL trypsin (Promega V5280) at room temperature for 10 min. Afterwards, 25 μL of 50 mM ammonium bicarbonate was added,

and the gel slices were incubated at 37°C for 16 h. The peptide-containing supernatants were dried at 56°C by SpeedVac for 30 min, then resuspended in 20 μL of solution containing 0.05% formic acid and 3% acetonitrile for mass spectrometry experiments. The resulting peptides were analyzed with a nano-HPLC (Agilent Technologies 1200) directly coupled to an ion-trap mass spectrometer (Bruker 6300 series) equipped with a nano-electrospray source. The separation gradient was 7 min from 5% to 90% acetonitrile. The fragmentation voltage was 1.3 V. The analysis of the spectra was performed with the DataAnalysis for the 6300 Series Ion Trap LC/MS Version 3.4 software package. The samples were run in two different modes. For peptide identification, the ion trap acquired successive sets of 4 scan modes consisting of: full scan MS over the ranges of 200-2000 m/z, followed by 3 data-dependent MS/MS scans on the 3 most abundant ions in the full scan. Alternatively, for the confirmation and quantification of the presence of a particular peptide in the sample, the ion trap was set in MRM mode. The sample was separated using the same nanoLC gradient, and the ion trap was set to isolate, fragment and MS/MS scan six parental ions having predetermined M/Z ratios.

For quantitative comparisons of various histone PTMs in different samples extracted ion chromatograms (EICs) representing each fragment peptide of a particular light or heavy parent peptide were generated in Agilent's "Data Analysis" software. The peak intensities of light fragment peptides and their respective heavy fragment peptides were summed up and used for H/L ratio calculation. A similar ratio was also calculated for the unmodified version of the same peptide. The proportions were calculated using these two ratios, i.e. H/L (modified) / H/L (unmodified) and represented in the form of graphs in the results section.

6.5 Immunoflourescent Microscopy

MRC-5 foetal lung fibroblasts were grown on cover slips in 6-well plates in DMEM high glucose containing dialyzed FBS. Cells were co-transfected with the plasmid constructs BirA-Rad18 (0.1 μg) and BAP-H2AZ (1 μg) using TurboFect (Cat. R0531, Fermentas). 24h after transfection, cells were treated with UVC (20 J/m^2, Philips TUV 15W/G 15 TB lamp). After 6h of irradiation, the DMEM containing dialyzed FBS was removed and the cells were pulse-labeled for 5 min with biotin (5 μg/mL final concentration). The biotin-containing medium was then removed, and cells were fixed either immediately (pulse) or 2h after the pulse (chase). For staining, cells were washed once with cold PBS and treated with CSK buffer (100 mM NACl, 300 mM Sucrose, 3 mM $MgCl_2$, 10 mM tris-HCl pH 7.5, 1 mM EGTA, Triton 0.2%) for 5 min with light agitation. Cells were then washed twice with cold PBS. Fixation was done in 4% formaldehyde for 20 min at room temperature (RT). After fixation, cells were rinsed with ice-cold methanol at -20°C for 10 sec. Cells were then washed 3 times for 5 min with PBS (1X) at RT. Blocking was done for 30 min using 3% BSA and 0.5% Tween20 in PBS. Cells were then incubated with either mouse monoclonal anti-PCNA antibody (Santa Cruz; PC10; sc-56) for 1h (1/500 in blocking solution) or mouse monoclonal anti-Rad18 antibody (Abcam; ab57447) followed by 3 washes in PBS (1X) at RT. Cells were then incubated with goat anti-mouse secondary antibody conjugated with Alexa-488 (Invitrogen; A11017) for 1h

(1/1000). Streptavidin-Cy3 conjugate (Sigma; S-6402) was also added to the same incubation mix (1/500). After 3 washes in PBS (1X) at RT, cover slips were mounted on glass slides using VectaShield mounting medium (Vector Laboratories; H-1000). The cells were observed under a Zeiss LSM 510 Meta confocal microscope, using a Plan-Apochromat 63x 1.4 oil immersion objective. Imaging was performed with sequential multitrack scanning using the 488 and 543 nm wavelengths lasers separately. The colocalization analysis was performed with LSM Examiner software. Pearson's correlation coefficient was calculated with setting the threshold signal common for all images compared.

For HEK293T cells a slightly different preparation method was used as these cells adhere much less strongly to the surfaces and can be easily detached. Briefly, after biotin pulse cells were washed two times with cold PBS and were fixed by immersion in -20°C methanol for 5 min at -20°C without agitation. Cells were washed again with cold PBS and were then permeabilized with 0.2% Triton X-100 in PBS for 10 min at 4°C in mild agitation. Fixation was done in 4% formaldehyde for 20 min at room temperature with mild agitation. Cells were washed twice with PBS at room (RT) without agitation. Blocking was done with 3% BSA for 30 min under mild agitation. Cells were then incubated with mouse monoclonal anti-Rad18 antibody (Abcam; ab57447) followed by 3 washes in PBS (1X) at RT. Cells were then incubated with goat anti-mouse secondary antibody conjugated with Alexa-488 (Invitrogen; A11017) for 1h (1/1000). Streptavidin-Cy3 conjugate (Sigma; S-6402) and ToPro (Invitrogen; T3605) was also added (1/500 and 1/1000 respectively) to the same incubation mix. After 3 washes in PBS (1X) at RT, cover slips were mounted on glass slides using VectaShield mounting medium (Vector Laboratories; H-1000). Cells were visualized under Zeiss LSM 510 Meta confocal microscope as previously. For histones H2A and H3 staining, cells were incubated with anti-H2A antibody (Cell Signaling; 2572) and anti-H3 antibody (Abcam; ab1791) respectively.

6.6 Proteome Fractionation

- **Preparation of Nuclei**

  Prepare CSK buffer with Triton 0.5% final, PMSF 1 mM final, and Na-Butyrate 5 mM final. Centrifuge cells at 700 g for 5 min. Wash cells twice with PBS in 10 times the volume of the pellet. Spin at 700 rcf for 5 min. Wash cells in CSK buffer, 10 times the volume of the pellet, without Triton. Spin at 700 g for 5 min. Incubate in CSK buffer with Triton in 10 times the volume of the pellet for 5 min at room temperature [with inhibitors including protease inhibitor cocktail (Roche; 1697498), 10 mM Na-Butyrate, 2 mM PMSF, 5 mM Nicotinamide (Sigma; N5535), 5 mM Na-Orthovanadate (Sigma; S6508)]. Observe under microscope if the nuclei are released. Spin nuclei at 4000 rpm for 20 min. Mark the supernatant as cytoplasmic extract and keep it frozen in -80°C. Measure the volume of the pellet.

- **Preparation of Nuclear Extract**

  Add 1/2 volume of the pellet of LSB + PMSF, and resuspend. Add drop wise 1/2 the volume of the pellet of HSB + PMSF (as earlier) and 0.07% β-mercaptoethanol (β-ME). Rotate in cold room for 30 min. Spin down at 13000 rpm for 30 min. Keep the supernatant (nuclear extract) and freeze in -80°C. Measure the volume of the pellet.

- **Preparation of Histones (Acid Extraction)**

  Prepare 1 M H_2SO_4 and put it in ice box. Also take 100% acetone in glass bottle and put it in the ice box. Now resuspend the pellet in H_2SO_4 0,4 M final. Add PMSF and Na-butyrate as usuall. Put the tubes in cold room in rotation for 1h. Centrifuge at 2500 g for 10 min and collect the supernatant containing histones. Keep the pellet which has DNA and matrix proteins and freeze it in -80°C. Add 100% acetone 6 times the volume of the supernatant. Incubate overnight in rotation at 4°C for histone precipitation. Next day, centrifuge at 4000 rpm for 20 min. Remove the supernatant and speed vac to remove the acetone. Then 200 μL of distilled water was added to each tube in two steps and try to resuspend the histones. They were then kept in -20°C till further use.

6.7 Calcium Phosphate Transfections

One day (20-24h) before the experiment 293T cells were seeded at a concentration of 0.5 M/well in a 6-well plate in 1.5 ml DMEM containing dialyzed FBS, 4 mM glutamine and 1% pen/strep. Two hours before the transfection, the medium was removed and cells were supplemented with 1.5 ml of fresh DMEM containing dialyzed FBS. For plasmid transfections, add the following in sterile 1.5 mL tubes, in the order listed,:

- 1 μL DNA (1 μg in sterile molecular biology water)
- 54 μL Molecular Biology Water
- 6 μL 2.5 M $CaCl_2$

The above mixture was mixed gently by pipetting. In a second sterile 1.5 ml micro-centrifuge tube, 60 μL of 2 X HEPES-Buffered Saline (HeBS) pH 7.05 was added. The $CaCl_2$/DNA solution was added to the HeBS buffer dropwise with a sterile pipette tip. Vortex for 2-4 seconds. The precipitate was allowed to sit undisturbed for 20 min at room temperature and later distributed over the cells in the culture dish and gently agitate to mix the precipitate and medium. The cells were incubated for 16h in a cell culture incubator (37°C). After 16h, the medium was removed and 1,5 ml of fresh DMEM supplemented with 10% dialyzed serum, 4 mM glutamine, 1% pen/strep was added to the cells. The cells were incubated for 24h in a cell culture incubator (37°C). Cells were pulsed with biotin (when needed) for 5-15 min, depending upon the experiment and were then harvested.

6.8 Native Chromatin Immunoprecipitation

- **Preparation of Nuclei**

  Cells were washed with PBS in 10 times the volume of the pellet [with inhibitors including protease inhibitor cocktail (Roche; 1697498), 10 mM Na-Butyrate, 2 mM PMSF, 5 mM Nicotinamide (Sigma; N5535), 5 mM Na-Orthovanadate (Sigma; S6508)]. After centrifuging at 700 rcf for 5 min, cells were lysed in 1 mL of CSK buffer (Sucrose 2.052 g, $MgCl_2$, 3 mM, Tris-HCl 10 mM pH 7.5, EGTA 1 mM, 0.5% Triton with inhibitors) for 5 min at RT. Nuclei were centrifuged at 4000 rpm for 10 min and kept at 4°C if the next step was performed immediately or kept at 20°C if not used immediately.

- **Micrococcal Nuclease Digestion**

  Nuclei were resuspended in 500 mL of TM2 buffer (10mMTris-HCl, 2 mM $MgCl_2$, 0.1% Triton, inhibitors). In case of aggregate formation, nuclei were sonicated very briefly to dissolve the aggregates. 2.5 μL of 0.5 M $CaCl_2$ was added to the suspended nuclei and then 3 μL of micrococcal nuclease (MNase) (1 U/μL) (Sigma; N3755) was added and incubated at 37°C for 10 min. The reaction was stopped by adding 15 μL of 0.1 M EGTA. Nuclei were collected by spinning at 400 rcf (2000 rpm on tabletop eppendorf centrifuge) for 10 min at 4°C.

- **Nucleosome Extraction**

  After micrococcal digestion, the nuclei pellet was resuspended in 500 μL of pre-chilled 0.4 M salt extraction buffer (385 mM NaCl, 10 mM Tris-HCl pH 7.4, 2 mM $MgCl_2$, 2 mM EGTA, 0.1% Triton X-100 with inhibitors). The tubes were rotated at 4°C for 30 min. Nuclei were collected by spinning at 400 rcf (2000 rpm on tabletop eppendorf centrifuge) for 10 min at 4°C. Supernatant with digested chromatin was transferred to a new tube and kept at 4°C for further processing.

- **Streptavidin Pulldown**

  An equal volume of TM2 buffer was added to the extracted chromatin to make 0.2 M final salt concentration. The sample was centrifuged at maximum speed (13,000 g/16,000 rpm) on Eppendorf centrifuge) for 5 min at 4°C. 200 μL of suspension of Sepaharose-streptavidin beads (GE Healthcare; 17-5113-01) were washed 3 times with 0.2M salt extraction buffer with Triton and finally resuspended in 100 μL of the same buffer. Beads suspension was combined with chromatin and rotated for 3h at 4°C. Afterwards, the beads were washed twice in 500 mL of 0.4 M salt extraction buffer with Triton for 5 min at 4°C.

- **Elution**

  For elution of biotinylated proteins from the streptavidin-sepharose beads, 100 μL of 1X LDS sample buffer (with reducing agent + 2% SDS) to each sample and heat at 99°C for 10 min. Eluant was collected in a fresh tube after spinning the beads very quickly on table-top centrifuge for 10 s and kept at 20°C. Elution was repeated again in 100 μL as before and pool the two eluates.

- **Western Blot and DNA analysis**

  Aliquots were taken after each step and the quality of preparation of DNA and proteins was verified by running these samples on 1.5% agarose gels and 15% SDS-PAGE gels respectively. In order to verify the efficiency of biotinylation, samples were run on 15% SDS-PAGE gels, tranfered on PVDF membrane and blotted with streptavidin-HRP.

6.9 RNA Preparation and Microarray Analysis

- **Total RNA extraction**

  FA-D2 corrected and mutated cells were grown in DMEM high glucose medium and treated with either MOP/UVA or IR (doses described earlier) and incubated for either 1h or 6h. Total RNA was isolated from the cells by lyzing the cells in Trizol reagent (SigmaAldrich), and later RNA was purified on RNeasy columns (Qiagen) according to the manufacturer's protocols.

  Briefly cells are mixed with 350 μL RLT buffer with β-Mercaptoethanol 1%. Then add 350 μL ethanol 70% and mix thoroughly by vortexing. The mixture is transferred to a RNA-binding spin cup and centrifuged at 8,000 g for 20 seconds. The spin cup is washed with 350 μL 1x low-salt wash buffer, incubated with RNase-free DNase-I at 37°C for 15 min, washed with 350 μL 1x high-salt wash buffer, then washed with 500 μL 1x low-salt wash buffer and finally with 500 μL ethanol 80%. Total RNA is eluted by addition of 15 μL nuclease free water to the RNA-binding spin cup and centrifugation at 20,000 g for 1 min. 1 μL of the eluted RNA is used for determination of concentration using Nanodrop spectrophotometer. Quality of RNA preparations is assessed using Lab-on-a-chip Bioanalyser 2000 technology (Agilent technologies), based on the 28S/18S ribosomal RNAs ratio. All samples included in this study displayed a ratio of ribosomal RNAs between 1.5 and 2. Concentration of all RNAs is adjusted at 100 ng/μL, and verified with a second measurement on Nanodrop spectrophotometer.

- **Labeling Protocol Cy3/Labeling Protocol Cy5**

  Agilent Low RNA Input Linear Amplification kit (ref 5188-5340) adapted for small amount of total RNA (500 ng total RNA per reaction) was used for probe labeling. In 1,5 ml tubes add 500 ng total RNA from sample. Add 1,2 μL T7 promoter primer, 2 μL spike in dilution 1/3200 (ref 5188-5279), use Spike in A to label Cy3 target and Spike In B to label Cy5 target, then add nuclease free water (Invitrogen ref:10977-015) to bring the volume up to 11,5 μL. Denature by incubating at 65°C for 10 minutes. Place the reactions on ice and incubate 5 min. Spin briefly. Prepare a Reverse Transcription master mix, adding for one reaction 4 μL first strand buffer 5X, 2 μL DTT 0,1 M, 1 μL dNTP 10 mM mix, 0.5 μL random hexamer, 1 μL MMLV Reverse Transcriptase (200 U/μL), 0.5 μL RNAse out (40 U/μL). Master mix is prepared in batch for all the samples included in the study (vol. per one reaction multiplied by number of samples. In each reaction tube containing denaturated RNA and T7 promoter primer in a volume

of 11,5 μL, add 9 μL of Reverse Transcriptase master mix, and mix by gently pipetting. Incubate at 40°C in a circulating water bath for 2 h. Move samples to 65°C for 15 min to inactivate MMLV RT, and incubate on ice for 5 min. Spin briefly. Prepare a in vitro transcription master mix, adding for one reaction : 20,1 μL nuclease free water, 20 μL Transcription buffer 4X, 6 μL DTT 0.1 M, 8 μL NTP mix, 0,5 μL RNAse out, 0.6 μL inorganic pyrophosphatase, 0,8 μL T7 RNA Polymerase. Transcription master mix is prepared in batch for all the samples included in the study. (vol. per one reaction multiplied by number of samples). Master mix is splited in two aliquots, one for Cy5 and one for Cy3. Add 1,6 μL (multiplied by reactions number) CTP-Cy5 25 mM (Perkin Elmer ref NEL 581) or 2,4 μL CTP-Cy3 (multiplied by reactions number) (Perkin Elmer, ref NEL 582) to a total volume of 60 μL/ reaction. To each RT reaction, add 60 μL Transcription master mix. Incubate at 40°C for 2 h. Add 20 μL nuclease free water and freeze at -20°C. Labeled probes are purified using Qiagen RNeasy mini kit and protocol provided by Agilent. For each probe, add 350 μL RLT buffer, and 250 μL ethanol 100°C. Mix by gently vortexing. Apply 700 μL on RNeasy columns and spin at 13,000 g for 30 s at 4°C. Discard flow-through. Wash twice with RPE buffer. Dry the column and elute in 60 μL nuclease free water. Measure concentration and Cy5/3 incorporation using a Nanodrop spectrophotometer. Adjust concentration at 100 ng/μL. Freeze at -20°C until hybridization.

- **Hybridization and washing protocol**

  Agilent Hybridization Protocol (Gene expression Hyb kit Large (ref 5188-5280), Chamber type: Agilent SureHyb Chamber; Quantity of labeled extract: 825 ng; duration: 17 h; volume: 110 μL ; Temperature in °C: 65)

  Add the following components in clean 1.5 ml tubes: 825 ng linearly amplified cRNA labeled with Cy5 for each sample, and 825 ng linearly amplified cRNA labeled with Cy3 for the control. Conversely, for dye swapped arrays, mix 825 ng linearly amplified cRNA labelled with Cy3 for each sample, and 825 ng linearly amplified cRNA labeled with Cy5 for the control. In each case, add 11 μL 2X Blocking Agent, 2,2 μL of fragmentation buffer 25X and nuclease free water up to 110 μL. Mix by vortexing. Incubate at 60°C for 30 min in dark. Spin briefly, add 110 μL of 2X hybridization buffer. Mix gently. Assembly the Sure-Hyb hybridization chamber from Agilent. Place a backing side with the plastic inner on upper side. Gently add 650 μL of 1X hybridization solution and cover with the Agilent 44k array properly oriented (active surface in contact with liquid). Finish to assemble the chamber and tight the screw. Hybridization carry out for 17 h at 65°C in a rotating oven (Robbins Scientific, Mountain View, CA) at 4 rpm. The arrays are disassembled at room temperature in SSPE Wash 1 Buffer (ref 5188-5327), then wash 1 min in a glass dish (Wheathon) at room temperature in wash 1 Buffer, then 1 minute in SSPE Wash 2 Buffer (ref 5188-5327) at 37°C and finally acetonitryl wash (sigma aldrich), then 1 min at room temperature.

- **Scanning Protocol**

  Scanning perform with a Agilent G2505C DNA Microarray scanner using default pa-

rameters : 20 bits mode, 5 μm resolution, at 20°C in low ozone concentration environment.

- **Normalization protocol**

  Microarray images are analysed by using Feature extraction software version 10.5.1.1 from Agilent technologies and Agilent normalization protocol. Default settings are used.

- **Transformation Protocol**

  Raw data files from the Agilent Feature Extraction software 10.5.1.1 for image analysis are then imported into Resolver (TM) system (version 7.2.1.1.9) for gene expression data analysis from Rosetta Inpharmatics. Then combined experiments are generated, to obtain average values from the dye swap experiments in order to avoid dye incorporation bias.

Publications

- PUB-NChIP - "In-vivo biotinylation" approach to study chromatin in proximity of a protein of interest.
  Shoaib M, Kulyyassov A, Robin C, Winczura K, Kannouche P, Ramankulov E, Lipinski M, Ogryzko V.
  In preparation

- PUB-MS - a mass-spectrometry-based method to monitor protein-protein proximity in-vivo. (2011)
  Kulyyassov A, **Shoaib M**, Pichugin A, Kannouche P, Ramankulov E, Lipinski M, Ogryzko V.
  J Proteome Res Sept 2 [Epub ahead of print]

- Use of In-Vivo Biotinylation for Chromatin Immunoprecipitation. (2011)
  Kulyyassov A, **Shoaib M**, Ogryzko V.
  Curr Protoc Cell Biol. *2011 Jun; Chapter 17: Unit17.12*

- Multiple Displacement Amplification for Complex Mixtures of DNA Fragments. (2008)
  Shoaib M, Baconnais S., Mechold U., Le Cam. E., Lipinski M., and Ogryzko V.
  BMC. Genomics 9: 415.

Bibliography

[Abbott et al., 2001] Abbott, D. W., Ivanova, V. S., Wang, X., Bonner, W. M. and Ausio, J. (2001). Characterization of the stability and folding of H2A.Z chromatin particles: implications for transcriptional activation. J Biol Chem *276*, 41945–41949.

[Adam et al., 2001] Adam, M., Robert, F., Larochelle, M. and Gaudreau, L. (2001). H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions. Mol Cell Biol *21*, 6270–6279.

[Adams-Cioaba and Min, 2009] Adams-Cioaba, M. A. and Min, J. (2009). Structure and function of histone methylation binding proteins. Biochem Cell Biol *87*, 93–105.

[Agalioti et al., 2002] Agalioti, T., Chen, G. and Thanos, D. (2002). Deciphering the transcriptional histone acetylation code for a human gene. Cell *111*, 381–392.

[Aguilar et al., 2001] Aguilar, C., Hofte, A. J., Tjaden, U. R. and van der Greef, J. (2001). Analysis of histones by on-line capillary zone electrophoresis-electrospray ionisation mass spectrometry. J Chromatogr A *926*, 57–67.

[Ahmad and Henikoff, 2002a] Ahmad, K. and Henikoff, S. (2002a). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol Cell *9*, 1191–1200.

[Ahmad and Henikoff, 2002b] Ahmad, K. and Henikoff, S. (2002b). Histone H3 variants specify modes of chromatin assembly. Proc Natl Acad Sci U S A *99 Suppl 4*, 16477–16484.

[Ahmed et al., 2007] Ahmed, S., Dul, B., Qiu, X. and Walworth, N. C. (2007). Msc1 acts through histone H2A.Z to promote chromosome stability in Schizosaccharomyces pombe. Genetics *177*, 1487–1497.

[Akkari et al., 2000] Akkari, Y. M., Bateman, R. L., Reifsteck, C. A., Olson, S. B. and Grompe, M. (2000). DNA replication is required To elicit cellular responses to psoralen-induced DNA interstrand cross-links. Mol Cell Biol *20*, 8283–8289.

[Allfrey et al., 1964] Allfrey, V. G., Faulkner, R. and Mirsky, A. E. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc Natl Acad Sci U S A *51*, 786–794.

[Allis et al., 1980] Allis, C. D., Bowen, J. K., Abraham, G. N., Glover, C. V. and Gorovsky, M. A. (1980). Proteolytic processing of histone H3 in chromatin: a physiologically regulated event in Tetrahymena micronuclei. Cell *20*, 55–64.

[Allis et al., 2007] Allis, C. D., Jenuwein, T. and Reinberg, D. (2007). Epigenetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

[Allis and Wiggins, 1984] Allis, C. D. and Wiggins, J. C. (1984). Proteolytic processing of micronuclear H3 and histone phosphorylation during conjugation in Tetrahymena thermophila. Exp Cell Res *153*, 287–298.

[Alpi et al., 2008] Alpi, A. F., Pace, P. E., Babu, M. M. and Patel, K. J. (2008). Mechanistic insight into site-restricted monoubiquitination of FANCD2 by Ube2t, FANCL, and FANCI. Mol Cell *32*, 767–777.

[Altaf et al., 2007] Altaf, M., Saksouk, N. S. and Cote, J. (2007). Histone modifications in response to DNA damage. Mutat Res *618*, 81–90.

[Angelov et al., 2003] Angelov, D., Molla, A., Perche, P.-Y., Hans, F., Cote, J., Khochbin, S., Bouvet, P. and Dimitrov, S. (2003). The histone variant macroH2A interferes with transcription factor binding and SWI/SNF nucleosome remodeling. Mol Cell *11*, 1033–1041.

[Arents and Moudrianakis, 1995] Arents, G. and Moudrianakis, E. N. (1995). The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. Proc Natl Acad Sci U S A *92*, 11170–11174.

[Armirotti and Damonte, 2010] Armirotti, A. and Damonte, G. (2010). Achievements and perspectives of top-down proteomics. Proteomics *10*, 3566–3576.

[Bannister and Kouzarides, 1996] Bannister, A. J. and Kouzarides, T. (1996). The CBP co-activator is a histone acetyltransferase. Nature *384*, 641–643.

[Bannister and Kouzarides, 2011] Bannister, A. J. and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res *21*, 381–395.

[Bannister et al., 2002] Bannister, A. J., Schneider, R. and Kouzarides, T. (2002). Histone methylation: dynamic or static? Cell *109*, 801–806.

[Bannister et al., 2001] Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C. and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120–124.

[Bantscheff et al., 2007] Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. and Kuster, B. (2007). Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem *389*, 1017–1031.

[Barlesi et al., 2007] Barlesi, F., Giaccone, G., Gallegos-Ruiz, M. I., Loundou, A., Span, S. W., Lefesvre, P., Kruyt, F. A. E. and Rodriguez, J. A. (2007). Global histone modifications predict prognosis of resected non small-cell lung cancer. J Clin Oncol *25*, 4358–4364.

[Barnidge et al., 2004] Barnidge, D. R., Hall, G. D., Stocker, J. L. and Muddiman, D. C. (2004). Evaluation of a cleavable stable isotope labeled synthetic peptide for absolute protein quantification using LC-MS/MS. J Proteome Res *3*, 658–661.

[Barski et al., 2007] Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

[Bartke et al., 2010] Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S. C., Mann, M. and Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell *143*, 470–484.

[Beck et al., 2006] Beck, H. C., Nielsen, E. C., Matthiesen, R., Jensen, L. H., Sehested, M., Finn, P., Grauslund, M., Hansen, A. M. and Jensen, O. N. (2006). Quantitative proteomic analysis of post-translational modifications of human histones. Mol Cell Proteomics *5*, 1314–1325.

[Bedford and Clarke, 2009] Bedford, M. T. and Clarke, S. G. (2009). Protein arginine methylation in mammals: who, what, and why. Mol Cell *33*, 1–13.

[Begley and Samson, 2004] Begley, T. J. and Samson, L. D. (2004). Network responses to DNA damaging agents. DNA Repair (Amst) *3*, 1123–1132.

[Ben-Yehoyada et al., 2009] Ben-Yehoyada, M., Wang, L. C., Kozekov, I. D., Rizzo, C. J., Gottesman, M. E. and Gautier, J. (2009). Checkpoint signaling from a single DNA interstrand crosslink. Mol Cell *35*, 704–715.

[Benson et al., 2006] Benson, L. J., Gu, Y., Yakovleva, T., Tong, K., Barrows, C., Strack, C. L., Cook, R. G., Mizzen, C. A. and Annunziato, A. T. (2006). Modifications of H3 and H4 during chromatin replication, nucleosome assembly, and histone exchange. J Biol Chem *281*, 9287–9296.

[Bergink et al., 2006] Bergink, S., Salomons, F. A., Hoogstraten, D., Groothuis, T. A. M., de Waard, H., Wu, J., Yuan, L., Citterio, E., Houtsmuller, A. B., Neefjes, J., Hoeijmakers, J. H. J., Vermeulen, W. and Dantuma, N. P. (2006). DNA damage triggers nucleotide excision repair-dependent monoubiquitylation of histone H2A. Genes Dev *20*, 1343–1352.

[Bhaskara et al., 2008] Bhaskara, S., Chyla, B. J., Amann, J. M., Knutson, S. K., Cortez, D., Sun, Z.-W. and Hiebert, S. W. (2008). Deletion of histone deacetylase 3 reveals critical roles in S phase progression and DNA damage control. Mol Cell *30*, 61–72.

[Birchler et al., 2000] Birchler, J. A., Bhadra, M. P. and Bhadra, U. (2000). Making noise about silence: repression of repeated genes in animals. Curr Opin Genet Dev *10*, 211–216.

[Blagoev et al., 2003] Blagoev, B., Kratchmarova, I., Ong, S.-E., Nielsen, M., Foster, L. J. and Mann, M. (2003). A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. Nat Biotechnol *21*, 315–318.

[Blagoev et al., 2004] Blagoev, B., Ong, S.-E., Kratchmarova, I. and Mann, M. (2004). Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. Nat Biotechnol *22*, 1139–1145.

[Blower et al., 2002] Blower, M. D., Sullivan, B. A. and Karpen, G. H. (2002). Conserved organization of centromeric chromatin in flies and humans. Dev Cell *2*, 319–330.

[Bogliolo et al., 2007] Bogliolo, M., Lyakhovich, A., Callen, E., Castella, M., Cappelli, E., Ramirez, M. J., Creus, A., Marcos, R., Kalb, R., Neveling, K., Schindler, D. and Surralles, J. (2007). Histone H2AX and Fanconi anemia FANCD2 function in the same pathway to maintain chromosome stability. EMBO J *26*, 1340–1351.

[Bonaldi et al., 2004] Bonaldi, T., Imhof, A. and Regula, J. T. (2004). A combination of different mass spectroscopic techniques for the analysis of dynamic changes of histone modifications. Proteomics *4*, 1382–1396.

[Bonenfant et al., 2007] Bonenfant, D., Towbin, H., Coulot, M., Schindler, P., Mueller, D. R. and van Oostrum, J. (2007). Analysis of dynamic changes in post-translational modifications of human histones during cell cycle by mass spectrometry. Mol Cell Proteomics *6*, 1917–1932.

[Bonnefoy et al., 2007] Bonnefoy, E., Orsi, G. A., Couble, P. and Loppin, B. (2007). The essential role of Drosophila HIRA for de novo assembly of paternal chromatin at fertilization. PLoS Genet *3*, 1991–2006.

[Bonner et al., 1980] Bonner, W. M., West, M. H. and Stedman, J. D. (1980). Two-dimensional gel analysis of histones in acid extracts of nuclei, cells, and tissues. Eur J Biochem *109*, 17–23.

[Boschetti and Righetti, 2008a] Boschetti, E. and Righetti, P. G. (2008a). Hexapeptide combinatorial ligand libraries: the march for the detection of the low-abundance proteome continues. Biotechniques *44*, 663–665.

[Boschetti and Righetti, 2008b] Boschetti, E. and Righetti, P. G. (2008b). The ProteoMiner in the proteomic arena: a non-depleting tool for discovering low-abundance species. J Proteomics *71*, 255–264.

[Botuyan et al., 2006] Botuyan, M. V., Lee, J., Ward, I. M., Kim, J.-E., Thompson, J. R., Chen, J. and Mer, G. (2006). Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. Cell *127*, 1361–1373.

[Boyne et al., 2006] Boyne, M. T., Pesavento, J. J., Mizzen, C. A. and Kelleher, N. L. (2006). Precise characterization of human histones in the H2A gene family by top down mass spectrometry. J Proteome Res *5*, 248–253.

[Brasacchio et al., 2009] Brasacchio, D., Okabe, J., Tikellis, C., Balcerczyk, A., George, P., Baker, E. K., Calkin, A. C., Brownlee, M., Cooper, M. E. and El-Osta, A. (2009). Hyperglycemia induces a dynamic cooperativity of histone methylase and demethylase en-

zymes associated with gene-activating epigenetic marks that coexist on the lysine tail. Diabetes *58*, 1229–1236.

[Brendel and Ruhland, 1984] Brendel, M. and Ruhland, A. (1984). Relationships between functionality and genetic toxicology of selected DNA-damaging agents. Mutat Res *133*, 51–85.

[Bridge et al., 2005] Bridge, W. L., Vandenberg, C. J., Franklin, R. J. and Hiom, K. (2005). The BRIP1 helicase functions independently of BRCA1 in the Fanconi anemia pathway for DNA crosslink repair. Nat Genet *37*, 953–957.

[Brown et al., 2001] Brown, C. E., Howe, L., Sousa, K., Alley, S. C., Carrozza, M. J., Tan, S. and Workman, J. L. (2001). Recruitment of HAT complexes by direct activator interactions with the ATM-related Tra1 subunit. Science *292*, 2333–2337.

[Brownell and Allis, 1995] Brownell, J. E. and Allis, C. D. (1995). An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei. Proc Natl Acad Sci U S A *92*, 6364–6368.

[Brownell et al., 1996] Brownell, J. E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D. G., Roth, S. Y. and Allis, C. D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell *84*, 843–851.

[Buschbeck et al., 2009] Buschbeck, M., Uribesalgo, I., Wibowo, I., Rue, P., Martin, D., Gutierrez, A., Morey, L., Guigo, R., Lopez-Schier, H. and Croce, L. D. (2009). The histone variant macroH2A is an epigenetic regulator of key developmental genes. Nat Struct Mol Biol *16*, 1074–1079.

[Caldecott, 2008] Caldecott, K. W. (2008). Single-strand break repair and genetic disease. Nat Rev Genet *9*, 619–631.

[Camporeale et al., 2004] Camporeale, G., Shubert, E. E., Sarath, G., Cerny, R. and Zempleni, J. (2004). K8 and K12 are biotinylated in human histone H4. Eur J Biochem *271*, 2257–2263.

[Capetillo et al., 2004] Capetillo, O. F., Allis, C. D. and Nussenzweig, A. (2004). Phosphorylation of histone H2B at DNA double-strand breaks. J Exp Med *199*, 1671–1677.

[Carey et al., 2006] Carey, M., Li, B. and Workman, J. L. (2006). RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. Mol Cell *24*, 481–487.

[Celeste et al., 2003] Celeste, A., Fernandez-Capetillo, O., Kruhlak, M. J., Pilch, D. R., Staudt, D. W., Lee, A., Bonner, R. F., Bonner, W. M. and Nussenzweig, A. (2003). Histone H2AX phosphorylation is dispensable for the initial recognition of DNA breaks. Nat Cell Biol *5*, 675–679.

[Celeste et al., 2002] Celeste, A., Petersen, S., Romanienko, P. J., Fernandez-Capetillo, O., Chen, H. T., Sedelnikova, O. A., Reina-San-Martin, B., Coppola, V., Meffre, E., Difilippantonio, M. J., Redon, C., Pilch, D. R., Olaru, A., Eckhaus, M., Camerini-Otero, R. D., Tessarollo, L., Livak, F., Manova, K., Bonner, W. M., Nussenzweig, M. C. and Nussenzweig, A. (2002). Genomic instability in mice lacking histone H2AX. Science *296*, 922–927.

[Ceol et al., 2011] Ceol, C. J., Houvras, Y., Jane-Valbuena, J., Bilodeau, S., Orlando, D. A., Battisti, V., Fritsch, L., Lin, W. M., Hollmann, T. J., FerrÃ©, F., Bourque, C., Burke, C. J., Turner, L., Uong, A., Johnson, L. A., Beroukhim, R., Mermel, C. H., Loda, M., Ait-Si-Ali, S., Garraway, L. A., Young, R. A. and Zon, L. I. (2011). The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. Nature *471*, 513–517.

[Chadwick and Willard, 2001] Chadwick, B. P. and Willard, H. F. (2001). A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. J Cell Biol *152*, 375–384.

[Chait, 2006] Chait, B. T. (2006). Chemistry. Mass spectrometry: bottom-up or top-down? Science *314*, 65–66.

[Chakalova et al., 2005] Chakalova, L., Debrand, E., Mitchell, J. A., Osborne, C. S. and Fraser, P. (2005). Replication and transcription: shaping the landscape of the genome. Nat Rev Genet *6*, 669–677.

[Chakravarthy et al., 2005] Chakravarthy, S., Park, Y.-J., Chodaparambil, J., Edayathumangalam, R. S. and Luger, K. (2005). Structure and dynamic properties of nucleosome core particles. FEBS Lett *579*, 895–898.

[Chalfie and Kain, 2005] Chalfie, M. and Kain, S. R. (2005). Green Fluorescent Protein: Properties, Applications, and Protocols. John Wiley & Sons, Inc.

[Champagne and Kutateladze, 2009] Champagne, K. S. and Kutateladze, T. G. (2009). Structural insight into histone recognition by the ING PHD fingers. Curr Drug Targets *10*, 432–441.

[Chang et al., 2007] Chang, B., Chen, Y., Zhao, Y. and Bruick, R. K. (2007). JMJD6 is a histone arginine demethylase. Science *318*, 444–447.

[Chapman and Jackson, 2008] Chapman, J. R. and Jackson, S. P. (2008). Phospho-dependent interactions between NBS1 and MDC1 mediate chromatin retention of the MRN complex at sites of DNA damage. EMBO Rep *9*, 795–801.

[Cheadle and Sampson, 2007] Cheadle, J. P. and Sampson, J. R. (2007). MUTYH-associated polyposis–from defect in base excision repair to clinical genetic testing. DNA Repair (Amst) *6*, 274–279.

[Chen et al., 2008] Chen, C.-C., Carson, J. J., Feser, J., Tamburini, B., Zabaronick, S., Linger, J. and Tyler, J. K. (2008). Acetylated lysine 56 on histone H3 drives chromatin assembly after repair and signals for the completion of repair. Cell *134*, 231–243.

[Chen et al., 2007] Chen, Y., Sprung, R., Tang, Y., Ball, H., Sangras, B., Kim, S. C., Falck, J. R., Peng, J., Gu, W. and Zhao, Y. (2007). Lysine propionylation and butyrylation are novel post-translational modifications in histones. Mol Cell Proteomics *6*, 812–819.

[Cheung, 2004] Cheung, P. (2004). Generation and characterization of antibodies directed against di-modified histones, and comments on antibody and epitope recognition. Methods Enzymol *376*, 221–234.

[Cheung et al., 2003] Cheung, W. L., Ajiro, K., Samejima, K., Kloc, M., Cheung, P., Mizzen, C. A., Beeser, A., Etkin, L. D., Chernoff, J., Earnshaw, W. C. and Allis, C. D. (2003). Apoptotic phosphorylation of histone H2B is mediated by mammalian sterile twenty kinase. Cell *113*, 507–517.

[Cheung et al., 2005] Cheung, W. L., Turner, F. B., Krishnamoorthy, T., Wolner, B., Ahn, S.-H., Foley, M., Dorsey, J. A., Peterson, C. L., Berger, S. L. and Allis, C. D. (2005). Phosphorylation of histone H4 serine 1 during DNA damage requires casein kinase II in S. cerevisiae. Curr Biol *15*, 656–660.

[Chi et al., 2007a] Chi, A., Bai, D. L., Geer, L. Y., Shabanowitz, J. and Hunt, D. F. (2007a). Analysis of intact proteins on a chromatographic time scale by electron transfer dissociation tandem mass spectrometry. Int J Mass Spectrom *259*, 197–203.

[Chi et al., 2007b] Chi, A., Huttenhower, C., Geer, L. Y., Coon, J. J., Syka, J. E. P., Bai, D. L., Shabanowitz, J., Burke, D. J., Troyanskaya, O. G. and Hunt, D. F. (2007b). Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. Proc Natl Acad Sci U S A *104*, 2193–2198.

[Chistiakov et al., 2008] Chistiakov, D. A., Voronova, N. V. and Chistiakov, P. A. (2008). Genetic variations in DNA repair genes, radiosensitivity to cancer and susceptibility to acute tissue reactions in radiotherapy-treated cancer patients. Acta Oncol *47*, 809–824.

[Choudhary et al., 2009] Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V. and Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science *325*, 834–840.

[Chowdhury et al., 2005] Chowdhury, D., Keogh, M.-C., Ishii, H., Peterson, C. L., Buratowski, S. and Lieberman, J. (2005). gamma-H2AX dephosphorylation by protein phosphatase 2A facilitates DNA double-strand break repair. Mol Cell *20*, 801–809.

[Chowdhury et al., 2008] Chowdhury, D., Xu, X., Zhong, X., Ahmed, F., Zhong, J., Liao, J., Dykxhoorn, D. M., Weinstock, D. M., Pfeifer, G. P. and Lieberman, J. (2008). A PP4-phosphatase complex dephosphorylates gamma-H2AX generated during DNA replication. Mol Cell *31*, 33–46.

[Chuikov et al., 2004] Chuikov, S., Kurash, J. K., Wilson, J. R., Xiao, B., Justin, N., Ivanov, G. S., McKinney, K., Tempst, P., Prives, C., Gamblin, S. J., Barlev, N. A. and Reinberg, D. (2004). Regulation of p53 activity through lysine methylation. Nature *432*, 353–360.

[Clapier and Cairns, 2009] Clapier, C. R. and Cairns, B. R. (2009). The biology of chromatin remodeling complexes. Annu Rev Biochem *78*, 273–304.

[Clarkson et al., 1999] Clarkson, M. J., Wells, J. R., Gibson, F., Saint, R. and Tremethick, D. J. (1999). Regions of variant histone His2AvD required for Drosophila development. Nature *399*, 694–697.

[Collins et al., 2005] Collins, R. E., Tachibana, M., Tamaru, H., Smith, K. M., Jia, D., Zhang, X., Selker, E. U., Shinkai, Y. and Cheng, X. (2005). In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. J Biol Chem *280*, 5563–5570.

[Contreras et al., 2003] Contreras, A., Hale, T. K., Stenoien, D. L., Rosen, J. M., Mancini, M. A. and Herrera, R. E. (2003). The dynamic mobility of histone H1 is regulated by cyclin/CDK phosphorylation. Mol Cell Biol *23*, 8626–8636.

[Cook et al., 2009] Cook, P. J., Ju, B. G., Telese, F., Wang, X., Glass, C. K. and Rosenfeld, M. G. (2009). Tyrosine dephosphorylation of H2AX modulates apoptosis and survival decisions. Nature *458*, 591–596.

[Coon et al., 2005] Coon, J. J., Ueberheide, B., Syka, J. E. P., Dryhurst, D. D., Ausio, J., Shabanowitz, J. and Hunt, D. F. (2005). Protein identification using sequential ion/ion reactions and tandem mass spectrometry. Proc Natl Acad Sci U S A *102*, 9463–9468.

[Copeland et al., 2009] Copeland, R. A., Solomon, M. E. and Richon, V. M. (2009). Protein methyltransferases as a target class for drug discovery. Nat Rev Drug Discov *8*, 724–732.

[Cosgrove et al., 2004] Cosgrove, M. S., Boeke, J. D. and Wolberger, C. (2004). Regulated nucleosome mobility and the histone code. Nat Struct Mol Biol *11*, 1037–1043.

[Couldrey et al., 1999] Couldrey, C., Carlton, M. B., Nolan, P. M., Colledge, W. H. and Evans, M. J. (1999). A retroviral gene trap insertion into the histone 3.3A gene causes partial neonatal lethality, stunted growth, neuromuscular deficits and male sub-fertility in transgenic mice. Hum Mol Genet *8*, 2489–2495.

[Cremer and Cremer, 2001] Cremer, T. and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet *2*, 292–301.

[Crick, 1970] Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561–563.

[Crick, 1984] Crick, F. (1984). Memory and molecular turnover. Nature *312*, 101.

[Cuomo et al., 2011] Cuomo, A., Moretti, S., Minucci, S. and Bonaldi, T. (2011). SILAC-based proteomic analysis to dissect the "histone modification signature" of human breast cancer cells. Amino Acids *41*, 387–399.

[Cuthbert et al., 2004] Cuthbert, G. L., Daujat, S., Snowden, A. W., Erdjument-Bromage, H., Hagiwara, T., Yamada, M., Schneider, R., Gregory, P. D., Tempst, P., Bannister, A. J. and Kouzarides, T. (2004). Histone deimination antagonizes arginine methylation. Cell *118*, 545–553.

[Dalgliesh et al., 2010] Dalgliesh, G. L., Furge, K., Greenman, C., Chen, L., Bignell, G., Butler, A., Davies, H., Edkins, S., Hardy, C., Latimer, C., Teague, J., Andrews, J., Barthorpe, S., Beare, D., Buck, G., Campbell, P. J., Forbes, S., Jia, M., Jones, D., Knott, H., Kok, C. Y., Lau, K. W., Leroy, C., Lin, M.-L., McBride, D. J., Maddison, M., Maguire, S., McLay, K., Menzies, A., Mironenko, T., Mulderrig, L., Mudie, L., O'Meara, S., Pleasance, E., Rajasingham, A., Shepherd, R., Smith, R., Stebbings, L., Stephens, P., Tang, G., Tarpey, P. S., Turrell, K., Dykema, K. J., Khoo, S. K., Petillo, D., Wondergem, B., Anema, J., Kahnoski, R. J., Teh, B. T., Stratton, M. R. and Futreal, P. A. (2010). Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature *463*, 360–363.

[D'Andrea and Grompe, 2003] D'Andrea, A. D. and Grompe, M. (2003). The Fanconi anaemia/BRCA pathway. Nat Rev Cancer *3*, 23–34.

[Dardalhon et al., 2007] Dardalhon, M., Lin, W., Nicolas, A. and Averbeck, D. (2007). Specific transcriptional responses induced by 8-methoxypsoralen and UVA in yeast. FEMS Yeast Res *7*, 866–878.

[Das et al., 2009] Das, C., Lucia, M. S., Hansen, K. C. and Tyler, J. K. (2009). CBP/p300-mediated acetylation of histone H3 on lysine 56. Nature *459*, 113–117.

[Daujat et al., 2005] Daujat, S., Zeissler, U., Waldmann, T., Happel, N. and Schneider, R. (2005). HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding. J Biol Chem *280*, 38090–38095.

[de Ruijter et al., 2003] de Ruijter, A. J. M., van Gennip, A. H., Caron, H. N., Kemp, S. and van Kuilenburg, A. B. P. (2003). Histone deacetylases (HDACs): characterization of the classical HDAC family. Biochem J *370*, 737–749.

[Delahunty and Yates, 2007] Delahunty, C. M. and Yates, J. R. (2007). MudPIT: multidimensional protein identification technology. Biotechniques *43*, 563, 565, 567 passim.

[Dhillon and Kamakaka, 2000] Dhillon, N. and Kamakaka, R. T. (2000). A histone variant, Htz1p, and a Sir1p-like protein, Esc2p, mediate silencing at HMR. Mol Cell *6*, 769–780.

[Doil et al., 2009] Doil, C., Mailand, N., Bekker-Jensen, S., Menard, P., Larsen, D. H., Pepperkok, R., Ellenberg, J., Panier, S., Durocher, D., Bartek, J., Lukas, J. and Lukas, C. (2009). RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins. Cell *136*, 435–446.

[Douglas et al., 2010] Douglas, P., Zhong, J., Ye, R., Moorhead, G. B. G., Xu, X. and Lees-Miller, S. P. (2010). Protein phosphatase 6 interacts with the DNA-dependent protein kinase catalytic subunit and dephosphorylates gamma-H2AX. Mol Cell Biol *30*, 1368–1381.

[Dovey et al., 2010] Dovey, O. M., Foster, C. T. and Cowley, S. M. (2010). Histone deacetylase 1 (HDAC1), but not HDAC2, controls embryonic stem cell differentiation. Proc Natl Acad Sci U S A *107*, 8242–8247.

[Downs et al., 2000] Downs, J. A., Lowndes, N. F. and Jackson, S. P. (2000). A role for Saccharomyces cerevisiae histone H2A in DNA repair. Nature *408*, 1001–1004.

[Drogaris et al., 2008] Drogaris, P., Wurtele, H., Masumoto, H., Verreault, A. and Thibault, P. (2008). Comprehensive profiling of histone modifications using a label-free approach and its applications in determining structure-function relationships. Anal Chem *80*, 6698–6707.

[Dronkert and Kanaar, 2001] Dronkert, M. L. and Kanaar, R. (2001). Repair of DNA interstrand cross-links. Mutat Res *486*, 217–247.

[Duncan et al., 2008] Duncan, E. M., Muratore-Schroeder, T. L., Cook, R. G., Garcia, B. A., Shabanowitz, J., Hunt, D. F. and Allis, C. D. (2008). Cathepsin L proteolytically processes histone H3 during mouse embryonic stem cell differentiation. Cell *135*, 284–294.

[Dunleavy et al., 2009] Dunleavy, E. M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., Daigo, Y., Nakatani, Y. and Almouzni-Pettinotti, G. (2009). HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. Cell *137*, 485–497.

[Dunn et al., 2006] Dunn, J., Potter, M., Rees, A. and Runger, T. M. (2006). Activation of the Fanconi anemia/BRCA pathway and recombination repair in the cellular response to solar ultraviolet light. Cancer Res *66*, 11140–11147.

[Eberl et al., 2011] Eberl, H. C., Mann, M. and Vermeulen, M. (2011). Quantitative proteomics for epigenetics. Chembiochem *12*, 224–234.

[Ehrenhofer-Murray, 2004] Ehrenhofer-Murray, A. E. (2004). Chromatin dynamics at DNA replication, transcription and repair. Eur J Biochem *271*, 2335–2349.

[Eirin-Lopez et al., 2008] Eirin-Lopez, J. M., Ishibashi, T. and Ausio, J. (2008). H2A.Bbd: a quickly evolving hypervariable mammalian histone that destabilizes nucleosomes in an acetylation-independent way. FASEB J *22*, 316–326.

[El-Mahdy et al., 2006] El-Mahdy, M. A., Zhu, Q., en Wang, Q., Wani, G., Praetorius-Ibba, M. and Wani, A. A. (2006). Cullin 4A-mediated proteolysis of DDB2 protein at DNA damage sites regulates in vivo lesion recognition by XPC. J Biol Chem *281*, 13404–13411.

[El-Osta et al., 2008] El-Osta, A., Brasacchio, D., Yao, D., Pocai, A., Jones, P. L., Roeder, R. G., Cooper, M. E. and Brownlee, M. (2008). Transient high glucose causes persistent epigenetic changes and altered gene expression during subsequent normoglycemia. J Exp Med *205*, 2409–2417.

[Elkon et al., 2005] Elkon, R., Rashi-Elkeles, S., Lerenthal, Y., Linhart, C., Tenne, T., Amariglio, N., Rechavi, G., Shamir, R. and Shiloh, Y. (2005). Dissection of a DNA-damage-induced transcriptional network using a combination of microarrays, RNA interference and computational promoter analysis. Genome Biol *6*, R43.

[Escargueil et al., 2008] Escargueil, A. E., Soares, D. G., Salvador, M., Larsen, A. K. and Henriques, J. A. P. (2008). What histone code for DNA repair? Mutat Res *658*, 259–270.

[Esteve et al., 2006] Esteve, P.-O., Chin, H. G., Smallwood, A., Feehery, G. R., Gangisetty, O., Karpf, A. R., Carey, M. F. and Pradhan, S. (2006). Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. Genes Dev *20*, 3089–3103.

[Faast et al., 2001] Faast, R., Thonglairoam, V., Schulz, T. C., Beall, J., Wells, J. R., Taylor, H., Matthaei, K., Rathjen, P. D., Tremethick, D. J. and Lyons, I. (2001). Histone variant H2A.Z is required for early mammalian development. Curr Biol *11*, 1183–1187.

[Fan et al., 2002] Fan, J. Y., Gordon, F., Luger, K., Hansen, J. C. and Tremethick, D. J. (2002). The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states. Nat Struct Biol *9*, 172–176.

[Feng et al., 2002] Feng, Q., Wang, H., Ng, H. H., Erdjument-Bromage, H., Tempst, P., Struhl, K. and Zhang, Y. (2002). Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. Curr Biol *12*, 1052–1058.

[Fenn et al., 1989] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. and Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science *246*, 64–71.

[Fernandez-Suarez et al., 2008] Fernandez-Suarez, M., Chen, T. S. and Ting, A. Y. (2008). Protein-protein interaction detection in vitro and in cells by proximity biotinylation. J Am Chem Soc *130*, 9251–9253.

[Finch and Klug, 1976] Finch, J. T. and Klug, A. (1976). Solenoidal model for superstructure in chromatin. Proc Natl Acad Sci U S A *73*, 1897–1901.

[Fischle et al., 2005] Fischle, W., Tseng, B. S., Dormann, H. L., Ueberheide, B. M., Garcia, B. A., Shabanowitz, J., Hunt, D. F., Funabiki, H. and Allis, C. D. (2005). Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. Nature *438*, 1116–1122.

[Fischle et al., 2003] Fischle, W., Wang, Y. and Allis, C. D. (2003). Binary switches and modification cassettes in histone biology and beyond. Nature *425*, 475–479.

[Foltz et al., 2009] Foltz, D. R., Jansen, L. E. T., Bailey, A. O., Yates, J. R., Bassett, E. A., Wood, S., Black, B. E. and Cleveland, D. W. (2009). Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. Cell *137*, 472–484.

[Foltz et al., 2006] Foltz, D. R., Jansen, L. E. T., Black, B. E., Bailey, A. O., Yates, 3rd, J. R. and Cleveland, D. W. (2006). The human CENP-A centromeric nucleosome-associated complex. Nat Cell Biol *8*, 458–469.

[Fraga et al., 2005] Fraga, M. F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Ropero, S., Petrie, K., Iyer, N. G., PÃ©rez-Rosado, A., Calvo, E., Lopez, J. A., Cano, A., Calasanz, M. J., Colomer, D., Piris, M. A., Ahn, N., Imhof, A., Caldas, C., Jenuwein, T. and Esteller, M. (2005). Loss of acetylation at Lys16

and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nat Genet *37*, 391–400.

[Galvani et al., 2008] Galvani, A., Courbeyrette, R., Agez, M., Ochsenbein, F., Mann, C. and Thuret, J.-Y. (2008). In vivo study of the nucleosome assembly functions of ASF1 histone chaperones in human cells. Mol Cell Biol *28*, 3672–3685.

[Garcia et al., 2005] Garcia, B. A., Barber, C. M., Hake, S. B., Ptak, C., Turner, F. B., Busby, S. A., Shabanowitz, J., Moran, R. G., Allis, C. D. and Hunt, D. F. (2005). Modifications of human histone H3 variants during mitosis. Biochemistry *44*, 13202–13213.

[Garcia et al., 2004] Garcia, B. A., Busby, S. A., Barber, C. M., Shabanowitz, J., Allis, C. D. and Hunt, D. F. (2004). Characterization of phosphorylation sites on histone H1 isoforms by tandem mass spectrometry. J Proteome Res *3*, 1219–1227.

[Garcia et al., 2005] Garcia, B. A., Busby, S. A., Shabanowitz, J., Hunt, D. F. and Mishra, N. (2005). Resetting the epigenetic histone code in the MRL-lpr/lpr mouse model of lupus by histone deacetylase inhibition. J Proteome Res *4*, 2032–2042.

[Garcia et al., 2007] Garcia, B. A., Hake, S. B., Diaz, R. L., Kauer, M., Morris, S. A., Recht, J., Shabanowitz, J., Mishra, N., Strahl, B. D., Allis, C. D. and Hunt, D. F. (2007). Organismal differences in post-translational modifications in histones H3 and H4. J Biol Chem *282*, 7641–7655.

[Garcia et al., 2006] Garcia, B. A., Joshi, S., Thomas, C. E., Chitta, R. K., Diaz, R. L., Busby, S. A., Andrews, P. C., Loo, R. R. O., Shabanowitz, J., Kelleher, N. L., Mizzen, C. A., Allis, C. D. and Hunt, D. F. (2006). Comprehensive phosphoprotein analysis of linker histone H1 from Tetrahymena thermophila. Mol Cell Proteomics *5*, 1593–1609.

[Garcia et al., 2007a] Garcia, B. A., Mollah, S., Ueberheide, B. M., Busby, S. A., Muratore, T. L., Shabanowitz, J. and Hunt, D. F. (2007a). Chemical derivatization of histones for facilitated analysis by mass spectrometry. Nat Protoc *2*, 933–938.

[Garcia et al., 2007b] Garcia, B. A., Pesavento, J. J., Mizzen, C. A. and Kelleher, N. L. (2007b). Pervasive combinatorial modification of histone H3 in human cells. Nat Methods *4*, 487–489.

[Gautier et al., 2004] Gautier, T., Abbott, D. W., Molla, A., Verdel, A., Ausio, J. and Dimitrov, S. (2004). Histone variant H2ABbd confers lower stability to the nucleosome. EMBO Rep *5*, 715–720.

[Geiger et al., 2010] Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R. and Mann, M. (2010). Super-SILAC mix for quantitative proteomics of human tumor tissue. Nat Methods *7*, 383–385.

[Geiger et al., 2011] Geiger, T., Wisniewski, J. R., Cox, J., Zanivan, S., Kruger, M., Ishihama, Y. and Mann, M. (2011). Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. Nat Protoc *6*, 147–157.

[Gerlich et al., 2003] Gerlich, D., Beaudouin, J., Kalbfuss, B., Daigle, N., Eils, R. and Ellenberg, J. (2003). Global chromosome positions are transmitted through mitosis in mammalian cells. Cell *112*, 751–764.

[Gershey et al., 1968] Gershey, E. L., Vidali, G. and Allfrey, V. G. (1968). Chemical studies of histone acetylation. The occurrence of epsilon-N-acetyllysine in the f2a1 histone. J Biol Chem *243*, 5018–5022.

[Giannasca et al., 1993] Giannasca, P. J., Horowitz, R. A. and Woodcock, C. L. (1993). Transitions between in situ and isolated chromatin. J Cell Sci *105 (Pt 2)*, 551–561.

[Giannattasio et al., 2005] Giannattasio, M., Lazzaro, F., Plevani, P. and Muzi-Falconi, M. (2005). The DNA damage checkpoint response requires histone H2B ubiquitination by Rad6-Bre1 and H3 methylation by Dot1. J Biol Chem *280*, 9879–9886.

[Glonek and Solomon, 2004] Glonek, G. F. V. and Solomon, P. J. (2004). Factorial and time course designs for cDNA microarray experiments. Biostatistics *5*, 89–111.

[Goldberg et al., 2010] Goldberg, A. D., Banaszynski, L. A., Noh, K.-M., Lewis, P. W., Elsaesser, S. J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., Wen, D., Chapgier, A., DeKelver, R. C., Miller, J. C., Lee, Y.-L., Boydston, E. A., Holmes, M. C., Gregory, P. D., Greally, J. M., Rafii, S., Yang, C., Scambler, P. J., Garrick, D., Gibbons, R. J., Higgs, D. R., Cristea, I. M., Urnov, F. D., Zheng, D. and Allis, C. D. (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell *140*, 678–691.

[Gonzalez-Romero et al., 2008] Gonzalez-Romero, R., Mendez, J., Ausio, J. and Eirin-Lopez, J. M. (2008). Quickly evolving histones, nucleosome stability and chromatin folding: all about histone H2A.Bbd. Gene *413*, 1–7.

[Good et al., 2007] Good, D. M., Wirtala, M., McAlister, G. C. and Coon, J. J. (2007). Performance characteristics of electron transfer dissociation mass spectrometry. Mol Cell Proteomics *6*, 1942–1951.

[Gorisch et al., 2005] Gorisch, S. M., Wachsmuth, M., Toth, K. F., Lichter, P. and Rippe, K. (2005). Histone acetylation increases chromatin accessibility. J Cell Sci *118*, 5825–5834.

[Govin et al., 2005] Govin, J., Caron, C., Rousseaux, S. and Khochbin, S. (2005). Testis-specific histone H3 expression in somatic cells. Trends Biochem Sci *30*, 357–359.

[Graumann et al., 2008] Graumann, J., Hubner, N. C., Kim, J. B., Ko, K., Moser, M., Kumar, C., Cox, J., Scholer, H. and Mann, M. (2008). Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. Mol Cell Proteomics *7*, 672–683.

[Green and Almouzni, 2003] Green, C. M. and Almouzni, G. (2003). Local action of the chromatin assembly factor CAF-1 at sites of nucleotide excision repair in vivo. EMBO J *22*, 5163–5174.

[Gregoretti et al., 2004] Gregoretti, I. V., Lee, Y.-M. and Goodson, H. V. (2004). Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis. J Mol Biol *338*, 17–31.

[Greil et al., 2006] Greil, F., Moorman, C. and van Steensel, B. (2006). DamID: mapping of in vivo protein-genome interactions using tethered DNA adenine methyltransferase. Methods Enzymol *410*, 342–359.

[Groth et al., 2007] Groth, A., Rocha, W., Verreault, A. and Almouzni, G. (2007). Chromatin challenges during DNA replication and repair. Cell *128*, 721–733.

[Gruhler et al., 2005] Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., Faergeman, N. J., Mann, M. and Jensen, O. N. (2005). Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. Mol Cell Proteomics *4*, 310–327.

[Guccione et al., 2007] Guccione, E., Bassi, C., Casadio, F., Martinato, F., Cesaroni, M., Schuchlautz, H., Luscher, B. and Amati, B. (2007). Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive. Nature *449*, 933–937.

[Gupta et al., 2005] Gupta, A., Sharma, G. G., Young, C. S. H., Agarwal, M., Smith, E. R., Paull, T. T., Lucchesi, J. C., Khanna, K. K., Ludwig, T. and Pandita, T. K. (2005). Involvement of human MOF in ATM function. Mol Cell Biol *25*, 5292–5305.

[Gurley et al., 1991] Gurley, L. R., London, J. E. and Valdez, J. G. (1991). High-performance capillary electrophoresis of histones. J Chromatogr *559*, 431–443.

[Gygi et al., 1999] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H. and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol *17*, 994–999.

[Haigis and Guarente, 2006] Haigis, M. C. and Guarente, L. P. (2006). Mammalian sirtuins–emerging roles in physiology, aging, and calorie restriction. Genes Dev *20*, 2913–2921.

[Hake and Allis, 2006] Hake, S. B. and Allis, C. D. (2006). Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis". Proc Natl Acad Sci U S A *103*, 6428–6435.

[Hake et al., 2007] Hake, S. B., Xiao, A. and Allis, C. D. (2007). Linking the epigenetic 'language' of covalent histone modifications to cancer. Br J Cancer *96 Suppl*, R31–R39.

[Hale et al., 2006] Hale, T. K., Contreras, A., Morrison, A. J. and Herrera, R. E. (2006). Phosphorylation of the linker histone H1 by CDK regulates its binding to HP1alpha. Mol Cell *22*, 693–699.

[Hall, 1992] Hall, B. (1992). Evolutionary Developmental Biology. Chapman & Hall; London.

[Hansen et al., 2008] Hansen, K. H., Bracken, A. P., Pasini, D., Dietrich, N., Gehani, S. S., Monrad, A., Rappsilber, J., Lerdrup, M. and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nat Cell Biol *10*, 1291–1300.

[Hassan et al., 2002] Hassan, A. H., Prochasson, P., Neely, K. E., Galasinski, S. C., Chandy, M., Carrozza, M. J. and Workman, J. L. (2002). Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. Cell *111*, 369–379.

[Hebbes et al., 1989] Hebbes, T. R., Turner, C. H., Thorne, A. W. and Crane-Robinson, C. (1989). A "minimal epitope" anti-protein antibody that recognises a single modified amino acid. Mol Immunol *26*, 865–873.

[Heck and Krijgsveld, 2004] Heck, A. J. R. and Krijgsveld, J. (2004). Mass spectrometry-based quantitative proteomics. Expert Rev Proteomics *1*, 317–326.

[Heinloth et al., 2003a] Heinloth, A. N., Shackelford, R. E., Innes, C. L., Bennett, L., Li, L., Amin, R. P., Sieber, S. O., Flores, K. G., Bushel, P. R. and Paules, R. S. (2003a). Identification of distinct and common gene expression changes after oxidative stress and gamma and ultraviolet radiation. Mol Carcinog *37*, 65–82.

[Heinloth et al., 2003b] Heinloth, A. N., Shackelford, R. E., Innes, C. L., Bennett, L., Li, L., Amin, R. P., Sieber, S. O., Flores, K. G., Bushel, P. R. and Paules, R. S. (2003b). ATM-dependent and -independent gene expression changes in response to oxidative stress, gamma irradiation, and UV irradiation. Radiat Res *160*, 273–290.

[Heitz, 1928] Heitz, E. (1928). Das heterochromatin der moose. I Jahrb Wiss Botanik *69*, 762–818.

[Henikoff and Dalal, 2005] Henikoff, S. and Dalal, Y. (2005). Centromeric chromatin: what makes it unique? Curr Opin Genet Dev *15*, 177–184.

[Hodawadekar and Marmorstein, 2007] Hodawadekar, S. C. and Marmorstein, R. (2007). Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design. Oncogene *26*, 5528–5540.

[Hodl and Basler, 2009] Hodl, M. and Basler, K. (2009). Transcription in the absence of histone H3.3. Curr Biol *19*, 1221–1226.

[Hoeijmakers, 2007] Hoeijmakers, J. H. J. (2007). Genome maintenance mechanisms are critical for preventing cancer as well as other aging-associated diseases. Mech Ageing Dev *128*, 460–462.

[Hoeijmakers, 2009] Hoeijmakers, J. H. J. (2009). DNA damage, aging, and cancer. N Engl J Med *361*, 1475–1485.

[Holliday, 1987] Holliday, R. (1987). The inheritance of epigenetic defects. Science *238*, 163–170.

[Holliday, 1990] Holliday, R. (1990). Mechanisms for the control of gene activity during development. Biol Rev Camb Philos Soc *65*, 431–471.

[Holliday, 1994] Holliday, R. (1994). Epigenetics: an overview. Dev Genet *15*, 453–457.

[Hon et al., 2009] Hon, G. C., Hawkins, R. D. and Ren, B. (2009). Predictive chromatin signatures in the mammalian genome. Hum Mol Genet *18*, R195–R201.

[Hopfield, 1974] Hopfield, J. J. (1974). Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. Proc Natl Acad Sci U S A *71*, 4135–4139.

[Houlard et al., 2006] Houlard, M., Berlivet, S., Probst, A. V., Quivy, J.-P., Hery, P., Almouzni, G. and Gerard, M. (2006). CAF-1 is essential for heterochromatin organization in pluripotent embryonic cells. PLoS Genet *2*, e181.

[Houston et al., 2008] Houston, S. I., McManus, K. J., Adams, M. M., Sims, J. K., Carpenter, P. B., Hendzel, M. J. and Rice, J. C. (2008). Catalytic function of the PR-Set7 histone H4 lysine 20 monomethyltransferase is essential for mitotic entry and genomic stability. J Biol Chem *283*, 19478–19488.

[Howlett et al., 2002] Howlett, N. G., Taniguchi, T., Olson, S., Cox, B., Waisfisz, Q., Die-Smulders, C. D., Persky, N., Grompe, M., Joenje, H., Pals, G., Ikeda, H., Fox, E. A. and D'Andrea, A. D. (2002). Biallelic inactivation of BRCA2 in Fanconi anemia. Science *297*, 606–609.

[Howman et al., 2000] Howman, E. V., Fowler, K. J., Newson, A. J., Redward, S., MacDonald, A. C., Kalitsis, P. and Choo, K. H. (2000). Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. Proc Natl Acad Sci U S A *97*, 1148–1153.

[Huang et al., 2009a] Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res *37*, 1–13.

[Huang et al., 2009b] Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44–57.

[Huang et al., 2010] Huang, J., Dorsey, J., Chuikov, S., Perez-Burgos, L., Zhang, X., Jenuwein, T., Reinberg, D. and Berger, S. L. (2010). G9a and Glp methylate lysine 373 in the tumor suppressor p53. J Biol Chem *285*, 9636–9641.

[Huang et al., 2006] Huang, Y., Fang, J., Bedford, M. T., Zhang, Y. and Xu, R.-M. (2006). Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. Science *312*, 748–751.

[Huen et al., 2007] Huen, M. S. Y., Grant, R., Manke, I., Minn, K., Yu, X., Yaffe, M. B. and Chen, J. (2007). RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly. Cell *131*, 901–914.

[Huyen et al., 2004] Huyen, Y., Zgheib, O., Ditullio, R. A., Gorgoulis, V. G., Zacharatos, P., Petty, T. J., Sheston, E. A., Mellert, H. S., Stavridi, E. S. and Halazonetis, T. D. (2004).

Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. Nature *432*, 406–411.

[ichi Tsukada et al., 2006] ichi Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P. and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811–816.

[Ikura and Ogryzko, 2003] Ikura, T. and Ogryzko, V. V. (2003). Chromatin dynamics and DNA repair. Front Biosci *8*, s149–s155.

[Ikura et al., 2000] Ikura, T., Ogryzko, V. V., Grigoriev, M., Groisman, R., Wang, J., Horikoshi, M., Scully, R., Qin, J. and Nakatani, Y. (2000). Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. Cell *102*, 463–473.

[Ikura et al., 2007] Ikura, T., Tashiro, S., Kakino, A., Shima, H., Jacob, N., Amunugama, R., Yoder, K., Izumi, S., Kuraoka, I., Tanaka, K., Kimura, H., Ikura, M., Nishikubo, S., Ito, T., Muto, A., Miyagawa, K., Takeda, S., Fishel, R., Igarashi, K. and Kamiya, K. (2007). DNA damage-dependent acetylation and ubiquitination of H2AX enhances chromatin dynamics. Mol Cell Biol *27*, 7028–7040.

[Ishihama et al., 2005] Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J. and Mann, M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol Cell Proteomics *4*, 1265–1272.

[Jackson and Gorovsky, 2000] Jackson, J. D. and Gorovsky, M. A. (2000). Histone H2A.Z has a conserved function that is distinct from that of the major H2A sequence variants. Nucleic Acids Res *28*, 3811–3816.

[Jackson, 1988] Jackson, V. (1988). Deposition of newly synthesized histones: hybrid nucleosomes are not tandemly arranged on daughter DNA strands. Biochemistry *27*, 2109–2120.

[Janicki et al., 2004] Janicki, S. M., Tsukamoto, T., Salghetti, S. E., Tansey, W. P., Sachidanandam, R., Prasanth, K. V., Ried, T., Shav-Tal, Y., Bertrand, E., Singer, R. H. and Spector, D. L. (2004). From silencing to gene expression: real-time analysis in single cells. Cell *116*, 683–698.

[Jenuwein and Allis, 2001] Jenuwein, T. and Allis, C. D. (2001). Translating the histone code. Science *293*, 1074–1080.

[Jeppesen and Turner, 1993] Jeppesen, P. and Turner, B. M. (1993). The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. Cell *74*, 281–289.

[Jiang et al., 2007a] Jiang, L., Smith, J. N., Anderson, S. L., Ma, P., Mizzen, C. A. and Kelleher, N. L. (2007a). Global assessment of combinatorial post-translational modification of core histones in yeast using contemporary mass spectrometry. LYS4 trimethylation correlates with degree of acetylation on the same H3 tail. J Biol Chem *282*, 27923–27934.

[Jiang et al., 2007b] Jiang, T., Zhou, X., Taghizadeh, K., Dong, M. and Dedon, P. C. (2007b). N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. Proc Natl Acad Sci U S A *104*, 60–65.

[Jin et al., 2009] Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K. and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet *41*, 941–945.

[Jiricny, 2006] Jiricny, J. (2006). The multifaceted mismatch-repair system. Nat Rev Mol Cell Biol *7*, 335–346.

[Johnson et al., 2004] Johnson, L., Mollah, S., Garcia, B. A., Muratore, T. L., Shabanowitz, J., Hunt, D. F. and Jacobsen, S. E. (2004). Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. Nucleic Acids Res *32*, 6511–6518.

[Jones and Gelbart, 1993] Jones, R. S. and Gelbart, W. M. (1993). The Drosophila Polycomb-group gene Enhancer of zeste contains a region with sequence similarity to trithorax. Mol Cell Biol *13*, 6357–6366.

[Jung et al., 2010] Jung, H. R., Pasini, D., Helin, K. and Jensen, O. N. (2010). Quantitative mass spectrometry of histones H3.2 and H3.3 in Suz12-deficient mouse embryonic stem cells reveals distinct, dynamic post-translational modifications at Lys-27 and Lys-36. Mol Cell Proteomics *9*, 838–850.

[Kalocsay et al., 2009] Kalocsay, M., Hiller, N. J. and Jentsch, S. (2009). Chromosome-wide Rad51 spreading and SUMO-H2A.Z-dependent chromosome fixation in response to a persistent DNA double-strand break. Mol Cell *33*, 335–343.

[Kanaar et al., 2008] Kanaar, R., Wyman, C. and Rothstein, R. (2008). Quality control of DNA break metabolism: in the 'end', it's a good thing. EMBO J *27*, 581–588.

[Kannouche and Lehmann, 2004] Kannouche, P. L. and Lehmann, A. R. (2004). Ubiquitination of PCNA and the polymerase switch in human cells. Cell Cycle *3*, 1011–1013.

[Karas and Hillenkamp, 1988] Karas, M. and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem *60*, 2299–2301.

[Kennedy and D'Andrea, 2005] Kennedy, R. D. and D'Andrea, A. D. (2005). The Fanconi Anemia/BRCA pathway: new faces in the crowd. Genes Dev *19*, 2925–2940.

[Keogh et al., 2006] Keogh, M.-C., Kim, J.-A., Downey, M., Fillingham, J., Chowdhury, D., Harrison, J. C., Onishi, M., Datta, N., Galicia, S., Emili, A., Lieberman, J., Shen, X., Buratowski, S., Haber, J. E., Durocher, D., Greenblatt, J. F. and Krogan, N. J. (2006). A phosphatase complex that dephosphorylates gammaH2AX regulates DNA damage checkpoint recovery. Nature *439*, 497–501.

[Khorasanizadeh, 2004] Khorasanizadeh, S. (2004). The nucleosome: from genomic organization to genomic regulation. Cell *116*, 259–272.

[Kim et al., 2006] Kim, J., Daniel, J., Espejo, A., Lake, A., Krishna, M., Xia, L., Zhang, Y. and Bedford, M. T. (2006). Tudor, MBT and chromo domains gauge the degree of lysine methylation. EMBO Rep *7*, 397–403.

[Kim et al., 2009] Kim, J., Guermah, M., McGinty, R. K., Lee, J.-S., Tang, Z., Milne, T. A., Shilatifard, A., Muir, T. W. and Roeder, R. G. (2009). RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. Cell *137*, 459–471.

[Kimura and Horikoshi, 1998] Kimura, A. and Horikoshi, M. (1998). Tip60 acetylates six lysines of a specific class in core histones in vitro. Genes Cells *3*, 789–800.

[Kimura and Cook, 2001] Kimura, H. and Cook, P. R. (2001). Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. J Cell Biol *153*, 1341–1353.

[Kirmizis et al., 2007] Kirmizis, A., Santos-Rosa, H., Penkett, C. J., Singer, M. A., Vermeulen, M., Mann, M., Bahler, J., Green, R. D. and Kouzarides, T. (2007). Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. Nature *449*, 928–932.

[Kislinger et al., 2005] Kislinger, T., Gramolini, A. O., MacLennan, D. H. and Emili, A. (2005). Multidimensional protein identification technology (MudPIT): technical overview of a profiling method optimized for the comprehensive proteomic investigation of normal and diseased heart tissue. J Am Soc Mass Spectrom *16*, 1207–1220.

[Kleff et al., 1995] Kleff, S., Andrulis, E. D., Anderson, C. W. and Sternglanz, R. (1995). Identification of a gene encoding a yeast histone H4 acetyltransferase. J Biol Chem *270*, 24674–24677.

[Klose and Zhang, 2007] Klose, R. J. and Zhang, Y. (2007). Regulation of histone methylation by demethylimination and demethylation. Nat Rev Mol Cell Biol *8*, 307–318.

[Knapp et al., 2007] Knapp, A. R., Ren, C., Su, X., Lucas, D. M., Byrd, J. C., Freitas, M. A. and Parthun, M. R. (2007). Quantitative profiling of histone post-translational modifications by stable isotope labeling. Methods *41*, 312–319.

[Kolas et al., 2007] Kolas, N. K., Chapman, J. R., Nakada, S., Ylanko, J., Chahwan, R., Sweeney, F. D., Panier, S., Mendez, M., Wildenhain, J., Thomson, T. M., Pelletier, L., Jackson, S. P. and Durocher, D. (2007). Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase. Science *318*, 1637–1640.

[Kornberg, 1974] Kornberg, R. D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868–871.

[Kouzarides, 2007a] Kouzarides, T. (2007a). SnapShot: Histone-modifying enzymes. Cell *131*, 822.

[Kouzarides, 2007b] Kouzarides, T. (2007b). Chromatin modifications and their function. Cell *128*, 693–705.

[Kraus, 2009] Kraus, W. L. (2009). New functions for an ancient domain. Nat Struct Mol Biol *16*, 904–907.

[Krishnan et al., 2009] Krishnan, N., Jeong, D. G., Jung, S.-K., Ryu, S. E., Xiao, A., Allis, C. D., Kim, S. J. and Tonks, N. K. (2009). Dephosphorylation of the C-terminal tyrosyl residue of the DNA damage-related histone H2A.X is mediated by the protein phosphatase eyes absent. J Biol Chem *284*, 16066–16070.

[Kruger et al., 2008] Kruger, M., Moser, M., Ussar, S., Thievessen, I., Luber, C. A., Forner, F., Schmidt, S., Zanivan, S., Fassler, R. and Mann, M. (2008). SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. Cell *134*, 353–364.

[Kruhlak et al., 2006] Kruhlak, M. J., Celeste, A., Dellaire, G., Fernandez-Capetillo, O., Muller, W. G., McNally, J. G., Bazett-Jones, D. P. and Nussenzweig, A. (2006). Changes in chromatin structure and mobility in living cells at sites of DNA double-strand breaks. J Cell Biol *172*, 823–834.

[Kulis and Esteller, 2010] Kulis, M. and Esteller, M. (2010). DNA methylation and cancer. Adv Genet *70*, 27–56.

[Kysela et al., 2005] Kysela, B., Chovanec, M. and Jeggo, P. A. (2005). Phosphorylation of linker histones by DNA-dependent protein kinase is required for DNA ligase IV-dependent ligation in the presence of histone H1. Proc Natl Acad Sci U S A *102*, 1877–1882.

[Lachner et al., 2001] Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116–120.

[Lamond and Earnshaw, 1998] Lamond, A. I. and Earnshaw, W. C. (1998). Structure and function in the nucleus. Science *280*, 547–553.

[Lan and Shi, 2009] Lan, F. and Shi, Y. (2009). Epigenetic regulation: methylation of histone and non-histone proteins. Sci China C Life Sci *52*, 311–322.

[Lavelle and Prunell, 2007] Lavelle, C. and Prunell, A. (2007). Chromatin polymorphism and the nucleosome superfamily: a genealogy. Cell Cycle *6*, 2113–2119.

[Leach et al., 2000] Leach, T. J., Mazzeo, M., Chotkowski, H. L., Madigan, J. P., Wotring, M. G. and Glaser, R. L. (2000). Histone H2A.Z is widely but nonrandomly distributed in chromosomes of Drosophila melanogaster. J Biol Chem *275*, 23267–23272.

[Lee et al., 2010] Lee, H.-S., Park, J.-H., Kim, S.-J., Kwon, S.-J. and Kwon, J. (2010). A cooperative activation loop among SWI/SNF, gamma-H2AX and H3 acetylation for DNA double-strand break repair. EMBO J *29*, 1434–1445.

[Lee et al., 2007] Lee, J.-S., Shukla, A., Schneider, J., Swanson, S. K., Washburn, M. P., Florens, L., Bhaumik, S. R. and Shilatifard, A. (2007). Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. Cell *131*, 1084–1096.

[Lee et al., 2005] Lee, M. G., Wynder, C., Cooch, N. and Shiekhattar, R. (2005). An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. Nature *437*, 432–435.

[Lee et al., 2008] Lee, S., Lee, J., Lee, S.-K. and Lee, J. W. (2008). Activating signal cointegrator-2 is an essential adaptor to recruit histone H3 lysine 4 methyltransferases MLL3 and MLL4 to the liver X receptors. Mol Endocrinol *22*, 1312–1319.

[Li et al., 2007] Li, B., Carey, M. and Workman, J. L. (2007). The role of chromatin during transcription. Cell *128*, 707–719.

[Li et al., 2008] Li, Q., Zhou, H., Wurtele, H., Davies, B., Horazdovsky, B., Verreault, A. and Zhang, Z. (2008). Acetylation of histone H3 lysine 56 regulates replication-coupled nucleosome assembly. Cell *134*, 244–255.

[Lindner et al., 1992] Lindner, H., Helliger, W., Dirschlmayer, A., Jaquemar, M. and Puschendorf, B. (1992). High-performance capillary electrophoresis of core histones and their acetylated modified derivatives. Biochem J *283 (Pt 2)*, 467–471.

[Lindner et al., 1996] Lindner, H., Sarg, B., Meraner, C. and Helliger, W. (1996). Separation of acetylated core histones by hydrophilic-interaction liquid chromatography. J Chromatogr A *743*, 137–144.

[Liu et al., 2004] Liu, H., Sadygov, R. G. and Yates, J. R. (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem *76*, 4193–4201.

[Liu et al., 1996] Liu, X., Bowen, J. and Gorovsky, M. A. (1996). Either of the major H2A genes but not an evolutionarily conserved H2A.F/Z variant of Tetrahymena thermophila can function as the sole H2A gene in the yeast Saccharomyces cerevisiae. Mol Cell Biol *16*, 2878–2887.

[Lobrich and Jeggo, 2007] Lobrich, M. and Jeggo, P. A. (2007). The impact of a negligent G2/M checkpoint on genomic instability and cancer induction. Nat Rev Cancer *7*, 861–869.

[Loppin et al., 2005] Loppin, B., Bonnefoy, E., Anselme, C., Laurencon, A., Karr, T. L. and Couble, P. (2005). The histone H3.3 chaperone HIRA is essential for chromatin assembly in the male pronucleus. Nature *437*, 1386–1390.

[Loyola and Almouzni, 2007] Loyola, A. and Almouzni, G. (2007). Marking histone H3 variants: how, when and why? Trends Biochem Sci *32*, 425–433.

[Loyola et al., 2006] Loyola, A., Bonaldi, T., Roche, D., Imhof, A. and Almouzni, G. (2006). PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. Mol Cell *24*, 309–316.

[Loyola et al., 2009] Loyola, A., Tagami, H., Bonaldi, T., Roche, D., Quivy, J. P., Imhof, A., Nakatani, Y., Dent, S. Y. R. and Almouzni, G. (2009). The HP1alpha-CAF1-SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. EMBO Rep *10*, 769–775.

[Lucio-Eterovic et al., 2008] Lucio-Eterovic, A. K. B., Cortez, M. A. A., Valera, E. T., Motta, F. J. N., Queiroz, R. G. P., Machado, H. R., Carlotti, C. G., Neder, L., Scrideli, C. A. and Tone, L. G. (2008). Differential expression of 12 histone deacetylase (HDAC) genes in astrocytomas and normal brain tissue: class II and IV are hypoexpressed in glioblastomas. BMC Cancer *8*, 243.

[Luger et al., 1997] Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251–260.

[Luger and Richmond, 1998] Luger, K. and Richmond, T. J. (1998). DNA binding within the nucleosome core. Curr Opin Struct Biol *8*, 33–40.

[Luijsterburg et al., 2009] Luijsterburg, M. S., Dinant, C., Lans, H., Stap, J., Wiernasz, E., Lagerwerf, S., Warmerdam, D. O., Lindh, M., Brink, M. C., Dobrucki, J. W., Aten, J. A., Fousteri, M. I., Jansen, G., Dantuma, N. P., Vermeulen, W., Mullenders, L. H. F., Houtsmuller, A. B., Verschure, P. J. and van Driel, R. (2009). Heterochromatin protein 1 is recruited to various types of DNA damage. J Cell Biol *185*, 577–586.

[Lundby and Olsen, 2011] Lundby, A. and Olsen, J. V. (2011). GeLCMS for in-depth protein characterization and advanced analysis of proteomes. Methods Mol Biol *753*, 143–155.

[Mace et al., 2007] Mace, G., Briot, D., Guervilly, J.-H. and Rosselli, F. (2007). [Fanconi anemia: cellular and molecular features]. Pathol Biol (Paris) *55*, 19–28.

[Macek et al., 2009] Macek, B., Mann, M. and Olsen, J. V. (2009). Global and site-specific quantitative phosphoproteomics: principles and applications. Annu Rev Pharmacol Toxicol *49*, 199–221.

[Macht, 2009] Macht, M. (2009). Mass spectrometric top-down analysis of proteins. Bioanalysis *1*, 1131–1148.

[Macurek et al., 2010] Macurek, L., Lindqvist, A., Voets, O., Kool, J., Vos, H. R. and Medema, R. H. (2010). Wip1 phosphatase is associated with chromatin and dephosphorylates gammaH2AX to promote checkpoint inhibition. Oncogene *29*, 2281–2291.

[Mahadevaiah et al., 2001] Mahadevaiah, S. K., Turner, J. M., Baudat, F., Rogakou, E. P., de Boer, P., Blanco-Rodriguez, J., Jasin, M., Keeney, S., Bonner, W. M. and Burgoyne, P. S. (2001). Recombinational DNA double-strand breaks in mice precede synapsis. Nat Genet *27*, 271–276.

[Malik and Henikoff, 2003] Malik, H. S. and Henikoff, S. (2003). Phylogenomics of the nucleosome. Nat Struct Biol *10*, 882–891.

[Malik and Henikoff, 2009] Malik, H. S. and Henikoff, S. (2009). Major evolutionary transitions in centromere complexity. Cell *138*, 1067–1082.

[Mann, 2006] Mann, M. (2006). Functional and quantitative proteomics using SILAC. Nat Rev Mol Cell Biol *7*, 952–958.

[Mann and Jensen, 2003] Mann, M. and Jensen, O. N. (2003). Proteomic analysis of post-translational modifications. Nat Biotechnol *21*, 255–261.

[Mannironi et al., 1989] Mannironi, C., Bonner, W. M. and Hatch, C. L. (1989). H2A.X. a histone isoprotein with a conserved C-terminal sequence, is encoded by a novel mRNA with both DNA replication type and polyA 3' processing signals. Nucleic Acids Res *17*, 9113–9126.

[Margueron et al., 2009] Margueron, R., Justin, N., Ohno, K., Sharpe, M. L., Son, J., Drury, W. J., Voigt, P., Martin, S. R., Taylor, W. R., Marco, V. D., Pirrotta, V., Reinberg, D. and Gamblin, S. J. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762–767.

[Marino-Ramirez et al., 2006] Marino-Ramirez, L., Jordan, I. K. and Landsman, D. (2006). Multiple independent evolutionary solutions to core histone gene regulation. Genome Biol *7*, R122.

[Marteijn et al., 2009] Marteijn, J. A., Bekker-Jensen, S., Mailand, N., Lans, H., Schwertman, P., Gourdin, A. M., Dantuma, N. P., Lukas, J. and Vermeulen, W. (2009). Nucleotide excision repair-induced H2A ubiquitination is dependent on MDC1 and RNF8 and reveals a universal DNA damage response. J Cell Biol *186*, 835–847.

[Martin and Zhang, 2005] Martin, C. and Zhang, Y. (2005). The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol *6*, 838–849.

[Martin et al., 2006] Martin, D. G. E., Baetz, K., Shi, X., Walter, K. L., MacDonald, V. E., Wlodarski, M. J., Gozani, O., Hieter, P. and Howe, L. (2006). The Yng1p plant homeodomain finger is a methyl-histone binding module that recognizes lysine 4-methylated histone H3. Mol Cell Biol *26*, 7871–7879.

[Martinez et al., 2008] Martinez, A., Hinz, J. M., Gomez, L., Molina, B., Acuna, H., Jones, I. M., Frias, S. and Coleman, M. A. (2008). Differential expression of TP53 associated genes in Fanconi anemia cells after mitomycin C and hydroxyurea treatment. Mutat Res *656*, 1–7.

[Marzluff and Pandey, 1988] Marzluff, W. F. and Pandey, N. B. (1988). Multiple regulatory steps control histone mRNA concentrations. Trends Biochem Sci *13*, 49–52.

[Mateescu et al., 2004] Mateescu, B., England, P., Halgand, F., Yaniv, M. and Muchardt, C. (2004). Tethering of HP1 proteins to chromatin is relieved by phosphoacetylation of histone H3. EMBO Rep *5*, 490–496.

[Mathew, 2006] Mathew, C. G. (2006). Fanconi anaemia genes and susceptibility to cancer. Oncogene *25*, 5875–5884.

[Matsuoka et al., 2007] Matsuoka, S., Ballif, B. A., Smogorzewska, A., McDonald, E. R., Hurov, K. E., Luo, J., Bakalarski, C. E., Zhao, Z., Solimini, N., Lerenthal, Y., Shiloh, Y., Gygi, S. P. and Elledge, S. J. (2007). ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science *316*, 1160–1166.

[Maurer-Stroh et al., 2003] Maurer-Stroh, S., Dickens, N. J., Hughes-Davies, L., Kouzarides, T., Eisenhaber, F. and Ponting, C. P. (2003). The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. Trends Biochem Sci *28*, 69–74.

[McKittrick et al., 2004] McKittrick, E., Gafken, P. R., Ahmad, K. and Henikoff, S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. Proc Natl Acad Sci U S A *101*, 1525–1530.

[Medzihradszky et al., 2004] Medzihradszky, K. F., Zhang, X., Chalkley, R. J., Guan, S., McFarland, M. A., Chalmers, M. J., Marshall, A. G., Diaz, R. L., Allis, C. D. and Burlingame, A. L. (2004). Characterization of Tetrahymena histone H2B variants and posttranslational populations by electron capture dissociation (ECD) Fourier transform ion cyclotron mass spectrometry (FT-ICR MS). Mol Cell Proteomics *3*, 872–886.

[Melander et al., 2008] Melander, F., Bekker-Jensen, S., Falck, J., Bartek, J., Mailand, N. and Lukas, J. (2008). Phosphorylation of SDT repeats in the MDC1 N terminus triggers retention of NBS1 at the DNA damage-modified chromatin. J Cell Biol *181*, 213–226.

[Mello et al., 2002] Mello, J. A., Sillje, H. H. W., Roche, D. M. J., Kirschner, D. B., Nigg, E. A. and Almouzni, G. (2002). Human Asf1 and CAF-1 interact and synergize in a repair-coupled nucleosome assembly pathway. EMBO Rep *3*, 329–334.

[Mellone and Allshire, 2003] Mellone, B. G. and Allshire, R. C. (2003). Stretching it: putting the CEN(P-A) in centromere. Curr Opin Genet Dev *13*, 191–198.

[Mendez-Acuna et al., 2010] Mendez-Acuna, L., Tomaso, M. V. D., Palitti, F. and Martinez-Lopez, W. (2010). Histone post-translational modifications in DNA damage response. Cytogenet Genome Res *128*, 28–36.

[Mermoud et al., 1999] Mermoud, J. E., Costanzi, C., Pehrson, J. R. and Brockdorff, N. (1999). Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. J Cell Biol *147*, 1399–1408.

[Metzger et al., 2005] Metzger, E., Wissmann, M., Yin, N., Muller, J. M., Schneider, R., Peters, A. H. F. M., Gunther, T., Buettner, R. and Schule, R. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature *437*, 436–439.

[Miller et al., 2010] Miller, K. M., Tjeertes, J. V., Coates, J., Legube, G., Polo, S. E., Britton, S. and Jackson, S. P. (2010). Human HDAC1 and HDAC2 function in the DNA-damage

response to promote DNA nonhomologous end-joining. Nat Struct Mol Biol *17*, 1144–1151.

[Mito et al., 2005] Mito, Y., Henikoff, J. G. and Henikoff, S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. Nat Genet *37*, 1090–1097.

[Mito et al., 2007] Mito, Y., Henikoff, J. G. and Henikoff, S. (2007). Histone replacement marks the boundaries of cis-regulatory domains. Science *315*, 1408–1411.

[Miyagi and Rao, 2007] Miyagi, M. and Rao, K. C. S. (2007). Proteolytic 18O-labeling strategies for quantitative proteomics. Mass Spectrom Rev *26*, 121–136.

[Mizzen and McLachlan, 2000] Mizzen, C. A. and McLachlan, D. R. (2000). Capillary electrophoresis of histone H1 variants at neutral pH in dynamically modified fused- silica tubing. Electrophoresis *21*, 2359–2367.

[Mizzen et al., 1996] Mizzen, C. A., Yang, X. J., Kokubo, T., Brownell, J. E., Bannister, A. J., Owen-Hughes, T., Workman, J., Wang, L., Berger, S. L., Kouzarides, T., Nakatani, Y. and Allis, C. D. (1996). The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. Cell *87*, 1261–1270.

[Moldovan and D'Andrea, 2009] Moldovan, G.-L. and D'Andrea, A. D. (2009). How the fanconi anemia pathway guards the genome. Annu Rev Genet *43*, 223–249.

[Monetti et al., 2011] Monetti, M., Nagaraj, N., Sharma, K. and Mann, M. (2011). Large-scale phosphosite quantification in tissues by a spike-in SILAC method. Nat Methods *8*, 655–658.

[Morillo-Huesca et al., 2010] Morillo-Huesca, M., Clemente-Ruiz, M., Andujar, E. and Prado, F. (2010). The SWR1 histone replacement complex causes genetic instability and genome-wide transcription misregulation in the absence of H2A.Z. PLoS One *5*, e12143.

[Morris et al., 2007] Morris, S. A., Rao, B., Garcia, B. A., Hake, S. B., Diaz, R. L., Shabanowitz, J., Hunt, D. F., Allis, C. D., Lieb, J. D. and Strahl, B. D. (2007). Identification of histone H3 lysine 36 acetylation as a highly conserved histone modification. J Biol Chem *282*, 7632–7640.

[Mosammaparast and Shi, 2010] Mosammaparast, N. and Shi, Y. (2010). Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. Annu Rev Biochem *79*, 155–179.

[Mujtaba et al., 2007] Mujtaba, S., Zeng, L. and Zhou, M.-M. (2007). Structure and acetyl-lysine recognition of the bromodomain. Oncogene *26*, 5521–5527.

[Mukherjee et al., 2006] Mukherjee, B., Kessinger, C., Kobayashi, J., Chen, B. P. C., Chen, D. J., Chatterjee, A. and Burma, S. (2006). DNA-PK phosphorylates histone H2AX during apoptotic DNA fragmentation in mammalian cells. DNA Repair (Amst) *5*, 575–590.

[Munro et al., 2010] Munro, S., Khaire, N., Inche, A., Carr, S. and Thangue, N. B. L. (2010). Lysine methylation regulates the pRb tumour suppressor protein. Oncogene *29*, 2357–2367.

[Murr et al., 2006] Murr, R., Loizou, J. I., Yang, Y.-G., Cuenin, C., Li, H., Wang, Z.-Q. and Herceg, Z. (2006). Histone acetylation by Trrap-Tip60 modulates loading of repair proteins and repair of DNA double-strand breaks. Nat Cell Biol *8*, 91–99.

[Nakayama et al., 2007] Nakayama, T., Nishioka, K., Dong, Y.-X., Shimojima, T. and Hirose, S. (2007). Drosophila GAGA factor directs histone H3.3 replacement that prevents the heterochromatin spreading. Genes Dev *21*, 552–561.

[Navajas et al., 2011] Navajas, R., Paradela, A. and Albar, J. P. (2011). Immobilized metal affinity chromatography/reversed-phase enrichment of phosphopeptides and analysis by CID/ETD tandem mass spectrometry. Methods Mol Biol *681*, 337–348.

[NEELIN and CONNELL, 1959] NEELIN, J. M. and CONNELL, G. E. (1959). Zone electrophoresis of chicken-erythrocyte histone in starch gel. Biochim Biophys Acta *31*, 539–541.

[NEELIN and NEELIN, 1960] NEELIN, J. M. and NEELIN, E. M. (1960). Zone electrophoresis of calf thymus histone in starch gel. Can J Biochem Physiol *38*, 355–363.

[Nelson et al., 2006] Nelson, C. J., Santos-Rosa, H. and Kouzarides, T. (2006). Proline isomerization of histone H3 regulates lysine methylation and gene expression. Cell *126*, 905–916.

[Ng et al., 2002] Ng, H. H., Feng, Q., Wang, H., Erdjument-Bromage, H., Tempst, P., Zhang, Y. and Struhl, K. (2002). Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. Genes Dev *16*, 1518–1527.

[Ng et al., 2009] Ng, S. S., Yue, W. W., Oppermann, U. and Klose, R. J. (2009). Dynamic protein methylation in chromatin biology. Cell Mol Life Sci *66*, 407–422.

[Nicklay et al., 2009] Nicklay, J. J., Shechter, D., Chitta, R. K., Garcia, B. A., Shabanowitz, J., Allis, C. D. and Hunt, D. F. (2009). Analysis of histones in Xenopus laevis. II. mass spectrometry reveals an index of cell type-specific modifications on H3 and H4. J Biol Chem *284*, 1075–1085.

[Niedernhofer et al., 2005] Niedernhofer, L. J., Lalai, A. S. and Hoeijmakers, J. H. J. (2005). Fanconi anemia (cross)linked to DNA repair. Cell *123*, 1191–1198.

[Niedzwiedz et al., 2004] Niedzwiedz, W., Mosedale, G., Johnson, M., Ong, C. Y., Pace, P. and Patel, K. J. (2004). The Fanconi anaemia gene FANCC promotes homologous recombination and error-prone DNA repair. Mol Cell *15*, 607–620.

[Niida and Nakanishi, 2006] Niida, H. and Nakanishi, M. (2006). DNA damage checkpoints in mammals. Mutagenesis *21*, 3–9.

[Nishioka et al., 2002a] Nishioka, K., Chuikov, S., Sarma, K., Erdjument-Bromage, H., Allis, C. D., Tempst, P. and Reinberg, D. (2002a). Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. Genes Dev *16*, 479–489.

[Nishioka et al., 2002b] Nishioka, K., Rice, J. C., Sarma, K., Erdjument-Bromage, H., Werner, J., Wang, Y., Chuikov, S., Valenzuela, P., Tempst, P., Steward, R., Lis, J. T., Allis, C. D. and Reinberg, D. (2002b). PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. Mol Cell *9*, 1201–1213.

[Niu et al., 2010] Niu, N., Qin, Y., Fridley, B. L., Hou, J., Kalari, K. R., Zhu, M., Wu, T.-Y., Jenkins, G. D., Batzler, A. and Wang, L. (2010). Radiation pharmacogenomics: a genome-wide association approach to identify radiation response biomarkers using human lymphoblastoid cell lines. Genome Res *20*, 1482–1492.

[Oda et al., 1999] Oda, Y., Huang, K., Cross, F. R., Cowburn, D. and Chait, B. T. (1999). Accurate quantitation of protein expression and site-specific phosphorylation. Proc Natl Acad Sci U S A *96*, 6591–6596.

[Ogryzko, 2001] Ogryzko, V. V. (2001). Mammalian histone acetyltransferases and their complexes. Cell Mol Life Sci *58*, 683–692.

[Ogryzko, 2008] Ogryzko, V. V. (2008). Erwin Schroedinger, Francis Crick and epigenetic stability. Biol Direct *3*, 15.

[Ogryzko et al., 1996] Ogryzko, V. V., Schiltz, R. L., Russanova, V., Howard, B. H. and Nakatani, Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell *87*, 953–959.

[Okuwaki et al., 2005] Okuwaki, M., Kato, K., Shimahara, H., ichi Tate, S. and Nagata, K. (2005). Assembly and disassembly of nucleosome core particles containing histone variants by human nucleosome assembly protein I. Mol Cell Biol *25*, 10639–10651.

[Olsen et al., 2006] Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P. and Mann, M. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell *127*, 635–648.

[Olsen et al., 2010] Olsen, J. V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M. L., Jensen, L. J., Gnad, F., Cox, J., Jensen, T. S., Nigg, E. A., Brunak, S. and Mann, M. (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci Signal *3*, ra3.

[Ong and Corces, 2009] Ong, C.-T. and Corces, V. G. (2009). Insulators as mediators of intra- and inter-chromosomal interactions: a common evolutionary theme. J Biol *8*, 73.

[Ong et al., 2002] Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics *1*, 376–386.

[Ong et al., 2003] Ong, S.-E., Kratchmarova, I. and Mann, M. (2003). Properties of 13C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). J Proteome Res *2*, 173–181.

[Ong and Mann, 2005] Ong, S.-E. and Mann, M. (2005). Mass spectrometry-based proteomics turns quantitative. Nat Chem Biol *1*, 252–262.

[Ong and Mann, 2006] Ong, S.-E. and Mann, M. (2006). A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). Nat Protoc *1*, 2650–2660.

[Ong et al., 2004] Ong, S.-E., Mittler, G. and Mann, M. (2004). Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. Nat Methods *1*, 119–126.

[Ooi et al., 2010] Ooi, S. L., Henikoff, J. G. and Henikoff, S. (2010). A native chromatin purification system for epigenomic profiling in Caenorhabditis elegans. Nucleic Acids Res *38*, e26.

[Ooi et al., 2006] Ooi, S. L., Priess, J. R. and Henikoff, S. (2006). Histone H3.3 variant dynamics in the germline of Caenorhabditis elegans. PLoS Genet *2*, e97.

[Ouararhni et al., 2006] Ouararhni, K., Hadj-Slimane, R., Ait-Si-Ali, S., Robin, P., Mietton, F., Harel-Bellan, A., Dimitrov, S. and Hamiche, A. (2006). The histone variant mH2A1.1 interferes with transcription by down-regulating PARP-1 enzymatic activity. Genes Dev *20*, 3324–3336.

[Owen et al., 2000] Owen, D. J., Ornaghi, P., Yang, J. C., Lowe, N., Evans, P. R., Ballario, P., Neuhaus, D., Filetici, P. and Travers, A. A. (2000). The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. EMBO J *19*, 6141–6149.

[Ozdag et al., 2006] Ozdag, H., Teschendorff, A. E., Ahmed, A. A., Hyland, S. J., Blenkiron, C., Bobrow, L., Veerakumarasivam, A., Burtt, G., Subkhankulova, T., Arends, M. J., Collins, V. P., Bowtell, D., Kouzarides, T., Brenton, J. D. and Caldas, C. (2006). Differential expression of selected histone modifier genes in human solid cancers. BMC Genomics *7*, 90.

[Palmer et al., 1991] Palmer, D. K., O'Day, K., Trong, H. L., Charbonneau, H. and Margolis, R. L. (1991). Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc Natl Acad Sci U S A *88*, 3734–3738.

[Park et al., 2006a] Park, J.-H., Park, E.-J., Lee, H.-S., Kim, S. J., Hur, S.-K., Imbalzano, A. N. and Kwon, J. (2006a). Mammalian SWI/SNF complexes facilitate DNA double-strand break repair by promoting gamma-H2AX induction. EMBO J *25*, 3986–3997.

[Park et al., 2006b] Park, K.-S., Mohapatra, D. P., Misonou, H. and Trimmer, J. S. (2006b). Graded regulation of the Kv2.1 potassium channel by variable phosphorylation. Science *313*, 976–979.

[Park et al., 2004] Park, Y.-J., Dyer, P. N., Tremethick, D. J. and Luger, K. (2004). A new fluorescence resonance energy transfer approach demonstrates that the histone variant H2AZ stabilizes the histone octamer within the nucleosome. J Biol Chem *279*, 24274–24282.

[Parks et al., 2007] Parks, B. A., Jiang, L., Thomas, P. M., Wenger, C. D., Roth, M. J., Boyne, M. T., Burke, P. V., Kwast, K. E. and Kelleher, N. L. (2007). Top-down proteomics on a chromatographic time scale using linear ion trap fourier transform hybrid mass spectrometers. Anal Chem *79*, 7984–7991.

[Parthun, 2007] Parthun, M. R. (2007). Hat1: the emerging cellular roles of a type B histone acetyltransferase. Oncogene *26*, 5319–5328.

[Paull et al., 2000] Paull, T. T., Rogakou, E. P., Yamazaki, V., Kirchgessner, C. U., Gellert, M. and Bonner, W. M. (2000). A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. Curr Biol *10*, 886–895.

[Pawson et al., 2001] Pawson, T., Gish, G. D. and Nash, P. (2001). SH2 domains, interaction modules and cellular wiring. Trends Cell Biol *11*, 504–511.

[Pehrson et al., 1997] Pehrson, J. R., Costanzi, C. and Dharia, C. (1997). Developmental and tissue expression patterns of histone macroH2A1 subtypes. J Cell Biochem *65*, 107–113.

[Pehrson and Fuji, 1998] Pehrson, J. R. and Fuji, R. N. (1998). Evolutionary conservation of histone macroH2A subtypes and domains. Nucleic Acids Res *26*, 2837–2842.

[Perkel, 2009] Perkel, J. M. (2009). iTRAQ gets put to the test. J Proteome Res *8*, 4885.

[Perpelescu and Fukagawa, 2011] Perpelescu, M. and Fukagawa, T. (2011). The ABCs of CENPs. Chromosoma *412*.

[Pesavento et al., 2008] Pesavento, J. J., Bullock, C. R., LeDuc, R. D., Mizzen, C. A. and Kelleher, N. L. (2008). Combinatorial modification of human histone H4 quantitated by two-dimensional liquid chromatography coupled with top down mass spectrometry. J Biol Chem *283*, 14927–14937.

[Pesavento et al., 2007] Pesavento, J. J., Garcia, B. A., Streeky, J. A., Kelleher, N. L. and Mizzen, C. A. (2007). Mild performic acid oxidation enhances chromatographic and top down mass spectrometric analyses of histones. Mol Cell Proteomics *6*, 1510–1526.

[Pesavento et al., 2004] Pesavento, J. J., Kim, Y.-B., Taylor, G. K. and Kelleher, N. L. (2004). Shotgun annotation of histone modifications: a new approach for streamlined characterization of proteins by top down mass spectrometry. J Am Chem Soc *126*, 3386–3387.

[Pesavento et al., 2006] Pesavento, J. J., Mizzen, C. A. and Kelleher, N. L. (2006). Quantitative analysis of modified proteins and their positional isomers by tandem mass spectrometry: human histone H4. Anal Chem *78*, 4271–4280.

[Pesavento et al., 2008] Pesavento, J. J., Yang, H., Kelleher, N. L. and Mizzen, C. A. (2008). Certain and progressive methylation of histone H4 at lysine 20 during the cell cycle. Mol Cell Biol *28*, 468–486.

[Peters et al., 2003] Peters, A. H. F. M., Kubicek, S., Mechtler, K., O'Sullivan, R. J., Derijck, A. A. H. A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., Martens, J. H. A. and Jenuwein, T. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol Cell *12*, 1577–1589.

[Petersen et al., 2001] Petersen, S., Casellas, R., Reina-San-Martin, B., Chen, H. T., Difilippantonio, M. J., Wilson, P. C., Hanitsch, L., Celeste, A., Muramatsu, M., Pilch, D. R., Redon, C., Ried, T., Bonner, W. M., Honjo, T., Nussenzweig, M. C. and Nussenzweig, A. (2001). AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. Nature *414*, 660–665.

[Petrie et al., 2003] Petrie, K., Guidez, F., Howell, L., Healy, L., Waxman, S., Greaves, M. and Zelent, A. (2003). The histone deacetylase 9 gene encodes multiple protein isoforms. J Biol Chem *278*, 16059–16072.

[Phanstiel et al., 2008] Phanstiel, D., Brumbaugh, J., Berggren, W. T., Conard, K., Feng, X., Levenstein, M. E., McAlister, G. C., Thomson, J. A. and Coon, J. J. (2008). Mass spectrometry identifies and quantifies 74 unique histone H4 isoforms in differentiating human embryonic stem cells. Proc Natl Acad Sci U S A *105*, 4093–4098.

[PHILLIPS, 1963] PHILLIPS, D. M. (1963). The presence of acetyl groups of histones. Biochem J *87*, 258–263.

[Pina and Suau, 1987] Pina, B. and Suau, P. (1987). Changes in histones H2A and H3 variant composition in differentiating and mature rat brain cortical neurons. Dev Biol *123*, 51–58.

[Plans et al., 2006] Plans, V., Scheper, J., Soler, M., Loukili, N., Okano, Y. and Thomson, T. M. (2006). The RING finger protein RNF8 recruits UBC13 for lysine 63-based self polyubiquitylation. J Cell Biochem *97*, 572–582.

[Plazas-Mayorca et al., 2009] Plazas-Mayorca, M. D., Zee, B. M., Young, N. L., Fingerman, I. M., LeRoy, G., Briggs, S. D. and Garcia, B. A. (2009). One-pot shotgun quantitative mass spectrometry characterization of histones. J Proteome Res *8*, 5367–5374.

[Polo et al., 2006] Polo, S. E., Roche, D. and Almouzni, G. (2006). New histone incorporation marks sites of UV repair in human cells. Cell *127*, 481–493.

[Popanda et al., 2009] Popanda, O., Marquardt, J. U., Chang-Claude, J. and Schmezer, P. (2009). Genetic variation in normal tissue toxicity induced by ionizing radiation. Mutat Res *667*, 58–69.

[Prakash et al., 2005] Prakash, S., Johnson, R. E. and Prakash, L. (2005). Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. Annu Rev Biochem *74*, 317–353.

[Prokhorova et al., 2009] Prokhorova, T. A., Rigbolt, K. T. G., Johansen, P. T., Henningsen, J., Kratchmarova, I., Kassem, M. and Blagoev, B. (2009). Stable isotope labeling by amino acids in cell culture (SILAC) and quantitative comparison of the membrane proteomes of

self-renewing and differentiating human embryonic stem cells. Mol Cell Proteomics *8*, 959–970.

[Pugh et al., 2009] Pugh, T. J., Keyes, M., Barclay, L., Delaney, A., Krzywinski, M., Thomas, D., Novik, K., Yang, C., Agranovich, A., McKenzie, M., Morris, W. J., Olive, P. L., Marra, M. A. and Moore, R. A. (2009). Sequence variant discovery in DNA repair genes from radiosensitive and radiotolerant prostate brachytherapy patients. Clin Cancer Res *15*, 5008–5016.

[Ramos-Fernandez et al., 2007] Ramos-Fernandez, A., Lopez-Ferrer, D. and Vazquez, J. (2007). Improved method for differential expression proteomics using trypsin-catalyzed 18O labeling with a correction for labeling efficiency. Mol Cell Proteomics *6*, 1274–1286.

[Ramus et al., 2006] Ramus, C., de Peredo, A. G., Dahout, C., Gallagher, M. and Garin, J. (2006). An optimized strategy for ICAT quantification of membrane proteins. Mol Cell Proteomics *5*, 68–78.

[Rangasamy et al., 2003] Rangasamy, D., Berven, L., Ridgway, P. and Tremethick, D. J. (2003). Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development. EMBO J *22*, 1599–1607.

[Ransom et al., 2010] Ransom, M., Dennehey, B. K. and Tyler, J. K. (2010). Chaperoning histones during DNA replication and repair. Cell *140*, 183–195.

[Raschle et al., 2008] Raschle, M., Knipscheer, P., Enoiu, M., Angelov, T., Sun, J., Griffith, J. D., Ellenberger, T. E., Scharer, O. D. and Walter, J. C. (2008). Mechanism of replication-coupled DNA interstrand crosslink repair. Cell *134*, 969–980.

[Rashi-Elkeles et al., 2011] Rashi-Elkeles, S., Elkon, R., Shavit, S., Lerenthal, Y., Linhart, C., Kupershtein, A., Amariglio, N., Rechavi, G., Shamir, R. and Shiloh, Y. (2011). Transcriptional modulation induced by ionizing radiation: p53 remains a central player. Mol Oncol *5*.

[Rea et al., 2000] Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D. and Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593–599.

[Redon et al., 2002] Redon, C., Pilch, D., Rogakou, E., Sedelnikova, O., Newrock, K. and Bonner, W. (2002). Histone H2A variants H2AX and H2AZ. Curr Opin Genet Dev *12*, 162–169.

[Reese et al., 2003] Reese, B. E., Bachman, K. E., Baylin, S. B. and Rountree, M. R. (2003). The methyl-CpG binding protein MBD1 interacts with the p150 subunit of chromatin assembly factor 1. Mol Cell Biol *23*, 3226–3236.

[Reitsema et al., 2005] Reitsema, T., Klokov, D., Banath, J. P. and Olive, P. L. (2005). DNA-PK is responsible for enhanced phosphorylation of histone H2AX under hypertonic conditions. DNA Repair (Amst) *4*, 1172–1181.

[Ren and Gorovsky, 2001] Ren, Q. and Gorovsky, M. A. (2001). Histone H2A.Z acetylation modulates an essential charge patch. Mol Cell *7*, 1329–1335.

[Ren and Gorovsky, 2003] Ren, Q. and Gorovsky, M. A. (2003). The nonessential H2A N-terminal tail can function as an essential charge patch on the H2A.Z variant N-terminal tail. Mol Cell Biol *23*, 2778–2789.

[Reynolds et al., 2002] Reynolds, K. J., Yao, X. and Fenselau, C. (2002). Proteolytic 18O labeling for comparative proteomics: evaluation of endoprotease Glu-C as the catalytic agent. J Proteome Res *1*, 27–33.

[Ridgway et al., 2004] Ridgway, P., Brown, K. D., Rangasamy, D., Svensson, U. and Tremethick, D. J. (2004). Unique residues on the H2A.Z containing nucleosome surface are important for Xenopus laevis development. J Biol Chem *279*, 43815–43820.

[Rigbolt et al., 2011] Rigbolt, K. T. G., Prokhorova, T. A., Akimov, V., Henningsen, J., Johansen, P. T., Kratchmarova, I., Kassem, M., Mann, M., Olsen, J. V. and Blagoev, B. (2011). System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. Sci Signal *4*, rs3.

[Robinson et al., 2006] Robinson, P. J. J., Fairall, L., Huynh, V. A. T. and Rhodes, D. (2006). EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. Proc Natl Acad Sci U S A *103*, 6506–6511.

[Robinson and Rhodes, 2006] Robinson, P. J. J. and Rhodes, D. (2006). Structure of the '30 nm' chromatin fibre: a key role for the linker histone. Curr Opin Struct Biol *16*, 336–343.

[Rogakou et al., 1998] Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. and Bonner, W. M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J Biol Chem *273*, 5858–5868.

[Rose et al., 1983] Rose, K., Simona, M. G., Offord, R. E., Prior, C. P., Otto, B. and Thatcher, D. R. (1983). A new mass-spectrometric C-terminal sequencing technique finds a similarity between gamma-interferon and alpha 2-interferon and identifies a proteolytically clipped gamma-interferon that retains full antiviral activity. Biochem J *215*, 273–277.

[Ross et al., 2004] Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A. and Pappin, D. J. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics *3*, 1154–1169.

[Rosselli et al., 2003] Rosselli, F., Briot, D. and Pichierri, P. (2003). The Fanconi anemia pathway and the DNA interstrand cross-links repair. Biochimie *85*, 1175–1184.

[Ryan et al., 1998] Ryan, M. P., Jones, R. and Morse, R. H. (1998). SWI-SNF complex participation in transcriptional activation at a step subsequent to activator binding. Mol Cell Biol *18*, 1774–1782.

[Sakai et al., 2009] Sakai, A., Schwartz, B. E., Goldstein, S. and Ahmad, K. (2009). Transcriptional and developmental functions of the H3.3 histone variant in Drosophila. Curr Biol *19*, 1816–1820.

[Santaguida and Musacchio, 2009] Santaguida, S. and Musacchio, A. (2009). The life and miracles of kinetochores. EMBO J *28*, 2511–2531.

[Santisteban et al., 2000] Santisteban, M. S., Kalashnikova, T. and Smith, M. M. (2000). Histone H2A.Z regulats transcription and is partially redundant with nucleosome remodeling complexes. Cell *103*, 411–422.

[Santos-Rosa et al., 2009] Santos-Rosa, H., Kirmizis, A., Nelson, C., Bartke, T., Saksouk, N., Cote, J. and Kouzarides, T. (2009). Histone H3 tail clipping regulates gene expression. Nat Struct Mol Biol *16*, 17–22.

[Santos-Rosa et al., 2002] Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C. T., Schreiber, S. L., Mellor, J. and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407–411.

[Sarg et al., 2006] Sarg, B., Helliger, W., Talasz, H., FÃ¶rg, B. and Lindner, H. H. (2006). Histone H1 phosphorylation occurs site-specifically during interphase and mitosis: identification of a novel phosphorylation site on histone H1. J Biol Chem *281*, 6573–6580.

[Sarg et al., 2004] Sarg, B., Helliger, W., Talasz, H., Koutzamani, E. and Lindner, H. H. (2004). Histone H4 hyperacetylation precludes histone H4 lysine 20 trimethylation. J Biol Chem *279*, 53458–53464.

[Sarg et al., 2002] Sarg, B., Koutzamani, E., Helliger, W., Rundquist, I. and Lindner, H. H. (2002). Postsynthetic trimethylation of histone H4 at lysine 20 in mammalian tissues is associated with aging. J Biol Chem *277*, 39195–39201.

[Sarraf and Stancheva, 2004] Sarraf, S. A. and Stancheva, I. (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Mol Cell *15*, 595–605.

[Savic et al., 2009] Savic, V., Yin, B., Maas, N. L., Bredemeyer, A. L., Carpenter, A. C., Helmink, B. A., Yang-Iott, K. S., Sleckman, B. P. and Bassing, C. H. (2009). Formation of dynamic gamma-H2AX domains along broken DNA strands is distinctly regulated by ATM and MDC1 and dependent upon H2AX densities in chromatin. Mol Cell *34*, 298–310.

[Schalch et al., 2005] Schalch, T., Duda, S., Sargent, D. F. and Richmond, T. J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature *436*, 138–141.

[Scharer, 2005] Scharer, O. D. (2005). DNA interstrand crosslinks: natural and drug-induced DNA adducts that induce unique cellular responses. Chembiochem *6*, 27–32.

[Schmidt et al., 2005] Schmidt, A., Kellermann, J. and Lottspeich, F. (2005). A novel strategy for quantitative proteomics using isotope-coded protein labels. Proteomics *5*, 4–15.

[Schneider et al., 2004a] Schneider, R., Bannister, A. J., Myers, F. A., Thorne, A. W., Crane-Robinson, C. and Kouzarides, T. (2004a). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. Nat Cell Biol *6*, 73–77.

[Schneider et al., 2004b] Schneider, R., Bannister, A. J., Weise, C. and Kouzarides, T. (2004b). Direct binding of INHAT to H3 tails disrupted by modifications. J Biol Chem *279*, 23859–23862.

[Scott et al., 1976] Scott, B. R., Pathak, M. A. and Mohn, G. R. (1976). Molecular and genetic basis of furocoumarin reactions. Mutat Res *39*, 29–74.

[Seligson et al., 2009] Seligson, D. B., Horvath, S., McBrian, M. A., Mah, V., Yu, H., Tze, S., Wang, Q., Chia, D., Goodglick, L. and Kurdistani, S. K. (2009). Global levels of histone modifications predict prognosis in different cancers. Am J Pathol *174*, 1619–1628.

[Seligson et al., 2005] Seligson, D. B., Horvath, S., Shi, T., Yu, H., Tze, S., Grunstein, M. and Kurdistani, S. K. (2005). Global histone modification patterns predict risk of prostate cancer recurrence. Nature *435*, 1262–1266.

[Shevchenko et al., 1996] Shevchenko, A., Wilm, M., Vorm, O. and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem *68*, 850–858.

[Shi et al., 2006] Shi, X., Hong, T., Walter, K. L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Pena, P., Lan, F., Kaadige, M. R., Lacoste, N., Cayrou, C., Davrazou, F., Saha, A., Cairns, B. R., Ayer, D. E., Kutateladze, T. G., Shi, Y., Jacques Cote, J., Chua, K. F. and Gozani, O. (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature *442*, 96–99.

[Shi et al., 2004] Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A. and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941–953.

[Shibahara and Stillman, 1999] Shibahara, K. and Stillman, B. (1999). Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. Cell *96*, 575–585.

[Shiio and Aebersold, 2006] Shiio, Y. and Aebersold, R. (2006). Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry. Nat Protoc *1*, 139–145.

[Shim et al., 2007] Shim, E. Y., Hong, S. J., Oum, J.-H., Yanez, Y., Zhang, Y. and Lee, S. E. (2007). RSC mobilizes nucleosomes to improve accessibility of repair machinery to the damaged chromatin. Mol Cell Biol *27*, 1602–1613.

[Shogren-Knaak et al., 2006] Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M. J., Davie, J. R. and Peterson, C. L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science *311*, 844–847.

[Shroff et al., 2004] Shroff, R., Arbel-Eden, A., Pilch, D., Ira, G., Bonner, W. M., Petrini, J. H., Haber, J. E. and Lichten, M. (2004). Distribution and dynamics of chromatin modification induced by a defined DNA double-strand break. Curr Biol *14*, 1703–1711.

[Shuaib et al., 2010] Shuaib, M., Ouararhni, K., Dimitrov, S. and Hamiche, A. (2010). HJURP binds CENP-A via a highly conserved N-terminal domain and mediates its deposition at centromeres. Proc Natl Acad Sci U S A *107*, 1349–1354.

[Shukla and Futrell, 2000] Shukla and Futrell (2000). Tandem mass spectrometry: dissociation of ions by collisional activation. J Mass Spectrom *35*, 1069–1090.

[Sims et al., 2005] Sims, R. J., Chen, C.-F., Santos-Rosa, H., Kouzarides, T., Patel, S. S. and Reinberg, D. (2005). Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. J Biol Chem *280*, 41789–41792.

[Siuti and Kelleher, 2010] Siuti, N. and Kelleher, N. L. (2010). Efficient readout of posttranslational codes on the 50-residue tail of histone H3 by high-resolution MS/MS. Anal Biochem *396*, 180–187.

[Siuti et al., 2006] Siuti, N., Roth, M. J., Mizzen, C. A., Kelleher, N. L. and Pesavento, J. J. (2006). Gene-specific characterization of human histone H2B by electron capture dissociation. J Proteome Res *5*, 233–239.

[Sivolob and Prunell, 2003] Sivolob, A. and Prunell, A. (2003). Linker histone-dependent organization and dynamics of nucleosome entry/exit DNAs. J Mol Biol *331*, 1025–1040.

[Smith and Denu, 2009] Smith, B. C. and Denu, J. M. (2009). Chemical mechanisms of histone lysine and arginine modifications. Biochim Biophys Acta *1789*, 45–57.

[Smith et al., 2003] Smith, C. M., Gafken, P. R., Zhang, Z., Gottschling, D. E., Smith, J. B. and Smith, D. L. (2003). Mass spectrometric quantification of acetylation at specific lysines within the amino-terminal tail of histone H4. Anal Biochem *316*, 23–33.

[Smith et al., 2002] Smith, C. M., Haimberger, Z. W., Johnson, C. O., Wolf, A. J., Gafken, P. R., Zhang, Z., Parthun, M. R. and Gottschling, D. E. (2002). Heritable chromatin structure: mapping "memory" in histones H3 and H4. Proc Natl Acad Sci U S A *99 Suppl 4*, 16454–16461.

[Smogorzewska et al., 2007] Smogorzewska, A., Matsuoka, S., Vinciguerra, P., McDonald, E. R., Hurov, K. E., Luo, J., Ballif, B. A., Gygi, S. P., Hofmann, K., D'Andrea, A. D. and Elledge, S. J. (2007). Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. Cell *129*, 289–301.

[Sobel et al., 1994] Sobel, R. E., Cook, R. G. and Allis, C. D. (1994). Non-random acetylation of histone H4 by a cytoplasmic histone acetyltransferase as determined by novel methodology. J Biol Chem *269*, 18576–18582.

[Soulas-Sprauel et al., 2007] Soulas-Sprauel, P., Rivera-Munoz, P., Malivert, L., Le Guyader, G., Abramowski, V., Revy, P. and de Villartay, J.-P. (2007). V(D)J and immunoglobulin class switch recombinations: a paradigm to study the regulation of DNA end-joining. Oncogene *26*, 7780–7791.

[Spector, 2001] Spector, D. L. (2001). Nuclear domains. J Cell Sci *114*, 2891–2893.

[Spycher et al., 2008] Spycher, C., Miller, E. S., Townsend, K., Pavic, L., Morrice, N. A., Janscak, P., Stewart, G. S. and Stucki, M. (2008). Constitutive phosphorylation of MDC1 physically links the MRE11-RAD50-NBS1 complex to damaged chromatin. J Cell Biol *181*, 227–240.

[Staller, 2010] Staller, P. (2010). Genetic heterogeneity and chromatin modifiers in renal clear cell carcinoma. Future Oncol *6*, 897–900.

[Stargell et al., 1993] Stargell, L. A., Bowen, J., Dadd, C. A., Dedon, P. C., Davis, M., Cook, R. G., Allis, C. D. and Gorovsky, M. A. (1993). Temporal and spatial association of histone H2A variant hv1 with transcriptionally competent chromatin during nuclear development in Tetrahymena thermophila. Genes Dev *7*, 2641–2651.

[Stassen et al., 1995] Stassen, M. J., Bailey, D., Nelson, S., Chinwalla, V. and Harte, P. J. (1995). The Drosophila trithorax proteins contain a novel variant of the nuclear receptor type DNA binding domain and an ancient conserved motif found in other chromosomal proteins. Mech Dev *52*, 209–223.

[Steward et al., 2006] Steward, M. M., Lee, J.-S., O'Donovan, A., Wyatt, M., Bernstein, B. E. and Shilatifard, A. (2006). Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. Nat Struct Mol Biol *13*, 852–854.

[Stewart et al., 2009] Stewart, G. S., Panier, S., Townsend, K., Al-Hakim, A. K., Kolas, N. K., Miller, E. S., Nakada, S., Ylanko, J., Olivarius, S., Mendez, M., Oldreive, C., Wildenhain, J., Tagliaferro, A., Pelletier, L., Taubenheim, N., Durandy, A., Byrd, P. J., Stankovic, T., Taylor, A. M. R. and Durocher, D. (2009). The RIDDLE syndrome protein mediates a ubiquitin-dependent signaling cascade at sites of DNA damage. Cell *136*, 420–434.

[Stiff et al., 2004] Stiff, T., O'Driscoll, M., Rief, N., Iwabuchi, K., Lobrich, M. and Jeggo, P. A. (2004). ATM and DNA-PK function redundantly to phosphorylate H2AX after exposure to ionizing radiation. Cancer Res *64*, 2390–2396.

[Strahl and Allis, 2000] Strahl, B. D. and Allis, C. D. (2000). The language of covalent histone modifications. Nature *403*, 41–45.

[Strahl et al., 2001] Strahl, B. D., Briggs, S. D., Brame, C. J., Caldwell, J. A., Koh, S. S., Ma, H., Cook, R. G., Shabanowitz, J., Hunt, D. F., Stallcup, M. R. and Allis, C. D. (2001). Methylation of histone H4 at arginine 3 occurs in vivo and is mediated by the nuclear receptor coactivator PRMT1. Curr Biol *11*, 996–1000.

[Stucki et al., 2005] Stucki, M., Clapperton, J. A., Mohammad, D., Yaffe, M. B., Smerdon, S. J. and Jackson, S. P. (2005). MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. Cell *123*, 1213–1226.

[Su et al., 2009] Su, X., Jacob, N. K., Amunugama, R., Hsu, P.-H., Fishel, R. and Freitas, M. A. (2009). Enrichment and characterization of histones by two-dimensional hydroxyapatite/reversed-phase liquid chromatography-mass spectrometry. Anal Biochem *388*, 47–55.

[Su et al., 2007a] Su, X., Jacob, N. K., Amunugama, R., Lucas, D. M., Knapp, A. R., Ren, C., Davis, M. E., Marcucci, G., Parthun, M. R., Byrd, J. C., Fishel, R. and Freitas, M. A. (2007a). Liquid chromatography mass spectrometry profiling of histones. J Chromatogr B Analyt Technol Biomed Life Sci *850*, 440–454.

[Su et al., 2007b] Su, X., Zhang, L., Lucas, D. M., Davis, M. E., Knapp, A. R., Green-Church, K. B., Marcucci, G., Parthun, M. R., Byrd, J. C. and Freitas, M. A. (2007b). Histone H4 acetylation dynamics determined by stable isotope labeling with amino acids in cell culture and mass spectrometry. Anal Biochem *363*, 22–34.

[Subramanian et al., 2008] Subramanian, K., Jia, D., Kapoor-Vazirani, P., Powell, D. R., Collins, R. E., Sharma, D., Peng, J., Cheng, X. and Vertino, P. M. (2008). Regulation of estrogen receptor alpha by the SET7 lysine methyltransferase. Mol Cell *30*, 336–347.

[Sugasawa et al., 2005] Sugasawa, K., Okuda, Y., Saijo, M., Nishi, R., Matsuda, N., Chu, G., Mori, T., Iwai, S., Tanaka, K., Tanaka, K. and Hanaoka, F. (2005). UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex. Cell *121*, 387–400.

[Sullivan, 2002] Sullivan, B. A. (2002). Centromere round-up at the heterochromatin corral. Trends Biotechnol *20*, 89–92.

[Sullivan and Karpen, 2004] Sullivan, B. A. and Karpen, G. H. (2004). Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. Nat Struct Mol Biol *11*, 1076–1083.

[Sung and Klein, 2006] Sung, P. and Klein, H. (2006). Mechanism of homologous recombination: mediators and helicases take on regulatory functions. Nat Rev Mol Cell Biol *7*, 739–750.

[Sury et al., 2010] Sury, M. D., Chen, J.-X. and Selbach, M. (2010). The SILAC fly allows for accurate protein quantification in vivo. Mol Cell Proteomics *9*, 2173–2183.

[Suto et al., 2000] Suto, R. K., Clarkson, M. J., Tremethick, D. J. and Luger, K. (2000). Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. Nat Struct Biol *7*, 1121–1124.

[Svotelis et al., 2009] Svotelis, A., Gevry, N. and Gaudreau, L. (2009). Regulation of gene expression and cellular proliferation by histone H2A.Z. Biochem Cell Biol *87*, 179–188.

[Syka et al., 2004a] Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J. and Hunt, D. F. (2004a). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A *101*, 9528–9533.

[Syka et al., 2004b] Syka, J. E. P., Marto, J. A., Bai, D. L., Horning, S., Senko, M. W., Schwartz, J. C., Ueberheide, B., Garcia, B., Busby, S., Muratore, T., Shabanowitz, J. and Hunt, D. F. (2004b). Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. J Proteome Res *3*, 621–626.

[Tagami et al., 2004] Tagami, H., Ray-Gallet, D., Almouzni, G. and Nakatani, Y. (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell *116*, 51–61.

[Talbert and Henikoff, 2010] Talbert, P. B. and Henikoff, S. (2010). Histone variants–ancient wrap artists of the epigenome. Nat Rev Mol Cell Biol *11*, 264–275.

[Tamaru et al., 2003] Tamaru, H., Zhang, X., McMillen, D., Singh, P. B., ichi Nakayama, J., Grewal, S. I., Allis, C. D., Cheng, X. and Selker, E. U. (2003). Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in Neurospora crassa. Nat Genet *34*, 75–79.

[Tamburini and Tyler, 2005] Tamburini, B. A. and Tyler, J. K. (2005). Localized histone acetylation and deacetylation triggered by the homologous recombination pathway of double-strand DNA repair. Mol Cell Biol *25*, 4903–4913.

[Tamura et al., 2009] Tamura, T., Smith, M., Kanno, T., Dasenbrock, H., Nishiyama, A. and Ozato, K. (2009). Inducible deposition of the histone variant H3.3 in interferon-stimulated genes. J Biol Chem *284*, 12217–12225.

[Taniguchi et al., 2002] Taniguchi, T., Garcia-Higuera, I., Andreassen, P. R., Gregory, R. C., Grompe, M. and D'Andrea, A. D. (2002). S-phase-specific interaction of the Fanconi anemia protein, FANCD2, with BRCA1 and RAD51. Blood *100*, 2414–2420.

[Taunton et al., 1996] Taunton, J., Hassig, C. A. and Schreiber, S. L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science *272*, 408–411.

[Taverna et al., 2006] Taverna, S. D., Ilin, S., Rogers, R. S., Tanny, J. C., Lavender, H., Li, H., Baker, L., Boyle, J., Blair, L. P., Chait, B. T., Patel, D. J., Aitchison, J. D., Tackett, A. J. and Allis, C. D. (2006). Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs. Mol Cell *24*, 785–796.

[Taverna et al., 2007] Taverna, S. D., Ueberheide, B. M., Liu, Y., Tackett, A. J., Diaz, R. L., Shabanowitz, J., Chait, B. T., Hunt, D. F. and Allis, C. D. (2007). Long-distance combinatorial

linkage between methylation and acetylation on histone H3 N termini. Proc Natl Acad Sci U S A *104*, 2086–2091.

[Thomas et al., 2006] Thomas, C. E., Kelleher, N. L. and Mizzen, C. A. (2006). Mass spectrometric characterization of human histone H3: a bird's eye view. J Proteome Res *5*, 240–247.

[Thompson et al., 2003] Thompson, A., SchÃfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A. K. A. and Hamon, C. (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem *75*, 1895–1904.

[Thompson and Hinz, 2009] Thompson, L. H. and Hinz, J. M. (2009). Cellular and molecular consequences of defective Fanconi anemia proteins in replication-coupled DNA repair: mechanistic insights. Mutat Res *668*, 54–72.

[Thompson and Schild, 2002] Thompson, L. H. and Schild, D. (2002). Recombinational DNA repair and human disease. Mutat Res *509*, 49–78.

[Thoms et al., 2007] Thoms, K.-M., Kuschal, C. and Emmert, S. (2007). Lessons learned from DNA repair defective syndromes. Exp Dermatol *16*, 532–544.

[Tian et al., 2011] Tian, R., Wang, S., Elisma, F., Li, L., Zhou, H., Wang, L. and Figeys, D. (2011). Rare cell proteomic reactor applied to stable isotope labeling by amino acids in cell culture (SILAC)-based quantitative proteomics study of human embryonic stem cell differentiation. Mol Cell Proteomics *10*, M110.000679.

[Tjeertes et al., 2009] Tjeertes, J. V., Miller, K. M. and Jackson, S. P. (2009). Screen for DNA-damage-responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells. EMBO J *28*, 1878–1889.

[Torres-Padilla et al., 2006] Torres-Padilla, M.-E., Bannister, A. J., Hurd, P. J., Kouzarides, T. and Zernicka-Goetz, M. (2006). Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos. Int J Dev Biol *50*, 455–461.

[Tripputi et al., 1986] Tripputi, P., Emanuel, B. S., Croce, C. M., Green, L. G., Stein, G. S. and Stein, J. L. (1986). Human histone genes map to multiple chromosomes. Proc Natl Acad Sci U S A *83*, 3185–3188.

[Trojer and Reinberg, 2007] Trojer, P. and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? Mol Cell *28*, 1–13.

[Truglio et al., 2006] Truglio, J. J., Croteau, D. L., Houten, B. V. and Kisker, C. (2006). Prokaryotic nucleotide excision repair: the UvrABC system. Chem Rev *106*, 233–252.

[Tschiersch et al., 1994] Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G. and Reuter, G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. EMBO J *13*, 3822–3831.

[Turner, 2000] Turner, B. M. (2000). Histone acetylation and an epigenetic code. Bioessays *22*, 836–845.

[Turner, 2002] Turner, B. M. (2002). Cellular memory and the histone code. Cell *111*, 285–291.

[Turner et al., 1992] Turner, B. M., Birley, A. J. and Lavender, J. (1992). Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in Drosophila polytene nuclei. Cell *69*, 375–384.

[Unal et al., 2004] Unal, E., Arbel-Eden, A., Sattler, U., Shroff, R., Lichten, M., Haber, J. E. and Koshland, D. (2004). DNA damage response pathway uses histone modification to assemble a double-strand break-specific cohesin domain. Mol Cell *16*, 991–1002.

[Urban and Zweidler, 1983] Urban, M. K. and Zweidler, A. (1983). Changes in nucleosomal core histone variants during chicken development and maturation. Dev Biol *95*, 421–428.

[Utley et al., 2005] Utley, R. T., Lacoste, N., Jobin-Robitaille, O., Allard, S. and Cote, J. (2005). Regulation of NuA4 histone acetyltransferase activity in transcription and DNA repair by phosphorylation of histone H4. Mol Cell Biol *25*, 8179–8190.

[van Attikum and Gasser, 2005] van Attikum, H. and Gasser, S. M. (2005). ATP-dependent chromatin remodeling and DNA double-strand break repair. Cell Cycle *4*, 1011–1014.

[van Attikum and Gasser, 2009] van Attikum, H. and Gasser, S. M. (2009). Crosstalk between histone modifications during the DNA damage response. Trends Cell Biol *19*, 207–217.

[Van Hooser et al., 2001] Van Hooser, A. A., Ouspenski, I. I., Gregson, H. C., Starr, D. A., Yen, T. J., Goldberg, M. L., Yokomori, K., Earnshaw, W. C., Sullivan, K. F. and Brinkley, B. R. (2001). Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. J Cell Sci *114*, 3529–3542.

[van Leeuwen et al., 2002] van Leeuwen, F., Gafken, P. R. and Gottschling, D. E. (2002). Dot1p modulates silencing in yeast by methylation of the nucleosome core. Cell *109*, 745–756.

[van Steensel and Henikoff, 2000] van Steensel, B. and Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. Nat Biotechnol *18*, 424–428.

[Vasen et al., 2007] Vasen, H. F. A., MÃ¶slein, G., Alonso, A., Bernstein, I., Bertario, L., Blanco, I., Burn, J., Capella, G., Engel, C., Frayling, I., Friedl, W., Hes, F. J., Hodgson, S., Mecklin, J.-P., MÃ¸ller, P., Nagengast, F., Parc, Y., Renkonen-Sinisalo, L., Sampson, J. R., Stormorken, A. and Wijnen, J. (2007). Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer). J Med Genet *44*, 353–362.

[Vaz et al., 2010] Vaz, F., Hanenberg, H., Schuster, B., Barker, K., Wiek, C., Erven, V., Neveling, K., Endt, D., Kesterton, I., Autore, F., Fraternali, F., Freund, M., Hartmann, L., Grimwade, D., Roberts, R. G., Schaal, H., Mohammed, S., Rahman, N., Schindler, D. and Mathew, C. G. (2010). Mutation of the RAD51C gene in a Fanconi anemia-like disorder. Nat Genet *42*, 406–409.

[Vermeulen et al., 2010] Vermeulen, M., Eberl, H. C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K. K., Olsen, J. V., Hyman, A. A., Stunnenberg, H. G. and Mann, M. (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell *142*, 967–980.

[Vermeulen and Selbach, 2009] Vermeulen, M. and Selbach, M. (2009). Quantitative proteomics: a tool to assess cell differentiation. Curr Opin Cell Biol *21*, 761–766.

[Vidali et al., 1968] Vidali, G., Gershey, E. L. and Allfrey, V. G. (1968). Chemical studies of histone acetylation. The distribution of epsilon-N-acetyllysine in calf thymus histones. J Biol Chem *243*, 6361–6366.

[Viens et al., 2006] Viens, A., Mechold, U., Brouillard, F., Gilbert, C., Leclerc, P. and Ogryzko, V. (2006). Analysis of human histone H2AZ deposition in vivo argues against its direct role in epigenetic templating mechanisms. Mol Cell Biol *26*, 5325–5335.

[Viens et al., 2004] Viens, A., Mechold, U., Lehrmann, H., Harel-Bellan, A. and Ogryzko, V. (2004). Use of protein biotinylation in vivo for chromatin immunoprecipitation. Anal Biochem *325*, 68–76.

[Vire et al., 2006] Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Eynde, A. V., Bernard, D., Vanderwinden, J.-M., Bollen, M., Esteller, M., Croce, L. D., de Launoit, Y. and Fuks, F. (2006). The Polycomb group protein EZH2 directly controls DNA methylation. Nature *439*, 871–874.

[Waddington, 1942] Waddington, C. (1942). The epigenotype. Endeavour *1*, 18–20.

[Waddington, 1957] Waddington, C. H. (1957). The Strategy of the Genes. Geo Allen & Unwin, London.

[Walsh, 2005] Walsh, C. (2005). Posttranslational Modifications of Proteins. Roberts and Company Publishers.

[Walther and Mann, 2010] Walther, T. C. and Mann, M. (2010). Mass spectrometry-based proteomics in cell biology. J Cell Biol *190*, 491–500.

[Wang and Elledge, 2007] Wang, B. and Elledge, S. J. (2007). Ubc13/Rnf8 ubiquitin ligases control foci formation of the Rap80/Abraxas/Brca1/Brcc36 complex in response to DNA damage. Proc Natl Acad Sci U S A *104*, 20759–20763.

[Wang et al., 2011] Wang, D., Hincapie, M., Rejtar, T. and Karger, B. L. (2011). Ultrasensitive characterization of site-specific glycosylation of affinity-purified haptoglobin

from lung cancer patient plasma using 10 $\hat{\mathrm{I}}\frac{1}{4}$m i.d. porous layer open tubular liquid chromatography-linear ion trap collision-induced dissociation/electron transfer dissociation mass spectrometry. Anal Chem *83*, 2029–2037.

[Wang et al., 2007] Wang, G. G., Allis, C. D. and Chi, P. (2007). Chromatin remodeling and cancer, Part II: ATP-dependent chromatin remodeling. Trends Mol Med *13*, 373–380.

[Wang et al., 2006] Wang, H., Zhai, L., Xu, J., Joo, H.-Y., Jackson, S., Erdjument-Bromage, H., Tempst, P., Xiong, Y. and Zhang, Y. (2006). Histone H3 and H4 ubiquitylation by the CUL4-DDB-ROC1 ubiquitin ligase facilitates cellular response to DNA damage. Mol Cell *22*, 383–394.

[Wang et al., 2010] Wang, Q., Zhang, Y., Yang, C., Xiong, H., Lin, Y., Yao, J., Li, H., Xie, L., Zhao, W., Yao, Y., Ning, Z.-B., Zeng, R., Xiong, Y., Guan, K.-L., Zhao, S. and Zhao, G.-P. (2010). Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux. Science *327*, 1004–1007.

[Wang, 2008] Wang, W. (2008). A major switch for the Fanconi anemia DNA damage-response pathway. Nat Struct Mol Biol *15*, 1128–1130.

[Wang et al., 2004a] Wang, X., Andreassen, P. R. and D'Andrea, A. D. (2004a). Functional interaction of monoubiquitinated FANCD2 and BRCA2/FANCD1 in chromatin. Mol Cell Biol *24*, 5850–5862.

[Wang et al., 2004b] Wang, Y., Wysocka, J., Sayegh, J., Lee, Y.-H., Perlin, J. R., Leonelli, L., Sonbuchner, L. S., McDonald, C. H., Cook, R. G., Dou, Y., Roeder, R. G., Clarke, S., Stallcup, M. R., Allis, C. D. and Coonrod, S. A. (2004b). Human PAD4 regulates histone arginine methylation levels via demethylimination. Science *306*, 279–283.

[Wang et al., 2009] Wang, Z., Zang, C., Cui, K., Schones, D. E., Barski, A., Peng, W. and Zhao, K. (2009). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. Cell *138*, 1019–1031.

[Ward and Chen, 2001] Ward, I. M. and Chen, J. (2001). Histone H2AX is phosphorylated in an ATR-dependent manner in response to replicational stress. J Biol Chem *276*, 47759–47762.

[Waterborg, 1993] Waterborg, J. H. (1993). Histone synthesis and turnover in alfalfa. Fast loss of highly acetylated replacement histone variant H3.2. J Biol Chem *268*, 4912–4917.

[Waterborg and Matthews, 1983] Waterborg, J. H. and Matthews, H. R. (1983). Patterns of histone acetylation in the cell cycle of Physarum polycephalum. Biochemistry *22*, 1489–1496.

[Watson, 2003] Watson, J. D. (2003). Celebrating the genetic jubilee: a conversation with James D. Watson. Interviewed by John Rennie. Sci Am *288*, 66–69.

[Wei et al., 1998] Wei, Y., Mizzen, C. A., Cook, R. G., Gorovsky, M. A. and Allis, C. D. (1998). Phosphorylation of histone H3 at serine 10 is correlated with chromosome condensation during mitosis and meiosis in Tetrahymena. Proc Natl Acad Sci U S A *95*, 7480–7484.

[Whetstine et al., 2006] Whetstine, J. R., Nottke, A., Lan, F., Huarte, M., Smolikov, S., Chen, Z., Spooner, E., Li, E., Zhang, G., Colaiacovo, M. and Shi, Y. (2006). Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. Cell *125*, 467–481.

[Widom, 1992] Widom, J. (1992). A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells. Proc Natl Acad Sci U S A *89*, 1095–1099.

[Williams et al., 1986] Williams, S. P., Athey, B. D., Muglia, L. J., Schappe, R. S., Gough, A. H. and Langmore, J. P. (1986). Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length. Biophys J *49*, 233–248.

[Wilson and Bohr, 2007] Wilson, D. M. and Bohr, V. A. (2007). The mechanics of base excision repair, and its relationship to aging and disease. DNA Repair (Amst) *6*, 544–559.

[Wolf, 2009] Wolf, S. S. (2009). The protein arginine methyltransferase family: an update about function, new perspectives and the physiological role in humans. Cell Mol Life Sci *66*, 2109–2121.

[Wolffe, 1992] Wolffe, A. P. (1992). New insights into chromatin function in transcriptional control. FASEB J *6*, 3354–3361.

[Wong et al., 2010] Wong, L. H., McGhie, J. D., Sim, M., Anderson, M. A., Ahn, S., Hannan, R. D., George, A. J., Morgan, K. A., Mann, J. R. and Choo, K. H. A. (2010). ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. Genome Res *20*, 351–360.

[Woodcock et al., 1993] Woodcock, C. L., Grigoryev, S. A., Horowitz, R. A. and Whitaker, N. (1993). A chromatin folding model that incorporates linker variability generates fibers resembling the native structures. Proc Natl Acad Sci U S A *90*, 9021–9025.

[Wu et al., 2009] Wu, J., Huen, M. S. Y., Lu, L.-Y., Ye, L., Dou, Y., Ljungman, M., Chen, J. and Yu, X. (2009). Histone ubiquitination associates with BRCA1-dependent DNA damage response. Mol Cell Biol *29*, 849–860.

[Wysocki et al., 2005] Wysocki, R., Javaheri, A., Allard, S., Sha, F., Cote, J. and Kron, S. J. (2005). Role of Dot1-dependent histone H3 methylation in G1 and S phase DNA damage checkpoint functions of Rad9. Mol Cell Biol *25*, 8430–8443.

[Xhemalce and Kouzarides, 2010] Xhemalce, B. and Kouzarides, T. (2010). A chromodomain switch mediated by histone H3 Lys 4 acetylation regulates heterochromatin assembly. Genes Dev *24*, 647–652.

[Xiao et al., 2009] Xiao, A., Li, H., Shechter, D., Ahn, S. H., Fabrizio, L. A., Erdjument-Bromage, H., Ishibe-Murakami, S., Wang, B., Tempst, P., Hofmann, K., Patel, D. J., Elledge, S. J. and Allis, C. D. (2009). WSTF regulates the H2A.X DNA damage response via a novel tyrosine kinase activity. Nature *457*, 57–62.

[Xiao et al., 2003] Xiao, B., Jing, C., Wilson, J. R., Walker, P. A., Vasisht, N., Kelly, G., Howell, S., Taylor, I. A., Blackburn, G. M. and Gamblin, S. J. (2003). Structure and catalytic mechanism of the human histone methyltransferase SET7/9. Nature *421*, 652–656.

[Yan et al., 2009] Yan, Q., Dutt, S., Xu, R., Graves, K., Juszczynski, P., Manis, J. P. and Shipp, M. A. (2009). BBAP monoubiquitylates histone H4 at lysine 91 and selectively modulates the DNA damage response. Mol Cell *36*, 110–120.

[Yang et al., 1996a] Yang, W. M., Inouye, C., Zeng, Y., Bearss, D. and Seto, E. (1996a). Transcriptional repression by YY1 is mediated by interaction with a mammalian homolog of the yeast global regulator RPD3. Proc Natl Acad Sci U S A *93*, 12845–12850.

[Yang et al., 1996b] Yang, X. J., Ogryzko, V. V., Nishikawa, J., Howard, B. H. and Nakatani, Y. (1996b). A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. Nature *382*, 319–324.

[Yang and Seto, 2007] Yang, X.-J. and Seto, E. (2007). HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. Oncogene *26*, 5310–5318.

[Yang and Seto, 2008] Yang, X.-J. and Seto, E. (2008). The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. Nat Rev Mol Cell Biol *9*, 206–218.

[Yang and Speed, 2002] Yang, Y. H. and Speed, T. (2002). Design issues for cDNA microarray experiments. Nat Rev Genet *3*, 579–588.

[Yao et al., 2001] Yao, X., Freas, A., Ramirez, J., Demirev, P. A. and Fenselau, C. (2001). Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. Anal Chem *73*, 2836–2842.

[Ye et al., 2005] Ye, J., Ai, X., Eugeni, E. E., Zhang, L., Carpenter, L. R., Jelinek, M. A., Freitas, M. A. and Parthun, M. R. (2005). Histone H4 lysine 91 acetylation a core domain modification associated with chromatin assembly. Mol Cell *18*, 123–130.

[Ye et al., 2009] Ye, X., Luke, B., Andresson, T. and Blonder, J. (2009). 18O stable isotope labeling in MS-based proteomics. Brief Funct Genomic Proteomic *8*, 136–144.

[Young et al., 2009] Young, N. L., DiMaggio, P. A., Plazas-Mayorca, M. D., Baliban, R. C., Floudas, C. A. and Garcia, B. A. (2009). High throughput characterization of combinatorial histone codes. Mol Cell Proteomics *8*, 2266–2284.

[Zabrouskov and Whitelegge, 2007] Zabrouskov, V. and Whitelegge, J. P. (2007). Increased coverage in the transmembrane domain with activated-ion electron capture dissociation for top-down Fourier-transform mass spectrometry of integral membrane proteins. J Proteome Res *6*, 2205–2210.

[Zanier et al., 2004] Zanier, R., Briot, D., du Villard, J.-A. D., Sarasin, A. and Rosselli, F. (2004). Fanconi anemia C gene product regulates expression of genes involved in differentiation and inflammation. Oncogene *23*, 5004–5013.

[Zegerman et al., 2002] Zegerman, P., Canas, B., Pappin, D. and Kouzarides, T. (2002). Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex. J Biol Chem *277*, 11621–11624.

[Zeng et al., 2010] Zeng, L., Zhang, Q., Li, S., Plotnikov, A. N., Walsh, M. J. and Zhou, M.-M. (2010). Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. Nature *466*, 258–262.

[Zhang et al., 2009] Zhang, K., Chen, Y., Zhang, Z. and Zhao, Y. (2009). Identification and verification of lysine propionylation and butyrylation in yeast core histones using PTMap software. J Proteome Res *8*, 900–906.

[Zhang et al., 2004] Zhang, K., Siino, J. S., Jones, P. R., Yau, P. M. and Bradbury, E. M. (2004). A mass spectrometric "Western blot" to evaluate the correlations between histone methylation and histone acetylation. Proteomics *4*, 3765–3775.

[Zhang and Tang, 2003] Zhang, K. and Tang, H. (2003). Analysis of core histones by liquid chromatography-mass spectrometry and peptide mapping. J Chromatogr B Analyt Technol Biomed Life Sci *783*, 173–179.

[Zhang et al., 2002a] Zhang, K., Tang, H., Huang, L., Blankenship, J. W., Jones, P. R., Xiang, F., Yau, P. M. and Burlingame, A. L. (2002a). Identification of acetylation and methylation sites of histone H3 from chicken erythrocytes by high-accuracy matrix-assisted laser desorption ionization-time-of-flight, matrix-assisted laser desorption ionization-postsource decay, and nanoelectrospray ionization tandem mass spectrometry. Anal Biochem *306*, 259–269.

[Zhang et al., 2002b] Zhang, K., Williams, K. E., Huang, L., Yau, P., Siino, J. S., Bradbury, E. M., Jones, P. R., Minch, M. J. and Burlingame, A. L. (2002b). Histone acetylation and deacetylation: identification of acetylation and methylation sites of HeLa histone H4 by mass spectrometry. Mol Cell Proteomics *1*, 500–508.

[Zhang and Freitas, 2004] Zhang, L. and Freitas, M, A. (2004). Comparison of peptide mass mapping and electron capture dissociation as assays for histone posttranslational modifications. Int J Mass Spectr *234(1â3)*, 213â225.

[Zhang et al., 2007] Zhang, L., Su, X., Liu, S., Knapp, A. R., Parthun, M. R., Marcucci, G. and Freitas, M. A. (2007). Histone H4 N-terminal acetylation in Kasumi-1 cells treated with

depsipeptide determined by acetic acid-urea polyacrylamide gel electrophoresis, amino acid coded mass tagging, and mass spectrometry. J Proteome Res *6*, 81–88.

[Zhang et al., 2003] Zhang, X., Yang, Z., Khan, S. I., Horton, J. R., Tamaru, H., Selker, E. U. and Cheng, X. (2003). Structural basis for the product specificity of histone lysine methyltransferases. Mol Cell *12*, 177–185.

[Zhao et al., 2007] Zhao, G. Y., Sonoda, E., Barber, L. J., Oka, H., Murakawa, Y., Yamada, K., Ikura, T., Wang, X., Kobayashi, M., Yamamoto, K., Boulton, S. J. and Takeda, S. (2007). A critical role for the ubiquitin-conjugating enzyme Ubc13 in initiating homologous recombination. Mol Cell *25*, 663–675.

[Zhao et al., 2010] Zhao, S., Xu, W., Jiang, W., Yu, W., Lin, Y., Zhang, T., Yao, J., Zhou, L., Zeng, Y., Li, H., Li, Y., Shi, J., An, W., Hancock, S. M., He, F., Qin, L., Chin, J., Yang, P., Chen, X., Lei, Q., Xiong, Y. and Guan, K.-L. (2010). Regulation of cellular metabolism by protein lysine acetylation. Science *327*, 1000–1004.

[Zheng et al., 2006] Zheng, H., Wang, X., Legerski, R. J., Glazer, P. M. and Li, L. (2006). Repair of DNA interstrand cross-links: interactions between homology-dependent and homology-independent pathways. DNA Repair (Amst) *5*, 566–574.

[Zhou et al., 2010] Zhou, B., Xu, W., Herndon, D., Tompkins, R., Davis, R., Xiao, W., Wong, W. H., Inflammation, to Injury Program, H. R., Toner, M., Warren, H. S., Schoenfeld, D. A., Rahme, L., McDonald-Smith, G. P., Hayden, D., Mason, P., Fagan, S., Yu, Y.-M., Cobb, J. P., Remick, D. G., Mannick, J. A., Lederer, J. A., Gamelli, R. L., Silver, G. M., West, M. A., Shapiro, M. B., Smith, R., Camp, D. G., Qian, W., Storey, J., Mindrinos, M., Tibshirani, R., Lowry, S., Calvano, S., Chaudry, I., West, M. A., Cohen, M., Moore, E. E., Johnson, J., Moldawer, L. L., Baker, H. V., Efron, P. A., Balis, U. G. J., Billiar, T. R., Ochoa, J. B., Sperry, J. L., Miller-Graziano, C. L., De, A. K., Bankey, P. E., Finnerty, C. C., Jeschke, M. G., Minei, J. P., Arnoldo, B. D., Hunt, J. L., Horton, J., Cobb, J. P., Brownstein, B., Freeman, B., Maier, R. V., Nathens, A. B., Cuschieri, J., Gibran, N., Klein, M. and O'Keefe, G. (2010). Analysis of factorial time-course microarrays with application to a clinical study of burn injury. Proc Natl Acad Sci U S A *107*, 9923–9928.

[Zhou et al., 2007] Zhou, J., Fan, J. Y., Rangasamy, D. and Tremethick, D. J. (2007). The nucleosome surface regulates chromatin compaction and couples it with transcriptional repression. Nat Struct Mol Biol *14*, 1070–1076.

[Zhu and Wani, 2010] Zhu, Q. and Wani, A. A. (2010). Histone modifications: crucial elements for damage response and chromatin restoration. J Cell Physiol *223*, 283–288.

[Zhu et al., 2009] Zhu, Q., Wani, G., Arab, H. H., El-Mahdy, M. A., Ray, A. and Wani, A. A. (2009). Chromatin restoration following nucleotide excision repair involves the incorporation of ubiquitinated H2A at damaged genomic sites. DNA Repair (Amst) *8*, 262–273.

[Zippo et al., 2009] Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., Krepelova, A. and Oliviero, S. (2009). Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. Cell *138*, 1122–1136.

[Zlatanova et al., 2009] Zlatanova, J., Bishop, T. C., Victor, J.-M., Jackson, V. and van Holde, K. (2009). The nucleosome family: dynamic and growing. Structure *17*, 160–171.

[Zubarev, 2006] Zubarev, R. (2006). Protein primary structure using orthogonal fragmentation techniques in Fourier transform mass spectrometry. Expert Rev Proteomics *3*, 251–261.

[Zubarev et al., 2000] Zubarev, R. A., Horn, D. M., Fridriksson, E. K., Kelleher, N. L., Kruger, N. A., Lewis, M. A., Carpenter, B. K. and McLafferty, F. W. (2000). Electron capture dissociation for structural characterization of multiply charged protein cations. Anal Chem *72*, 563–573.

7

Annex - I

The representative MS/MS spectra of various modified and unmodified histone peptides along with the lists of theoretical masses of fragment peptides, corresponding to various histone H3 and H4 modifications described in [Chapter 2], are presented here.

Histone H3 residues 9-17 (KSTGGKAPR)

Heavy →	191	278	379	436	493	669	740	837	-
Light →	185	272	373	430	487	657	728	825	-
	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9
	K-Pr	**S**	**T**	**G**	**G**	**K-Ac**	**A**	**P**	**R**
	y_9	y_8	y_7	y_6	y_5	y_4	y_3	y_2	y_1
Light →	-	815	728	627	570	513	343	272	175
Heavy →	-	831	744	643	586	529	353	282	185

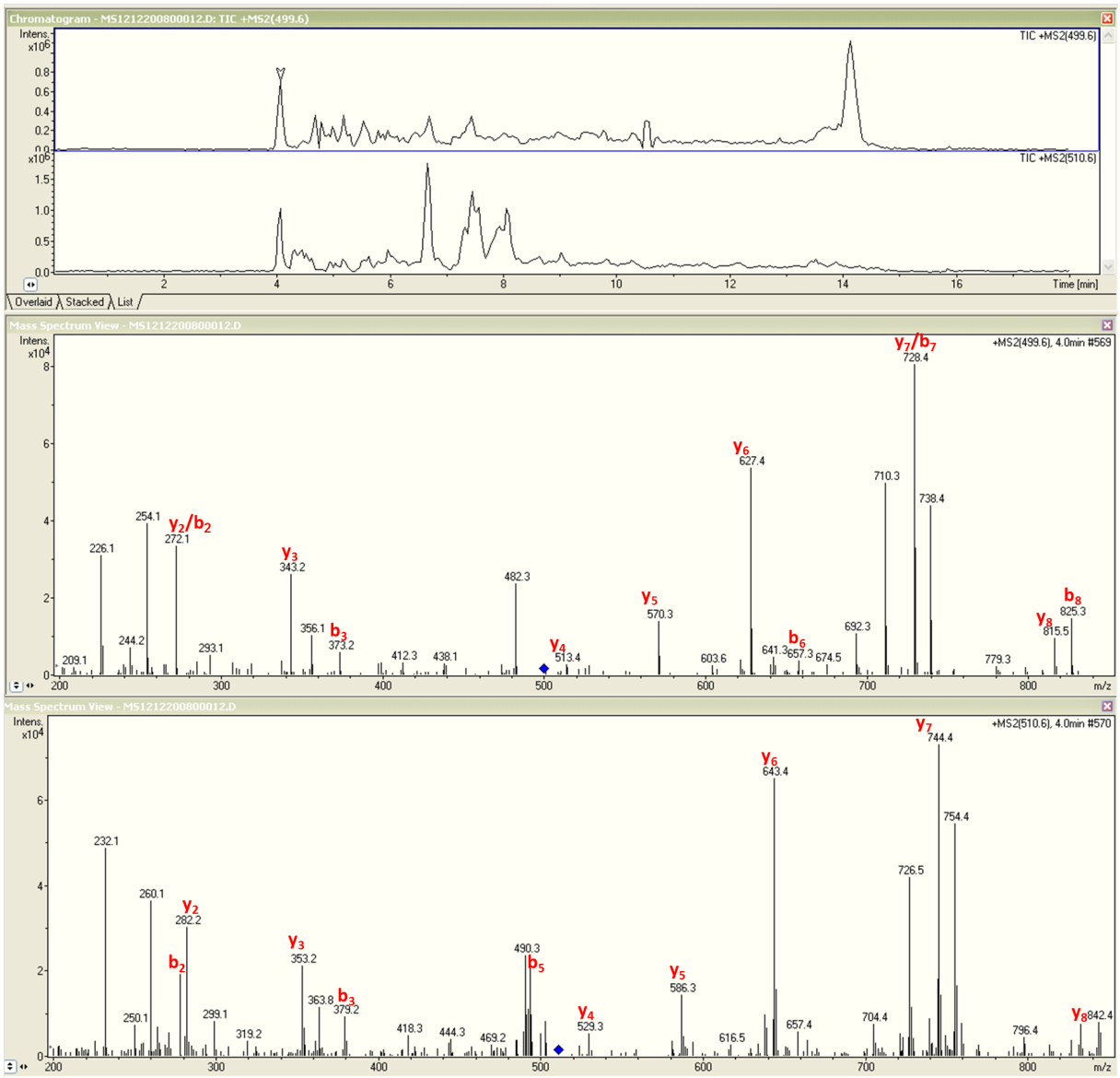


Figure 7.1: MS2 spectrum of Histone H3 residues 9-17 (K14ac)

Histone H3 residues 9-17 (KSTGGKAPR)

Heavy →	191	278	379	436	493	683	754	852	-
Light →	185	272	373	430	487	671	742	840	-
	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9
	K-Pr	**S**	**T**	**G**	**G**	**K-Pr**	**A**	**P**	**R**
	y_9	y_8	y_7	y_6	y_5	y_4	y_3	y_2	y_1
Light →	-	829	742	641	584	527	343	272	175
Heavy →	-	845	758	657	600	543	353	282	185

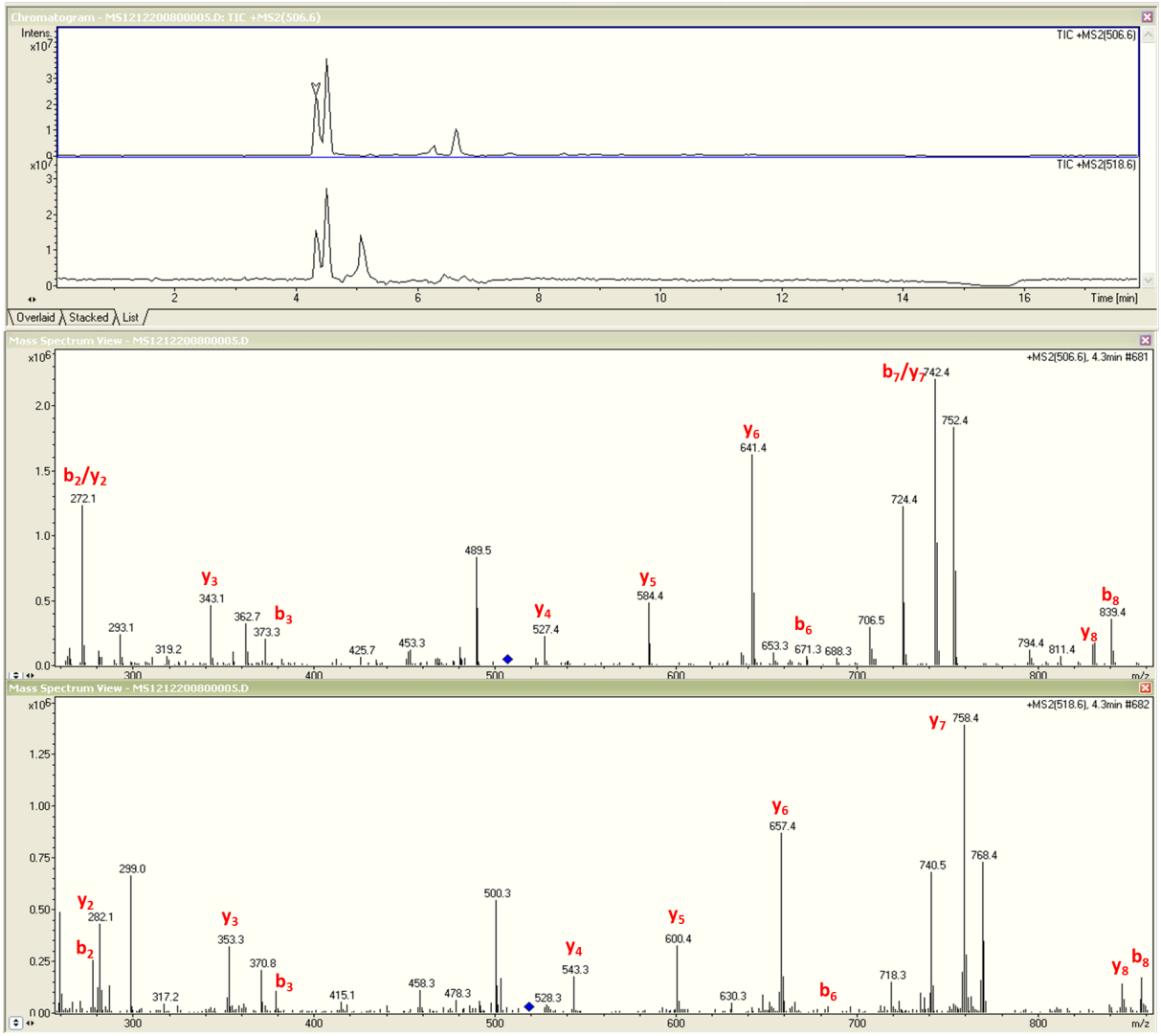


Figure 7.2: MS2 spectrum of Histone H3 residues 9-17 (K9UM + 14UM)

Histone H3 residues 9-17 (KSTGGKAPR)

Heavy ⟶	205	292	393	450	507	697	768	866	-
Light ⟶	199	286	387	444	501	685	756	854	-
	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9
	K-Me.Pr	**S**	**T**	**G**	**G**	**K-Pr**	**A**	**P**	**R**
	y_9	y_8	y_7	y_6	y_5	y_4	y_3	y_2	y_1
Light ⟶	-	829	742	641	584	527	343	272	175
Heavy ⟶	-	845	758	657	600	543	353	282	185

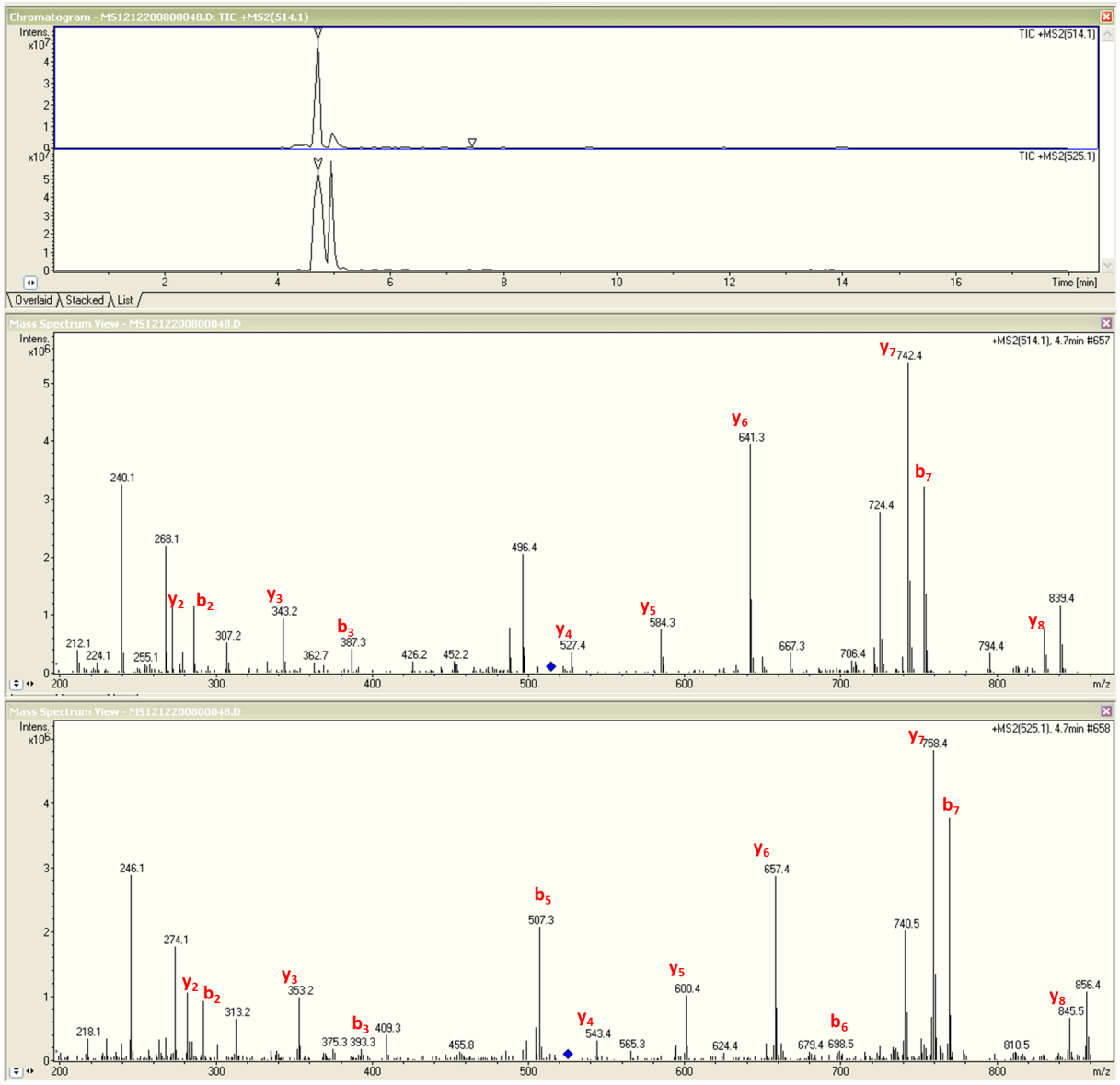


Figure 7.3: MS2 spectrum of Histone H3 residues 9-17 (K9me + K14UM)

Histone H3 residues 54-63 (YQKSTELLIR)

Heavy ⟶	164	292	482	569	670	799	913	1026	1139	-
Light ⟶	164	292	476	563	664	793	907	1020	1133	-
	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
	Y	**Q**	**K-Pr**	**S**	**T**	**E**	**L**	**L**	**I**	**R**
	y_{10}	y_9	y_8	y_7	y_6	y_5	y_4	y_3	y_2	y_1
Light ⟶	-	1144	1016	832	744	643	514	401	288	175
Heavy ⟶	-	1160	1032	842	754	653	524	411	298	185

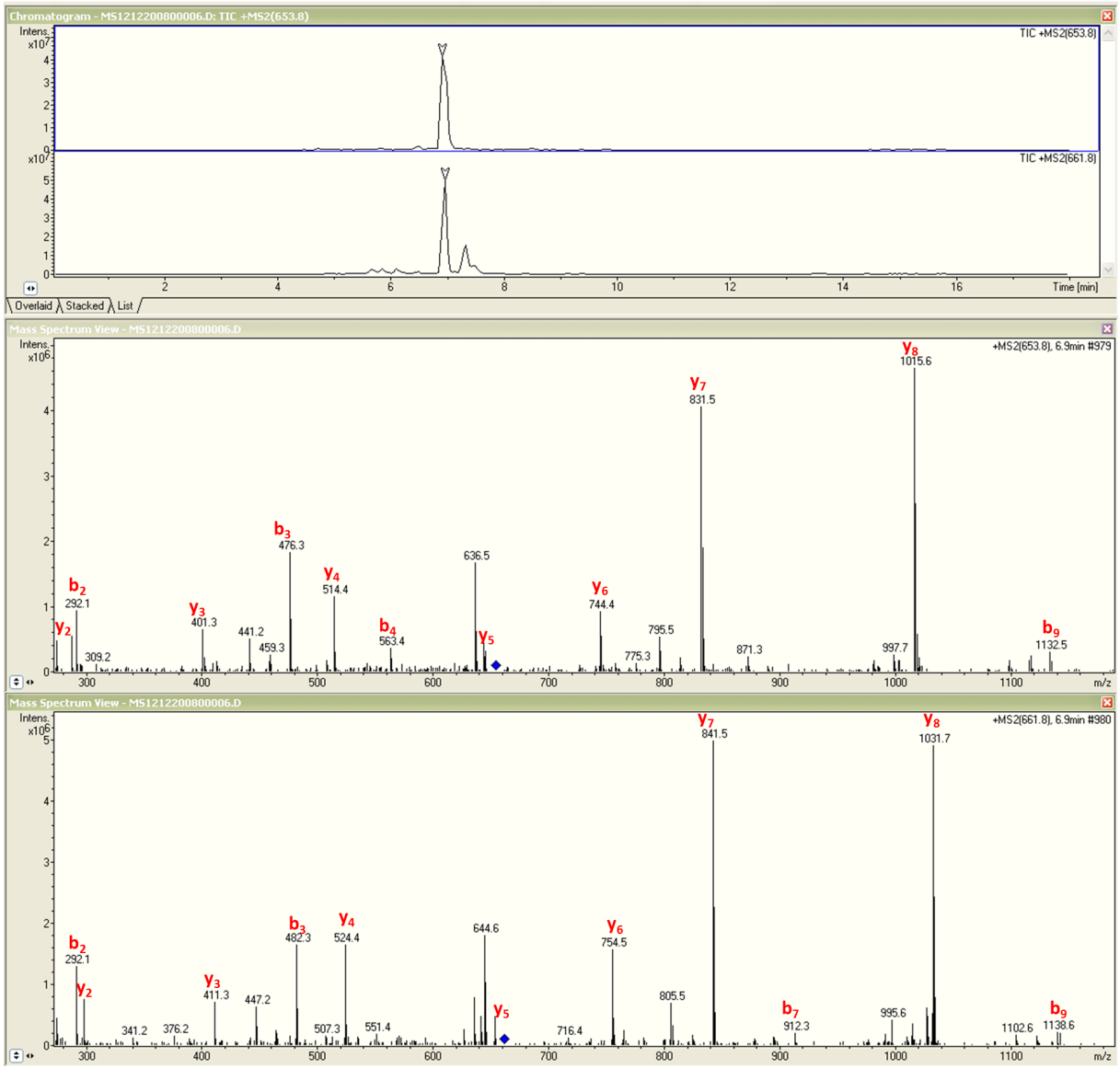


Figure 7.4: MS2 spectrum of Histone H3 residues 54-63 (K56UM)

Histone H3 residues 54-63 (YQKSTELLIR)

Heavy ⟶	164	292	496	583	684	813	927	1040	1153	-
Light ⟶	164	292	490	577	678	807	921	1034	1147	-
	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
	Y	**Q**	**K-Me.Pr**	**S**	**T**	**E**	**L**	**L**	**I**	**R**
	y_{10}	y_9	y_8	y_7	y_6	y_5	y_4	y_3	y_2	y_1
Light ⟶	-	1158	1030	832	744	643	514	401	288	175
Heavy ⟶	-	1174	1046	842	754	653	524	411	298	185

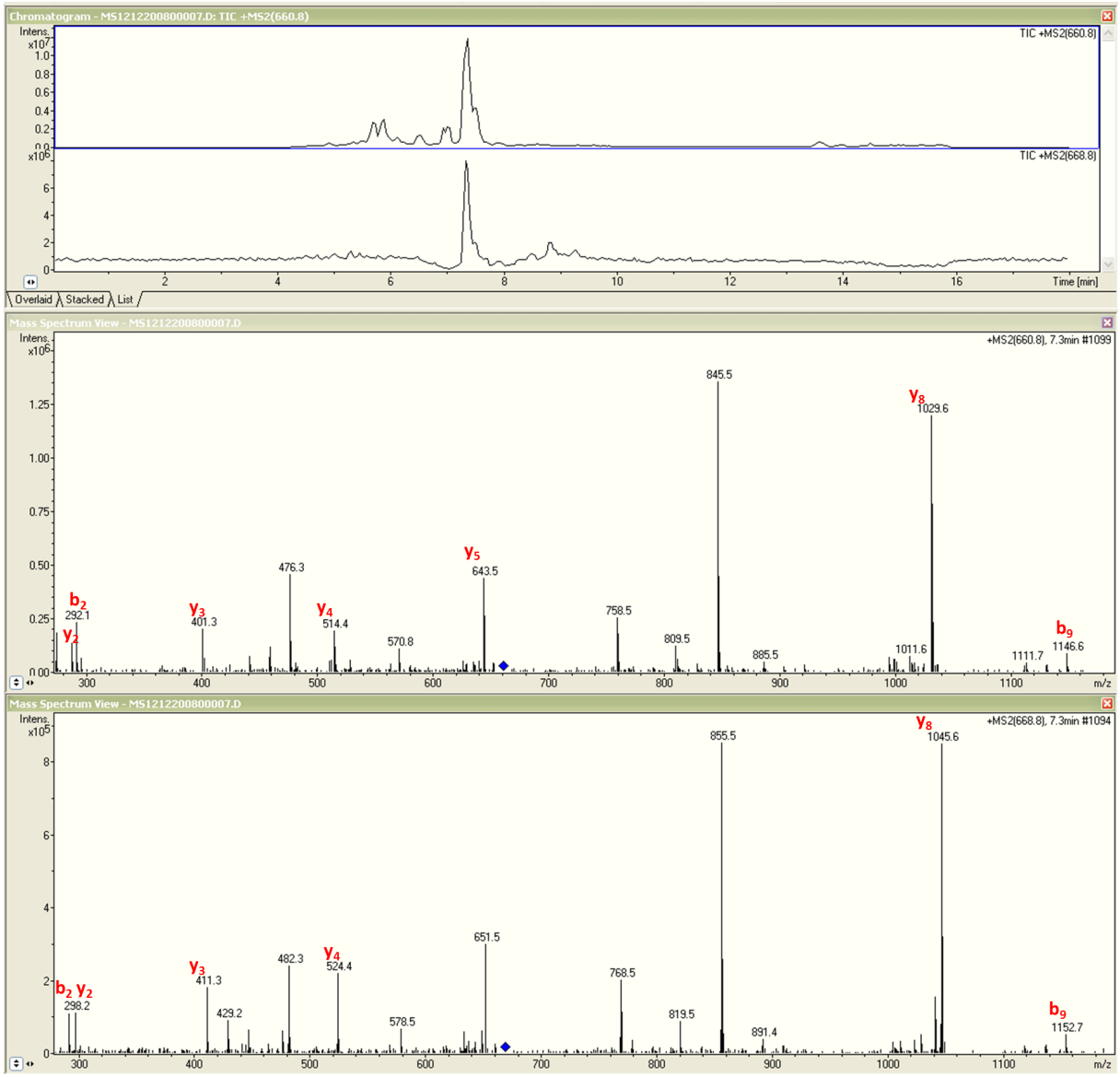


Figure 7.5: MS2 spectrum of Histone H3 residues 54-63 (K56me)

Histone H3 residues 73-83 (EIAQDFKTDLR)

Heavy →	130	243	314	442	557	704	895	996	1111	1224	-
Light →	130	243	314	442	557	704	889	990	1105	1218	-
	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}
	E	**I**	**A**	**Q**	**D**	**F**	**K-Pr**	**T**	**D**	**L**	**R**
	y_{11}	y_{10}	y_9	y_8	y_7	y_6	y_5	y_4	y_3	y_2	y_1
Light →	-	1263	1150	1079	951	836	688	504	403	288	175
Heavy →	-	1279	1166	1095	967	852	704	514	413	298	185

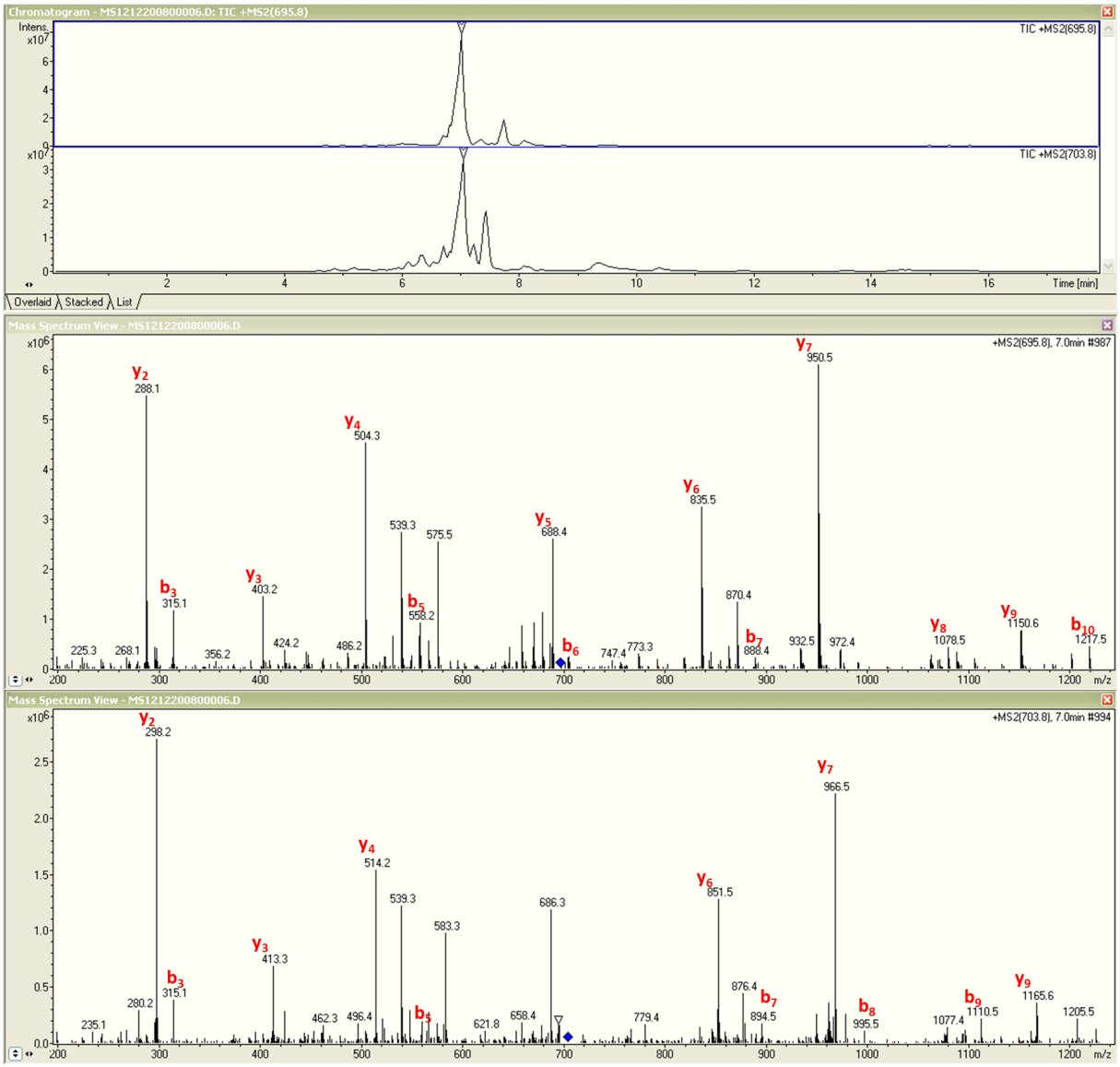


Figure 7.6: MS2 spectrum of Histone H3 residues 73-83 (K79UM)

Histone H3 residues 73-83 (EIAQDFKTDLR)

Heavy ⟶	130	243	314	442	557	704	909	1010	1125	1238	-
Light ⟶	130	243	314	442	557	704	903	1004	1119	1232	-
	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}
	E	**I**	**A**	**Q**	**D**	**F**	**K-Me.Pr**	**T**	**D**	**L**	**R**
	y_{11}	y_{10}	y_9	y_8	y_7	y_6	y_5	y_4	y_3	y_2	y_1
Light ⟶	-	1277	1164	1093	965	850	702	504	403	288	175
Heavy ⟶	-	1293	1180	1109	981	866	418	514	413	298	185

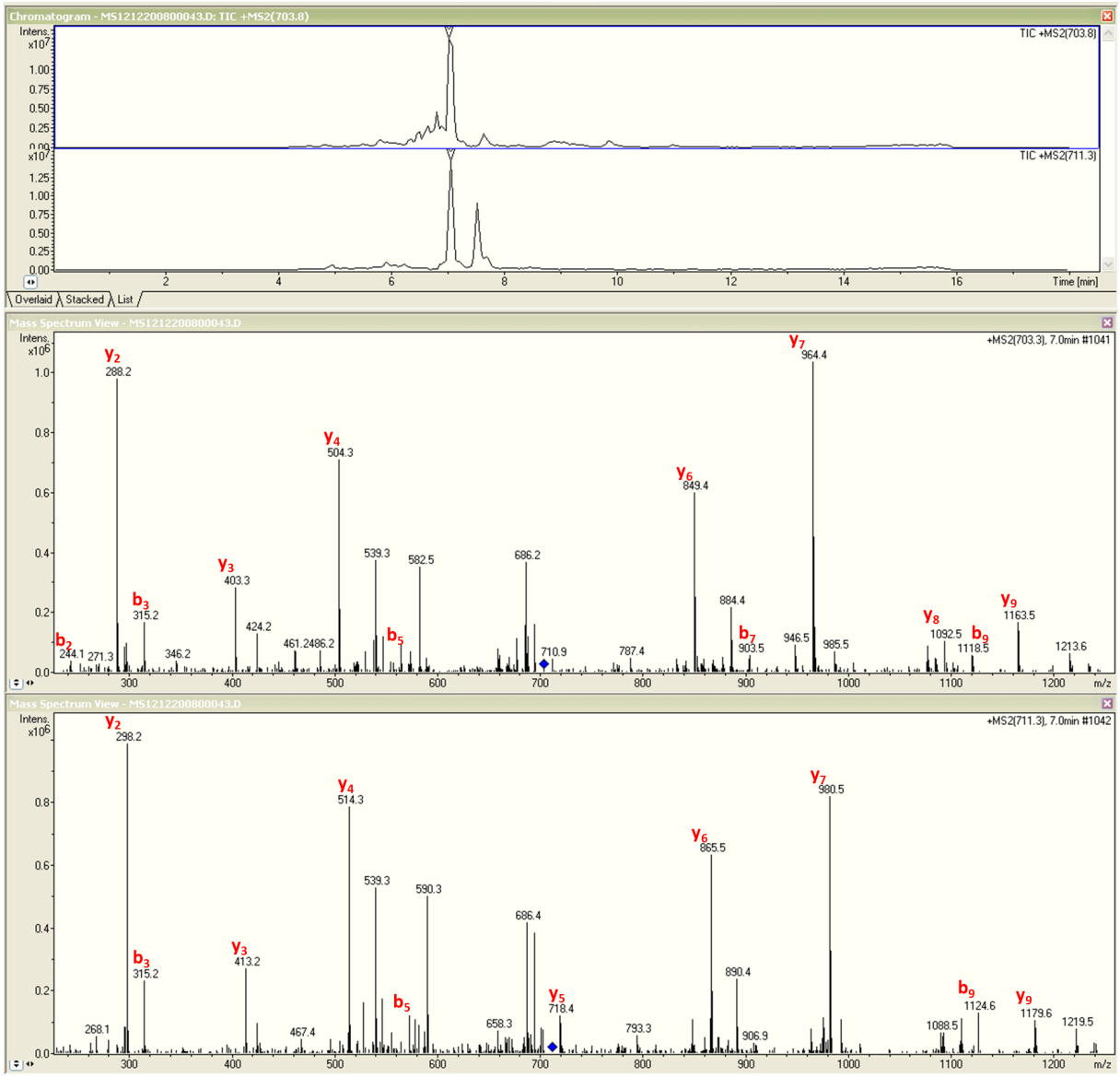


Figure 7.7: MS2 spectrum of Histone H3 residues 73-83 (K79me)

Histone H4 residues 68-78 (DAVTYTEHAKR)

Heavy ⟶	116	187	286	387	550	651	780	918	989	1179	-
Light ⟶	116	187	286	387	550	651	780	918	989	1173	-
	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}
	D	**A**	**V**	**T**	**Y**	**T**	**E**	**H**	**A**	**K-Pr**	**R**
	y_{11}	y_{10}	y_9	y_8	y_7	y_6	y_5	y_4	y_3	y_2	y_1
Light ⟶	-	1232	1161	1062	961	797	696	567	430	359	175
Heavy ⟶	-	1248	1177	1078	977	813	712	583	446	375	185

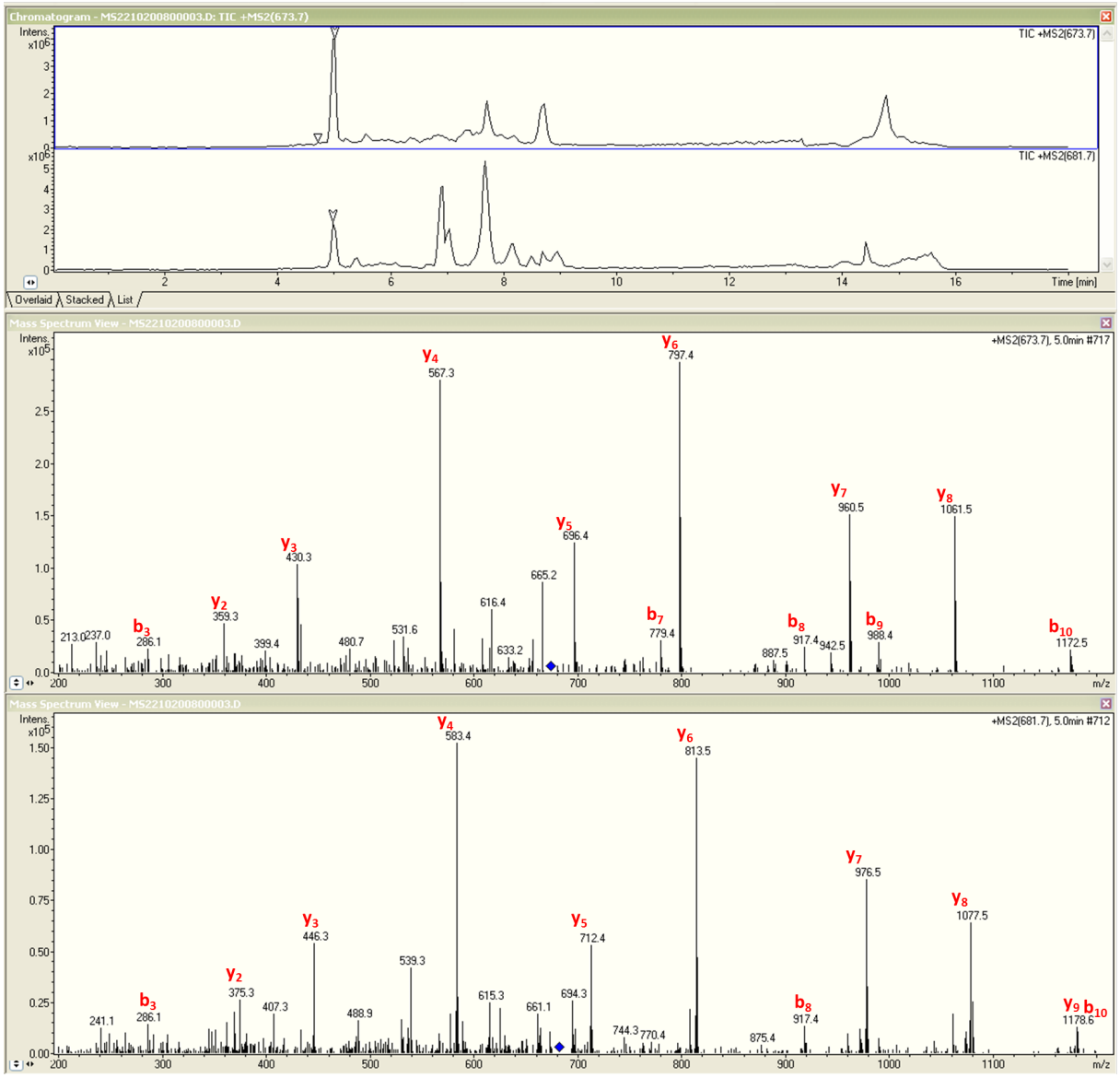


Figure 7.8: MS2 spectrum of Histone H4 residues 68-78 (K77UM)

Histone H4 residues 68-78 (DAVTYTEHAKR)

Heavy	⟶	116	187	286	387	550	651	780	918	989	1193	-
Light	⟶	116	187	286	387	550	651	780	918	989	1187	-
		b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	B_{10}	b_{11}
		D	**A**	**V**	**T**	**Y**	**T**	**E**	**H**	**A**	**K-Me.Pr**	**R**
		y_{11}	y_{10}	y_9	y_8	y_7	y_6	y_5	y_4	y_3	Y_2	y_1
Light	⟶	-	1246	1175	1076	975	811	710	581	444	373	175
Heavy	⟶	-	1262	1191	1092	991	827	726	597	460	389	185

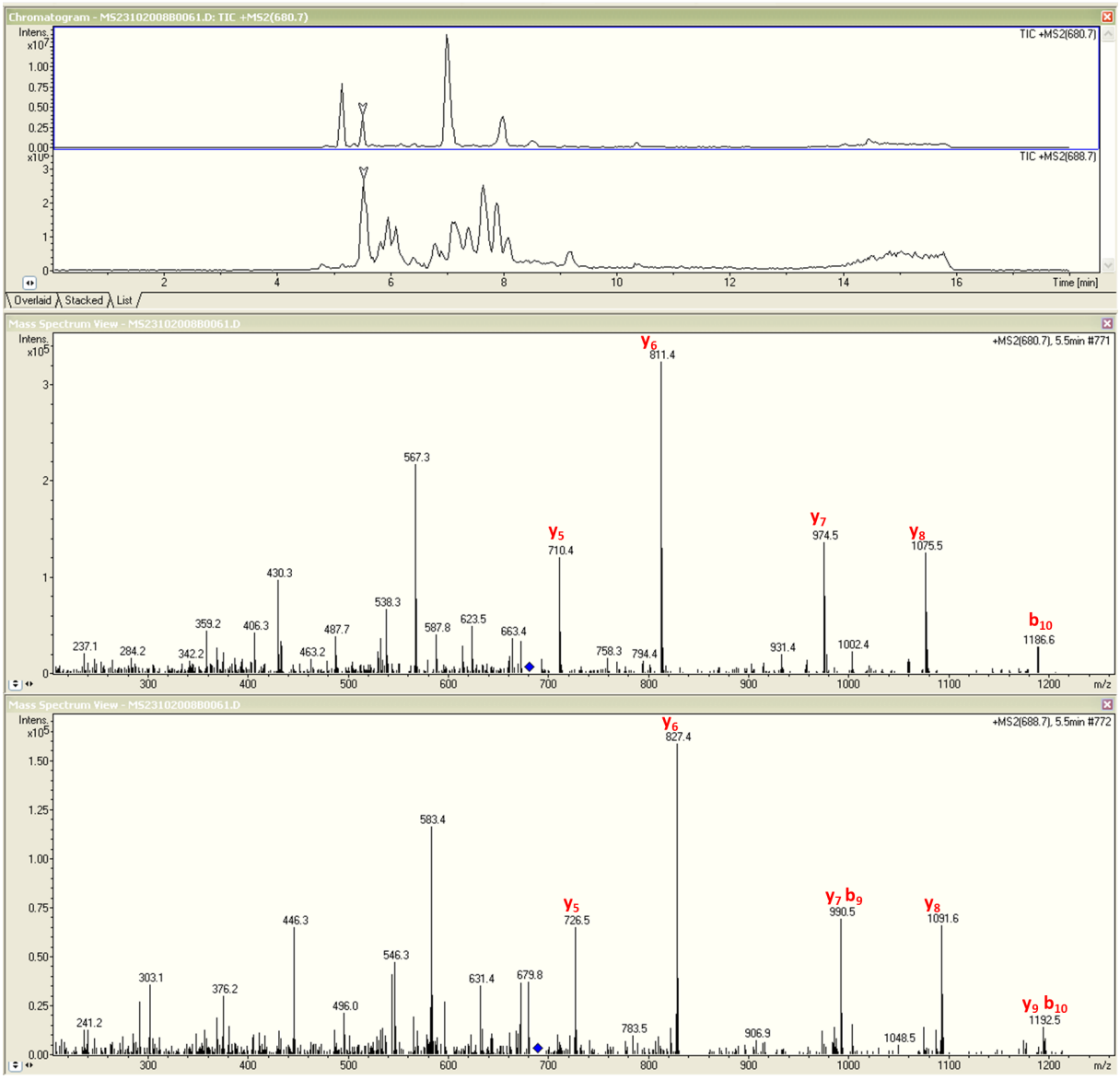


Figure 7.9: MS2 spectrum of Histone H4 residues 68-78 (K77me)

Histone H4 residues 24-35 (DNIQGITKPAIR)

Heavy ⟶	116	230	343	471	528	641	742	933	1030	1101	1214	-
Light ⟶	116	230	343	471	528	641	742	927	1024	1095	1208	-
	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}
	D	N	I	Q	G	I	T	K-Pr	P	A	I	R
	y_{12}	y_{11}	y_{10}	y_9	y_8	y_7	y_6	y_5	y_4	y_3	Y_2	y_1
Light ⟶	-	1267	1153	1040	912	855	741	640	456	359	288	175
Heavy ⟶	-	1283	1169	1056	928	871	757	656	466	369	298	185

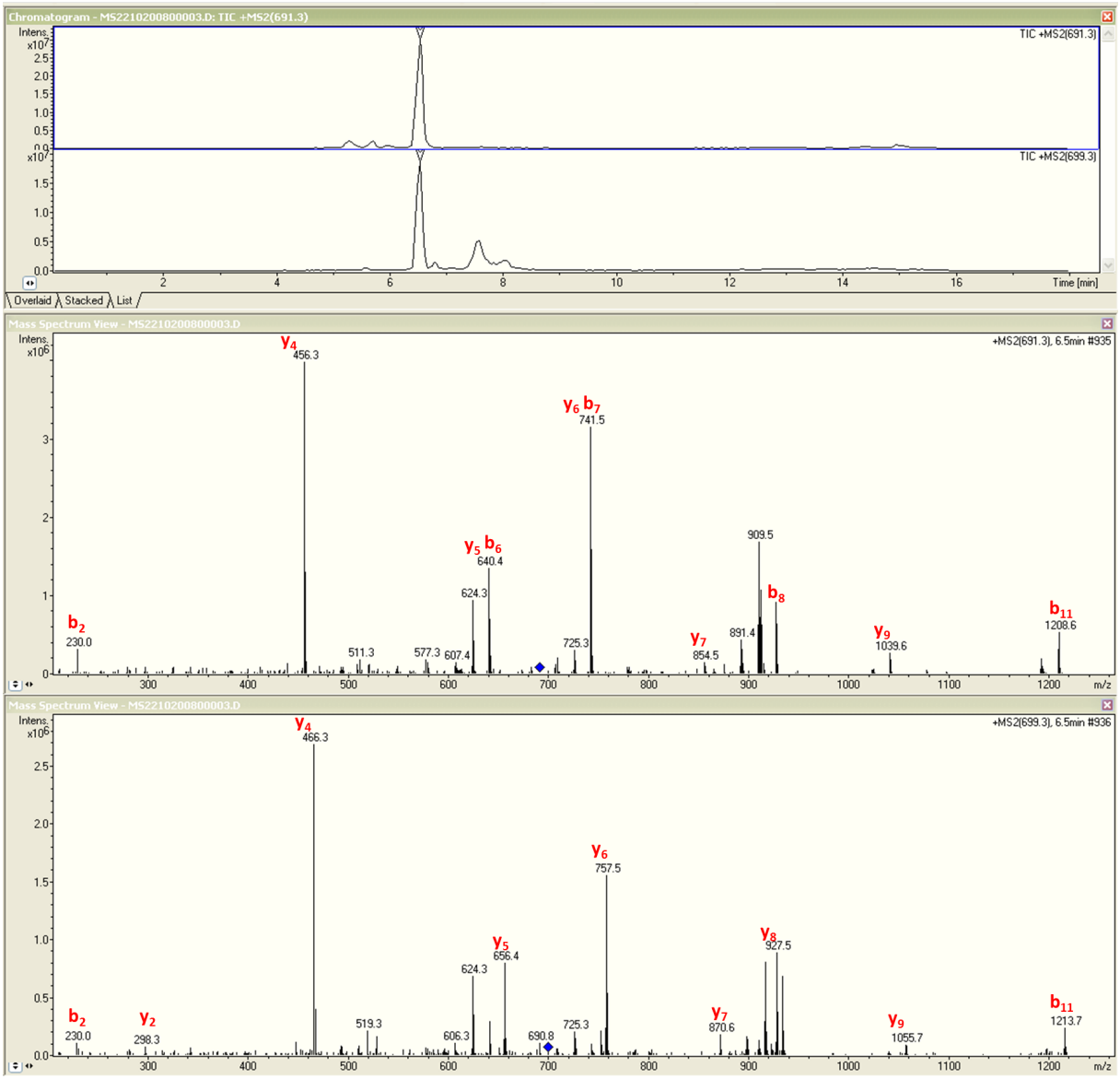


Figure 7.10: MS2 spectrum of Histone H4 residues 24-35 (K31UM)

Histone H4 residues 24-35 (DNIQGITKPAIR)

Heavy ⟶	116	230	343	471	528	641	742	947	1044	1115	1228	-
Light ⟶	116	230	343	471	528	641	742	941	1038	1109	1222	-
	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}
	D	**N**	**I**	**Q**	**G**	**I**	**T**	**K-Me.Pr**	**P**	**A**	**I**	**R**
	y_{12}	y_{11}	y_{10}	y_9	y_8	y_7	y_6	y_5	y_4	y_3	Y_2	y_1
Light ⟶	-	1281	1167	1054	926	869	755	654	456	359	288	175
Heavy ⟶	-	1297	1183	1070	942	885	771	670	466	369	298	185

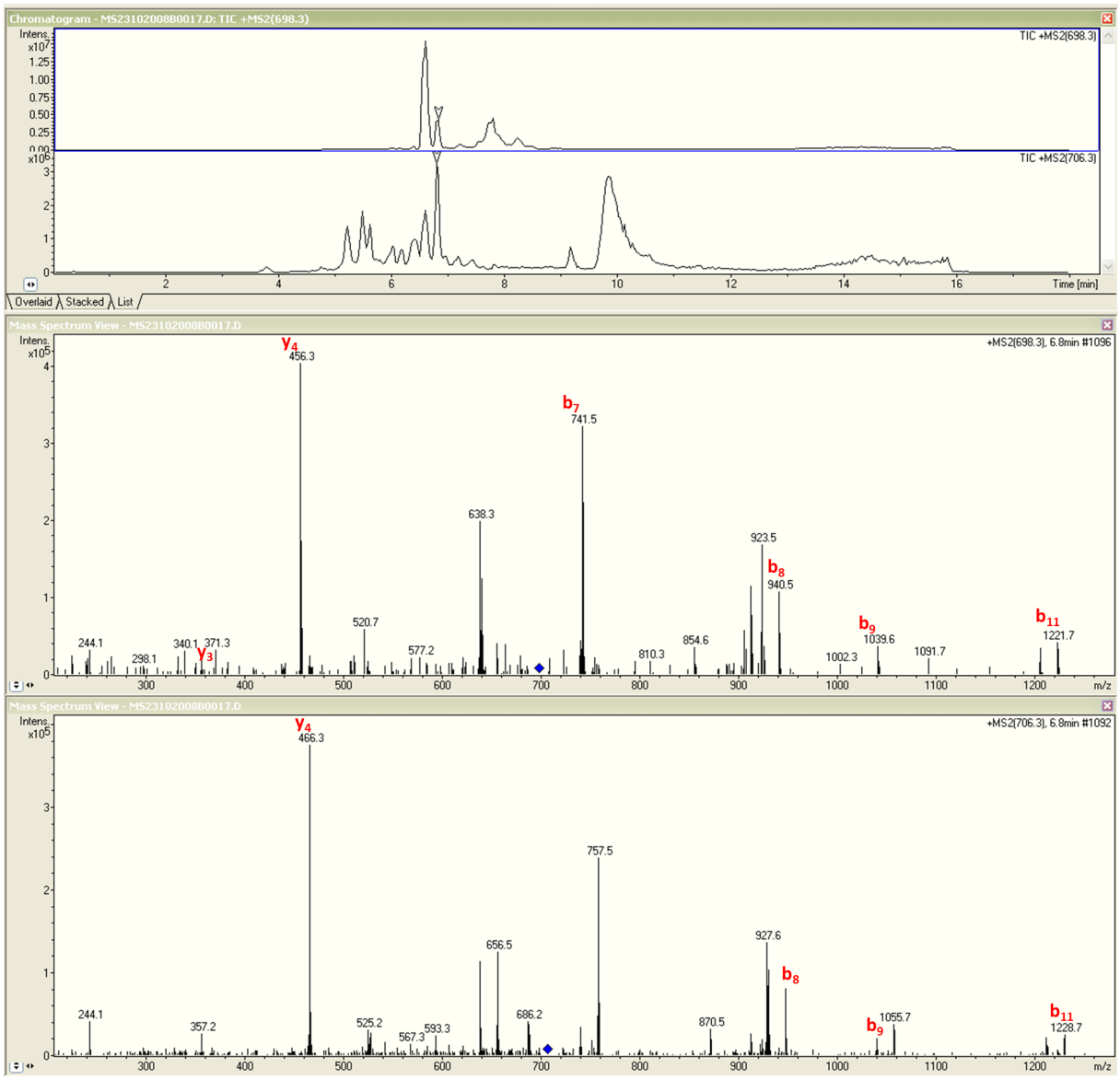


Figure 7.11: MS2 spectrum of Histone H4 residues 24-35 (K31me)

Histone H4 residues 4-17 (GKGGKGLGKGGAKR) 4ac

Heavy ⟶	58	234	291	348	524	581	694	751	928	985	1042	1113	1289	-
Light ⟶	58	228	285	342	512	569	682	739	910	967	1024	1095	1265	-
	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}	b_{14}
	G	**K-Ac**	**G**	**G**	**K-Ac**	**G**	**L**	**G**	**K-Ac**	**G**	**G**	**A**	**K-Ac**	**R**
	y_{14}	y_{13}	y_{12}	y_{11}	y_{10}	y_9	y_8	y_7	y_6	y_5	y_4	y_3	Y_2	y_1
Light ⟶	-	1382	1212	1155	1098	928	871	757	700	530	473	416	345	175
Heavy ⟶	-	1416	1240	1183	1126	950	893	779	722	546	489	432	361	185

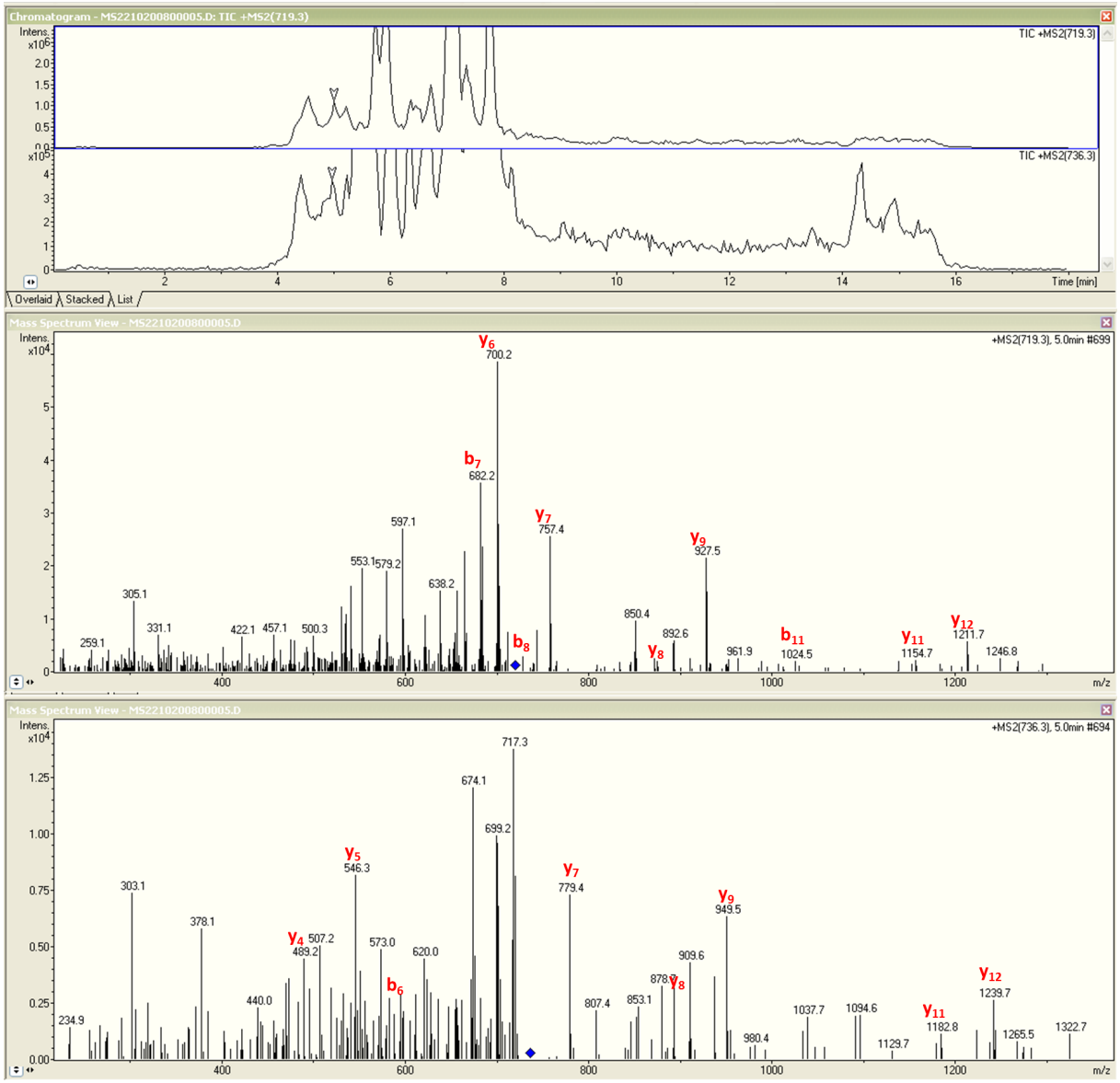


Figure 7.12: MS2 spectrum of Histone H4 residues 4-17 (4ac)

Histone H4 residues 4-17 (GKGGKGLGKGGAKR) 3ac

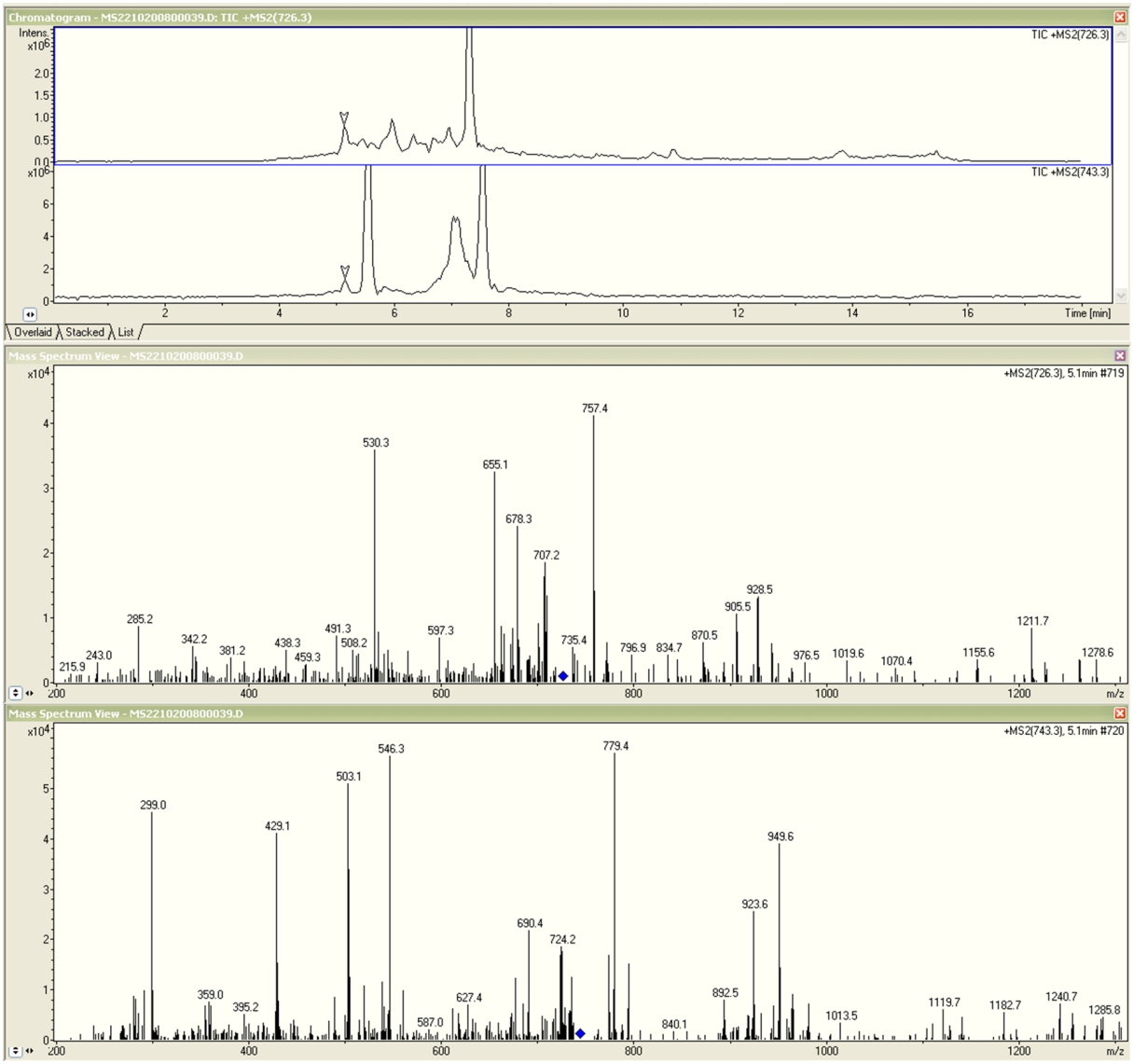


Figure 7.13: MS2 spectrum of Histone H4 residues 4-17 (3ac)

Histone H4 residues 4-17 (GKGGKGLGKGGAKR) 2ac

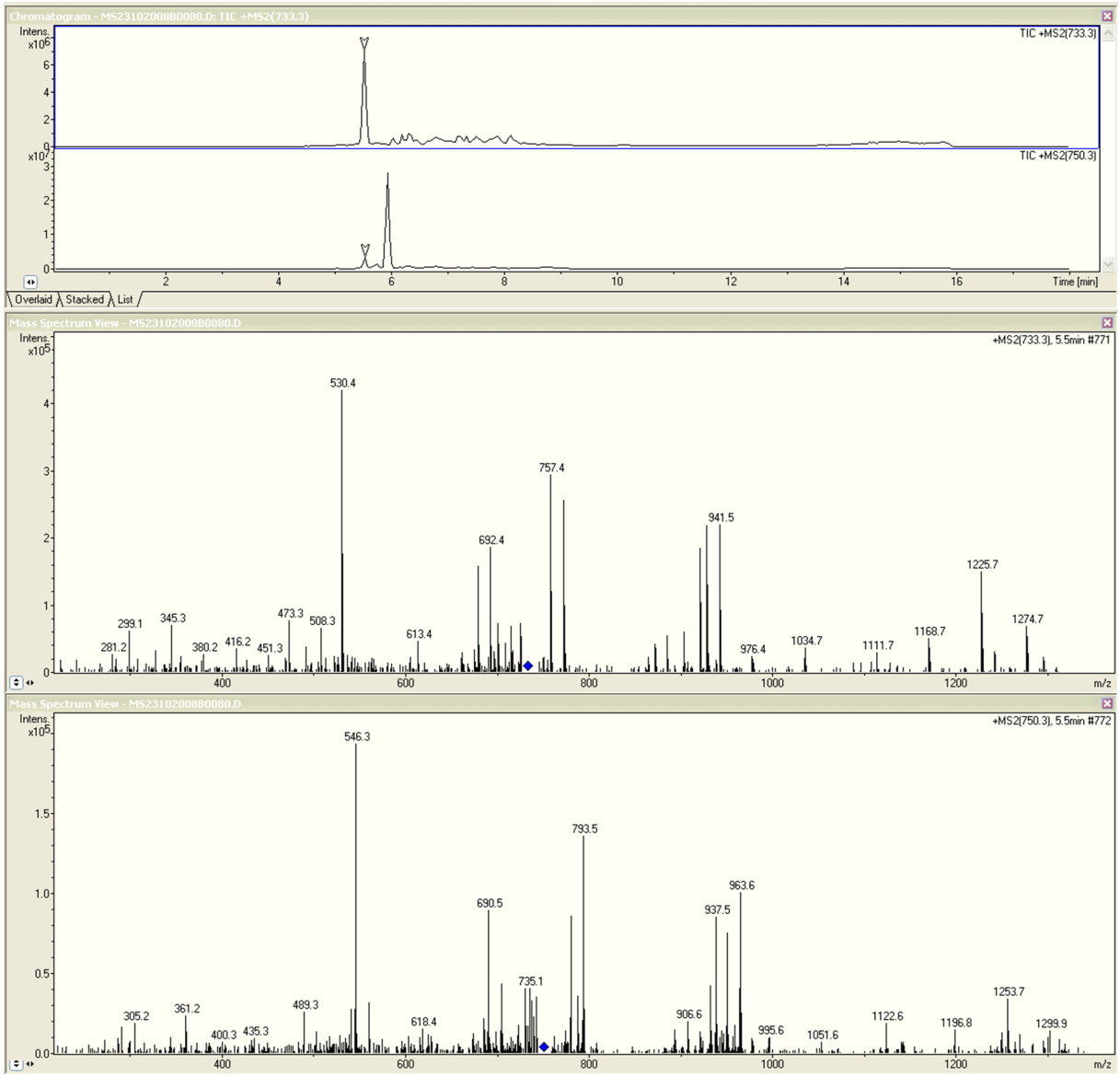


Figure 7.14: MS2 spectrum of Histone H4 residues 4-17 (2ac)

Histone H4 residues 4-17 (GKGGKGLGKGGAKR) 1ac

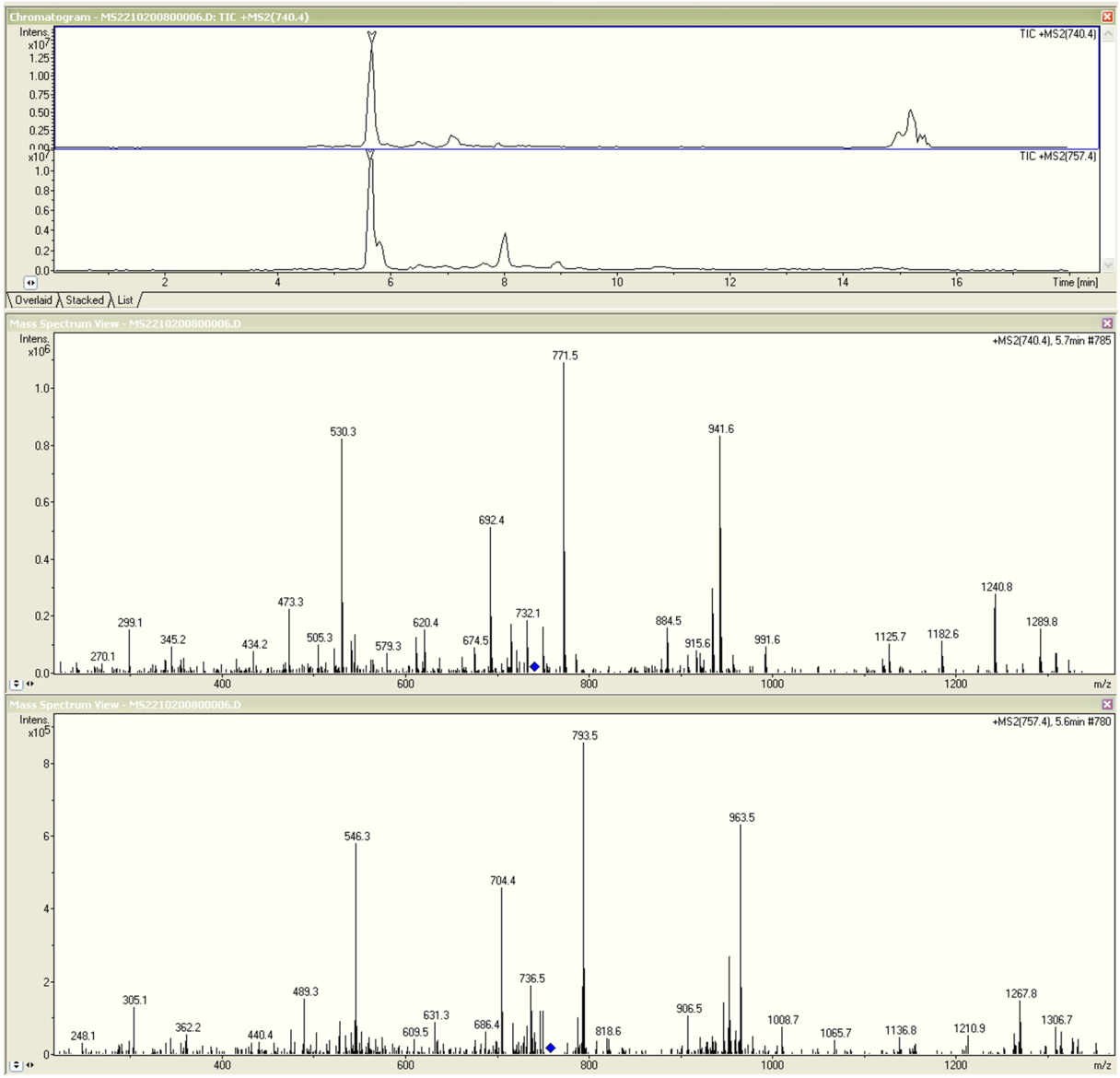


Figure 7.15: MS2 spectrum of Histone H4 residues 4-17 (1ac)

Histone H4 residues 4-17 (GKGGKGLGKGGAKR) UM

Heavy ⟶	58	248	305	362	542	609	722	779	970	1027	1084	1155	1345	-
Light ⟶	58	242	299	356	540	597	710	767	952	1009	1066	1137	1321	-
	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}	b_{14}
	G	**K-Pr**	**G**	**G**	**K-Pr**	**G**	**L**	**G**	**K-Pr**	**G**	**G**	**A**	**K-Pr**	**R**
	y_{14}	y_{13}	y_{12}	y_{11}	y_{10}	y_9	y_8	y_7	y_6	y_5	y_4	y_3	Y_2	y_1
Light ⟶	-	1438	1254	1197	1140	956	899	785	728	544	487	430	359	175
Heavy ⟶	-	1472	1282	1225	1168	978	921	807	750	560	503	446	375	185

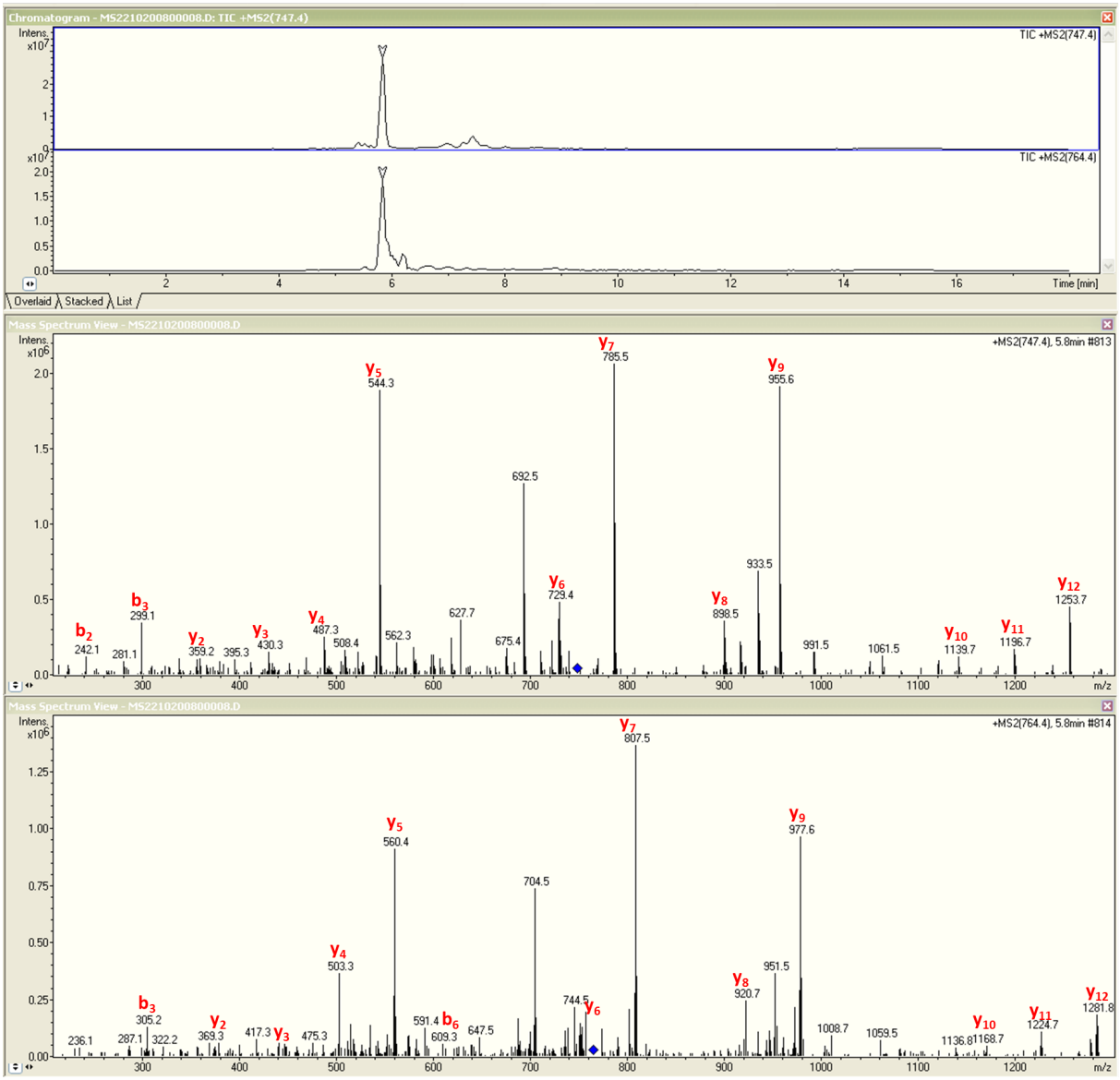


Figure 7.16: MS2 spectrum of Histone H4 residues 4-17 (K5, 8, 12, 16 UM)

8

Annex - II

PUB-MS - A Mass-Spectrometry-based Method to Monitor Protein-Protein Proximity *In-Vivo*. (2011)

Arman Kulyyassov [1,2], **Muhammad Shoaib** [1], Andrey Pichugin [1], Patricia Kannouche [1], Erlan Ramankulov [1,2], Marc Lipinski [1] and Vasily Ogryzko [1]

I CNRS UMR-8126, Université Paris-Sud 11, Institut de Cancérologie Gustave Roussy, 39, rue Camille Desmoulins, 94805, Villejuif, France.

II National Center for Biotechnology of the Republic of Kazakhstan, 43 Valikhanova Str., 010000, Astana, Republic of Kazakhstan.

J Proteome Res. Sep 2 [Epub ahead of print]

Journal of proteome research

ARTICLE

pubs.acs.org/jpr

PUB-MS: A Mass Spectrometry-based Method to Monitor Protein−Protein Proximity *in vivo*

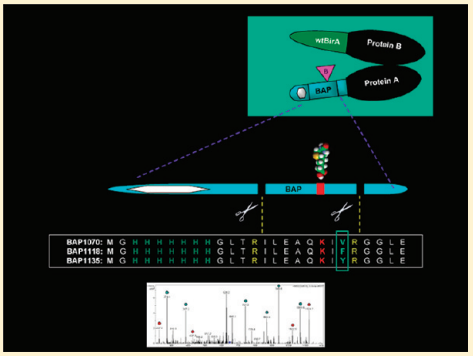
Arman Kulyyassov,[†,‡] Muhammad Shoaib,[†] Andrei Pichugin,[†] Patricia Kannouche,[†] Erlan Ramanculov,[‡] Marc Lipinski,[†] and Vasily Ogryzko[*,†]

[†]Institut Gustave Roussy, 39 Rue Camilles Desmoulin, 94805, Villejuif, France

[‡]National Center of Biotechnology, Valihanova 43, 01000, Astana, Kazakhstan

Supporting Information

ABSTRACT: The common techniques to study protein−protein proximity *in vivo* are not well adapted to the capabilities and the expertise of a standard proteomics laboratory, typically based on the use of mass spectrometry. With the aim of closing this gap, we have developed PUB-MS (for proximity utilizing biotinylation and mass spectrometry), an approach to monitor protein−protein proximity, based on biotinylation of a protein fused to a biotin-acceptor peptide (BAP) by a biotin-ligase, BirA, fused to its interaction partner. The biotinylation status of the BAP can be further detected by either Western analysis or mass spectrometry. The BAP sequence was redesigned for easy monitoring of the biotinylation status by LC−MS/MS. In several experimental models, we demonstrate that the biotinylation *in vivo* is specifically enhanced when the BAP- and BirA-fused proteins are in proximity to each other. The advantage of mass spectrometry is demonstrated by using BAPs with different sequences in a single experiment (allowing multiplex analysis) and by the use of stable isotopes. Finally, we show that our methodology can be also used to study a specific subfraction of a protein of interest that was in proximity with another protein at a predefined time before the analysis.



KEYWORDS: protein−protein interactions, biotinylation, proximity, multiplexing, pulse-chase, SILAC

INTRODUCTION

One of the ultimate goals of molecular biology is reconstruction of the spatiotemporal structure of a living cell at the molecular level. This task includes determination (and cataloguing) of proximities between different molecular components in the cell and monitoring their time- and physiological state-dependent changes. In many cases, proximity between macromolecules arises due to their interactions (either direct or else via intermediates); however, the contribution of dynamic self-organization in generation of spatiotemporal order is emerging as another viable possibility.[1,2] Specifically, in proteomics, this implies that the detection of protein−protein *proximity* is a more general task than gaining information about *physical interactions* between proteins, as it could detail aspects of spatial order *in vivo* that are challenging (at the very least) to reconstitute in binding experiments *in vitro*.

In recent years, several approaches have been developed to monitor protein−protein interactions inside living mammalian cells, and thus to provide the physiologically relevant information about the spatial organization of cell. They include fluorescence resonance energy transfer (FRET) with fluorescent proteins,[3,4] bioluminescence resonance energy transfer (BRET),[5,6] protein complementation assays (PCA),[7,8] and various two-hybrid systems.[9,10] However, they suffer from several limitations that we aimed to overcome in our work. For example, only relatively close proximity can be studied with these methods (for example, FRET provides intra- or intermolecular distance data in the range 1−10 nm). Second, although the study of protein−protein interactions falls within the scope of proteomics, the FRET/BRET and PCA methods require access to instrumentation (e.g., confocal microscope) that usually do not match the capabilities and the expertise of a standard proteomics laboratory.

Given that a mass spectrometer is the working horse of a core proteomics facility, it is desirable to develop a mass-spectrometry-oriented approach to monitor protein−protein interactions (and, more generally, their proximity) *in vivo*. Since accurate mass measurement is the main principle of this methodology, such an approach could be based on the proximity-dependent introduction of a modification in the tested protein, which can be further detected in a quantitative manner with the help of mass spectrometry. One potential advantage of mass spectrometry, as compared to other methods (such as FRET, BRET, PCA, two-hybrid), is the high resolution of mass spectrometer, that is, its ability to reliably distinguish between thousands of ions with different m/z ratios. This opens a possibility of multiplexing, that is, monitoring of several interactions in a single experiment.

Received: February 28, 2011

   dx.doi.org/10.1021/pr200189p | *J. Proteome Res.* XXXX, XXX, 000–000

Previously, we and others have developed a method for efficient biotinylation of proteins *in vivo*,[11,12] based on the coexpression within the same cell, of the biotin ligase BirA and the protein of interest fused to a short biotin acceptor peptide (BAP). Such *in vivo* biotinylated proteins have been used for protein complex purification,[11,13,14] chromatin immunoprecipitation,[12,13,15,16] immunoelectron microscopy[17] and other purposes. In this work, we have adapted our system to the task of monitoring protein−protein proximity *in vivo*, with the use of mass spectrometry. For this purpose, the BirA was fused to one of the interaction partners, whereas the BAP was modified to make the detection of its biotinylation possible by mass spectrometry. Using several experimental systems, we show that the biotinylation of BAP is interaction (and/or proximity) dependent. In addition, we demonstrate that BAP domains with different primary amino acid structures and thus with different molecular weights can be used in the same experiment, providing the possibility of multiplexing. Alternatively to the changes in primary amino acid structure, the stable isotope format can also be used, providing another way to perform multiplexing experiments.

An additional advantage of this technology resides in the fact that it is not limited to detection of short-range proximities, as FRET and PCA are. In particular, we demonstrate that sharing the same intracellular compartment can also give a signal above the background. Thus, with appropriate controls, our approach can provide a new kind of information, complementary to the short-range proximity data obtained by the FRET and PCA methodologies.

Finally, we also demonstrate that our system can help to overcome another limitation of current methodologies to detect protein−protein proximity. Given an introduction of a covalent mark in the protein of interest, one can ask questions that are impossible to address with FRET/BRET or PCA. For example, one can follow the state of a protein of interest at a defined time after its interaction with another protein has occurred. This application should be particularly useful for studying multistep intracellular processes, where the proximities between proteins (e.g., interaction partners) and protein properties typically change in a sequential manner. Thus, in addition to helping to reconstruct the cell topology in space, this approach has a promise to add the temporal dimension to such studies.

EXPERIMENTAL PROCEDURES

Recombinant DNA

For the pCMV enhancer vectors, the pCDNA3.1 (Invitrogen) backbone was used. For the MoMuLV enhancer vectors, the pOZ vector was used.[18] The BirA ORF was PCR amplified from the vector pBBHN.[12] The BAP domain fragments were prepared by designing oligonucleotides and sequential PCR, followed by restriction digest and insertion into the pCDNA3.1 vector. The insert sequences were confirmed by sequencing. The vectors generated were used to clone the ORFs of H2AZ, HP1γ, TAP54α, KAP1-HP1BD, Rad18, GFP, H2ABBD, macroH2A and H3.3.

Cell Culture and Cell Lines

HEK293T cells were grown in Dulbecco's modified Eagle's medium with high glucose (PAA) and 10% fetal bovine serum (FCS, PAA). For transient transfection and Western analysis, a standard calcium phosphate precipitation method was used, and the cells were analyzed one or two days after transfection, as indicated. For the biotin labeling *in vivo*, cells were grown for several days before transfection in the DMEM supplied with dialyzed FCS, and for the specified time of labeling, biotin (Sigma) was added to a final concentration of 5 μg/mL, while the pH was stabilized by addition of 50 mM HEPES (pH 7.35) to the medium. For SILAC experiments, the cells were grown in DMEM with $^{12}C_6$ L-lysine or $^{13}C_6$ L-lysine (Thermo Scientific, cat # 89983) for at least 5 divisions before transfection, and kept in the same medium until harvest. For the Western analysis, 3×10^5 of cells (corresponding to one well in a 6-well plate) were used for transfection. For the mass-spectrometry analysis, 3×10^6 of cells were transfected per data point.

Biochemistry and Western Blot Analysis

Except where indicated otherwise, cell nuclei were used for analysis. They were prepared by cell disruption in CSK buffer (100 mM NaCl, 300 mM Sucrose, 10 mM Tris pH 7.5, 3 mM $MgCl_2$, 1 mM EGTA, 1.2 mM PMSF, 0.5% Triton X-100) and centrifugation for 5 min at 4000 rpm. For the analysis of chromatin-associated histones (chromatin fraction), the nuclei were first incubated in CSK buffer containing 450 mM NaCl by 30 min rotation at 4 °C, then spun at 4000 rpm, and the supernatant containing soluble histones was discarded. For Western analysis, 1× NuPAGE LDS Sample buffer (Invitrogen) with DTT (10 mM) was added, and the nuclei were sonicated, boiled for 5 min at 96 °C and loaded on 4−12% gradient Novex Tris-Glycine precast gels. After separation, the proteins were transferred to nitrocellulose membranes and probed with HRP-conjugated streptavidin (Sigma, # S5512) or HRP conjugated α-PentaHis antibody (QiaGen, # 34460) according to the manufacturer's protocol, except that for the detection of biotinylated proteins by the HRP-conjugated streptavidin, 500 mM NaCl was added to the washing buffer (PBS + 0.1% Tween). For protein visualization, the gels were stained with PageBlue (Fermentas, # R0579). For the densitometric analysis of Western blots, the program ImageJ 1.42q (freely available online) was used. To compare the biotinylation levels between different samples, the value of the streptavidin-HRP signal was normalized by taking a ratio with the value of α-His signal, which reflects the amount of the transfected protein regardless of its biotinylation status. Every Western analysis was performed three times, with a representative figure shown.

For Ni^{2+}NTA purification of 6× His tagged proteins, the nuclei were solubilized in buffer A (10% glycerol, 250 mM NaCl, 6 M Guanidine-HCl, 20 mM TrisHCl (pH 8.0), 0.1% Tween) by rotation for 30 min at 4 °C. One-tenth of the volume of Ni^{2+}NTA agarose, prewashed in the same buffer, was added to the lysate, followed by rotation at 4 °C for 3 h. The beads with bound proteins were washed twice with buffer A and then twice with buffer B (10% glycerol, 250 mM NaCl, 20 mM TrisHCl (pH 8.0), 0.1% Tween, 0.2 mM PMSF and protease inhibitor cocktail complete (Roche)). The bound proteins were eluted by incubating the beads in 4 volumes of buffer C (10% glycerol, 250 mM NaCl, 20 mM Tris HCl (pH 8.0), 0.1% Tween, 0.2 mM PMSF, 300 mM Imidazole, 50 mM EDTA) and concentrating the sample by ultrafiltration with Microcon YM-10 (Millipore). Alternatively to elution, the beads were washed twice with 25 mM ammonium bicarbonate and direct on-beads digestion was performed.

Mass Spectrometry Analysis

The protein bands were excised from the gel and processed as in.[19] The gel slices were dehydrated with 300 μL of 50% acetonitrile followed by 300 μL of 100% acetonitrile, then

dx.doi.org/10.1021/pr200189p | J. Proteome Res. XXXX, XXX, 000–000

rehydrated with 300 μL of 50 mM ammonium bicarbonate. A final dehydration was performed with 2 washes of 300 μL of 50% acetonitrile, followed by 2 washes of 300 μL of 100% acetonitrile. Each wash was carried out for 10 min at 25 °C with shaking at 1400 rpm. The gel slices were dried in a SpeedVac at 35 °C for 10 min. For trypsin digestion, the gel slices were preincubated with 7 mL of 15 ng/mL trypsin (Promega # V5280) at room temperature for 10 min. Afterward, 25 μL of 50 mM ammonium bicarbonate was added, and the gel slices were incubated at 37 °C for 16 h. The peptide-containing supernatants were dried at 56 °C by SpeedVac for 30 min, then resuspended in 20 μL of solution containing 0.05% formic acid and 3% acetonitrile for mass spectrometry experiments. Alternatively, for the on-beads digestion, ammonium bicarbonate was added to the beads at a final concentration of 25 mM, trypsin digestion was performed overnight (12.5 ng/mL), and the peptide mixture was further purified on a Zip tip (Millipore # ZTC18S096). The resulting peptides were analyzed with a nano-HPLC (Agilent Technologies 1200) directly coupled to an ion-trap mass spectrometer (Bruker 6300 series) equipped with a nanoelectrospray source. The separation gradient was 7 min from 5 to 90% acetonitrile. The fragmentation voltage was 1.3 V. The analysis of the spectra was performed with the DataAnalysis for the 6300 Series Ion Trap LC−MS Version 3.4 software package. The samples were run in two different modes. For peptide identification, the ion trap acquired successive sets of 4 scan modes consisting of: full scan MS over the ranges of 200−2000 m/z, followed by 3 data-dependent MS/MS scans on the 3 most abundant ions in the full scan. Alternatively, for the confirmation and quantification of the presence of a particular peptide in the sample, the ion trap was set in MRM mode. The sample was separated using the same nanoLC gradient, and the ion trap was set to isolate, fragment, and MS/MS scan 5 parental ions having predetermined M/Z ratios. The relative quantity of each peptide in the different fractions was estimated by comparison between the peak areas in the Total Ion Chromatograms (TIC) for this peptide obtained from MRM analysis of these fractions.

Microscopy

MRC-5 fetal lung fibroblasts were grown on coverslips in 6-well plates in DMEM high glucose containing dialyzed FBS. Cells were cotransfected with the plasmid constructs BirA-Rad18 (1 μg) and BAP-H2AZ (1 μg) using TurboFect (Cat. #R0531, Fermentas). Forth-eight hours after transfection, cells were treated with UVC (20 J/m^2, Philips TUV 15W/G 15 TB lamp). After 6 h of irradiation, the DMEM containing dialyzed FBS was removed and the cells were pulse-labeled for 5 min with biotin (5 μg/mL final concentration). The biotin-containing medium was then removed, and cells were fixed either immediately (pulse) or 2 h after the pulse (chase). For staining, cells were washed once with cold PBS and treated with CSK buffer (100 mM NACl, 300 mM Sucrose, 3 mM $MgCl_2$, 10 mM Tris-HCl pH 7.5, 1 mM EGTA, Triton 0.2%) for 5 min with light agitation. Cells were then washed twice with cold PBS. Fixation was done in 4% formaldehyde for 20 min at room temperature (RT). After fixation, cells were rinsed with ice-cold methanol at −20 °C for 10 s. Cells were then washed 3 times for 5 min with PBS (1×) at RT. Blocking was done for 30 min using 3% BSA and 0.5% Tween20 in PBS. Cells were then incubated with either mouse monoclonal anti-PCNA antibody (Santa Cruz; PC10; sc-56) for 1 h (1/500 in blocking solution) or mouse monoclonal anti-Rad18 antibody (Abcam; ab57447) followed by 3 washes in PBS (1×) at RT. Cells were then incubated with goat anti-mouse secondary antibody conjugated with Alexa-488 (Invitrogen # A11017) for 1 h (1/1000). Streptavidin-Cy3 conjugate (Sigma; # S-6402) was also added to the same incubation mix (1/500). After 3 washes in PBS (1×) at RT, coverslips were mounted on glass slides using VectaShield mounting medium (Vector Laboratories; # H-1000). The cells were observed under a Zeiss LSM 510 Meta confocal microscope, using a Plan-Apochromat 63× 1.4 oil immersion objective. Imaging was performed with sequential multitrack scanning using the 488 and 543 nm wavelengths lasers separately. The colocalization analysis was performed with LSM Examiner software. Pearson's correlation coefficient was calculated with setting the threshold signal common for all images compared.

RESULTS

1. Design and Features of the System

The principle of the method is presented in Figure 1A. Two proteins to be tested for their proximity *in vivo* are coexpressed in the cells of interest, the first protein being expressed as a BirA fusion and the second one as a fusion to the Biotin Acceptor Peptide (BAP). More efficient biotinylation of the BAP is expected when the two proteins are in proximity to each other, for example, when interaction occurs (Figure 1A, bottom), as compared to the background biotinylation arising from random collisions between noninteracting proteins (Figure 1A, top). The biotinylation status of the BAP fusion can be further monitored by Western blot or mass spectrometry.

To implement this principle, we constructed two types of vectors: for expression of the BirA fusion and for expression of the BAP fusion, correspondingly (Figure 1B, top). We constructed each vector in two forms, with the protein expression regulated either by a strong CMV or a weaker MoMuLV enhancer. In this work, we used the MoMuLV enhancer for the expression of BirA fusions, while the BAP fusions were expressed from the vectors with the stronger (CMV) promoter. This setting typically allowed us to achieve an excess of the BAP fusion (biotinylation target) over the BirA fusion (biotinylating enzyme), which is essential for the further quantitative analysis.

In addition, we significantly modified the sequence of the BAP, by introducing two flanking arginines that yield a peptide of an "MS/MS-friendly" size after trypsin digestion of the BAP sequence. We also included a 7XHis tag in the BAP-domain, to have the option of purifying or/and monitoring the amounts of the target protein regardless of its biotinylation status (Figure 1B, bottom).

To test if the newly designed BAP could be biotinylated by BirA, we cotransfected 293T cells with BirA-GFP and BAP-GFP. Regardless of the fact that GFP can form dimers at high concentrations,[20] the main aim of this experiment was not to detect the GFP-GFP interaction but rather to test the BAP biotinylation, as even background biotinylation due to random collisions between noninteracting proteins could be sufficient to give detectable signal. As seen in Figure 1C, the biotinylation was indeed observed; moreover, as expected, the level of biotinylated BAP-GFP linearly increased with the time of incubation of the cells with biotin. Notably, in parallel contrasfections, the GFP carrying a different BAP sequence (MAGLNDIFEAQKIEWHE), used previously,[12] was biotinylated with significantly faster kinetics (the time of half-saturation was 15 min, contrasting with 680 min for the newly designed BAP). Because saturation in the BAP biotinylation levels obscures differences in biotinylation efficiency, the slower kinetics of biotinylation of the

dx.doi.org/10.1021/pr200189p | J. Proteome Res. XXXX, XXX, 000–000

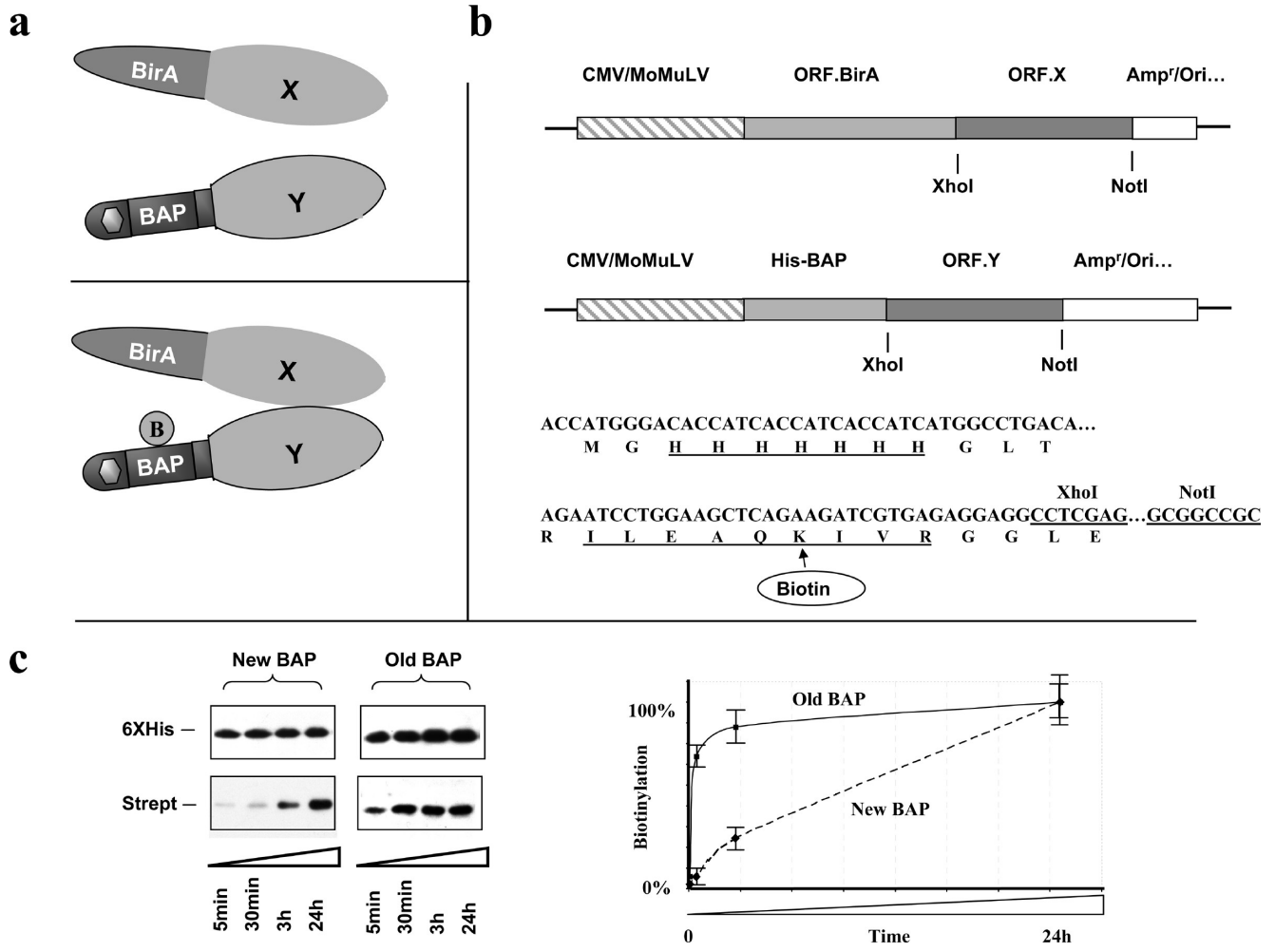


Figure 1. Design and features of the system. (A) Principle. Two candidate interaction proteins X and Y are coexpressed in the cells of interest, the first protein being expressed as a BirA fusion, and the second one as a fusion to the Biotin Acceptor Peptide (BAP). The cells are pulse-labeled with biotin. Where there is interaction (top), the BAP-X fusion should be more efficiently biotinylated. (B) Vector design. Top, the positions of the CMV/MoMuLV enhancers, and of the BAP and BirA sequences, relative to the cloned ORF, are indicated. Bottom, nucleotide and amino acid sequences encoding the newly designed His-BAP are shown. (C) New BAP is biotinylated with slower kinetics as compared to the old BAP. Left, Western analysis of biotinylation of the new BAP-GFP (ILEAQKIVR, left) versus previously used efficient old BAP-GFP (MAGLNDIFEAQKIEWHE, right) in the presence of BirA-GFP fusion. Top, α-6XHis-HRP signal; bottom, streptavidin-HRP signal. Right, quantification of the kinetics of biotinylation of the two BAPs. The streptavidin signal was first normalized by taking a ratio with the respective α-6XHis signal. For every individual experiment, the normalized value of the biotinylation level for each time point was divided by the respective value for the 24 h time point from the same experiment, which was taken as 100%. Plotted are averaged results from three independent experiments.

new BAP was welcome, given our ultimate goal to correlate BAP biotinylation with protein−protein proximity.

2. Biotinylation Levels Are Interaction and/or Proximity Dependent

To confirm that our method can detect specific protein−protein proximities, we used several experimental systems described below.

a. Protein Oligomerization (TAP54α vs HP1γ). TAP54α (RuvB-like 1) was shown to exist as oligomers, composed of several copies together with the closely related TAP54β (RuvB-like 2).[21,22] The heterochromatin proteins HP1 (α, β, γ) are also known to oligomerize through their chromoshadow domains,[23,24] and we typically detected various HP1 forms in our affinity pull-downs using epitope-tagged HP1 proteins (VO, in preparation). On the other hand, no interactions between the HP1 and the RuvB-like proteins had been reported; thus, the heterologous cotrasfection with these two proteins (e.g., BirA-HP1γ and BAP-TAP54α) is expected to give close to background levels of biotinylation. Figure 2A (top) shows the biotinylation status of the BAP-TAP54α and BAP-HP1γ, when these proteins were coexpressed with either the BirA-TAP54α or BirA-HP1γ fusions. Whereas the total levels of the BAP fusions are not affected, as judged by Western blotting with the anti-6XHis antibody (Figure 2A, top left), the BAP-TAP54α is appreciably biotinylated when coexpressed with the homologous construct BirA-TAP54α. On the other hand, BAP-HP1γ is significantly more strongly biotinylated when coexpressed with the BirA-HP1γ, as compared to the control cotransfection with heterologous BirA-TAP54α (Figure 2A, top right). In analogous experiments, the related proteins TAP54β and HP1α behaved similarly to the TAP54α and HP1γ, correspondingly (not shown).

In the previous experiment, we could not measure the levels of the BirA-TAP54α and BirA-HP1γ proteins. To rule out the

dx.doi.org/10.1021/pr200189p |J. Proteome Res. XXXX, XXX, 000–000

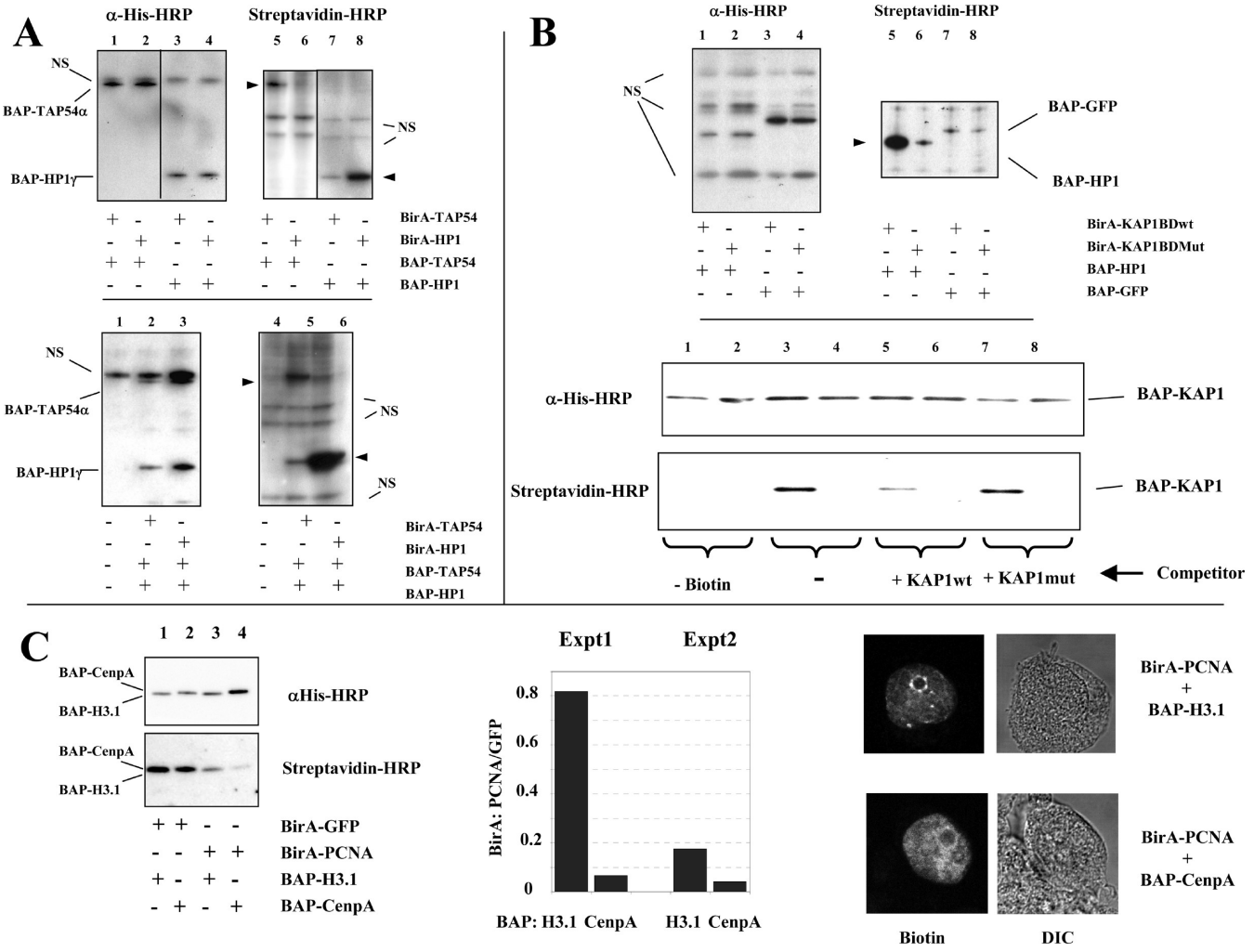


Figure 2. Biotinylation levels are interaction/proximity dependent. (A) HP1γ versus TAP54α oligomerization. Top, 4 combinations of BirA and BAP fusions were transfected separately into cells: 1,5, BirA-TAP54α + BAP-TAP54α; 2,6, BirA-HP1γ + BAP-TAP54α; 3,7, BirA-TAP54α + BAP-HP1γ; 4,8, BirA-HP1γ + BAP-HP1γ; 1−4, α-His-HRP Western; 5−8, streptavidin-HRP Western. The positions of the BAP-fusions and nonspecific signal (NS) are indicated by lines, and the biotinylated proteins are indicated by black triangles. Bottom, Biotinylation of BAP-HP1γ and BAP-TAP54α, cotransfected into the same cells, by either BirA-HP1γ or BirA-TAP54α. 1,4, untransfected cells; 2,5, BirA-TAP54α + BAP-TAP54α + BAP-HP1γ; 3,6, BirA-HP1γ + BAP-TAP54α + BAP-HP1γ. Left, α-His-HRP Western, Right, streptavidin-HRP Western. The positions of the BAP-fusions (triangles) and nonspecific signal (NS) are indicated. Importantly, the nonspecific signal (NS) is also detected in untransfected cells, thus it corresponds to cellular proteins that bind to the α-His antibodies or streptavidin, and do not represent nonspecific biotinylation by BirA-fusions. (B) Binary interaction between HP1γ and KAP1. Top, BirA-KAP1BD biotinylating BAP-HP1γ. 1,5, BirA-KAP1BDwt + BAP-HP1γ ; 2,6, BirA-KAP1BDmut + BAP-HP1γ; 3,7, BirA-KAP1BDwt + BAP-GFP; 4,8, BirA-KAP1Bdmut + BAP-GFP. Left, α-His-HRP Western. Right, Streptavidin-HRP Western. The positions of the BAP-fusions and nonspecific signal (NS) are indicated. Bottom, BirA-HP1γ biotinylating BAP-KAP1BD. 1,3,5,7, BirA-HP1γ + BAP-KAP1BDwt; 2,4,6,8, BirA-HP1γ + BAP-KAP1Bdmut. 1,2, No biotin was added to cells; 5,6, 10-fold excess of a plasmid expressing untagged KAP1BDwt was added; 7,8, the same as 5,6 but KAP1BDmut used. Top, α-His-HRP Western. Bottom, streptavidin-HRP Western. (C) Colocalization in the nucleus (PCNA+H3.1 vs PCNA+CenpA). Left, Western analysis of the biotinylation efficiency of BAP-H3.1 (lanes 1,3) as compared to BAP-CenpA (2,4) by BirA-PCNA fusion (3,4). BirA-GFP biotinylation (1,2) was used to control for the differences in biotinylability of BAP-peptide in context of H3.1 and CenpA fusion. Top, αHis. Bottom, Streptavidin signals. Middle, Quantification of the relative biotinylation efficiency. The signal intensities were first measured by densitometry, then the streptavidin signal for every BAP-fusion was normalized by dividing it by the α-His signal. Relative efficiency of biotinylation of a respective BAP-fusion by the BirA-PCNA as compared to BirA-GFP was obtained by dividing the normalized values of biotinylation for the lanes 3 and 1 for H3.1, and for lanes 4 and 2 for CenpA, respectively. Shown are results of calculations for two independent experiments. The lower biotinylation signal in the case of BirA-PCNA transfection is due to a significantly lower levels of BirA-PCNA expression levels as compared to BirA-GFP (not shown). Right, Replication-like foci of biotinylated BAP-H3.1 after cotransfection with BirA-PCNA. Confocal microscopy of cells cotransfected with "BAP-H3.1 + BirA-PCNA" (top) or "BAP-CenpA + BirA-PCNA" (bottom), labeled with biotin 48 h after transfection, and stained with straptavidin-Cy3 (left) or analyzed by Differential Interference Contrast (DIC, right).

remaining possibility that the observed differences in biotinylation are due to variations in the amounts of these proteins between different transfection samples, both BAP-TAP54α and BAP-HP1γ fusions were now coexpressed in the same cells, along with either the BirA-TAP54α or BirA-HP1γ fusions (Figure 2A, bottom). Despite the presence of comparable amounts of the BAP-TAP54α

dx.doi.org/10.1021/pr200189p |J. Proteome Res. XXXX, XXX, 000–000

and BAP-HP1γ fusions in each sample (Figure 2A, bottom left), each protein was more strongly biotinylated in the sample with the homologous BirA fusion (Figure 2A, bottom right).

b. Binary Protein−Protein Interaction (KAP1 and HP1). HP1 strongly interacts with the KAP1 protein, and the domain of KAP1 responsible for the interaction, including a crucial residue, has been identified.[25,26] We chose the HP1-binding domain of KAP1 (called here KAP1BDwt) together with the point mutant abrogating the interaction (KAP1BDmut) to test if our system can detect specific binary protein−protein interactions (Figure 2B). First, we coexpressed either the wild type or the mutant version of BirA-KAP1BD with BAP-HP1γ. BAP-GFP was used as negative control (Figure 2B, top). Strong biotinylation of BAP-HP1γ was observed only when it was coexpressed with BirA-KAP1BDwt (lane 5). The BirA-KAP1Bdmut construct yielded levels of biotinylation (lane 6) comparable to that of BAP-GFP (lanes 7, 8), most likely reflecting the background levels of labeling.

We also performed a reciprocal experiment, by expressing BAP-KAPBD1wt or BAP-KAP1BDmut, together with the BirA-HP1γ fusion (Figure 2B, bottom). Judging by the anti-6XHis Western, the expression levels of the mutant and wild type fragments of KAP1 were identical (Figure 2B, bottom, top). However, the BAP-KAPBD1wt was biotinylated to a significantly greater extent than the mutant protein (Figure 2, bottom, lane 3 versus 4). Notably, the coexpression of the competitor KAPBDwt (which contains no BAP or His-tag) suppressed this biotinylation (lane 5), whereas the mutant competitor KAPBDmut did not have an effect (lane 7).

c. Colocalization in the Nucleus ("PCNA + H3.1" vs "PCNA + CenpA"). Two proteins might not stably interact with each other and yet be colocalized in the cell, which could also increase the biotinylation efficiency. Replication processivity factor PCNA binds directly to p150, the largest subunit of CAF-1, and the two proteins colocalize at sites of DNA replication in cells.[27] CAF-1 is a histone chaperone, which assists in chromatin assembly on replicating DNA, binding directly to the H4/H3 histone dimer.[28] No direct interaction between PCNA and histones has been reported. On the other hand, a specialized H3 histone variant CenpA is deposited on centromeres, in DNA replication independent manner, and thus it is not expected to be present at replication sites.[29] Therefore, despite the absence of direct interaction, one would expect to observe higher proximity between PCNA and canonical H3.1 histone, as compared to PCNA and CenpA. Consistent with these expectations, BAP-H3.1 was biotinylated more efficiently by BirA-PCNA fusion, as compared to BAP-CenpA (Figure 2C). Using fluorescent microscopy, we have also observed that in some cells transfected with "BAP-H3 + BirA-PCNA", BAP.H3.1 is biotinylated in pattern of bright foci, closely resembling the pattern of replication foci. No such pattern was observed in the "BAP-CenpA + BirA-PCNA" transfected cells, which is consistent with the notion that increased biotinylation of BAP-H3.1 by BirA-PCNA is due to their colocalization at replication sites in nuclei.

d. Different Subnuclear Domains (macroH2A vs H2ABBD). As another model example of protein proximity due to similar localization in the cell, we used replacement variant histones. The replacement variant histone H2ABBD has been shown to mark active chromatin (i.e., to be excluded from heterochromatin, such as Barr bodies),[30] whereas the variant macroH2A has been shown to associate with repressed chromatin.[31,32] One should expect that, on average, different molecules of the histone of the same variant type would be closer to each other in the nucleus as compared to the molecules of different histone type. The BirA- and BAP- fusions of these two proteins were generated and cotransfected in different combinations. Both total nuclear lysate and the chromatin fraction were analyzed by Western blot. In both cases, the coexpression with a homologous BirA fusion significantly increased the efficiency of biotinylation of the BAP-histone fusion (Figure S1, left, and for quantification, Figure S1, right, Supporting Information), further supporting the notion that, regardless of physical interaction, similar localization could also be detected with proximity-mediated biotinylation.

e. Intracellular Compartmentalization (GFP and HP1). Biotinylation efficiency could also depend on compartmentalization, as sharing common compartment will increase the effective concentration of the BirA fusion near the BAP-target and promote the reaction, even when driven by random collisions between noninteracting proteins. Accordingly, we compared biotinylation of BAP-GFP and BAP-HP1γ when coexpressed with BirA-GFP or BirA-TAP54α fusions. Although TAP54α and HP1γ have not been reported to interact, they both are nuclear proteins, whereas GFP is found in both nucleus and cytoplasm (Figure S2C, Supporting Information). Thus, we expected that BAP-GFP will be less efficiently biotinylated by the BirA-TAP54α, as compared to BAP-HP1γ as a target. Indeed, despite the comparable total levels of BAP-GFP and BAP-HP1γ in the cell, (as α-6XHis Western shows, Figure S2A, top, lane 2, Supporting Information), BAP-HP1γ was consistently biotinylated with higher efficiency than BAP-GFP, when these two proteins were coexpressed together with BirA-TAP54α (Figure S2A, bottom, lane 2, Supporting Information). In control experiment with BirA-GFP fusion, BAP-GFP was biotinylated with higher efficiency than BAP-HP1γ (Figure S2A, bottom, lane 1, Supporting Information), consistent with the notion that the fraction of BirA-GFP excluded from the nucleus was not available for the BAP-HP1γ biotinylation, but could biotinylate the respective BAP-GFP fraction.

3. MS/MS Detection and Multiplexing

We next tested whether the biotinylated and nonbiotinylated forms of the BAP peptide could be detected by mass spectrometry instead of Western blot. We used propionylation to protect the nonbiotinylated BAP peptide from tryptic cleavage on the target lysine. This method has been widely used elsewhere—for example, in the analysis of histone modifications.[33,34] Such an approach allows one to obtain modified and nonmodified peptides of comparable sizes, facilitating the interpretation of results.

As seen on the Figure 3A, we can detect both the biotinylated and the propionylated (i.e., nonbiotinylated) versions of BAP by LC−MS/MS, as verified from the comparison between the detected mass spectra and the theoretically predicted spectra. Thus, as expected, in addition to the Western analysis, LC−MS/MS can also be used to monitor the biotinylation of the target proteins.

In the context of our approach, an important advantage of mass spectrometry is the ability to distinguish between versions of the same peptide that differ in amino acid sequence. One can use this feature by slightly modifying the sequence of the BAP peptide, in order to design a "multiplexing" approach that would allow one to test and compare the biotinylation levels of different proteins in the same experiment. To test this possibility, we generated two new versions of the BAP peptide, replacing Val with Phe or Tyr, which have different masses (Figure 3B). The ORF of H2AZ was cloned into these vectors, and the resulting plasmids were cotransfected with BirA-H2A in 293T cells.

dx.doi.org/10.1021/pr200189p |J. Proteome Res. XXXX, XXX, 000–000

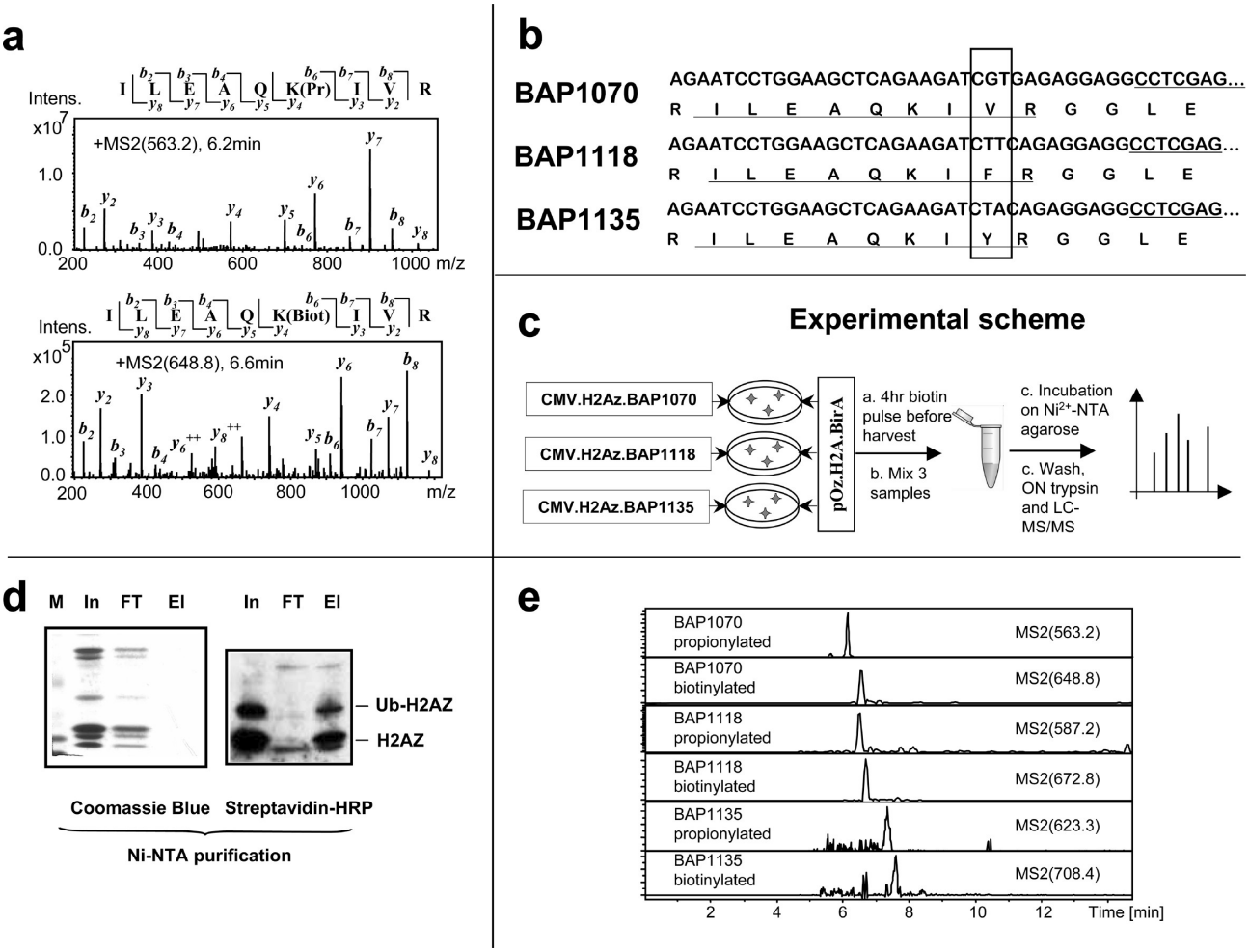


Figure 3. Multiplexing. (A) MS/MS detection of BAP. Shown are the MS/MS fragmentation spectra of the propionylated (top) and biotinylated (bottom) forms of BAP, with the detected y-series and b-series ions indicated. (B) New BAPs. Shown are the nucleotide and amino acid sequences of the new BAPs (BAP1118 and BAP1135). The site of the amino acid change, as compared to the original BAP1070, is indicated by a box. (C) Scheme of the multiplexing experiment. Cells were transfected with BirA.H2A and with a vector expressing different BAP fusions of H2AZ. After 4 h labeling, they were harvested, mixed, and the BAP-H2AZ proteins purified on Ni^{2+}-NTA agarose. (D) Ni^{2+}-NTA purification of the BAP-H2AZ fusions. M, Marker; In, input; FT, flowthrough; E, eluate. Left, Coomassie Brilliant Blue stain; right, streptavidin-HRP Western. The positions of BAP-H2AZ and the ubiquitinated BAP-H2AZ are indicated. (E) MRM detection of different BAPs. Shown are extracted ions chromatograms for the most intensive fragmentation ions present in the MS/MS spectra of the respective peptides. Note that the BAP1135 is propionylated on both K and Y, which affects the M/Z of the respective ions.

Afterward, the transfected cells were mixed, and the 6XHis-tagged histones were purified from the guanidine HCl lysate of the cells by immobilization on Ni-NTA beads. The proteins were digested by trypsin directly on the beads, after which the MRM analysis was used to detect, in the same sample, the biotinylated and propionylated forms of all three peptides (Figure 3C).

An aliquot of the Ni-NTA beads was used to elute the three BAP-H2AZ fusions. As can be seen from the comparison between the LC-MS/MS (Figure 3F) and Western (Figure 3D) analyses, whereas Western blotting fails to distinguish between the different biotinylated proteins because they contribute to the same signal on the autoradiograph, all biotinylated and non-biotinylated forms of the corresponding BAP peptides can be detected separately and distinguished from each other by mass spectrometry (see Supplementary Figures S3 and S4 for the MS/MS spectra of the corresponding ions, Supporting Information). Thus, it should be possible to use this approach to simultaneously detect and compare the interactions of different proteins (e.g., different point mutation variants of the same protein), carrying different versions of BAP, with a single BirA fused protein in the same cells.

Alterations in amino acid sequence can change affinity of a BAP to BirA and thus can affect their biotinylation efficiency regardless of protein−protein proximity. Accordingly, we compared biotinylation levels of different BAPs in the context of the same protein pair "BAP-H2AZ + BirA-GFP". As one can see from the Figure S5 (Supporting Information), the biotinylation efficiencies do vary for different BAP peptides. The resulting problems with interpretation could be controlled by swapping different BAPs between the tested proteins in reciprocal experiments.

4. Stable Isotopes

An alternative approach to multiplexing, free of the problem of variation in biotinylation efficiency between peptides with different sequence, is the use of stable isotopes, a strategy common in quantitative proteomics.[35,36] To test if the stable isotope

dx.doi.org/10.1021/pr200189p | J. Proteome Res. XXXX, XXX, 000–000

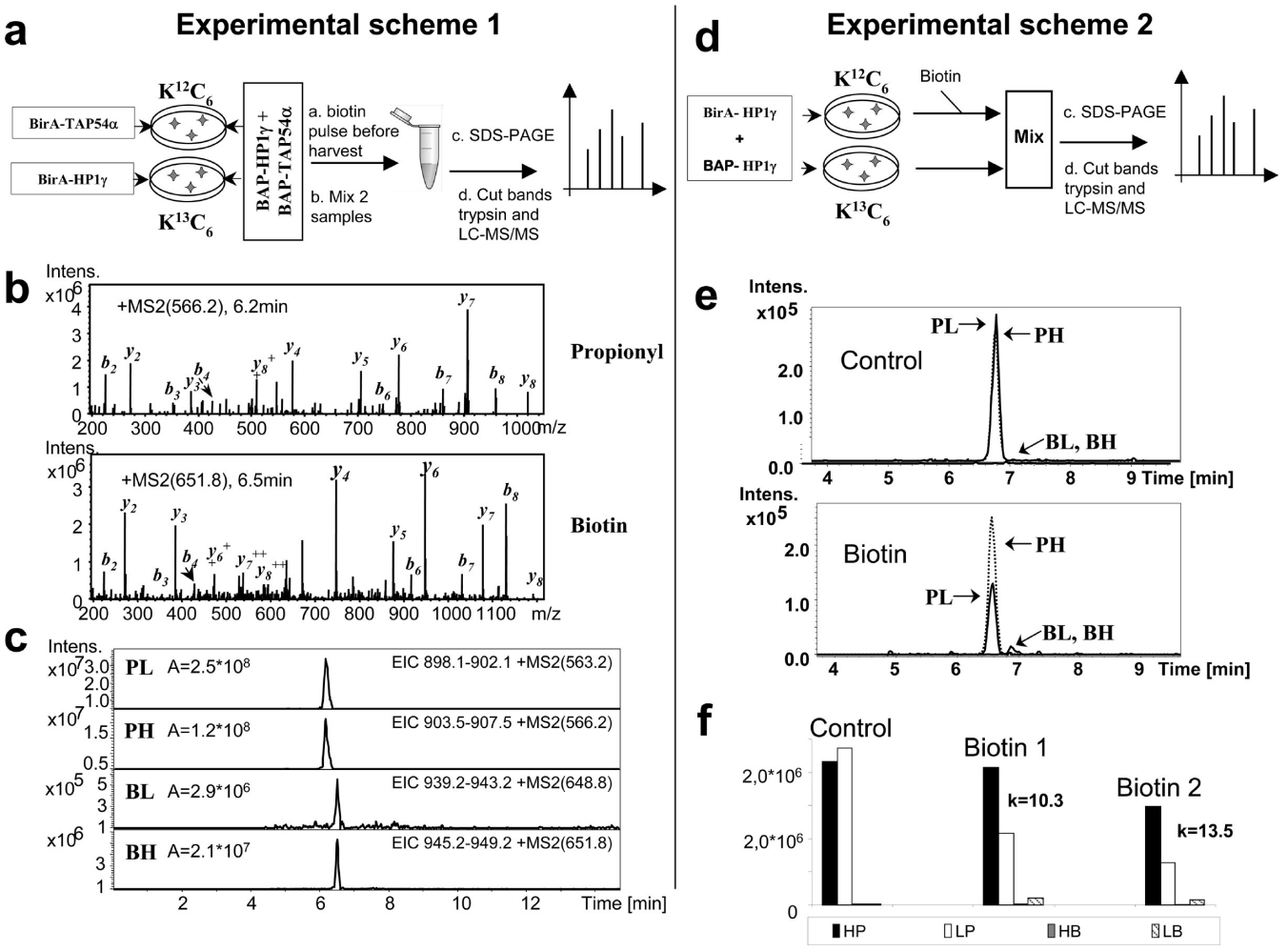


Figure 4. Stable isotopes. (A) Experimental scheme 1. Cells were grown in medium containing $K^{12}C_6$ or $K^{13}C_6$ and transfected, respectively, with BirA-TAP54α or with BirA-HP1γ, along with a mixture of BAP-TAP54α and BAP-HP1γ. After 4 h labeling, the cells were harvested and mixed, the proteins from the nuclei were fractionated on SDS-PAGE, and the gel bands containing the HP1γ and TAP54α were processed for MRM analysis on LC−MS/MS. (B) MS/MS spectra of heavy BAP1070. Shown are the MS/MS fragmentation spectra of the propionylated (top) and biotinylated (bottom) forms of $K^{13}C_6$-labeled BAP1070, with the detected y-series and b-series ions indicated. (C) MRM analysis of a SILAC experiment. Shown are extracted ions chromatograms (EIC) for the most intensive fragmentation ions present in the MS/MS spectra of the BAP1070. PL, propionylated light; PH, propionylated heavy; BL, biotinylated light; BH, biotinylated heavy. The areas of every peak (A) are indicated on the left of the corresponding chromatogram. PL and BL correspond to transfection with the BirA-TAP54α fusion, whereas PH and BH correspond to transfection with the BirA-HP1γ fusion; accordingly, the BAP1070 peptide is 17 times better biotinylated in the BirA-HP1γ−fusion-containing sample. (D) Experimental scheme 2. Cells were grown in medium with $K^{12}C_6$ or $K^{13}C_6$ and transfected with BirA-HP1γ and BAP-HP1γ. The $K^{12}C_6$ grown cells were labeled overnight with biotin before being harvested and mixed, in equal amounts, with the $K^{13}C_6$ cells grown in absence of biotin. (E) MRM quantification of different forms of BAP1070. Shown are extracted ion chromatograms for the different forms of BAP1070 from the above experiment. The positions of the peaks (labeled as in the legend to Figure 4c) are indicated by arrows. Control, the $K^{12}C_6$ grown cells were not labeled with biotin. Biotin, the $K^{12}C_6$ grown cells were labeled with biotin. (F) Determination of the coefficient of relative ionization of the propionylated and biotinylated forms of BAP1070. Shown are the histograms representing the areas of the peaks from the control experiment (as in E) and two independent biotinylation experiments (Biotin 1 and Biotin 2), and the relative ionization coefficient k between the biotinylated and propionylated BAP peptides for each experiment, estimated as $k = (HP - LP)/(LB - HB)$, where the areas of the peaks are labeled as in the legend to C.

labeling scheme is also applicable to our system, we repeated the experiment with TAP54α and HP1γ (Figure 4A). This time, however, the BirA-TAP54α−containing combination was grown in medium containing $^{12}C_6$ lysine, whereas the BirA-HP1γ was grown in medium containing $^{13}C_6$ lysine. After labeling with biotin, the cells were harvested and the two suspensions combined. All further procedures, including SDS-PAGE separation, gel band excision and trypsin digestion were performed with the mixture. The degree of BAP-HP1γ biotinylation by BirA-HP1γ and BirA-TAP54α was estimated by calculating the ratios between the biotinylated versus propionylated versions of the heavy (BirA-HP1γ) or light (BirA-TAP54α) peptides (See Figure 4B for the MS/MS data). As seen on the Figure 4C, a significant difference (about 17-fold) in biotinylation of BAP-HP1γ by the two BirA fusions was observed; the reciprocal difference was seen for the BAP-TAP54α (not shown).

The quantitative measurement of biotinylation should benefit from determining the molar amounts of biotin label on the BAP peptide, which is challenging to estimate with Western analysis, as the signal intensity can depend dramatically on the antibody

dx.doi.org/10.1021/pr200189p | J. Proteome Res. XXXX, XXX, 000–000

used. A comparison between the TIC signal of the biotinylated and propionylated BAP in LC−MS/MS also cannot be used directly, as the ionization efficiency generally depends on chemical structure and will thus be affected by the change from a propionyl to a biotin residue. The use of stable isotopes helped us to estimate relative ionization intensities between these two forms of BAP. Cells grown on $^{12}C_6$ lysine were cotransfected with BAP-HP1γ and BirA-HP1γ and then labeled overnight with biotin. In parallel, cells were grown on $^{13}C_6$ lysine medium and transfected with the same plasmids, but not labeled with biotin. The labeled cells were harvested and mixed, in equal amounts, with the corresponding sample of the heavy-isotope−grown cells (Figure 4D). The BAP-HP1γ protein, prepared from the cell mixture by Ni^{2+}-NTA pulldown and SDS-PAGE separation, was propionylated, digested by trypsin, and analyzed by LC−MS/MS (Figure 4E). The relative ionization coefficient k between the biotinylated and propionylated BAP peptides was estimated as $k = (HP - LP)/(LB - HB)$, where HP corresponds to the area of TIC of heavy propionylated, LP of light propionylated, LB of light biotinylated, and HB of heavy biotinylated BAP. K were calculated for two independent experiments, giving an average value 11.9 ± 1.6 (Figure 4F).

5. Monitoring the Protein at a Defined Time after the Interaction Has Occurred

The distinctive feature of our method is in leaving one of the tested proteins (BAP-fusion) with a permanent molecular mark, which can persist after the proximity between the two proteins has been lost. One can take advantage of this property to add kinetic dimension to the study of protein−protein proximity.

As one such possibility, we tested if a pulse-chase labeling with biotin can be used to trace the fate of a BAP-fused protein of interest that was interacting at the moment of labeling with its BirA-fused proximity partner. This approach is particularly interesting if one wishes to study multistep intracellular processes, where the interaction partners and protein properties (such as its localization and post-translational modifications) change. Thus, in addition to purifying and studying the properties of a protein of interest at the moment of its interaction/proximity with another protein, our method will enable one also to study the same cellular fraction of the protein at defined times after the interaction has occurred.

As a "proof of principle" model to explore this opportunity, we used the Rad18 protein, known to form characteristic foci in cell nuclei (these foci are believed to correspond to specialized compartments containing replication proteins, including PCNA). MRC fibroblasts were cotransfected with the BirA-Rad18 fusion together with BAP-H2A, a histone that is homogenously distributed over the nucleus, but, in our system, is expected to be biotinylated only in the neighborhood of the BirA-Rad18 containing foci. Two days after transfection, the cells were pulse-labeled with biotin (Figure 5A). The cells were either fixed immediately after the biotin pulse (pulse sample), or else intensively washed to eliminate biotin and then left in fresh medium for 2 h before fixation (chase sample). As demonstrated by Western analysis (Figure 5B), only the BAP-H2A histone is biotinylated under these conditions, hence immunofluorescent microscopy detection of the biotin signal should reveal a localization of the BirA-Rad18 in proximity to BAP-H2A at the moment of pulse-labeling.

Consistent with such expectation, staining of the pulse-labeled cells with streptavidin-Cy3 revealed foci of labeled BAP-H2A which colocalized with Rad18 (Figure 5C top) as well as with PCNA (Supplementary Figure S6, Supporting Information), detected by specific antibodies. However, the analysis of the chase sample revealed a different pattern of staining (Figure 5C middle, and Supplementary Figure S6, Supporting Information). Whereas both Rad18 and PCNA formed foci, the biotin signal appeared more diffuse, and we generally observed a smaller number of foci per cell, whereas the colocalization of the remaining biotinylated foci with the PCNA or Rad18 foci was considerably less pronounced (see Figure 5C bottom for an example).

We conclude that, as was expected, BAP-H2A was biotinylated in the proximity of BirA-Rad18 at the moment of the biotin pulse labeling, and that, with time, the Rad18 partially changed its location inside the nucleus and started to associate with different parts of chromatin, as could be seen by the loss of colocalization of the Rad18 (and PCNA) foci with the biotinylated histones. Regardless of its precise interpretation, this observation clearly demonstrates the possibility to monitor the protein of interest (in this case, BAP-H2A), which was proximity-biotinylated by the BirA fusion of another protein, at later times, that is, after the proximity between these two proteins could have been lost.

DISCUSSION

FRET/BRET, PCA and two-hybrid methods—the predominant approaches for studying protein−protein interactions *in vivo*—frequently lie outside the sphere of expertise of a typical proteomics laboratory. The development of more mass-spectrometry and biochemistry-oriented protocols would be a welcome addition to the application portfolio of a standard mass-spectrometry facility.

Also, the FRET/BRET and PCA data are usually interpreted in terms of protein−protein interactions. However, as far as *in vivo* data are concerned, the interpretation in terms of *proximity* between two proteins appears to be more adequate than that of "physical interaction". First, unlike the *in vitro* binding studies, where the environment (ion composition, cofactors, etc) can be controlled at will, the intracellular microenvironment is far more challenging to control (to say the least), and the contribution of local intracellular context in stabilizing an interaction can never be ruled out. Second, two proteins could interact via an intermediate, for example, be near each other in cell due to their binding to two closely located sites on DNA. Finally, the role of dynamic self-organization[1,2] in bringing molecules together, without their physical interaction, also cannot be ruled out. Thus, in the *in vivo* context, the notion of protein−protein proximity (and, accordingly, "PPP-networks"), as opposed to protein−protein interactions (PPI-networks), appears to be a more adequate term—as, in addition to the direct physical interactions, it can also cover interactions via intermediates, as well as the aspects of intracellular order arisen via dynamic self-organization. Regardless of physical interactions, any information about the protein proximity *in vivo* has a value in itself, as it could be useful in the ultimate task of reconstructing the spatial order in the cell.

In these terms, one shortcoming of the FRET/BRET, PCA and two hybrid approaches is that they measure short-range proximity only. The methodologies that could probe the protein proximities on larger scales could serve as a complement to these established methods, providing additional information about intracellular order. Given that it is sensitive to intracellular compartmentalization, the scope of proximity utilizing biotinylation approach is apparently not limited to the short-range proximities between proteins. We have to admit, however, that

dx.doi.org/10.1021/pr200189p |J. Proteome Res. XXXX, XXX, 000–000

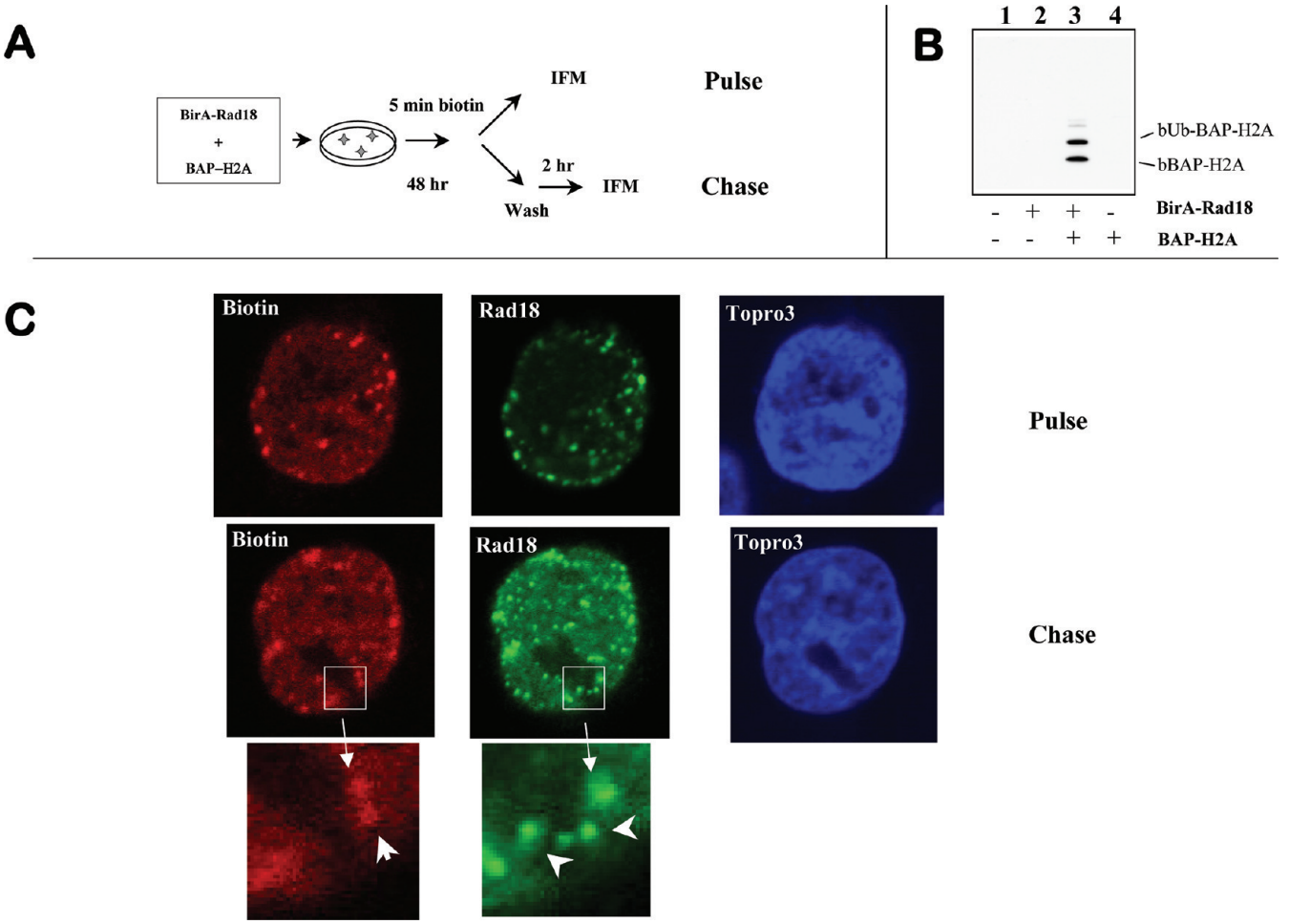


Figure 5. Pulse-chase experiment. (A) Experimental scheme. MRC fibroblasts are cotransfected by the BirA-Rad18 fusion together with BAP-H2A. Two days after transfection, the cells are pulse-labeled with biotin 6 h later. The cells are either fixed immediately after the biotin pulse (pulse sample), or else washed extensively to remove biotin, then left for 2 h before fixation (chase sample). (B) Specificity of biotinylation. Western blot analysis with streptavidin-HRP, showing that a specific signal appears only after BirA-Rad18 and BAP-H2A are coexpressed. One, control untransfected sample; 2, BirA-Rad18; 3, BirA-Rad18 + BAP-H2A; 4, BAP-H2A. The two forms of H2A (nonubiquitinated and ubiquitinated), both biotinylated in this experiment, are indicated by asterisks. (C) Colocalization analysis. Left Top, pulse sample. Middle, chase sample. Bottom, zoomed area from the chase sample showing an example of biotinylated foci that do not colocalize with the Rad18 foci and vice versa (indicated by asterisk).

the exact distance range in which the proximity can still be detected could depend on the pair of proteins studied and thus merits a further study. In particular, the distance range can depend on the mutual orientation and the mobility of proximity partners. As an illustration (Figure S7, Supporting Information), first consider two proteins (A_1 and A_2) that are close to each other, but their movement is strongly constrained (e.g., by being tethered to a rigid scaffold). In this case, the BirA and BAP moieties have a low probability to directly encounter each other. In the second case, the average distance between proteins (B_1 and B_2) is larger, but they move more freely. This will lead to higher chance of biotinylation, which, as a permanent modification, will accumulate on the BAP-tagged protein. This illustration, together with the effects of compartmentalization, also suggests that due to the many factors that can contribute to the differences in observed biotinylation levels, the correct interpretation of data requires careful choice of proper controls. For example, when testing proximity between nuclear proteins, one cannot use BAP-GFP as a negative control, as its extranuclear fraction will be inaccessible for BirA fused to a nuclear protein and thus contribute to an artificially low background biotinylation levels (and any pair of nuclear proteins will give a higher biotinylation signal). The proper negative control should be BAP-GFP with a nuclear localization signal. On the other hand, protein−protein interactions or proximities are often studied not for their own sake, but rather to access the network of causal relations in the cell. From this perspective, the increased ability of two PUB proteins to biotinylate one another (above the rate expected from the modeling of biotinylation rates based on random protein−protein collisions in a homogeneous solution) has direct relevance, regardless of the exact interpretation of what had caused the biotinylation increase.

To demonstrate the utility of our approach, we have chosen several examples of protein pairs that represent different instances of protein proximity. (a) Many proteins can form homooligomers, and we demonstrated, using the examples of TAP54α and HP1γ self-association, that such interactions can be detected with our method. (b) To show that this methodology is not limited to protein self-association, we also demonstrated that it can detect interaction between two different proteins (e.g., KAP1 and HP1γ).

dx.doi.org/10.1021/pr200189p |J. Proteome Res. XXXX, XXX, 000–000

(c) Two proteins do not have to interact directly, but can be in proximity in common locus (e.g., PCNA and histone H3.1, or the alternative histones H2ABB and macroH2A, enriched in active chromatin and repressed chromatin in the nucleus, respectively). (d) Finally, sharing common intracellular compartment will increase the effective concentration of the BirA-fusion near the BAP-target and promote the reaction, as we demonstrated using GFP and HP1 as examples of compartmentalization differences.

Another type of additional information that could be provided by PUB is based on the fact that biotinylation leaves one of the proteins with a permanent modification, which persist after the protein−protein proximity could have been lost. We presented a way to take advantage of this feature that adds kinetic dimension to the studies of the protein−protein proximity. Using BAP-fused histone H2A and BirA-Rad18 as a model, we demonstrated that one can use a pulse-chase labeling with biotin to trace the fate of a BAP-fused protein of interest that was interacting with a given BirA-fused protein at the moment of labeling. Among the many potential applications of such type of analysis are the study of protein transport (e.g., by using a fusion of BirA with a component of the membrane pore complex and labeling a protein transported to the nucleus by BAP), measurement of the stability and consequent modifications of a particular fraction of protein that has been a part of a complex with a specific protein partner, etc. Thus, in addition to helping to reconstruct the cell structure in space, our approach has a potential to add temporal dimension to such reconstruction studies.

We also need to mention a difference of PUB approach from the other biochemistry-friendly approaches to study protein−protein interactions, such as coimmunoprecipitation and TAP (tandem-affinity purification). Our method is intended for *monitoring* interactions/proximities between two given proteins, not for *identification* of new interaction partners of a given protein. One can easily envision many instances when such a task might be needed. A case in point is structure−functional analysis of a protein (deletion and/or site-specific mutagenesis) with the aim to map interaction domains (as we illustrated on the example of the KAP1 wild type and mutant interacting with HP1). This limitation notwithstanding, our method has the following appealing difference from TAP and related approaches to study protein−protein interactions. These methods work satisfactorily only for the sufficiently strong interactions that can withstand the procedures of solubilization and affinity purification. For example, the study of membrane complexes or the interactions between matrix proteins is challenging with these methods, given the severe limitations on the extraction and purification conditions compatible with the stable protein−protein interactions. On the contrary, in the proposed approach, the generation of a permanent covalent mark on one of the proteins studied will allow one to bypass the limitations imposed by the extraction and purification procedures. Thus, the method should prove useful for the study of interactions that are otherwise difficult to detect by the Co-IP and TAP methods.

In some cases, for example, when comparing interactions of two proteins between different cell lines with uncontrolled levels of their endogenous counterparts, a potential problem with interpretation could arise due to the competition of the endogenous proteins for the interaction. As seen in Figure 2B, we could efficiently decrease the biotinylation of BAP-KAP1 by overexpressing untagged KAP1. This problem should be taken into account when concluding from results obtained using this method. Notably, however, this valid concern equally applies to other methods to study protein−protein interactions in homologous systems in vivo, such as FRET, BRET and two hybrid (e.g., mammalian two hybrid used to study mammalian interaction partners).

As far as future development of this system is concerned, the vectors utilized in our work require "classic" cloning to be performed for each pair of proteins tested, which greatly increases the amount of work required per experiment, effectively preventing large-scale study. Making our vectors Gateway compatible will facilitate the molecular biology manipulations and benefit from large ORF collection compatible with this tagging system.

Finally, while this project was in progress, we learnt that a somewhat similar approach to detect protein−protein interactions *in vivo* has been developed by another group.[37] However, we wish to emphasize the crucial and distinguishing feature of our system: its mass-spectrometry-oriented format. In this respect, the sequence of the BAP domain used in ref 37 is not expected to yield a tryptic peptide that would be efficiently detected by mass spectrometry, thus precluding its use in techniques such as multiplexing with varying amino acid sequences or stable isotope labeling and quantification, which constitute one of the main advantages of the technique we describe here.

■ ASSOCIATED CONTENT

Supporting Information

Supplementary tables and figures. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: vogryzko@gmail.com. Phone: 33-1 42 11 65 25. Fax: 33-1 42 11 65 25.

■ ACKNOWLEDGMENT

We thank Dr. M. Lechner (Drexel University, Philadelphia) for the KAP1-HP1BDwt and KAP1-HP1BDwt plasmids, Drs H. Willard and M. Chadwick (Case Western Reserve University, Cleveland) for the ORFs of the H2A.BBD and macroH2A, and Dr. L. L. Pritchard for critical reading of the manuscript. We also thank Chloé Robin for the Figure 2C. This work was supported by grants from "La Ligue Contre le Cancer" (9ADO1217/1B1-BIOCE), the "Institut National du Cancer" (247343/1B1-BIOCE) and Centre National de la Recherche Scientifique [CNRS-INCA-MSHE Franco-Pologne #3037987) to VO, by NCB Kazakhstan (0103_00404) to AK and by a travel grant to A.K. from the Coopération Universitaire et Scientifique Department of the French Embassy in Astana, Kazakhstan.

■ ABBREVIATIONS:

BAP, biotin acceptor peptide; LC−MS/MS, liquid chromatography coupled with tandem mass-spectrometry; TIC, total ion current

■ REFERENCES

(1) Karsenti, E. Self-organization in cell biology: a brief history. *Nat. Rev. Mol. Cell. Biol.* **2008**, *9*, 255–262.

(2) Misteli, T. The concept of self-organization in cellular architecture. *J. Cell Biol.* **2001**, *155*, 181–185.

(3) Day, R. N.; Periasamy, A.; Schaufele, F. Fluorescence resonance energy transfer microscopy of localized protein interactions in the living cell nucleus. *Methods* **2001**, *25*, 4–18.

(4) Pollok, B. A.; Heim, R. Using GFP in FRET-based applications. *Trends Cell Biol.* **1999**, *9*, 57–60.

(5) Boute, N.; Jockers, R.; Issad, T. The use of resonance energy transfer in high-throughput screening: BRET versus FRET. *Trends Pharmacol. Sci.* **2002**, *23*, 351–354.

(6) Pfleger, K. D.; Eidne, K. A. Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET). *Nat. Methods* **2006**, *3*, 165–174.

(7) Remy, I.; Michnick, S. W. Clonal selection and in vivo quantitation of protein interactions with protein-fragment complementation assays. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 5394–5399.

(8) Remy, I; Michnick, S. W. Mapping biochemical networks with protein-fragment complementation assays. *Methods Mol. Biol.* **2004**, *261*, 411–426.

(9) Fields, S.; Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **1989**, *340*, 245–246.

(10) Lievens, S.; Lemmens, I.; Tavernier, J. Mammalian two-hybrids come of age. *Trends Biochem. Sci.* **2009**, *34*, 579–588.

(11) de Boer, E.; Rodriguez, P.; Bonte, E.; Krijgsveld, J.; Katsantoni, E.; Heck, A.; Grosveld, F.; Strouboulis, J. Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7480–7485.

(12) Viens, A.; Mechold, U.; Lehrmann, H.; Harel-Bellan, A.; Ogryzko, V. Use of protein biotinylation in vivo for chromatin immunoprecipitation. *Anal. Biochem.* **2004**, *325*, 68–76.

(13) Kim, J.; Cantor, A. B.; Orkin, S. H.; Wang, J. Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. *Nat. Protoc.* **2009**, *4*, 506–517.

(14) Rodriguez, P.; Braun, H.; Kolodziej, K. E.; de Boer, E.; Campbell, J.; Bonte, E.; Grosveld, F.; Philipsen, S.; Strouboulis, J. Isolation of transcription factor complexes by in vivo biotinylation tagging and direct binding to streptavidin beads. *Methods Mol. Biol.* **2006**, *338*, 305–323.

(15) Ooi, S. L.; Henikoff, J. G.; Henikoff, S. A native chromatin purification system for epigenomic profiling in Caenorhabditis elegans. *Nucleic Acids Res.* **2010**, *38* (4), e26.

(16) van Werven, F. J.; Timmers, H. T. The use of biotin tagging in Saccharomyces cerevisiae improves the sensitivity of chromatin immunoprecipitation. *Nucleic Acids Res.* **2006**, *34*, e33.

(17) Viens, A.; Harper, F.; Pichard, E.; Comisso, M.; Pierron, G.; Ogryzko, V. Use of protein biotinylation in vivo for immunoelectron microscopic localization of a specific protein isoform. *J. Histochem. Cytochem.* **2008**, *56*, 911–919.

(18) Viens, A.; Mechold, U.; Brouillard, F.; Gilbert, C.; Leclerc, P.; Ogryzko, V. Analysis of human histone H2AZ deposition in vivo argues against its direct role in epigenetic templating mechanisms. *Mol. Cell. Biol.* **2006**, *26*, 5325–5335.

(19) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **1996**, *68*, 850–858.

(20) Chalfie, M.; Kain, S. R. *Green Fluorescent Protein: Properties, Applications, and Protocols*; Wiley-Liss: New York, 1998.

(21) Ikura, T.; Ogryzko, V. V.; Grigoriev, M.; Groisman, R.; Wang, J.; Horikoshi, M.; Scully, R.; Qin, J.; Nakatani, Y. Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. *Cell* **2000**, *102*, 463–473.

(22) Puri, T.; Wendler, P.; Sigala, B.; Saibil, H.; Tsaneva, I. R. Dodecameric structure and ATPase activity of the human TIP48/TIP49 complex. *J. Mol. Biol.* **2007**, *366*, 179–192.

(23) Eissenberg, J. C.; Elgin, S. C. The HP1 protein family: getting a grip on chromatin. *Curr. Opin. Genet. Dev.* **2000**, *10*, 204–210.

(24) Lomberk, G.; Wallrath, L.; Urrutia, R. The Heterochromatin Protein 1 family. *Genome Biol.* **2006**, *7*, 228.

(25) Lechner, M. S.; Begg, G. E.; Speicher, D. W.; Rauscher, F. J., 3rd Molecular determinants for targeting heterochromatin protein 1-mediated gene silencing: direct chromoshadow domain-KAP-1 corepressor interaction is essential. *Mol. Cell. Biol.* **2000**, *20*, 6449–6465.

(26) Ryan, R. F.; Schultz, D. C.; Ayyanathan, K; Singh, P. B.; Friedman, J. R.; Fredericks, W. J.; Rauscher, F. J., 3rd KAP-1 corepressor protein interacts and colocalizes with heterochromatic and euchromatic HP1 proteins: a potential role for Kruppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing. *Mol. Cell. Biol.* **1999**, *19*, 4366–4378.

(27) Shibahara, K.; Stillman, B. Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. *Cell* **1999**, *96*, 575–585.

(28) Mello, J. A.; Almouzni, G. The ins and outs of nucleosome assembly. *Curr. Opin. Genet. De.* **2001**, *11*, 136–141.

(29) Furuyama, T.; Dalal, Y.; Henikoff, S. Chaperone-mediated assembly of centromeric chromatin in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 6172–6177.

(30) Chadwick, B. P.; Willard, H. F. A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. *J. Cell Biol.* **2001**, *152*, 375–384.

(31) Chadwick, B. P.; Valley, C. M.; Willard, H. F. Histone variant macroH2A contains two distinct macrochromatin domains capable of directing macroH2A to the inactive X chromosome. *Nucleic Acids Res.* **2001**, *29*, 2699–2705.

(32) Chadwick, B. P.; Willard, H. F. Histone H2A variants and the inactive X chromosome: identification of a second macroH2A variant. *Hum. Mol. Genet.* **2001**, *10*, 1101–1113.

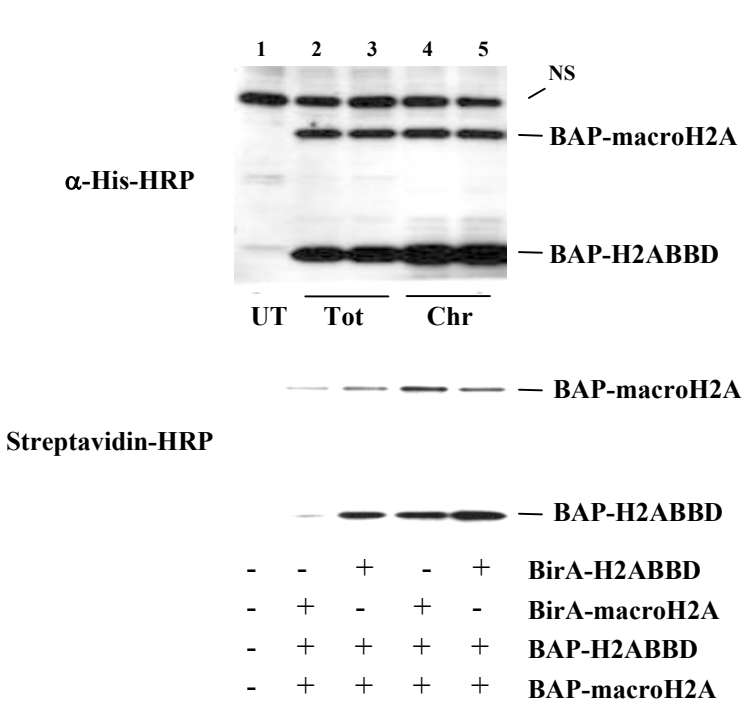
(33) Peters, A. H.; Kubicek, S.; Mechtler, K.; O'Sullivan, R. J.; Derijck, A. A.; Perez-Burgos, L.; Kohlmaier, A.; Opravil, S.; Tachibana, M.; Shinkai, Y.; et al. Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol. Cell* **2003**, *12*, 1577–1589.

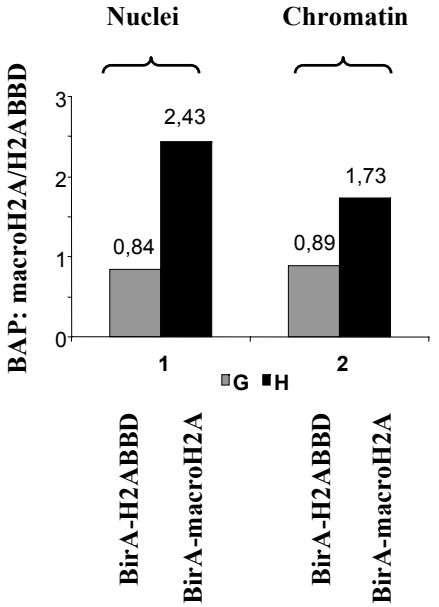
(34) Villar-Garea, A.; Imhof, A. The analysis of histone modifications. *Biochim. Biophys. Acta* **2006**, *1764*, 1932–1939.

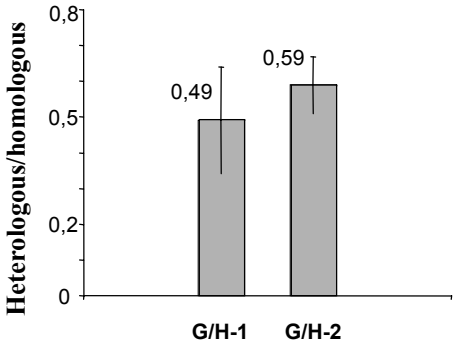
(35) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999**, *17*, 994–999.

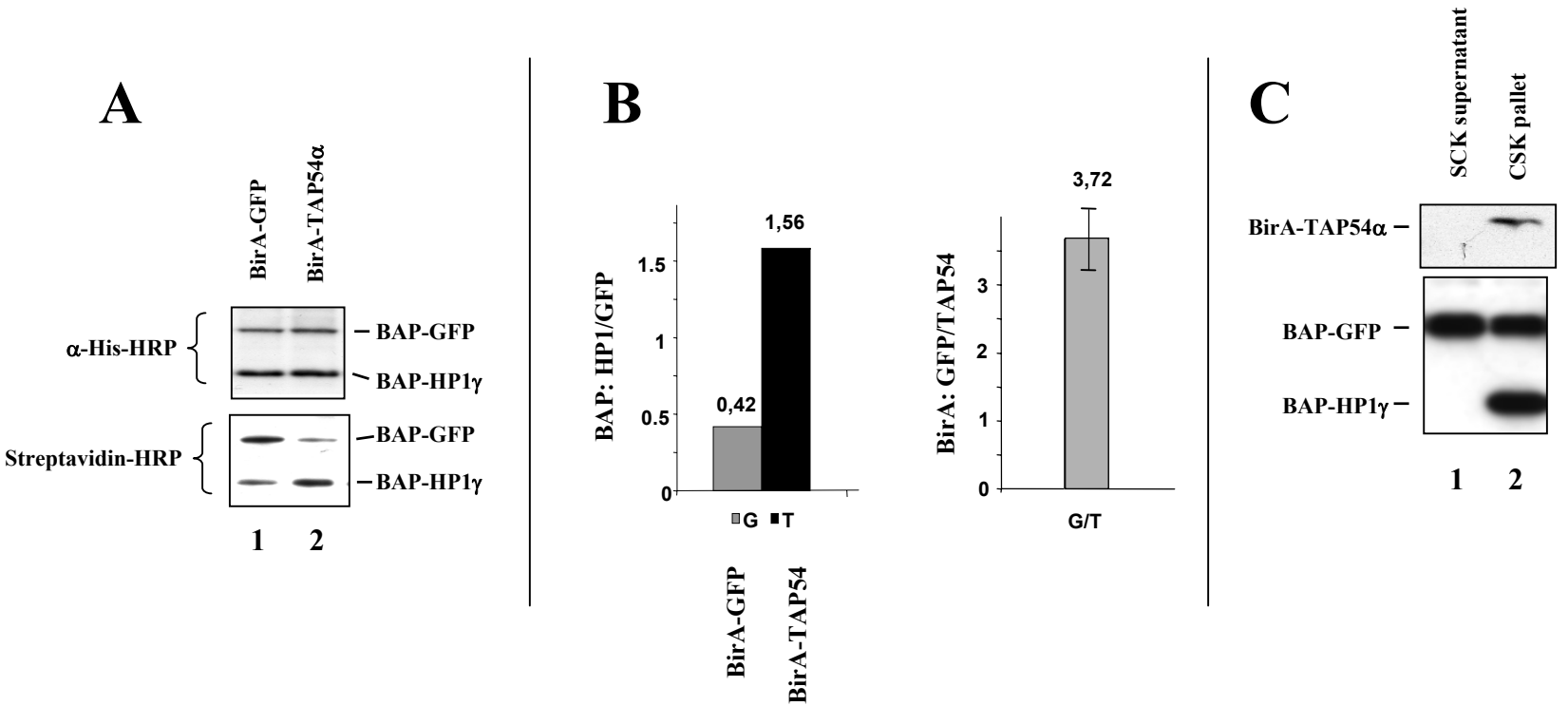
(36) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002**, *1*, 376–386.

(37) Fernandez-Suarez, M.; Chen, T. S.; Ting, A. Y. Protein-protein interaction detection in vitro and in cells by proximity biotinylation. *J. Am. Chem. Soc.* **2008**, *130*, 9251–9253.

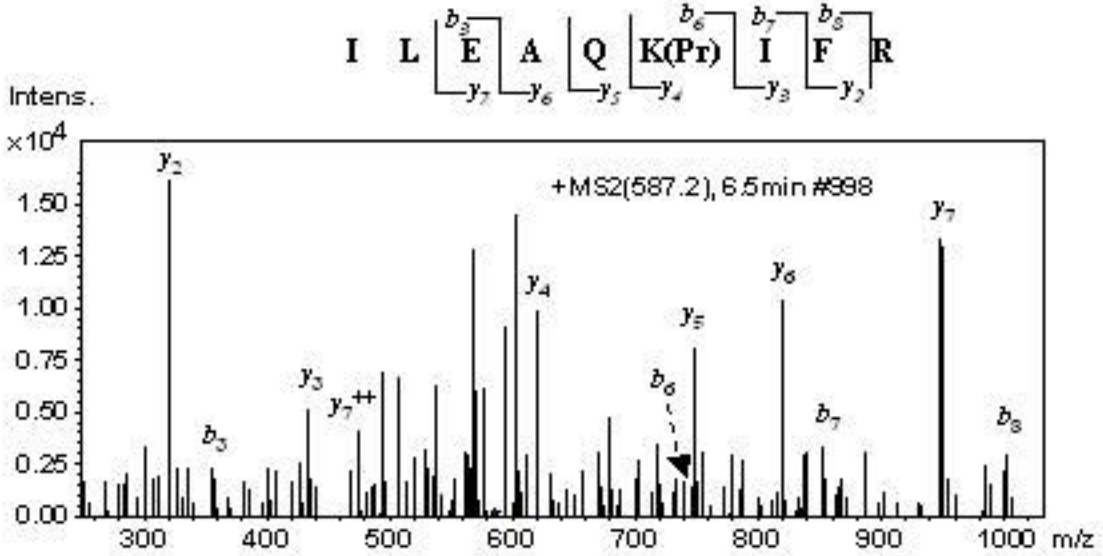

1 2 3 4 5
NS
BAP-macroH2A
α-His-HRP
BAP-H2ABBD
UT Tot Chr
BAP-macroH2A
Streptavidin-HRP
BAP-H2ABBD
- - + - + BirA-H2ABBD
- + - + - BirA-macroH2A
- + + + + BAP-H2ABBD
- + + + + BAP-macroH2A

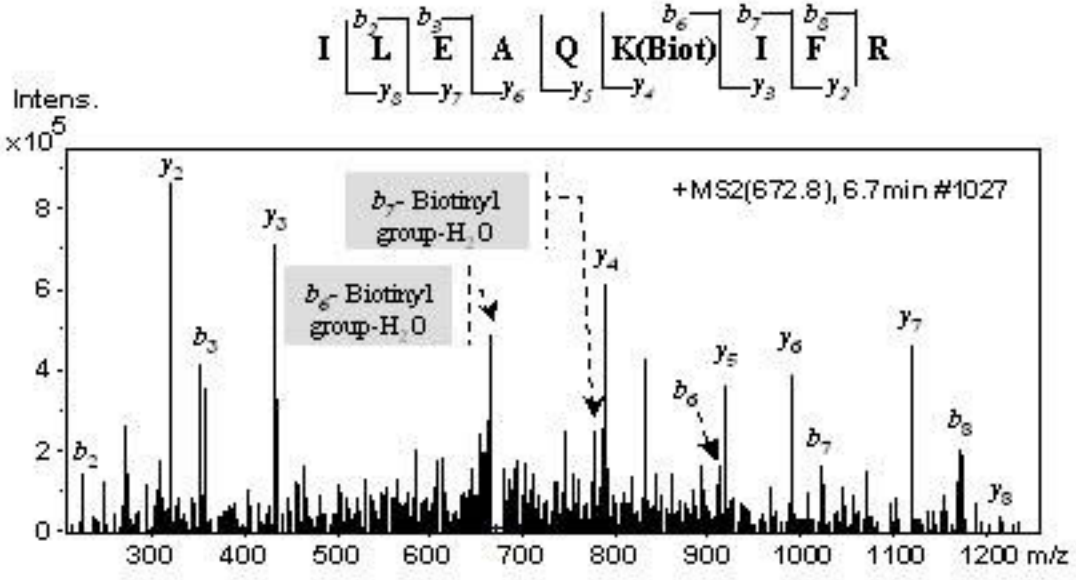

Nuclei
Chromatin
BAP: macroH2A/H2ABBD
3
2
1
0
2,43
1,73
0,84
0,89
1
2
G H
BirA-H2ABBD
BirA-macroH2A
BirA-H2ABBD
BirA-macroH2A


Heterologous/homologous
0,8
0,5
0,2
0
0,49
0,59
G/H-1
G/H-2

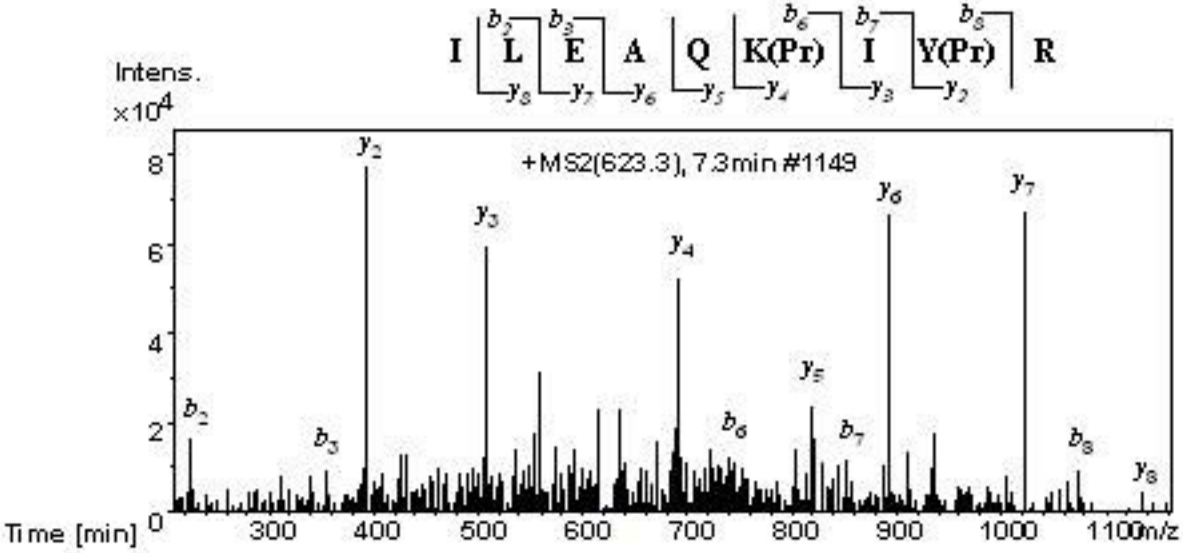

A
BirA-GFP
BirA-TAP54α
α-His-HRP
BAP-GFP
BAP-HP1γ
Streptavidin-HRP
BAP-GFP
BAP-HP1γ
1
2
B
BAP: HP1/GFP
1,56
0,42
1.5
1
0.5
0
G
T
BirA-GFP
BirA-TAP54
BirA: GFP/TAP54
3,72
3
2
1
0
G/T
C
SCK supernatant
CSK pallet
BirA-TAP54α
BAP-GFP
BAP-HP1γ
1
2

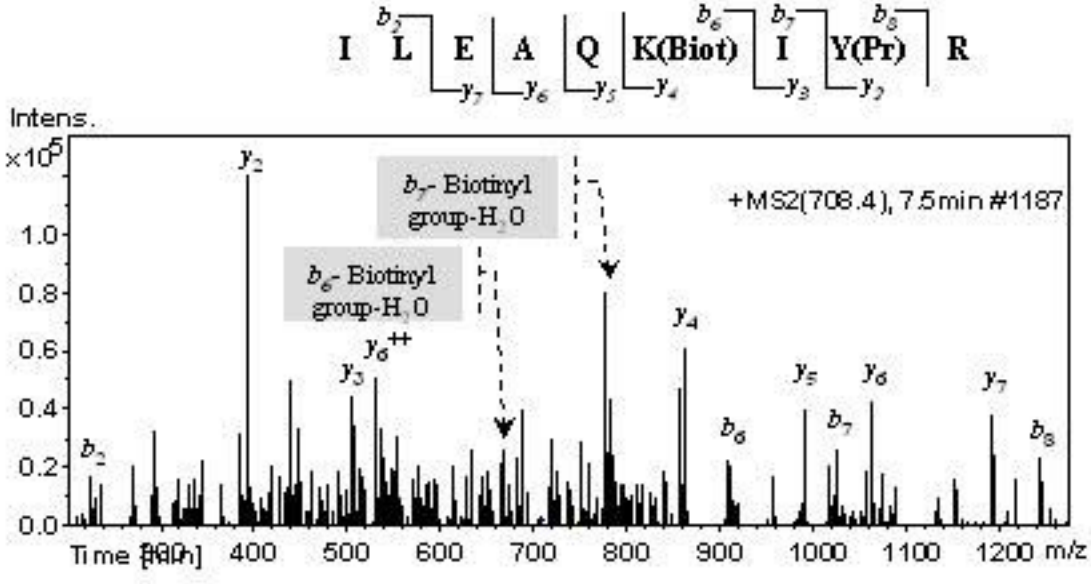
MS/MS OF BAP1118 (With V→F replacement) IN MULTIPLEX EXPERIMENT



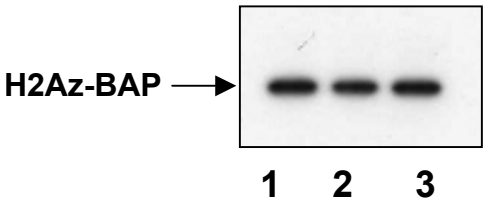


MS/MS OF BAP1135 (With V→Y replacement) IN MULTIPLEX EXPERIMENT





Penta-His-HRP



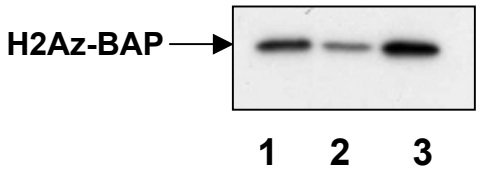
	Area
1	30578.39
2	26129.97
3	30200.78

Intensities ratios for normalisation

1/2=1.17

3/2=1.16

Streptavidin-HRP



	Area
1	22526.27
2	10888.05
3	31500.92

Normalized intensities on Streptavidin and corrected ratios

	Area
1	19253.20
2	10888.05
3	27155.96

1/2 = 1.77÷0.12

3/2 = 2.49÷0.31

1 – BAP1118 (RILEAQKIFRGGLE)
2 – BAP1070 (RILEAQKIVRGGLE)
3 – BAP1135 (RILEAQKIYRGGLE)

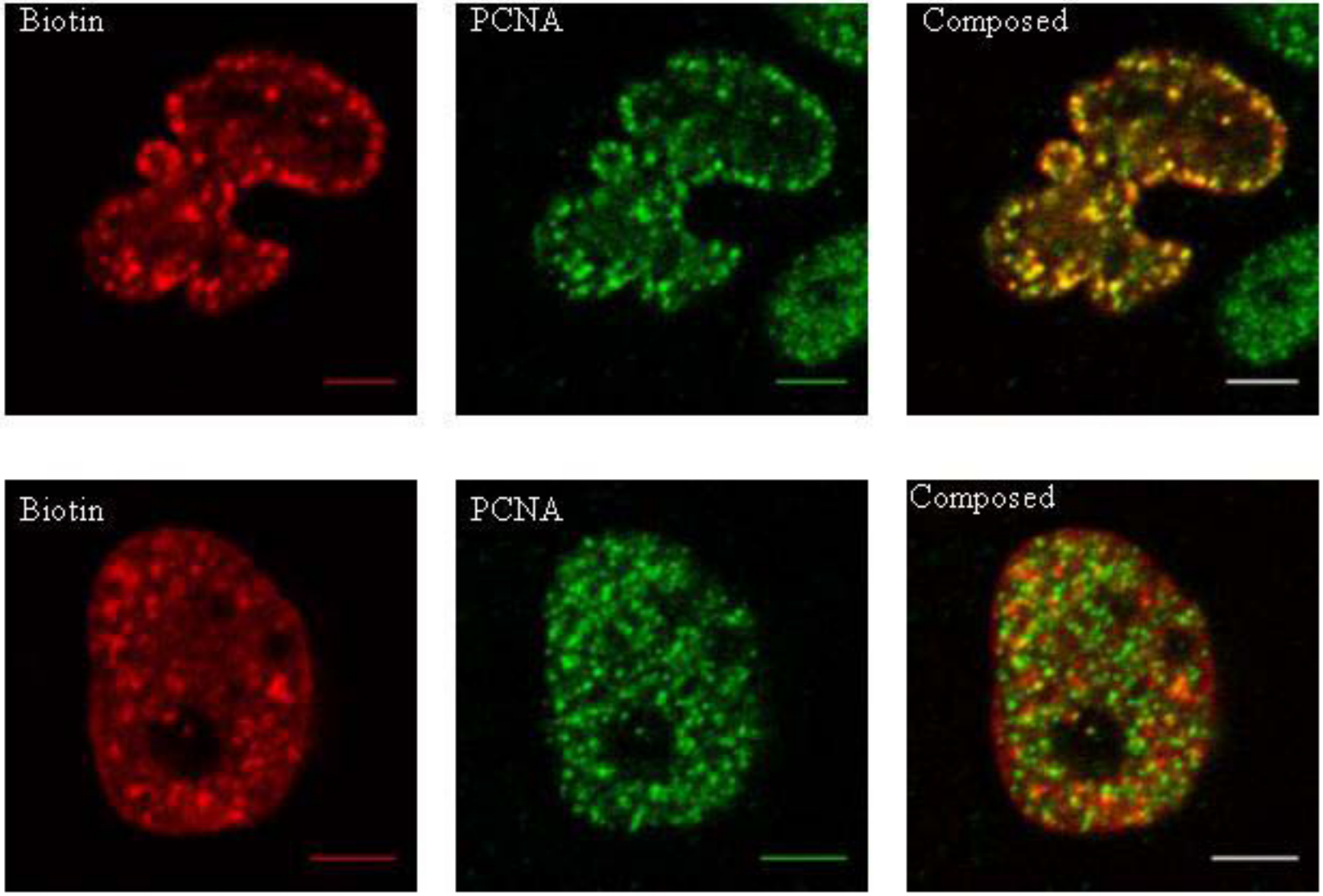
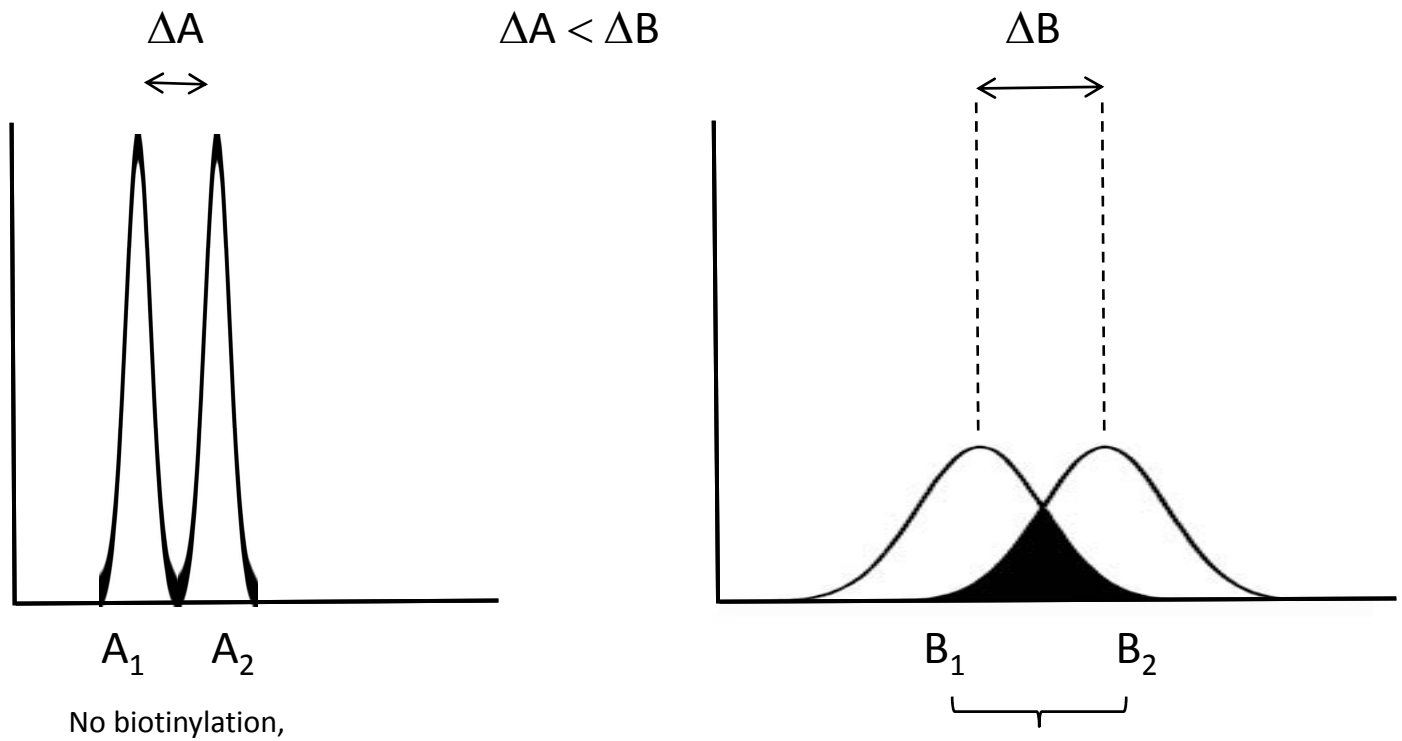


Figure S3. PCNA 6hr pulse and chase



No biotinylation,
due to the non-overlapping
distributions of A_1 and A_2

Biotinylation proportional
to the overlap between
the distributions of B_1 and B_2

Figure S1. macroH2A vs H2ABBD.

Left: 2 combinations of BirA and BAP fusions were transfected separately into cells: 1 – Untransfected cells; 2,4 – BirA-macroH2A + BAP-macroH2A + BAP-H2ABBD; 3,5 – BirA-H2ABBD + BAP-macroH2A + BAP-H2ABBD. 2,3 – total nuclear lysate (Tot), 4,5 – Chromatin associated histones (Chr). Top – α-His-HRP Western, Bottom – streptavidin-HRP Western. The positions of the BAP-fusions and nonspecific signal (NS) are indicated.
Right: Quantification of the biotinylation efficiencies. The signal intensities were first measured by densitometry, then the streptavidin signal for every BAP-fusion was normalized by dividing it to the α-His signal. G – the ratio between biotinylation of BAP-macroH2A and BAP-H2ABBD in the presence of BirA-H2ABBD; H – the ratio between biotinylation of BAP-macroH2A and BAP-H2ABBD in the presence of BirA-macroH2A. 1 – total nuclei, 2 – chromatin fraction. **Far right** – Ratio between the efficiencies of heterologous (H2ABBD vs macroH2A) and homologous (H2ABBD vs H2ABBD and macroH2A vs macroH2A) biotinylation. Shown are average values of G/H calculated for three experiments, for total nuclei (1) and chromatin fraction (2)

Figure S2. Detection of differences in compartmentalization

A. BAP-GFP and BAP-HP1γ where coexpressed either with BirA-GFP (left) or BirA-TAP54α (right) fusions. Top – α-His-HRP Western, Bottom – streptavidin-HRP Western. The positions of the BAP-fusions are indicated.

B. Quantification of biotinylation differences. The value of the streptavidin-HRP signal was normalized by taking a ratio with the value of α-His-HRP signal, which reflects the amount of the transfected protein regardless of its biotinylation status. Left: The ratios between the normalized signals of biotinylated BAP-HP1 and BAP-GFP was calculated for both BirA-GFP (G) and BirA-TAP54α (T) cotransfections. Right – the average ratio and standard deviation between the G and T values from 3 independent experiments.

C. Intracellular localization of the ectopically expressed HP1, GFP and TAP54. After cotransfection with BAP-GFP, BAP-HP1γ and BirA-TAP54α, the 293T cells were disrupted in CSK buffer, and nuclei pellets (right) were separated from supernatant (left). The volumes of both fractions corresponding to equal numbers of cells were loaded on SDS-PAGE gel, the proteins separated and the presence of BAP- and BirA fusions detected by α-His-HRP Western. Due to the lower expression levels of TAP54 protein (weak retroviral MoMuLV enhancer), significantly longer times of exposure were required to detect the respective signal.

Figure S3. MS/MS spectra of the biotinylated and propionylated BAP1118 (V→F)**.**

Shown are the MS/MS fragmentation spectra of the propionylated (top) and biotinylated (bottom) forms of BAP, with the detected y-series and b-series ions indicated.

Figure S4. MS/MS spectra of the biotinylated and propionylated BAP1135 (V→Y).
Shown are the MS/MS fragmentation spectra of the propionylated (top) and biotinylated (bottom) forms of BAP, with the detected y-series and b-series ions indicated. Note that, in this peptide, both Lysine (K) and thyrosine (Y) are propionylated, which affects the masses of the parental and daughter ions.

Figure S5. Relative biotinylation efficiencies of 3 BAP peptides.

Western analysis of biotinylation levels of three BAP-H2AZ fusions, cotransfected separately with BirA-GFP. Shown is one example of the analysis. Top – α–His signal, used for normalization. Bottom – Streptavidin signal. The intensities of the signals were measured by densitometry (left tables, top and bottom). First, the values of streptavidin signals were normalized by taking into account different α–His signal intensities (right table). Afterwards, the relative intensities were calculated by dividing the normalized values for BAP1118 or BAP1135 (1,3) by the normalized value for BAP1070 (2). The average values from three different experiments and standard deviations are shown on the right side.

Figure S6. Colocalization analysis of PCNA and Biotin label after pulse and 2 hr chase.

Top - pulse sample. Bottom - chase sample.

Figure S7. Partner mobility can contribute to the biotinylation efficiency

Left - The average positions of the BAP and BirA fusions (A_1 and A_2, respectively) are close to each other, but their motion is so constrained that they practically have no chance to encounter each other (as depicted by the absence of an overlap between the distributions of their positions).

Right - The average distance between two proteins (B_1 and B_2) is larger than that for the A pair, but their positions fluctuate more significantly, leading to the larger overlap in their positions and thus to accumulation of biotinylated BAD species.

9

Annex - III

Use of In Vivo Biotinylation for Chromatin Immunoprecipitation (2011)

Arman Kulyyassov [1,2], **Muhammad Shoaib** [1] and Vasily Ogryzko [1]

I CNRS UMR-8126, Université Paris-Sud 11, Institut de Cancérologie Gustave Roussy, 39, rue Camille Desmoulins, 94805, Villejuif, France.

II National Center for Biotechnology of the Republic of Kazakhstan, 43 Valikhanova Str., 010000, Astana, Republic of Kazakhstan.

Curr Protoc Cell Biol. 2011 Jun; Chapter 17: Unit17.12

Use of In Vivo Biotinylation for Chromatin Immunoprecipitation

Arman Kulyyassov,[1,2] Muhammad Shoaib,[1] and Vasily Ogryzko[1]

[1]CNRS, Université Paris, Villejuif, France
[2]National Center for Biotechnology of the Republic of Kazakhstan, Astana, Republic of Kazakhstan

ABSTRACT

This unit describes a system for expression of biotinylated proteins in mammalian cells in vivo, and its application to chromatin immunoprecipitation (ChIP). The system is based on co-expression of the target protein fused to a short biotin acceptor domain, together with the biotinylating enzyme BirA from *Escherichia coli*. The superior strength of the biotin-avidin interaction in the modified ChIP protocol presented here allows one to employ more stringent washing conditions, resulting in a better signal/noise ratio. Methods for interpreting the data obtained from ChIP samples analyzed by qPCR, and methods for testing the efficiency of biotinylation using a streptavidin gel-shift are also presented. In addition, a complementary method, based on isothermal multiple strand displacement amplification (IMDA) of circular concatemers generated from the DNA fragments obtained after ChIP, is described. This method helps to decrease bias in DNA amplification and is useful for the analysis of complex mixtures of DNA fragments typically generated in miniscale ChIP experiments. *Curr. Protoc. Cell Biol.* 51:17.12.1-17.12.22. © 2011 by John Wiley & Sons, Inc.

Keywords: chromatin immunoprecipitation • biotinylation in vivo • BirA • IMDA • amplification bias

INTRODUCTION

Chromatin immunoprecipitation (ChIP) is a method for analyzing protein-DNA interactions directly in cells. The principle of this procedure is illustrated in Figure 17.12.1A. Given that the in vivo binding between DNA and proteins is labile (with the exception of core histones, the typical rate of dissociation of proteins bound to DNA in vivo is only a few minutes; Phair and Misteli, 2000; Hager et al., 2002), intact cells are first subjected to cross-linking (typically by formaldehyde), in order to fix the chromatin proteins on the DNA. Cellular lysates are prepared from the cross-linked cells, which are then sonicated to shear chromatin into fragments of the size that allow accurate determination of the protein position, without compromising the efficiency of DNA detection afterwards. The chromatin fragments containing the protein of interest are purified from the cellular lysates by affinity chromatography.

In the original ChIP procedure, a specific antibody against the protein of interest was used for such purification. The method described in Basic Protocol 1 employs proteins biotinylated in vivo, which permits the use of immobilized streptavidin (avidin) instead of antibodies, allowing for use of more stringent washing conditions and resulting in a better signal/noise ratio. After reversal of the cross-links, the selectively enriched DNA is further analyzed, usually by PCR, DNA array methodology (Negre et al., 2006), or high-throughput sequencing (Park, 2009). Support Protocol 1 provides methods for interpreting the data obtained from ChIP samples analyzed by quantitative real-time PCR (qPCR), while Support Protocol 2 describes a procedure for testing the efficiency

Current Protocols in Cell Biology 17.12.1-17.12.22, June 2011
Published online June 2011 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/0471143030.cb1712s51

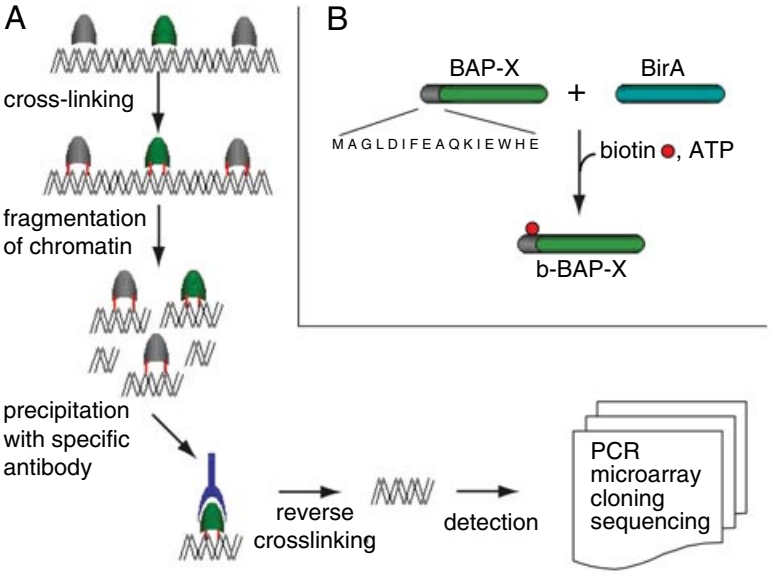


Figure 17.12.1 General principles of chromatin immunoprecipitation (ChIP) and in vivo biotinylation. (**A**) General scheme of ChIP. The proteins are cross-linked to DNA in vivo, the chromatin is fragmented, and the chromatin fragments with the protein of interest are affinity-purified. The cross-links are reversed, and the DNA is usually analyzed by PCR, DNA arrays, or high-throughput sequencing. (**B**) Epitope tagging by in vivo biotinylation. Two recombinant constructs are co-expressed in the cells: BAP-X, comprising the protein of interest X fused to a minimal biotin acceptor peptide (BAP, sequence shown), and bacterial biotin ligase BirA, specifically transferring the biotin moiety on the lysine residue of BAP. The reaction requires the presence of ATP and biotin in the cells.

of biotinylation using a streptavidin gel-shift. In addition, a complementary method, which helps to decrease bias in DNA amplification, is presented in Basic Protocol 2. It is based on isothermal multiple strand displacement amplification (IMDA) of circular concatemers generated from the DNA fragments obtained after ChIP. This method is useful for the analysis of complex mixtures of DNA fragments typically generated in miniscale ChIP experiments.

STRATEGIC PLANNING

Choice of Expression System

The principle of in vivo biotinylation is based on co-expression of the protein of interest fused to a biotin acceptor domain (BAD), or a shorter biotin acceptor peptide (BAP), together with the bacterial biotin ligase BirA. These two proteins can be expressed from different plasmids (see *trans* design in Fig. 17.12.2; de Boer et al., 2003; Kim et al., 2009), although this method is somewhat cumbersome. It can also decrease the biotinylation efficiency due to promoter competition and the need for the molecules of substrate and enzyme to meet in the confines of a large mammalian cell, in order for the biotinylation event to occur. Alternatively, a single bicistronic vector that expresses one mRNA encoding both BirA and the target BAP-fused protein can be used (see *cis* design in Fig 17.12.2). Such bicistronic design increases the local concentration of the enzyme in the vicinity of the target, leading to more efficient biotinylation (Viens et al., 2004). Although such design has been shown to work well for protein detection by immunoblotting or microscopy (electron or fluorescence; Viens et al., 2008), in our experience, the expression levels of the biotinylated proteins can sometimes be low. Therefore the choice of system depends on the protein of interest—its abundance and

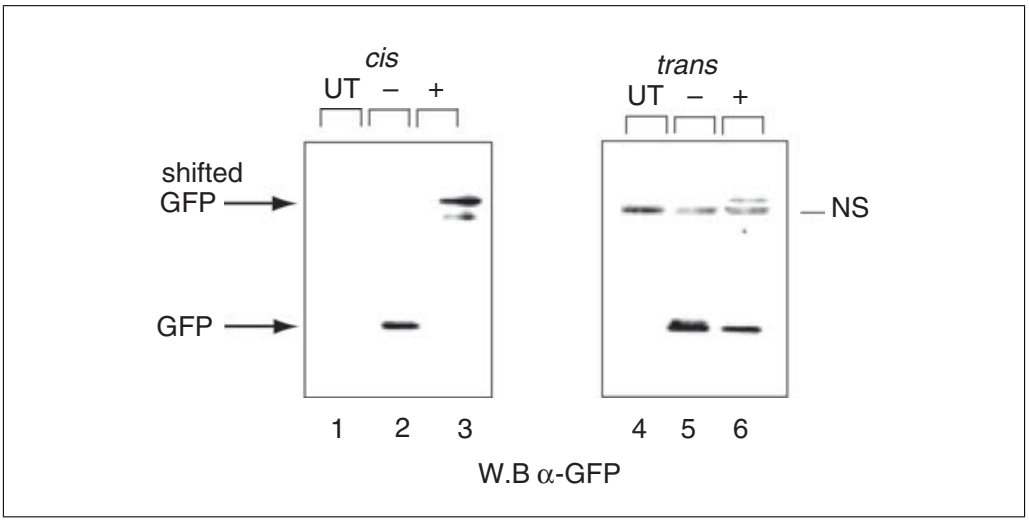


Figure 17.12.2 Streptavidin gel-shift. Immunoblot analysis with anti-GFP antibody of extracts from the HEK 293 cells expressing N-terminal BAP fusion of GFP. Streptavidin was added before SDS-PAGE in lanes 3 and 6 (+). Left panel: both BAP-GFP and BirA were expressed from the same mRNA (*cis*-biotinylation). Right panel: BAP-GFP and BirA were expressed from two different plasmids that were cotransfected into HEK 293 cells (*trans*-biotinylation). The positions of GFP and its streptavidin-shifted version are indicated by arrows. Abbreviations: GFP, green-fluorescent protein; NS, nonspecific signal corresponding to endogenously biotinylated proteins; UT, extract from untransfected or untransduced cells.

other characteristics. If expression from two separate plasmids is chosen, the use of human-optimized BirA helps to achieve a more efficient biotinylation of the target protein (Mechold et al., 2005).

Transient Transfection Versus Stable Cell Lines

Generation of cell lines that express the biotinylated protein of interest is time consuming and can become a limiting factor, especially if the two genes must be introduced separately into the cell line. Transient transfection typically yields less material; however, the use of unbiased DNA amplification after ChIP makes it a viable alternative. Accordingly, we provide a method for amplifying, without bias, DNA obtained in ChIP experiments (Basic Protocol 2).

Testing Biotinylation Efficiency with the Streptavidin Gel-Shift

If the antibodies against the protein of interest are available, two useful tests combined in one preliminary experiment can be performed. Their aims are (1) comparison of the levels of expression of the BAP-tagged protein and those of its endogenous counterpart, and (2) testing the efficiency of biotinylation of the BAP-fused protein. Both aims can be accomplished by using a streptavidin gel-shift (Viens et al., 2004; van Werven and Timmers, 2006; see Support Protocol 2 and Fig. 17.12.2). Given that the interaction between streptavidin and biotin is sufficiently strong to survive the conditions of SDS-PAGE separation, the addition of streptavidin to the sample before loading leads to an easily detectable decrease in the mobility of the biotinylated protein (~50 kDa). Comparing the intensities of the immunoblot signals corresponding to the shifted (streptavidin-bound) and nonshifted (unbound, and hence, unbiotinylated) forms of the protein provides an estimate of what percent of the total BAP-fusion protein was biotinylated. If the antibodies against native protein are not available, or if knowledge about the relative expression levels of the BAP-fusion and the endogenous protein are not essential for the experiment, antibodies for an additional tag, usually present in the expression vector (e.g., 6×*His*, Viens et al., 2004; FLAG, Kim et al., 2009) can be used for the immunoblot analysis.

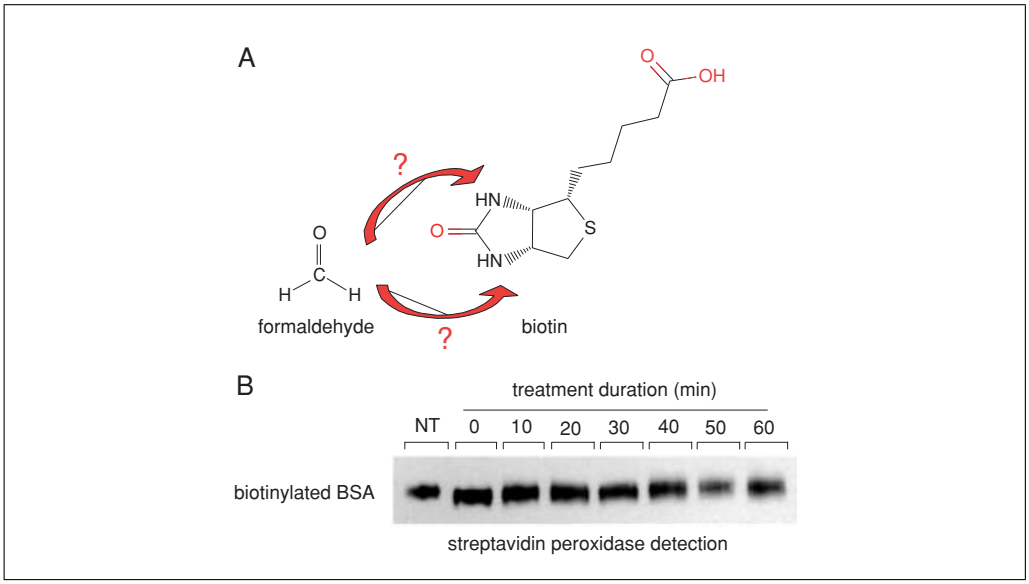


Figure 17.12.3 Use of biotinylated proteins is compatible with cross-linking by formaldehyde. Since biotin possesses two amino groups, it can react with formaldehyde, thus losing its affinity to streptavidin. To test whether formaldehyde treatment impairs the biotin-streptavidin interaction, biotinylated bovine serum albumin (BSA) was incubated in cell culture medium and treated with 1% formaldehyde for 10 to 60 min. The effect of the treatment on the ability of streptavidin to recognize biotinylated BSA was monitored by immunoblotting using streptavidin-conjugated peroxidase as a detection reagent. No effect of the formaldehyde treatment on the biotin-streptavidin interaction was observed, based on the intensity of the chemiluminescent signal, even for long incubations. Thus, the biotin tag is compatible with formaldehyde cross-linking. Abbreviation: NT, not treated.

Choice of Cross-Linking Method

The methods used to covalently link protein to DNA in vivo include ultraviolet (UV) (Gilmour and Lis, 1984; Pashev et al., 1991) and chemical cross-linking, most commonly by formaldehyde. Formaldehyde produces both protein-nucleic acid and protein-protein cross-links in vivo. A key feature of formaldehyde-induced cross-links is their reversibility, which allows one to purify the DNA for subsequent analyses (Orlando, 2000). As a result, formaldehyde has been the reagent of choice for cross-linking in ChIP applications. While it is possible that the use of formaldehyde can negatively affect epitope recognition. (e.g., biotin's nitrogen atoms could participate in a nucleophilic reaction with formaldehyde), we have shown that for the typical reaction times of formaldehyde treatment used in a ChIP protocol, the biotin-streptavidin affinity is not significantly affected (Fig. 17.12.3).

Another potential problem with formaldehyde is that it cross-links macromolecules within ~2 Å of each other; thus, it becomes less efficient when examining proteins that indirectly associate with DNA. To improve detection of such proteins, longer-range bifunctional cross-linkers have been used in addition to formaldehyde (Zeng et al., 2006). However, before using such a cross-linker in the biotin-ChIP protocol, we recommend first testing how the particular reagent affects the biotin moiety.

Finally, in contrast to cross-link ChIP (XChIP), native ChIP (NChIP) omits cross-linking (O'Neill and Turner, 2003). It is well suited for the analysis of histones because of their strong binding to DNA. Compared to biotin-ChIP, NChIP does not take advantage of stringent washing conditions (e.g., 2% SDS) because of the potential to disrupt the histone-DNA interactions. Nonetheless, biotin-NChIP has been used for histones (Ooi et al., 2010), as the high biotin-streptavidin affinity allows one to use smaller amounts of affinity resin. As a result, using regular washing conditions, one can achieve less nonspecific background compared to regular XChIP.

Choice of Chromatin Fragmentation Method

Two methods of fragmentation have been used for ChIP: (1) enzymatic fragmentation with micrococcal nuclease and (2) sonication. Enzymatic digestion is more acceptable for NChIP (Ooi et al., 2010), being more gentle and also providing information about nucleosome positioning. When cross-linked, chromatin becomes less accessible to nuclease, and DNA shearing by sonication becomes the method of choice.

The optimal size of chromatin fragments depends on the downstream analysis. Generally, it should be ~400 to 500 bp. This size covers two to three nucleosomes. Longer fragments will diminish resolution and increase nonspecific signal; therefore, they are not recommended for high-throughput analysis (microarray analysis or sequencing). Very short fragments are generally not good for PCR detection, as it becomes challenging to find suitable primers for amplification.

Choice of Reference Sample and Controls

Reference sample

The binding of the protein of interest to its target site is determined by the relative enrichment of ChIP signal over an unenriched reference sample. The choice of reference sample is a subtle issue. In most ChIP protocols, a mock ChIP sample (e.g., produced from the cells not expressing biotinylated protein, or expressing BirA alone or biotinylated GFP; Kim et al., 2009), or the input chromatin (i.e., the sample before the affinity pull-down; Dahl and Collas, 2008), is used as a reference.

However, one should be aware of certain caveats. In the case of an ideal mock purification, the sample is expected to contain very little DNA, complicating accurate measurement of amounts of different sequences in it. On the other hand, the use of input chromatin might be also problematic. During ChIP, some chromatin fragments could be particularly prone to aggregation and nonspecific binding, due to sequence, protein content, or higher-level structure. This introduces an artificial bias in the representation of different DNA sequences after ChIP, which cannot be accounted for when the input chromatin is used as a reference.

The use of cells expressing biotinylated histones is another attractive choice for a reference sample, as histones should be uniformly distributed over the genome. However, the histone density is lower in the chromatin with actively transcribed genes (Wolffe, 1999). One should be aware of this problem before deciding on what reference sample is best suited for a particular application. In this unit, we use input chromatin as a reference.

Controls

At least one positive control is recommended for quality control of ChIP. It should be a protein known to colocalize with an identified locus in the cell type examined.

We also recommend the use of two types of negative controls: (1) one with a tagged, irrelevant protein, known to have no interaction with the interrogated locus (GFP is a good choice, as the same sample can also serve to control the transfection efficiency), and (2) one with an irrelevant locus, a DNA sequence not expected to interact with the protein under question, for analysis of the same ChIP sample.

Note that keeping constant cell numbers between batches is important, as the cell number affects the precipitation efficiency. Triplicate ChIPs should be performed to assess variation between replicates.

EXPRESSION OF BIOTINYLATED PROTEINS IN MAMMALIAN CELLS IN VIVO AND ITS APPLICATION TO CHROMATIN IMMUNOPRECIPITATION (ChIP)

To take advantage of the strong interaction between biotin and streptavidin (K_d of 10^{-15} M; Lindqvist and Schneider, 1996) and commercial availability of many reagents for purification and detection of biotinylated macromolecules, we (Viens et al., 2004) and others (de Boer et al., 2003) have developed a variant of epitope tagging, based on protein biotinylation in vivo. Protein biotinylation in vivo has been used to study protein-protein interactions (de Boer et al., 2003) and electron microscopic localization of proteins of interest (Viens et al., 2008). In this protocol, the applications of these methods to ChIP are described, where the main advantage is the ability to use more stringent washing conditions to reduce background noise, as well as more efficient (i.e., more quantitative) recovery of the protein-DNA complexes (Fig. 17.12.4). Such an approach, in its various forms, has been applied to a broad variety of experimental systems, including mammalian cells, yeast, insects, and worms (van Werven and Timmers 2006; Kim et al., 2009; Kolodziej et al., 2009; Lausen et al., 2010; Ooi et al., 2010).

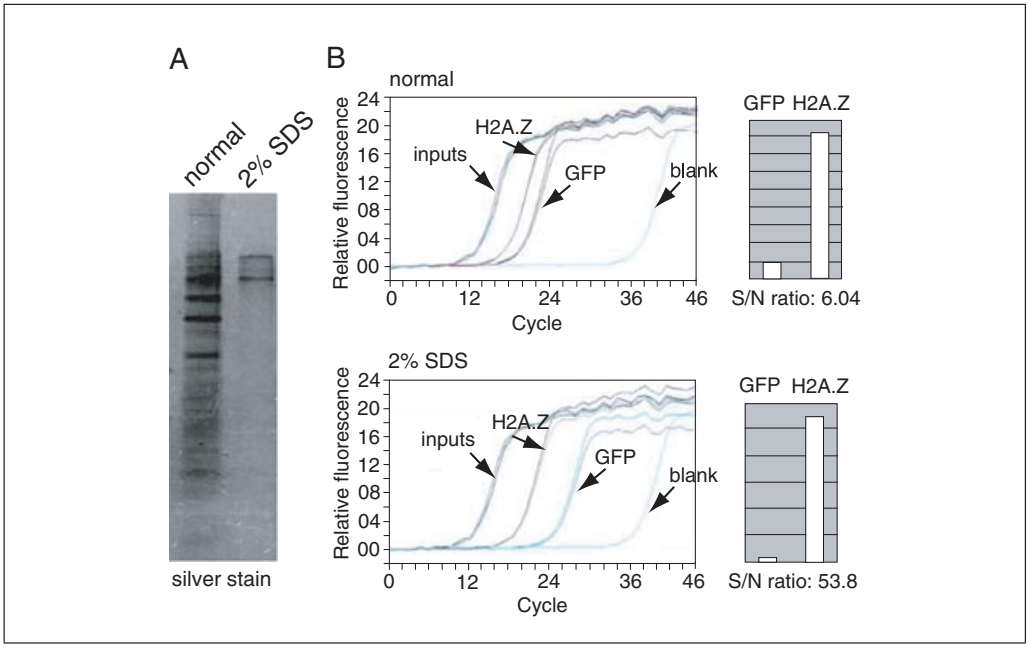


Figure 17.12.4 Use of stringent washing conditions in ChIP experiments significantly improves specificity (signal-to-noise ratio). (**A**) Silver stain of biotinylated protein expressed in HEK 293 cells, separated by SDS-PAGE and washed with either standard buffer or 2% SDS. (**B**) Interaction of histone H2A.Z with the total genome of NIH3T3 cells. NIH3T3 cells were transiently transfected with biotinylated H2A.Z- or eGFP-expressing plasmids, and biotin ChIP was performed. Quantitative PCR analysis was performed using genomic DNA present in the immunoprecipitates from the cells. Primers B1 5′: gcc ggg tgt ggt ggc gca cac ctt t and B1 3′: gag aca ggg ttt ctc tgt gta gcc ct were used to amplify short, abundant (~106 copies per genome), randomly distributed, repetitive B1 sequences, allowing for very sensitive detection of mouse genomic DNA in the sample. Shown are results representative of four completely separate experiments. A comparison of the ratio between amounts of DNA pulled down from an H2A.Z-transfected sample and an eGFP-transfected sample (nonspecific background) shows that with the more stringent washing conditions possible with the biotin-ChIP method, the signal-to-noise ratio is significantly increased and the specificity of interaction detection is improved.

Materials

Human embryonic kidney (HEK) 293 cells (ATCC)
Dulbecco's modified Eagle's medium (*APPENDIX 2A*)/10% (v/v) fetal bovine serum (DMEM-10)
2 M $CaCl_2$ (store up to 6 months At –20°C)
Plasmid DNA with expression system for target protein (see Strategic Planning) and controls (plasmid expressing BAP-GFP; Viens et al., 2004), checked for biotinylation efficiency (Support Protocol 2)
2×HeBS (*APPENDIX 2A*)
Biotin (Sigma cat. no. B4639)
37% (v/v) formaldehyde stock solution (Electron Microscopy Services)
1.25 M glycine stock (store up to 3 months at room temperature)
Phosphate-buffered saline (PBS, *APPENDIX 2A*), ice-cold
25× protease inhibitor stock (see recipe)
PBS + protease inhibitors, ice cold: dilute 25× protease inhibitor stock (see recipe) 1:1000 with PBS just before use
1% (w/v) agarose gel (see Voytas, 2001)
ChIP Buffer (see recipe)
70% (v/v) ethanol
5 M and 300 mM NaCl
10 mg/ml RNase A (e.g., Invitrogen)
Streptavidin-coupled magnetic beads (Dynabeads M-280 Streptavidin; Invitrogen)
Washing buffer 1: 2% (w/v) SDS; store up to 2 months at room temperature
Washing buffer 2: 10 mM Tris·Cl, pH 8 (see *APPENDIX 2A*)/1 mM EDTA (see *APPENDIX 2A*)/0.25 mM LiCl/1% (w/v) Nonidet P-40/1% (v/v)/sodium deoxycholic acid; store up to several months at 4°C
Washing buffer 3: 20 mM Tris·Cl, pH 7.6 (see *APPENDIX 2A*)/50 mM NaCl/1 mM EDTA (see *APPENDIX 2A*); store up to 2 months at room temperature
20 mg/ml proteinase K (Roche Applied Science) in 10 mM Tris·Cl, pH 7.5 (see *APPENDIX 2A*)/20 mM calcium chloride/5% (v/v) glycerol
5× proteinase K buffer: 50 mM Tris·Cl, pH 7.5 (see *APPENDIX 2A*)/25 mM EDTA (see *APPENDIX 2A*)/1.25% (w/v) SDS
3 M sodium acetate (NaOAc), pH 5 to 5.2
DNA purification kit for PCR (e.g., Qiagen): includes MiniElute columns, PBI buffer, PE buffer, and EB buffer

1.5-ml microcentrifuge tubes
14-ml round-bottom tube (Falcon 2059, or equivalent).
15-ml conical polystyrene tubes (Falcon)
Sonicator, with appropriate tube holder and accessories (e.g., Diagenode)
Magnetic separator for 1.5-ml tubes (e.g., Dynal MPC, Invitrogen)
Test-tube rotators, at room temperature and in a cold room
1.5-ml screw-cap tubes (VWR)
67°C heating block *or* thermal cycler

Additional reagents and equipment for culturing mammalian cells (*UNIT 1.1*) performing agarose gel electrophoresis (Voytas, 2001)

Transfect cells to generate biotinylated protein

1. Plate HEK 293 cells at the density of 6×10^6 per 100-mm tissue culture dish in 10 ml of DMEM-10, and grow overnight at 5% CO_2, 37°C (e.g., see *UNIT 1.1*).

   Transient transfection procedures (reagents, DNA amounts) depend on the cell line used. Here we describe transfection of HEK 293 with calcium phosphate.

2. Dilute 24.8 μl of 2 M $CaCl_2$ with water to a total volume of 200 μl in a 1.5-ml microcentrifuge tube, and add 5 μg of the plasmid DNA to be transfected.

3. Place 200 μl of 2× HeBS in a 14-ml round-bottom tube. Add the DNA/calcium mixture (from step 2) into the 2× HeBS dropwise, mixing after each addition.

   The pH of the HeBS buffer is crucial. The precipitate should be barely visible under microscope.

   For control of transfection efficiency, we recommend transfecting a separate cell well with BAP-GFP (Viens et al., 2004) or any other fluorescent protein. This sample can also be later used as a negative control in ChIP.

4. Add the transfection mixture from step 3 dropwise to the packaging cells in the dish (from step 1), stirring well after each addition, and incubate for 12 hr at 5% CO_2, 37°C.

5. Replace the medium with fresh DMEM-10 with 1 mg/ml biotin added. Incubate overnight.

Perform formaldehyde cross-linking

6. In a fume hood, add 270 μl of 37% formaldehyde stock to a final concentration of 1% (v/v) directly to the culture plate without changing the medium.

   Make sure that the reagent is mixed well with the medium.

7. Keep the plate at room temperature for 10 min, shaking occasionally.

   Time of cross-linking may vary, depending on the cells and the protein to be immunoprecipitated. For optimization, several samples with different cross-linking times should be tested. In most cases, 8 to 10 min for cross-linking is sufficient for ChIP with mammalian cells.

8. Add 1 ml of 1.25 M glycine stock (125 mM final concentration) to the culture plate to quench the formaldehyde, mix well, and incubate for 5 min at room temperature.

9. Remove the culture medium completely by aspiration, and rinse the cells with 5 ml ice-cold PBS + protease inhibitors. Remove the PBS + protease inhibitors by aspiration. Avoid drying the plates.

10. Scrape the adherent cells from the culture dish into a 15-ml conical polystyrene tube. Rinse the culture dish with 5 ml PBS + protease inhibitors, add the rinse to the same 15-ml tube, and mix.

    Trypsin can be used to detach the cells from the culture plate; however we have found that for cross-linked HEK 293 cells, a scraper works well.

11. Gently pipet up and down to break up cell aggregates. Centrifuge 5 min at 2,400 × *g*, 4°C.

    To prevent cell lysis during pipetting of cells, use a 1000- or 200-ml pipet tip cut off to increase the diameter of the opening.

12. Remove the supernatant by careful aspiration to avoid losing the cell pellet.

13. Add 10 ml cold PBS + protease inhibitors, wash the cells by gently pipetting, and centrifuge 5 min at 2,400 × *g*, 4°C.

14. Remove the supernatant by careful aspiration to avoid losing the cell pellet. Repeat steps 12 and 13.

15. Without adding any more solution, centrifuge the tube one more time at 5 min at 2,400 × *g*, 4°C.

16. Remove all of the supernatant by careful pipetting.

    The sample medium contains biotin, which can compete for the binding to streptavidin with the biotinylated chromatin in the downstream pull-down step. Thorough washing of the medium from the cells is therefore critical.

    The sample pellet can be snap frozen in liquid nitrogen and stored indefinitely in liquid nitrogen, or up to several months at −80°C.

Lyse cells and shear chromatin

Typically (Weinmann and Farnham, 2002), nuclei are prepared from the cross-linked cells before chromatin shearing. We found that, due to the strong affinity between biotin and streptavidin and our stringent washing conditions, this step is not necessary in our protocol.

17. Resuspend the pellet (kept cold on ice) in 500 µl of ice-cold ChIP buffer.

    Applying more ChIP buffer helps to shear the chromatin; however, too much dilution increases the volume of the reaction, which may require increased volume of streptavidin beads used for binding.

18. Sonicate the sample using a sonicator kept in cold room, as follows (for the Diagenode sonicator):

    a. Transfer the samples to appropriate tubes for sonication.

       The best choice is 15-ml polystyrene tubes.

    b. Remove the tube holder, and check that water level is at the blue mark.

    c. Use the appropriate tube holder and accessories for samples, and rinse with ethanol before use.

    d. Balance the tubes in sonicator.

    e. Input sonication settings of 5 pulses for 30 sec each (red needle), with 1-min rest intervals between pulses (green needle). Set the LMH dial to high.

       This results in chromatin fragments of 300 to 800 bp in length.

       The pulse duration, intensity, and number will vary, depending on the extent of cross-linking and cell type. You must optimize conditions for your experiment. Ideally, the least amount of input energy that gives satisfactory fragmentation should be used.

    f. Rinse the used tube holder and accessories with water, and then with 70% ethanol.

Check sonication efficiency and fragment size

19. Reverse the cross-link of a 10-µl aliquot, as follows:

    a. Add 40 µl 300 nM NaCl to 10 µl sample.

    b. Add 2 µl of 5 M NaCl (0.2 M NaCl final concentration).

    c. Boil for 15 min.

    d. Cool down to room temperature.

    e. Add 1 µl of 10 mg/ml RNase A.

20. Electrophorese the sample on a 1% agarose gel and visualize by staining (e.g., see Voytas, 2001).

    The average size of the sheared DNA fragments should be in the range 0.2 to 0.5 kb (Fig. 17.12.5). If the average size of sheared DNA fragments is too large, continue sonication as in step 18, and repeat steps 19 and 20 to check the results.

Isolate chromatin samples

21. Transfer the sonicated material to an ice-chilled, 1.5-ml microcentrifuge tube, and centrifuge 10 min at 12,000 × *g*, 4°C.

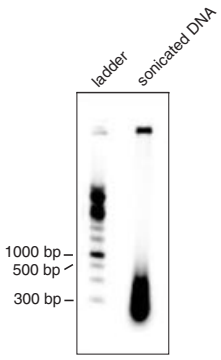


Figure 17.12.5 DNA shearing by sonication. Typical results after electrophoresis (1% agarose gel) of DNA sheared by sonication in a chromatin immunoprecipitation experiment. Left: DNA ladder, Right: sonicated DNA.

22. Carefully aspirate the supernatant (chromatin), and transfer it into a new ice-chilled, 1.5-ml microcentrifuge tube.

    To avoid aspirating the sedimented material, leave 30 μl of supernatant in the tube after aspiration.

23. Add 400 μl ChIP buffer to the remaining pellet, mix by vortexing, and centrifuge 10 min at 12,000 × *g*, 4°C.

24. Aspirate the supernatant, leaving 30 μl in the tube, and pool it with the first supernatant. Discard the pellets.

    Chromatin can be stored up to several months at −80°C.

Preclear sample before IP

The aim of the preclearing step is to get rid of nonspecifically binding material. An ideal preclearing medium for the biotin-ChIP is streptavidin-conjugated Dynabeads (or Sepharose) blocked with biotin, as it provides the surface that most closely mimics the affinity medium (the unblocked streptavidin-conjugated beads) used in the actual ChIP experiment, yet does not bind biotin. Typically, the time for incubation in the preclearing step should be at least as long as is the time of incubation in the ChIP experiment itself.

25. Resuspend the streptavidin-conjugated Dynabeads stock by shaking the vial to obtain a homogeneous suspension. Dispense 50 μl streptavidin-magnetic bead slurry per sample into a 1.5-ml microcentrifuge tube. Fix the tube on a Dynal MPC magnet for 2 min, and remove the supernatant by pipetting.

    Take care not to disrupt the streptavidin beads, i.e., avoid touching the inside wall of the tube (faced towards the magnet) with the pipe tip.

26. Add 1 ml of ChIP buffer to the tube and wash the beads well by gentle pipetting. Place the tube on the Dynal MPC magnet for 2 min, and collect the beads. Remove the buffer.

    When performing magnetic separation, always briefly spin the tubes in a microcentrifuge for 1 sec to bring down any liquid trapped in the lid, prior to positioning the tubes in the magnetic rack.

27. Add 1 ml of ChIP buffer with 1 μg/ml biotin added. Rotate for 10 min at room temperature. Place the tube on Dynal MPC magnet for 2 min, and collect the blocked beads. Remove the buffer.

28. Repeat wash twice as in step 26. Then add 50 μl ChIP buffer per sample, plus a small additional volume for pipetting error.

    For example, calculate the amount of buffer as (N + 0.5) × 50, where N is the number of samples.

29. Dispense in aliquots of 50 μl into prelabeled 1.5-ml microcentrifuge tubes. Collect the beads using the magnetic separator, and remove the buffer from each tube.

    Make sure the stock bead suspension is homogenous before pipetting.

30. Add the chromatin samples to the tubes, reserving an aliquot (at least 10 μl) as an input chromatin sample. Rotate 3 hr at 4°C.

31. Meanwhile, prepare the streptavidin-conjugated Dynabeads (affinity medium) for the ChIP procedure by following the steps 25 to 29, skipping the biotin blocking in step 27.

32. Magnetically separate the blocked beads and precleared chromatin samples (step 30), and transfer the precleared chromatin to the prelabeled tubes containing the washed affinity medium (unblocked streptavidin-conjugated beads; step 31).

Perform binding for immunoprecipitation

33. Mix the chromatin sample and streptavidin beads carefully with gentle pipetting. Incubate the tube on the test-tube rotator 3 hr at 4°C.

    A 3-hr incubation is the minimum time recommended for successful pull-down of biotinylated chromatin. Longer incubation times can also be used. In this case, the length of incubation in the preclearing step (step 30) should be increased accordingly.

34. Briefly (1 to 2 sec) microcentrifuge the tube to collect the beads into the bottoms of the tubes.

35. Mix the sample and beads well by gently pipetting, and transfer the supernatant and the beads to a fresh tube placed on the magnetic separator.

    The tube surface is a source of unspecific binding of chromatin. Transferring the ChIP material to a fresh tube enhances specificity of the ChIP procedure.

36. Wait for 2 min to collect beads to the side of the tube.

37. Remove the supernatant with a pipet, taking care not to disrupt the beads.

Wash beads

All of the following washing steps are carried out at room temperature; each wash should be performed quickly to prevent the beads from drying out.

38. Add 1 ml washing buffer 1 to the tube, gently resuspend beads by flicking the tube, and rotate the tube for 10 min at room temperature.

39. Collect the beads by magnet, and remove the washing buffer. Wash a second time by adding 1 ml washing buffer 1, resuspending and transferring the suspension to the new tube, rotating for 10 min at room temperature, collecting the beads by magnet, and removing the washing buffer.

    The advantage of biotin-ChIP protocol is that more stringent washing conditions can be used, in particular, 2% SDS. (Fig. 17.12.4).

40. Wash twice with 1 ml washing buffer 2, as described in steps 38 and 39, except for changing the tube.

    Changing the tube is not necessary at this point.

41. Wash twice with 1 ml washing buffer 3, as described in steps 38 and 39, except for changing the tube between the washes. After the last wash, transfer the sample to a 1.5-ml screw-cap tube.

Elute and de-cross-link samples

Due to strong affinity between streptavidin and biotin, the elution of the biotinylated molecule from streptavidin-coupled beads is not trivial. However, after de-cross-linking, DNA comes off the beads independently from the biotinylated protein. Accordingly, in our protocol, the elution and de-cross-linking steps are combined.

42. Add 60 µl of 300 mM NaCl to the beads contained in screw-cap tubes. Incubate overnight at 67°C.

    Using screw-cap tubes minimizes evaporation of elution buffer during overnight de-cross-linking. Alternatively, regular microcentrifuge tubes can be wrapped with Parafilm.

43. Briefly centrifuge, and add 3 µl of 20 mg/ml proteinase K and 15 µl of 5× proteinase K buffer. Incubate 2 hr at 67°C.

44. Meanwhile, prepare the input chromatin sample (reserved in step 30). To 10 µl of input chromatin samples, add 60 µl of ChIP buffer and 3 µl of 20 mg/ml proteinase K, vortex, and incubate on a heating block for 2 hr at 67°C.

    From this time point, process the input samples and the ChIP samples in parallel.

45. Briefly centrifuge all of the incubated tubes, and place them in the Dynal MPC magnet. Wait for 3 min for the beads to separate.

46. Transfer the supernatant by pipetting to a new tube, and discard the magnetic beads.

DNA purification

ChIP protocols designed for PCR analysis do not include an RNase step, but it can be introduced before step 47 of the DNA purification.

47. Add 500 µl PBI buffer to each sample, including the input sample.

48. Add 10 µl of 3 M NaOAc (pH 5 to 5.2), and lightly vortex.

    This makes the solution less basic.

49. Transfer to a DNA purification column.

50. Centrifuge 1 min at 2000 × *g*, 4°C, and discard the flow-through.

51. Place the column back into the catch tube and add 500 µl PE buffer to each tube. Centrifuge 1 min at 2000 × *g*, 4°C and discard flow through.

52. Repeat step 51.

53. Reinsert the column into the catch tube, and centrifuge 30 min at maximum speed. Make sure that no liquid is left at the rim of the tube.

54. Label fresh 1.5-ml microcentrifuge tubes, and place each column into a new microcentrifuge tube. Discard the catch tubes.

55. Add 30 µl EB buffer to the column membrane, and incubate 2 min at room temperature.

56. Microcentrifuge 1 min at maximum speed, room temperature, and discard the column.

    The DNA solution in the microcentrifuge tube can be stored up to several months at –80°C, if necessary, before analysis (e.g., by qPCR; see Support Protocol 1).

ANALYSIS AND DATA INTERPRETATION BY qPCR

The DNA from the ChIP samples isolated in Basic Protocol 1 can be analyzed in different ways, e.g., by PCR, DNA arrays, or sequencing; thus, the setup of the samples and interpretation of the results will vary accordingly. Here we discuss general principles of the analysis of ChIP samples using quantitative real-time polymerase chain reaction (qPCR). We more specifically address the calculation of immunoprecipitation efficiencies from the data obtained by qPCR, and their interpretation. See *UNIT 17.7* (particularly, Alternate Protocol 2) for more information about procedures used to carry out qPCR.

Make several dilutions of the reference sample (i.e., input chromatin from Basic Protocol 1, step 30) in the elution buffer (EB from the Qiagen DNA purification kit). Analyze the ChIP samples (obtained in Basic Protocol 1) and the reference sample with the same PCR program. Include DNA melting curve analysis to check the product specificity. Typical qPCR protocols include melting curves, and most qPCR machines have this option, as a quality control to check for correct PCR-fragment amplification and quantification during the qPCR procedure.

Typically, you will have to quantify the immunoprecipitation efficiency for the following cases:

P1 - presence of the gene of interest G1 in the ChIP of protein of interest
P2 - presence of the control gene G2 in the ChIP of protein of interest
C1 - presence of the gene of interest G1 in the mock ChIP (e.g., using Biotin-GFP expressing cells)
C2 - presence of the control gene G2 in the mock ChIP.

To determine the relative enrichment of a protein on a particular DNA site perform the following calculations:

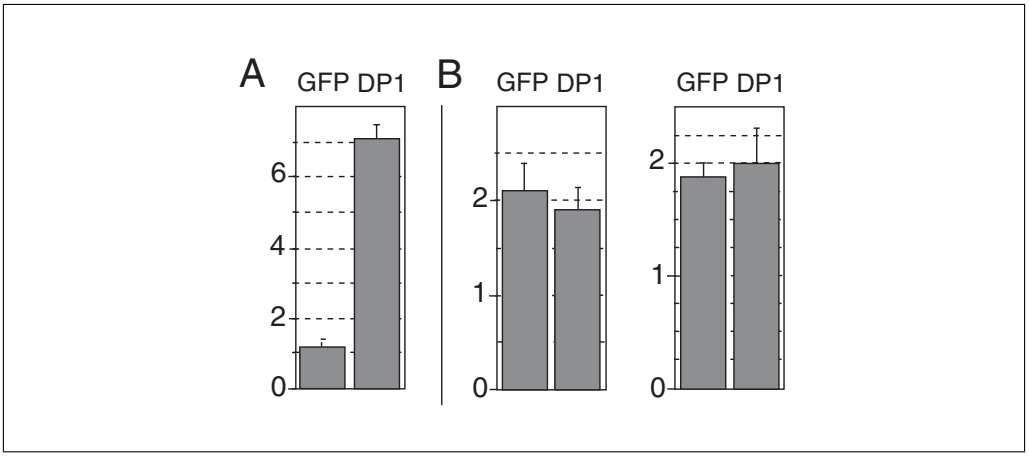


Figure 17.12.6 Typical result from a biotin-chromatin immunoprecipitation (biotin-ChIP) experiment analyzed by qPCR. (**A**) Interaction of transcription factor DP1 with its cognate DHFR promoter. The cells were transiently transfected with pBBHN.DP1 vector. Shown is the ratio between amounts of DNA pulled down from BAP-DP1 transfected sample and BAP-eGFP transfected sample, considered as a nonspecific background (mean of three experiments, standard deviations indicated by error bars). The sequences of primers for qPCR analysis are: DHFR 5′: gcg gag cct tag ctg cac aa, DHFR 3′: tac cag cct tca cgc tag ga. (**B**) DP1 interaction with irrelevant GAPDH (left) and β-actin (right) promoters. Immunoprecipitate from the same experiment as in (A) was tested with the GAPDH- and β-actin-specific primers. The sequences are: β-actin 5′: acc gag cgt ggc tac agc tt, b-actin 3′: tgg ccg tca ggc agc tca ta, GAPDH 5′: cca atg tgt ccg tcg tgg atc t, and GAPDH 3′: gtt gaa gtc gca gga gac acc.

1. Calculate the immunoprecipitation efficiency (IE) for each sample (i.e., P1, P2, C1 and C2) by dividing the signal obtained from immunoprecipitated material by the signal obtained from the input chromatin. Ideally, the mean and SD of each ChIP are calculated from at least three independent experiments, and two independent amplifications.
2. For each DNA fragment analyzed, subtract the IE value obtained for the nonrelevant protein (C1, C2) from the IE value obtained for the protein of interest (P1, P2), respectively. This subtraction removes background noise due to nonspecific binding of chromatin to beads.
3. Take the ratio between these data for the gene of interest G1 and control gene G2 by calculating the enrichment of a protein on a particular DNA site G1relative to a DNA site G2, according to the formula: (P1−C1)/(P2−C2). This formula gives us the relative (*x*-fold) enrichment of a protein on a particular site as compared to an irrelevant DNA site.

Typical results from the biotin-ChIP procedure and qPCR analysis are shown in Figure 17.12.6. The high ratio between the ChIP samples in Figure 17.12.6A signifies a specific interaction of the protein (DP1) with DNA (DHFR promoter).

SUPPORT PROTOCOL 2

TESTING THE EFFICIENCY OF BIOTINYLATION USING A STREPTAVIDIN GEL-SHIFT

Due to the strong interaction between streptavidin and biotin, the addition of streptavidin to the biotinylated protein of interest before loading onto an SDS-PAGE gel leads to an easily detectable decrease in the mobility of the biotinylated protein (∼50 kDa). Comparing the intensities of the immunoblot signals corresponding to the shifted (streptavidin-bound) and nonshifted (unbound, and hence, unbiotinylated) forms of the protein provides an estimate of the percentage of the total BAP-fusion protein that was biotinylated.

The estimation of biotinylation efficiency with a streptavidin gel-shift can be done with amounts of material typically much smaller than required for ChIP. In addition, the cross-linking is not performed. Therefore, we recommend performing the streptavidin gel-shift as a preliminary experiment with a small number of cells (transiently transfected or stable cell lines). Given that only nuclear proteins are concerned when ChIP is performed, we recommend using nuclei instead of the total lysate, purifying nuclei from specific cell line according to your preferred method.

This protocol describes preparation of nuclei from HEK 293 cells and estimation of the biotinylation efficiency of the BAP-fused protein. Typically, one well from a 6-well plate is sufficient for the experiment.

Materials

5×10^6 HEK 293 cells expressing the biotinylated protein of interest (see Basic Protocol 1, steps 1 to 5) grown in one well of a 6-well tissue culture plate
Phosphate-buffered saline (PBS, *APPENDIX 2A*)
CSK buffer (see recipe)
1× NuPAGE LDS sample buffer (Invitrogen)
5 mg/ml streptavidin (Sigma)
4 to 12% gradient Novex Tris·glycine precast gels (Invitrogen)
Antibody for detection of the protein of interest (e.g., anti-GFP)

1.5-ml microcentrifuge tube
Sonicator (e.g., Diagenode), with appropriate tube holder and accessories
Heating block set at 96°C

Additional reagents and equipment for performing immunoblot analysis (*UNIT 6.2*)

Prepare chromatin samples

1. Detach and collect ~5 × 10^5 cells expressing the biotinylated protein of interest, by pipetting in 1 ml of PBS, and transfer to a 1.5-ml microcentrifuge tube.

   HEK 293 cells can be easily detached by pipetting. However, most of the other cells lines will require trypsinization.

2. Centrifuge 5 min at 1000 × *g*, room temperature.
3. Aspirate and discard the supernatant.
4. Lyse the cells by adding 100 µl of CSK buffer. Incubate for 5 min on ice.
5. Centrifuge 5 min at 2000 × *g*, room temperature.
6. Aspirate and remove the supernatant.

   The pellet contains the nuclei. Store the nuclei up to 1 week at −20°C.

7. Add 100 µl of 1× NuPAGE LDS sample buffer with 10 mM DTT added just before use.
8. Heat the sample 5 min at 96°C, and vortex intensively for 10 sec.
9. If the viscous pellet remains, homogenize the sample by sonication.
10. Dispense two 10-µl aliquots of the lysate into 1.5-ml microcentrifuge tubes, and add 1 µl of 5 mg/ml streptavidin (5 µg) to one of the aliquots.

Carry out immunoblotting

11. Incubate the samples for 5 min at room temperature.
12. Load the samples on a 4 to 12% gradient Novex Tris·glycine precast gel, separate the proteins, and perform immunoblot analysis under the conditions of your choice (see *UNIT 6.2*).

    If the antibodies against the protein are not available, use HRP-conjugated PentaHis antibody (Qiagen cat. no. 34460) according to the manufacturer's protocol. The vector we use includes a 6×His tag.

13. Interpret results.

    Typical results are presented in Figure 17.12.2. Biotinylated GFP is shown in the presence and absence of streptavidin. Both a cis (both BirA and BAP-GFP expressed from one mRNA) and trans (expressed from two independent, cotransfected plasmids) biotinylation are shown.

    The addition of streptavidin to the sample before loading leads to an easily detectable decrease in the mobility of the biotinylated protein (~50 kDa). Comparing the intensities of the immunoblot signals corresponding to the shifted (streptavidin-bound) and nonshifted (unbound, and hence, unbiotinylated) forms of the protein provides an estimate of what percent of the total BAP-fusion protein was biotinylated.

UNBIASED AMPLIFICATION OF ChIP SAMPLE BY ISOTHERMAL MULTIPLE STRAND DISPLACEMENT AMPLIFICATION (IMDA)

BASIC PROTOCOL 2

Most of the currently available techniques for amplifying small DNA fragments introduce sequence-dependent bias in the amplified mixture. We describe here a method, useful for the analysis of complex mixtures of DNA fragments, typically generated in

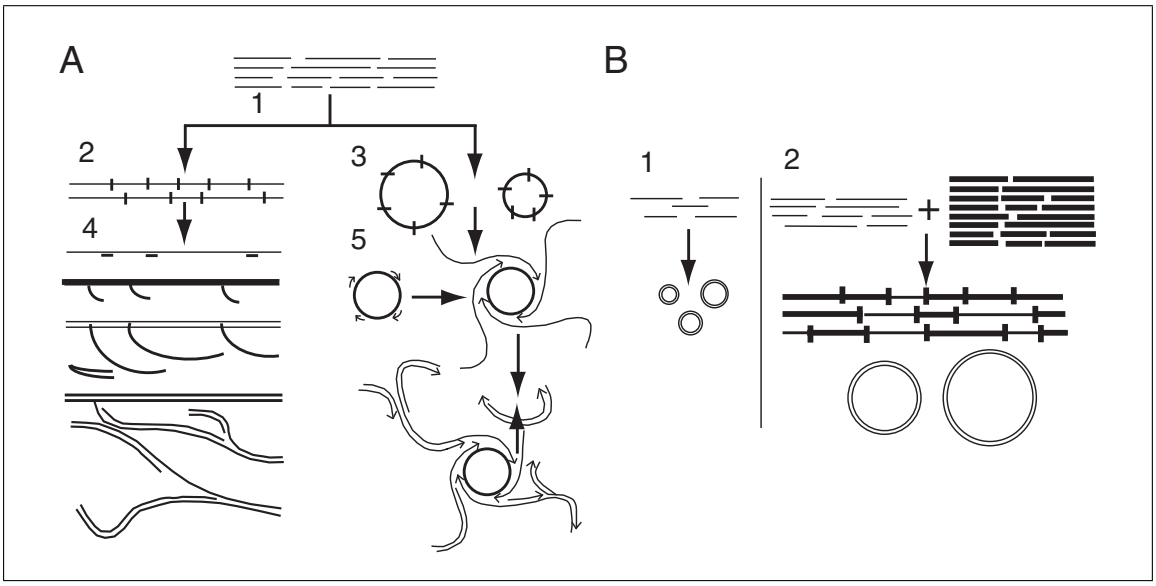


Figure 17.12.7 Amplification of complex mixtures of DNA fragments by isothermal multiple strand displacement amplification (IMDA.) (**A**) The principle of concatemer-mediated multiple displacement amplification. (1) Religation of DNA fragments with T4 DNA ligase leads to two types of products: (2) linear concatemers and (3) circular concatemers. Annealing of random hexamer primers and addition of phi29-DNA polymerase leads to concatemer-mediated multiple strand displacement amplification from (4) linear and (5) circular concatemers respectively. (**B**) The principle of stuffer DNA. (1) In the case of much-diluted samples, religation preferentially leads to self-circularization of fragments, which are then amplified as individual molecules. (2) The addition of stuffer DNA (derived from an evolutionarily distant source; indicated by heavy lines) to the sample DNA favors intermolecular ligation and leads to the formation of long concatemers, both linear and circular. These long concatemers can then be reliably amplified using the IMDA technique. The resulting DNA can be analyzed by PCR and DNA microarrays; however, high-throughput sequencing will produce mostly irrelevant information due to the excess of foreign DNA in the amplified mixture.

the miniscale ChIP experiments (Shoaib et al., 2008). If the DNA concentration in the sample exceeds 10 ng/μl, the DNA is blunt-ended and treated with ligase to form large concatemers (circles or linear molecules), which are further amplified using Φ29-based isothermal multiple-strand displacement amplification (IMDA). In these randomly composed concatemers, the differences in amplifiability between individual DNA fragments (due to their length, GC content, and secondary structure) are averaged out, helping to decrease the sequence-dependent bias in the amplified pool of DNA. If, however, the DNA concentration is less than 10 ng/μl, an evolutionarily distant "stuffer DNA" is added to the sample to adjust the total DNA concentration to the value sufficient for concatemerization. This procedure requires generation of long concatemers from the fragments present in the original ChIP samples, but the DNA concentration of $<$10 ng/μl will lead to self-circularization of the ChIP fragments, instead of concatemer formation (Shoaib et al., 2008). To prevent self-circularization, an irrelevant stuffer DNA should be added to the sample before ligation (Fig. 17.12.7). Therefore, the DNA concentration should first be measured to determine if the stuffer DNA is necessary.

Materials

ChIP DNA sample (Basic Protocol 1)
~500-bp fragments of sonicated *E. coli* DNA, optional (see Basic Protocol 1, step 18)
Takara Mighty Cloning Kit, blunt end (Takara), including
  10× Blunting Kination Buffer
  Blunting Kination Enzyme Mix
  Ligation Mighty Mix

25:24:1 (v/v/v) phenol/chloroform/isoamyl alcohol
24:1 (v/v) chloroform/isoamyl alcohol
3 M sodium acetate
20 mg/ml glycogen
70% (v/v) and 100% ethanol, cold
TE (Tris/EDTA) buffer (*APPENDIX 2A*)
5 U/μl Ø29-DNA polymerase: dilute 1000 U/μl Ø29-DNA polymerase (Epicentre Biotechnologies) 1:200 with the supplied buffer and store up to 3 months at −20°C
2× IMDA reaction buffer (see recipe)
200 μM random hexamer primers (NEB)
Mineral oil, biotechnology grade

NanoDrop spectrophotometer (NanoDrop)
1.5-ml microcentrifuge tubes
16°C, 30°C, 37°C, 65°C, and 70°C cooling/heating blocks *or* thermal cycler

Prepare sample DNA

1. Use 1 μl of each sample to measure the DNA concentration using a NanoDrop spectrophotometer according to the manufacturer's directions.
2. If the DNA concentration is lower than 10 ng/μl, add sonicated *E.coli* DNA (stuffer DNA) to a final concentration of 15 ng/μl.
3. Using components of the Takara Mighty Cloning Kit, prepare the following 20-μl reaction mixture in a 1.5-ml microcentrifuge tube:

   17 μl sample DNA
   2 μl of 10× Blunting Kination Buffer
   1 μl Blunting Kination Enzyme Mix.

   Incubate 10 min at 37°C.
4. Mix 80 μl of distilled water and 100 μl of phenol/chloroform/isoamyl alcohol. Centrifuge 5 min at 16,000 × *g*, room temperature, and transfer the upper layer to a new tube.
5. Add an equal amount of chloroform/isoamyl alcohol and mix. Centrifuge 5 min at 16,000 × *g*, room temperature, and transfer the upper layer to new a tube.
6. Add 10 μl of 3 M sodium acetate, 2 μl of 20 mg/ml glycogen, and 250 μl of chilled 100% ethanol, and mix. Incubate 20 min at −80°C.
7. Centrifuge at 10 min at 16,000 × *g*, 4°C.
8. Remove the supernatant. Wash the precipitate with chilled 70% ethanol. Centrifuge 5 min at 16,000 × *g*, 4°C.
9. Remove the supernatant, and dry the precipitate on the bench at room temperature (5 to 10 min).
10. Dissolve the precipitate in 10 μl TE buffer.

Perform ligation

11. Add 10 μl Ligation Mighty Mix, and mix gently. Incubate 1 hr at 16°C.
12. Inactivate the enzymes by heating 5 min at 65°C.

Perform IMDA procedure

13. Prepare the Ø29 reaction mix immediately before setting up the reactions. For a 10-μl of reaction mix, combine the following:

    6.3 μl of 2× buffer
    1.2 μl of 5 units/μl Ø29-DNA polymerase (0.6 units/μl final concentration)
    2.5 μl of 200 μM random hexamer primers (50 μM final concentration).

    The volume of the reaction mix may be adjusted as necessary, for the number of samples.

14. Combine 1 μl of DNA concatemers with 2 μl of Q29 reaction mix.

15. Overlay the sample with 10 μl of biotechnology grade mineral oil to prevent evaporation, and briefly microcentrifuge to make sure that the aqueous phase forms a small sphere at the bottom of the tube.

16. Incubate 6 hr at 30°C, and then heat for 10 min at 70°C to inactivate the enzyme.

17. Store the samples up to 3 months at −20°C or up to 1 year at −80°C, or use immediately for further analysis.

REAGENTS AND SOLUTIONS

Use deionized, distilled water in all recipes and protocol steps. For common stock solutions, see ***APPENDIX 2A****; for suppliers, see* ***SUPPLIERS APPENDIX****.*

ChIP buffer

0.3% (w/v) SDS
1% (v/v) Triton X-100
2 mM EDTA (*APPENDIX 2A*)
30 mM Tris·Cl, pH 8 (*APPENDIX 2A*)
150 mM NaCl
Store up to several months at 4°C. Just before use add 25× protease inhibitor stock (see recipe) to a final concentration of 1×.

CSK buffer

100 mM NaCl
300 mM sucrose
10 mM Tris·Cl, pH 7.5 (see *APPENDIX 2A*)
3 mM $MgCl_2$
1 mM EGTA
1.2 mM phenymethylsulfonyl fluoride (PMSF, *APPENDIX 2A*)
0.5% (v/v) Triton X-100
Store up to several months at 4°C. Just before use, add 1.2 mM (final concentration) phenylmethylsulfonyl fluoride (PMSF, *APPENDIX 2A*) and 25× protease inhibitor stock (see recipe) to a final concentration of 1×.

IMDA reaction buffer, 2×

Prepare 10× IMDA reaction buffer containing the following components:

370 mM Tris·Cl, pH 7.5 (see *APPENDIX 2A*)
100 mM $MgCl_2$
500 mM KCl
50 mM $(NH_4)_2SO_4$
2% (v/v) Tween 20

continued

1 μg/μl bovine serum albumin (BSA)
Store up to 6 months at −20°C.

Prepare 2× buffer from the 10× buffer, by adding 10 mM dNTP mix (Promega; 2 mM final concentration), DTT (2 mM final concentration), and water.

Protease inhibitor stock, 25×

Dissolve a 50-ml size of Complete Protease Cocktail Tablet (Roche) in 2 ml water, and dispense 100- to 200-μl aliquots into microcentrifuge tubes. Store up to 6 months at −20°C.

COMMENTARY

Background Information

All genomic processes in eukaryotes—transcription, replication, repair, and recombination—occur in the context of chromatin, the hierarchically organized and dynamic complex of DNA and histone and nonhistone proteins. The systematic study of the time-dependent association among these components of chromatin requires robust and effective approaches to characterizing protein-DNA interactions. Methods such as the electrophoretic gel mobility shift assay (EMSA; Revzin 1989; Molloy, 2000), systematic evolution of ligands by exponential enrichment (SELEX; Gold et al., 1995), and others (Veenstra, 1999) have been developed to study binding of protein to DNA. However, in these approaches, the interactions are studied in vitro, out of the context of chromatin and the nuclear environment, and thus an independent confirmation is required to assess the relevance of the data obtained. Chromatin immunoprecipitation (ChIP; Kuo and Allis 1999; Orlando, 2000) has emerged in the last decade as part of a standard toolbox for analyzing protein-DNA interactions in a more physiological context, as it detects them directly in their native environments. In the past several years, it has been widely used to study the in vivo association of a particular DNA sequence with regulatory proteins and various forms of histones (i.e., their post-translationally modified forms and replacement variants).

Despite its conceptual simplicity, the ChIP protocol requires fine tuning for each individual protein and cell type. Compared to affinity purifications from cellular and nuclear extracts, the most common problem in ChIP is high nonspecific background, caused by the easy aggregation of chromatin and its binding to surfaces. Accordingly, the success of ChIP crucially depends on the quality of antibodies used. ChIP grade antibodies have to tolerate stringent binding and washing conditions, and also recognize the parts of the protein that are (1) not involved in the interaction under question and (2) minimally affected by the formaldehyde cross-linking.

Some of the limitations of the native antibody approach are overcome by using epitope-tagging, i.e., by expression in the cells of the protein of interest fused to an epitope tag recognized by a well characterized antibody (Jarvik and Telmer, 1998; Wang, 2009). Among the advantages of this method are standardization and commercial availability of relatively inexpensive epitope-tagging systems. Unlike using antibody against a native protein, the epitope-tagging approach lends itself to an easy and efficient negative control, where the nonspecific signal is detected by a parallel purification from a sample that does not express an epitope-tagged protein (i.e., mock purification). Epitope tagging is a preferred method for comparing closely related proteins when the antibodies for distinguishing between the different protein forms are not readily available (e.g., mutant or alternative splicing variants of the same protein). Finally, the binding of the tag to antibody usually does not compete with the protein-DNA interaction under study, as compared to the native antibody. Nonetheless, given that epitope-tagging might affect the protein function, the use of several versions of epitope tagged protein of interest, typically with their N- or C- termini modified, is recommended.

The interaction between biotin and streptavidin (K_d of 10^{-15} M) is one of the strongest known noncovalent interactions (Lindqvist and Schneider, 1996). Many commercially available reagents for purification and detection of biotinylated macromolecules have been developed. Accordingly, to take advantage of this interaction, we (Viens et al., 2004) and others (van Werven and Timmers, 2006; Ooi et al., 2010) have developed biotin-ChIP, a

variant of the epitope-tagging approach to ChIP, based on protein biotinylation in vivo. Such an approach, in its various forms, has been applied to a broad variety of experimental systems, including mammalian cells, yeast, insects, and worms (van Werven and Timmers, 2006; Kim et al., 2009; Kolodziej et al., 2009; Lausen et al., 2010; Ooi et al., 2010). The main advantage of this method for ChIP is the more stringent washing conditions, as well as more efficient (i.e., more quantitative) recovery of the protein-DNA complexes (Figure 17.12.4). Note that the in vivo biotinylation of histones has been also used for native ChIP (NChIP) format (i.e., without cross-linking, Ooi et al., 2010), which takes advantage of the exceptionally strong and specific binding between biotin and streptavidin and allows one to reduce the volume of affinity medium, and thus decrease nonspecific background. Moreover, although not tested yet, one would expect that in vivo biotinylation can be equally useful for many modifications of ChIP developed to date, e.g., 3C-5C (Vassetzky et al., 2009) and RNAChIP (Gilbert and Svejstrup, 2006; Percipalle and Obrdlik 2009).

Critical Parameters

Transfection efficiency (Basic Protocol 1, steps 1 to 5)

When using calcium phosphate transfection, the pH of the HeBS buffer is crucial. The precipitate should be barely visible under microscope. For control of transfection efficiency, we recommend transfecting a separate cell well with BAP-GFP (or any other fluorescent protein); this sample can also be later used as a negative control in ChIP.

Cross-linking (Basic Protocol 1, steps 6 to 16)

Excessive cross-linking can negatively affect the efficiency of chromatin fragmentation. The empirically chosen dose of formaldehyde (1% final solution, 10 min) might require optimization, depending on the cell type and protein used. Several samples with different cross-linking times should be tested. In most cases, 8 to 10 min for cross-linking is sufficient for ChIP with mammalian cells.

Fragmentation (Basic Protocol 1, steps 17 to 20)

The desired size of chromatin fragments depends on the downstream analysis. Generally, it should be ~400 to 500 bp. Longer fragments will diminish resolution and increase nonspecific signal; therefore, they are not recommended for high-throughput analysis (microarray analysis or sequencing). Fragments that are too short are generally not good for PCR detection, as it becomes challenging to find suitable primers for amplification.

The sonication parameters (pulse duration, intensity, and number) vary, depending on the extent of cross-linking and cell type; thus, they must be optimized for each experiment. Ideally, the least amount of input energy that gives satisfactory fragmentation should be used.

Affinity purification (Basic Protocol 1, 33 to 41)

After cell lysis (steps 9 to 16), the sample will contain biotin, which will compete with the biotinylated chromatin for binding to streptavidin in the downstream pull-down step. Thorough washing of medium from the cells before cell lysis is therefore critical.

The affinity between biotinylated BAP and streptavidin (steps 33 to 41) can be weakened by steric hindrance, depending on whether the tagged part of the protein (e.g., its N or C terminus) is well exposed or buried. Accordingly, the incubation times and washing stringency could be adjusted. Typically, a 3-hr incubation is sufficient for successful pull-down of biotinylated chromatin. If longer incubation times are used, the length of the preclearing incubation step (step 30) should be increased accordingly.

IMDA (Basic Protocol 2)

A DNA concentration of <10 ng/μl will lead to self-circularization of the fragments obtained in ChIP, instead of the formation of concatemers. As the result, the fragments will be amplified individually, and their sequence and size-dependent bias in amplification efficiency will not be averaged out. Therefore, the first step in applying ligation-IMDA is measuring DNA concentration, which will determine if the stuffer DNA is necessary.

Troubleshooting

Insufficient fragmentation of DNA after sonication

When monitoring the size of sheared DNA after sonication, one can observe insufficient fragmentation. There could be several causes for this problem: (a) material is overcross-linked (Basic Protocol 1, steps 6 to 16), (b) cells are insufficiently lysed (step 17), or (c) sonication is inefficient (steps 18 to 20). Accordingly, take the following step: (a) Cross-linking time could be reduced.

(b) If there are reasons to suspect that too many cells in a small volume of ChIP buffer were used for lysis, additional ChIP buffer should be added. Alternatively, cells can be snap-frozen in liquid nitrogen and thawed. The success of lysis can be estimated visually by inspecting the sample under the microscope. (c) In the case of insufficient sonication, the duration, energy, and/or number of rounds of sonication should be increased.

Low or absent PCR signal after ChIP

Low or no PCR signal after ChIP could have several causes: (a) insufficient amount of chromatin for immunoprecipitation (Basic Protocol 1, steps 33 to 37), (b) failure of immunoprecipitation, or (c) failure of PCR. Accordingly, (a) the amount of cells or chromatin should be increased. Note that it may be difficult to extract all chromatin from certain primary cell types. (b) The biotinylation efficiency of the target protein should be tested, as well as the quality of streptavidin-beads. In addition, one should make sure that the cells were washed thoroughly before lysis, as biotin in the medium will compete with the biotinylated protein for binding (step 16). (c) One should set up a control qPCR with the same primers on genomic DNA, and optimize PCR conditions.

Elevated background signal

Instead of low PCR signal, one may observe an elevated background signal. This can be caused by (a) nonspecific immunoprecipitation or (b) insufficient washes. Accordingly, one should (a) reduce duration of incubation of streptavidin-beads with chromatin (step 33) and ensure the tube-switch step (steps 35 and 39) is performed and (b) increase the number and/or stringency of the washes after IP (steps 38 to 41).

Anticipated Results

We usually obtain ~15 to 20 ng/μl of final ChIP DNA from biotin-ChIP (total 30 μl). In total, 5% of biotin-ChIP material is used in PCR to test the quality of the sample. In most of the cases, we observe enrichment of 5 to 100-fold from the target loci.

The use of DNA amplification (e.g., IMDA) should allow the analysis of more genomic sites and of the whole genome.

Time Considerations

For the biotin-ChIP protocol, transient transfection takes 2 days before cell harvest. The cell lysis, preclearing, binding, and washing procedures take 1 full day, with the DNA eluted from beads and ready for PCR analysis the next day. If necessary, the preclearing and binding to streptavidin beads can take longer than 3 hr, e.g., overnight for both steps. With practice, ~10 samples can be processed in parallel.

The IMDA procedure takes 1 day, starting from DNA measurement and ligation. The longest part is amplification, which can take a minimum of 6 hr, but can also be extended overnight.

The streptavidin gel-shift experiment typically takes 1 day, as it includes SDS-PAGE separation, transfer to PVDF (or nitrocellulose) membrane, blocking, incubation with primary and secondary antibodies, and enhanced chemiluminescence (ECL) detection.

Literature Cited

Dahl, J.A. and Collas, P. 2008. A rapid micro chromatin immunoprecipitation assay (microChIP). *Nat. Protoc.* 3:1032-1045.

de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F., and Strouboulis, J. 2003. Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc. Natl. Acad. Sci. U.S.A.* 100:7480-7485.

Gilbert, C. and Svejstrup, J.Q. 2006. RNA immunoprecipitation for determining RNA-protein associations in vivo. *Curr. Protoc. Mol. Biol.* 75:27.4.1-27.4.11.

Gilmour, D.S. and Lis, J.T. 1984. Detecting protein-DNA interactions in vivo: Distribution of RNA polymerase on specific bacterial genes. *Proc. Natl. Acad. Sci. U.S.A.* 81:4275-4279.

Gold, L., Polisky, B., Uhlenbeck, O., and Yarus, M. 1995. Diversity of oligonucleotide functions. *Annu. Rev. Biochem.* 64:763-797.

Hager, G.L., Elbi, C., and Becker, M. 2002. Protein dynamics in the nuclear compartment. *Curr. Opin. Genet. Dev.* 12:137-141.

Jarvik, J.W. and Telmer, C.A. 1998. Epitope tagging. *Annu. Rev. Genet.* 32:601-618.

Kim, J., Cantor, A.B., Orkin, S.H., and Wang, J. 2009. Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. *Nat. Protoc.* 4:506-517.

Kolodziej, K.E., Pourfarzad, F., de Boer, E., Krpic, S., Grosveld, F., and Strouboulis, J. 2009. Optimal use of tandem biotin and V5 tags in ChIP assays. *BMC Mol. Biol.* 10:6.

Kuo, M.H. and Allis, C.D. 1999. In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods* 19:425-433.

Lausen, J., Pless, O., Leonard, F., Kuvardina, O.N., Koch, B., and Leutz, A. 2010. Targets of the Tal1 transcription factor in erythrocytes: E2 ubiquitin conjugase regulation by Tal1. *J. Biol. Chem.* 285:5338-5346.

Lindqvist, Y. and Schneider, G. 1996. Protein-biotin interactions. *Curr. Opin. Struct. Biol.* 6:798-803.

Mechold, U., Gilbert, C., and Ogryzko, V. 2005. Codon optimization of the BirA enzyme gene leads to higher expression and an improved efficiency of biotinylation of target proteins in mammalian cells. *J. Biotechnol.* 116:245-249.

Molloy, P.L. 2000. Electrophoretic mobility shift assays. *Methods Mol. Biol.* 130:235-246.

Negre, N., Lavrov, S., Hennetin, J., Bellis, M., and Cavalli, G. 2006. Mapping the distribution of chromatin proteins by ChIP on chip. *Methods Enzymol.* 410:316-341.

O'Neill, L.P. and Turner, B.M. 2003. Immunoprecipitation of native chromatin: NChIP. *Methods* 31:76-82.

Ooi, S.L., Henikoff, J.G., and Henikoff, S. 2010. A native chromatin purification system for epigenomic profiling in *Caenorhabditis elegans*. *Nucleic Acids Res*. 38:e26.

Orlando, V. 2000. Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem. Sci.* 25:99-104.

Park, P.J. 2009. ChIP-seq: Advantages and challenges of a maturing technology. *Nat. Rev. Genet.* 10:669-680.

Pashev, I.G., Dimitrov, S.I., and Angelov, D. 1991. Crosslinking proteins to nucleic acids by ultraviolet laser irradiation. *Trends Biochem. Sci.* 16:323-326.

Percipalle, P. and Obrdlik, A. 2009. Analysis of nascent RNA transcripts by chromatin RNA immunoprecipitation. *Methods Mol. Biol.* 567:215-235.

Phair, R.D. and Misteli, T. 2000. High mobility of proteins in the mammalian cell nucleus. *Nature* 404:604-609.

Revzin, A. 1989. Gel electrophoresis assays for DNA-protein interactions. *Biotechniques* 7:346-355.

Shoaib, M., Baconnais, S., Mechold, U., Le Cam, E., Lipinski, M., and Ogryzko, V. 2008. Multiple displacement amplification for complex mixtures of DNA fragments. *BMC Genomics* 9:415.

van Werven, F.J. and Timmers, H.T. 2006. The use of biotin tagging in *Saccharomyces cerevisiae* improves the sensitivity of chromatin immunoprecipitation. *Nucleic Acids Res*. 34:e33.

Vassetzky, Y., Gavrilov, A., Eivazova, E., Priozhkova, I., Lipinski, M., and Razin, S. 2009. Chromosome conformation capture (from 3C to 5C) and its ChIP-based modification. *Methods Mol. Biol.* 567:171-188.

Veenstra, T.D. 1999. Electrospray ionization mass spectrometry: A promising new technique in the study of protein/DNA noncovalent complexes. *Biochem. Biophys. Res. Commun.* 257:1-5.

Viens, A., Mechold, U., Lehrmann, H., Harel-Bellan, A., and Ogryzko, V. 2004. Use of protein biotinylation in vivo for chromatin immunoprecipitation. *Anal. Biochem.* 325:68-76.

Viens, A., Harper, F., Pichard, E., Comisso, M., Pierron, G., and Ogryzko, V. 2008. Use of protein biotinylation in vivo for immunoelectron microscopic localization of a specific protein isoform. *J. Histochem. Cytochem.* 56:911-919.

Voytas, D. 2001. Agarose gel electrophoresis. *Curr. Protoc. Mol. Biol*. 51:2.5A.1-2.5A.9.

Wang, Z. 2009. Epitope tagging of endogenous proteins for genome-wide chromatin immunoprecipitation analysis. *Methods Mol. Biol.* 567:87-98.

Weinmann, A.S. and Farnham, P.J. 2002. Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. *Methods* 26:37-47.

Wolffe, A. 1999. Chromatin: Structure and Function. Academic Press, London.

Zeng, P.Y., Vakoc, C.R., Chen, Z.C., Blobel, G.A., and Berger, S.L. 2006. In vivo dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation. *Biotechniques* 41:694-698.

10

Annex - IV

PUB-NChIP "In Vivo Biotinylation" Approach to Study Chromatin in Proximity of a Protein of Interest (2011)

Muhammad Shoaib [1], Arman Kulyyassov [1,2] , Chloé Robin [1], Kinga Winczura [1], Patricia Kannouche [1], Erlan Ramankulov [1,2], Marc Lipinski [1] and Vasily Ogryzko [1]

I CNRS UMR-8126, Université Paris-Sud 11, Institut de Cancérologie Gustave Roussy, 39, rue Camille Desmoulins, 94805, Villejuif, France.

II National Center for Biotechnology of the Republic of Kazakhstan, 43 Valikhanova Str., 010000, Astana, Republic of Kazakhstan.

In preparation

PUB-NChIP – “in vivo biotinylation” approach to study chromatin in proximity to a protein of interest.

Muhammad Shoaib[1], Arman Kulyyassov[1,3], Chloé Robin[1], Kinga Winczura[1], Patricia Kannouche[2], Erlan Ramanculov[3], Marc Lipinski[1] and Vasily Ogryzko[1]

1 - UMR8126, Univ. Paris-Sud 11, CNRS, Institut de cancérologie Gustave Roussy, 94805 Villejuif, France

2 - UMR8200, Univ. Paris-Sud 11, CNRS, Institut de cancérologie Gustave Roussy, 94805 Villejuif, France

3 - National Center of Biotechnology, Valihanova 43, 01000, Astana, Kazakhstan

Abstract

Chromatin ImmunoPrecipitation (ChIP) does not allow direct analysis of histone post-translational modifications in the vicinity of the chromatin protein of interest. This information has to be indirectly inferred from separate ChIP experiments using antibodies against the protein and specific histone PTMs. We have developed an alternative approach named PUB-NChIP (Proximity Utlilizing Biotinylation with Native ChIP) allowing one to circumvent this problem. It is based on coexpression of a) a protein of interest, fused with the bacterial biotin ligase BirA together with b) a histone fused to BAP biotin acceptor peptide, which is specifically biotinylated by BirA. We demonstrate that BAP-fused histones are labeled by biotin specifically in the locations of a model protein, Rad18. The biotinylated chromatin can be purified and the PTMs analyzed by western blot and/or mass spectrometry. Moreover, using BAP fusions with alternative histone variants, instead of canonical histones, one can analyze chromatin in the context of a particular functional state, marked by this histone. Finally, biotin pulse-chase experiments allowed us to analyze the histone PTMs at a determined time after they were in the proximity with the protein of interest.

Introduction

Chromatin Immunoprecipitation (ChIP) [1-3] has been widely used for over a decade to study the in vivo association of a particular DNA sequence with regulatory proteins. When combined with high throughput approaches in the form of ChIP-on-chip or ChIP-seq [4-7], it complements gene expression profiling as a way to reconstruct and analyze gene regulatory networks.

Although both DNA and protein components are purified during the ChIP procedure, only the DNA is typically analyzed, and the protein part is discarded. This is because the cross-linking procedure used during ChIP, which is necessary to fix the protein of interest on DNA, irreversibly damages the protein component of chromatin (targeting mostly lysines), rendering it difficult to analyze by mass spectrometry or western blot. On the other hand, given that it is increasingly recognized that epigenetic regulation is of utmost importance in both normal and disease states, the analysis of histone modifications and replacement histone variants could provide valuable information about the state of chromatin in the vicinity of a particular protein of interest.

With the aim of addressing this limitation of ChIP, we have developed a new technique called PUB-NChIP (Proximity Utilizing Biotinylation and Native Chromatin ImmunoPrecipitation). It is based on coexpression of a) the protein of interest (transcription factor, or any other nuclear protein), fused to the bacterial biotin ligase BirA, together with b) a histone fused to biotin acceptor peptide (BAP), which is specifically biotinylated by BirA. With the incorporation of BAP-histones into chromatin, the chromatin located in the proximity of the BirA-fusion protein of interest is preferentially biotinylated. The subsequent steps of the procedure take advantage of the strong interaction between core histones and DNA, which allows one to avoid cross-linking for the purification of the biotinylated chromatin (so-called NChIP [8, 9]). In the absence of cross-linking-induced modifications, analysis of the protein component of the chromatin (e.g., histone post-translational modifications) becomes feasible.

Another advantage of our technique is the possibility of using, instead of the canonical histones, their alternative variants expressed as BAP-fusions. Alternative variants of the canonical core histones share the same overall structure with the canonical histones, but differ in primary sequence from their canonical relatives. Most importantly, their presence was found to correlate with particular functional states of chromatin; for example, silenced chromatin is enriched in histone macroH2A and depleted of H2A.BBD [10-12]. These features make the variant histones a convenient tool for purification of alternative chromatin states. Using the BAP fusions of these histones instead of the canonical histones allows one to analyse the histone PMTs associated with the protein of interest, but now in the context of a particular functional state of chromatin – e.g., using BAP-macroH2A, which marks repressed chromatin, or BAP-H2A.BBD, which marks mostly active chromatin, etc.

Finally, we also demonstrate that given the introduction of a covalent mark into the protein of interest, one can follow the state of histone PTMs at a defined times after it came into close proximity with the BirA-fusion. Thus, in addition to helping in direct analysis of histone PTMs in the proximity to the nuclear protein of interest, our approach adds a temporal dimension to such analyses.

Materials and Methods

Recombinant DNA

The vectors for expression of BAP- and BirA- fusions, as well as BirA-GFP, BirA-Rad18, BAP-H2A, BAP-macroH2A2 and BAP-H2ABBD expression vectors have been described elsewhere [13]. The H2AZ and H3.1 ORFs were subcloned from pOZ.FHH vectors [14], whereas the PolH ORFs (wild type and mutant) were amplified by PCR and inserted to the BirA- fusion vector using XhoI and NotI restriction sites.

Cell culture

HEK 293T cells were grown in Dulbecco's modified Eagle's medium with high glucose (PAA) and 10% fetal bovine serum (FCS, PAA). For transient transfection and Western analysis, a standard calcium phosphate precipitation method was used, and the cells were analyzed one or two days after transfection, as indicated. For the biotin labeling *in vivo*, cells were grown for several days before transfection in the DMEM supplemented with dialyzed FCS, and for the specified time of labeling, biotin (Sigma) was added to a final concentration of 5 μg/ml, while the pH was stabilized by addition of 50 mM HEPES (pH 7.35) to the medium. For SILAC experiments, the cells were grown in DMEM with $^{12}C_6$ L-lysine and $^{12}C_6$ $^{14}N_4$ L-arginine or $^{13}C_6$ L-lysine $^{13}C_6$ $^{15}N_4$ L-arginine (Thermo Scientific, cat # 89983) for at least 5 divisions before transfection, and kept in the same medium until harvest. For the Western analysis, $3 \cdot 10^5$ cells (corresponding to one well in a 6-well plate) were used for transfection. For the mass-spectrometry analysis, $3 \cdot 10^6$ cells were transfected per data point. DNA damage was induced with UVC (20 J/m^2) or Ionizing radiation (10 Gy). The time of incubation after various treatments is indicated in the results section.

Biochemistry and Western analysis

Except where indicated otherwise, cell nuclei were used for analysis. They were prepared by cell disruption in CSK buffer (100 mM NaCl, 300 mM Sucrose, 10 mM Tris pH 7.5, 3 mM $MgCl_2$, 1 mM EGTA, 1.2 mM PMSF, 0.5% Triton X-100), and centrifugation for 5 min at 4000 rpm. For the analysis of chromatin-associated histones (chromatin fraction), the nuclei were first incubated in CSK buffer containing 450 mM NaCl by 30 min rotation at 4°C, then spun at 4000 rpm, and the supernatant containing soluble histones was discarded. For Western analysis, 1x NuPAGE LDS Sample buffer (Invitrogen, NP0007) with DTT (10 mM) was added, the nuclei were sonicated, boiled for 5 min at 96° C and loaded on 4-12% gradient Novex Tris-Glycine precast gels (Invitrogen, NP0315). After separation, the proteins were transferred to nitrocellulose membranes and probed with HRP-conjugated streptavidin (Sigma, # S5512) or HRP conjugated α-PentaHis antibody (QiaGen, # 34460) according to the manufacturer's protocol, except that for the detection of biotinylated proteins by the HRP-conjugated streptavidin, 500 mM NaCl was added to the washing buffer (PBS + 0.1% Tween). For protein visualization, the gels were stained with PageBlue (Fermentas, # R0579). For the densitometric analysis of Western blots, the program ImageJ 1.42q (freely available online) was used. To compare the biotinylation levels between different samples, the value of the streptavidin-HRP signal was normalized to the value of the α-PentaHis signal, which reflects the amount of the transfected protein regardless of its biotinylation status. Every Western analysis was performed three times; a representative figure is shown in each case.

For Ni^{2+}NTA purification of 6XHis tagged proteins, the nuclei were solubilized in buffer A (10% glycerol, 250 mM NaCl, 6 M Guanidine-HCl, 20 mM TrisHCl (pH 8.0), 0.1% Tween) by rotation for 30 min at 4° C. 1/10 of the volume of Ni^{2+}NTA agarose, prewashed in the same buffer, was added to the lysate, followed by rotation at 4° C for 3 h. The beads with bound proteins were washed twice with buffer A, then twice with buffer B (10% glycerol, 250 mM NaCl, 20 mM TrisHCl (pH 8.0), 0.1% Tween, 0.2 mM PMSF and cOmplete protease inhibitor cocktail (Roche)). The bound proteins were eluted by incubating the beads in 4 volumes of buffer C (10% glycerol, 250 mM NaCl, 20 mM Tris HCl (pH 8.0), 0.1% Tween, 0.2 mM PMSF, 300 mM imidazole, 50 mM EDTA) and concentrating the sample by ultrafiltration with Microcon YM-10 (Millipore).

NChIP

Cells were washed with PBS and lysed in 1 mL of CSK buffer (with protease inhibitor cocktail (Roche; 1697498), 10 mM Na-Butyrate, 2 mM PMSF, 5 mM nicotinamide (Sigma; N5535), 5 mM Na-orthovanadate (Sigma; S6508)) for 5 min at RT. Nuclei were centrifuged at 4000 rpm for 10 min and kept at 4°C or else at –20°C if not used immediately. For micrococcal digestion, the nuclei were resuspended in 500 µL of TM2 buffer (10 mM Tris-HCl, 2 mM $MgCl_2$, 0.1% Triton, inhibitors), and 2.5 µL of 0.5 M $CaCl_2$ and 3 µL of micrococcal nuclease (MNase) (1 U/µL) (Sigma; N3755) were added before incubation at 37°C for 10 min. The reaction was stopped by adding 15 µL of 0.1 M EGTA. Nuclei were collected by spinning at 400 ***g*** (2000 rpm in Eppendorf tabletop centrifuge) for 10 min at 4°C. After MNase digestion, the nuclei pellet was resuspended in 500 µL of pre-chilled 0.4 M salt extraction buffer (385 mM NaCl, 10 mM Tris-HCl pH 7.4, 2 mM $MgCl_2$, 2 mM EGTA, 0.1% Triton X-100 with inhibitors). The tubes were rotated at 4°C for 30 min. Supernatant containing digested chromatin was separated from the remaining material by spinning at 400 ***g*** (2000 rpm in Eppendorf tabletop centrifuge) for 10 min at 4°C. An equal volume of TM2 buffer was added to the extracted chromatin to arrive at a 0.2 M final salt concentration. The sample was centrifuged at maximum speed (13,000 ***g***, i.e., 16,000 rpm) in an Eppendorf centrifuge) for 5 min at 4°C. 200 µL of suspension of Sepharose-streptavidin beads (GE Healthcare; 17-5113-01) were washed 3 times with 0.2 M salt extraction buffer containing Triton and finally resuspended in 100 µL of the same buffer. The bead suspension was combined with chromatin and rotated for 3 h at 4°C. Afterwards, the beads were washed twice in 500 µL of 0.4 M salt extraction buffer containing Triton for 5' at 4°C. The biotinylated chromatin was eluted from the beads by adding 100 µL of 1X LDS sample buffer (with reducing agent + 2% SDS), heating at 99°C for 10 min, and separating the beads by a quick spin in the tabletop centrifuge for 10 sec. Elution was repeated a second time, and the two eluates were combined.

Mass spectrometry

The protein bands were excised from the gel and were chemically derivatized using propionic anhydride[15]. Briefly, 100 µL of freshly prepared 75% propionic anhydride in methanol was added to each tube and immediately supplemented by 40 µL of 50 mM Ammonium Bicarbonate. After 30 min at 37°C, the solution was discarded and propionic anhydride derivatization was repeated again. The peptides were digested by trypsin as in [16]. The gel slices were dehydrated with 300 µl of 50% acetonitrile followed by 300 µl of 100% acetonitrile, then re-hydrated with 300 µl of 50 mM ammonium bicarbonate. A final dehydration was performed with 2 washes of 300 µl of 50% acetonitrile, followed by 2 washes of 300 µl of 100% acetonitrile. Each wash was carried out for 10 min at 25° C with shaking at 1400 rpm. The gel slices were dried in a SpeedVac at 35° C for 10 min. For trypsin digestion,

the gel slices were pre-incubated with 7 ml of 15 ng/ml trypsin (Promega # V5280) at room temperature for 10 min. Afterwards, 25 µl of 50 mM ammonium bicarbonate was added, and the gel slices were incubated at 37° C for 16 h. The peptide-containing supernatants were dried at 56° C by SpeedVac for 30 min, then resuspended in 20 µl of solution containing 0.05% formic acid and 3% acetonitrile for mass spectrometry experiments. The resulting peptides were analyzed with a nano-HPLC (Agilent Technologies 1200) directly coupled to an ion-trap mass spectrometer (Bruker 6300 series) equipped with a nano-electrospray source. The separation gradient was 7 min from 5% to 90% acetonitrile. The fragmentation voltage was 1.3V. For quantification of the presence of a particular peptide in the sample, the ion trap was set in MRM mode, i.e., it was set to isolate, fragment and MS/MS scan several parental ions having predetermined M/Z ratios. The relative quantity of each peptide in the different fractions was estimated by comparison between the sum of peak areas in the Extracted Ion Chromatograms (EIC) for several most intensive daughter ions of this peptide obtained from MRM analysis of these fractions.

Immunofluorescent Microscopy

HEK 293T cells were washed two times with cold PBS and fixed by immersion in –20°C methanol for 5 min. These fixed cells were washed with cold PBS and then permeabilized with 0.2% Triton X-100 in PBS for 10 min at 4°C and submitted to additional fixation in 4% formaldehyde for 20 min at RT. Cells were then washed twice with PBS at room (RT). Blocking was done with 3% BSA for 30 min under mild agitation. Cells were then incubated with mouse monoclonal anti-Rad18 antibody (Abcam; ab57447) followed by 3 washes in PBS (1X) at RT. Cells were then incubated with goat anti-mouse secondary antibody conjugated with Alexa-488 (Invitrogen; A11017) for 1 h (1/1000). Streptavidin-Cy3 conjugate (Sigma; S-6402) and ToPro (Invitrogen) was also added (1/500 and 1/1000 respectively) to the same incubation mix. After 3 washes in PBS (1X) at RT, cover slips were mounted on glass slides using VectaShield mounting medium (Vector Laboratories; H-1000). Cells were visualized with a Zeiss LSM 510 Meta confocal scanning microscope, using a Plan-Apochromat 63x 1.4 oil immersion objective. Imaging was performed with sequential multitrack scanning using the 488 and 543 nm laser wavelengths separately. The colocalization analysis was performed with LSM Examiner software. Pearson's correlation coefficient (Figure 3A) was calculated with setting the threshold signal common for all images compared (15 cells for both experiment and control samples). For histones H2A and H3 staining, cells were incubated with α-H2A antibody (Cell Signaling; 2572) or γ-H2AX (Cell signaling; 2577S) and α-H3 antibody (Abcam; ab1791), respectively.

Results

1. Analysis of post-translational modifications of a specific protein fraction using Proximity Utilizing Biotinylation (PUB).

PUB-MS (Proximity Utilising Biotinylation) has been developed in our laboratory, originally as a new method to study protein-protein proximity (PPP-networks) in vivo using mass spectrometry (the vectors are described elsewhere, [[13]], Figure 1). A similar method, although not designed for mass-spectrometry analysis, was also developed in the laboratory of Dr. A.Y. Ting [17]. The general principle of the approach is based on coexpression of two proteins of interest fused to a biotin acceptor domain (BAP) in one case and a biotin ligase enzyme (BirA), in the other [13]. The BAP domain was modified to weaken its affinity to BirA and thus to decrease the background biotinylation levels. Proximity between the two proteins (e.g., due to their interaction) leads to a more efficient biotinylation of the BAP domain, which can be either detected by western blotting, or else quantified using proteomics approaches, such as combination of multiple reaction monitoring (MRM) analysis and stable isotope labeling in vivo (SILAC) [18].

The same protein can exist in several complexes in the cell, and its properties can vary depending on the proximity partners (PP). Given that PUB allows a PP-dependent labeling of a fraction of the BAP-fusion of the protein of interest, one can expand the scope of the PUB approach beyond the mere study of protein-protein proximity. Namely, one can purify a particular PP-dependent subfraction of a BAP-fused protein biochemically, and then study its properties, e.g., its post-translational modifications (PTMs).

To test this possibility experimentally, we have analyzed the ubiquitination status of the subset of replication processivity factor PCNA [19] that is located in proximity to the translesion polymerase PolH after UV irradiation of cells [20, 21]. Ubiquitination is a convenient model to study PTMs, because it is easy to detect as a shift in protein mobility on SDS-PAGE. UVC irradiation induces ubiquitination of PCNA, which then recruits PolH to the sites of DNA damage [22, 23]. The PolH has PCNA-binding and ubiquitin binding domains (PIP and UBZ, respectively), both believed to be involved in the interaction.

To analyze the ubiquitination status of PCNA located in proximity to PolH in vivo, the cells were transfected with plasmids expressing BAP-PCNA and BirA-PolH. Two days later, UV damage was induced; and 6 h after that, the cells were pulse-labeled with biotin. These cells were harvested, and the biotinylated BAP-PCNA was separated from the non-biotinylated BAP-PCNA using streptavidin-sepharose pulldown. The ubiquitinylation status of the two fractions was analyzed by western blotting with α–PCNA antibodies. As one can see from the Figure 1B (lane 5), a significant part of the biotinylated BAP-PCNA is present in a ubiquitinated form, whereas this form is absent from the non-biotinylated BAP-PCNA fraction remaining in the flowthrough (lane 3). A control experiment, performed in parallel with a BirA-GFP fusion, showed no such enrichment in the biotinylated BAP-PCNA fraction (lanes 2 and 4).

To confirm that the ubiquitinated fraction of BAP-PCNA is preferentially biotinylated due to its specific interaction with PolH, we used mutations in PolH lacking the PIP or/and UBZ domains in our next experiments. This time, the ubiquitinilation status of the biotinylated BAP-PCNA was analyzed directly in nuclear lysates by western blotting with streptavidin-HRP. As one can see in Figure 1C (middle), the higher-molecular-weight form of biotinylated PCNA

(lane 1) was decreased when we used the BirA-PolH$_{\Delta PIP}$ and BirA-PolH$_{\Delta UBZ}$, and it practically disappeared when the double mutant BirA-PolH$_{\Delta\Delta}$ was used (lanes 4, 5 and 3, respectively). Consistent with the notion that the higher-molecular-weight fraction of BAP-PCNA corresponds to ubiquitinated BAP-PCNA, a BAP-PCNA carrying a mutation in the ubiquitination site showed no such fraction when cotransfected with BirA-PolH in a parallel experiment.

We conclude from these experiments that purification of the biotinylated fraction of a BAP-fusion of a protein of interest allowed us to study PTMs specific for the particular fraction of this protein that is located in proximity to another protein of interest, fused to BirA. Whereas in the case of ubiquitination, the PTM analysis could be accomplished by western blot analysis of cellular extracts with streptavidin-HRP, in most other cases it will require biochemical purification of the biotinylated fraction of the protein.

2. PUB-NChIP reveals a specific pattern of H4 acetylation in the Rad18-proximal chromatin.

Next, we applied the same principle for the analysis of chromatin in close proximity to a protein of interest fused to BirA. For this purpose, we fused the BAP domain to the core histone proteins H3.1 and H2A. After their incorporation into chromatin, the biotinylation of the BAP-histone fusions by a protein of interest fused to BirA is expected to lead to specific labeling of chromatin located in the proximity of the BirA-fusion protein. Purification and analysis of such biotinylated chromatin is the principle behind the PUB-NChIP approach.

To verify that mainly the BAP-histones that are located in proximity to a BirA fusion are biotinylated under such conditions, we used repair-related Rad18 protein as a convenient model. This protein is known to form characteristic foci in the nuclei, clearly seen after UV treatment of cells (these foci are believed to correspond to specialized compartments in the nuclei, containing repair and replication proteins, including PCNA). HEK 293T cells were cotransfected with the BirA-Rad18 fusion together with BAP-H2A. Two days after transfection, the cells were irradiated with UVC and pulse-labeled with biotin 6 h later, i.e., at the time when Rad18 forms distinct foci in the nuclei. The cells were fixed immediately after the biotin pulse. Staining of the pulse-labeled cells with streptavidin-Cy3 revealed foci of biotinylation signal, which colocalized with Rad18 (Figure 2A, top), detected by specific antibodies. Importantly, staining with α–H2A antibodies that detect H2A regardless of its biotinylation status, gave a very different pattern, consistent with the known homogeneous distribution of H2A histone throughout the nucleus (Figure 2A, bottom). Cotransfection of BAP-H2A with control BirA-GFP construct also produced a homogeneous biotinylation pattern (Figure S1A). Western blot analysis (Figure 2B, bottom left) shows that only the BAP-H2A histone is biotinylated in cells under these conditions, hence the Cy3 signal that we observe in microscopy should reflect the localization of biotinylated BAP-H2A. On the other hand, labeling with biotin for different lengths of time shows that during the 15 min of exposure to biotin, only a small portion of the BAP-H2A is biotinylated (Figure 2B, left). Overall, these data clearly demonstrate that mainly the subfraction of BAP-H2A containing chromatin located in proximity to Rad18 was biotinylated in our experiments. Similar results were obtained when we used BAP-H3.1 fusion instead of BAP-H2A (Figure 2B, supplementary figure Figure S1B).

The biotinylated chromatin was purified on streptavidin sepharose without prior crosslinking (See **Materials and Methods**, **NChIP**). Judging by the DNA content of the input, flow-through and eluate samples, only a small fraction of total chromatin was purified in this

experiment (Figure 2C, top); however, western blot analysis of the same fractions with streptavidin-HRP (Figure 2C, middle), shows that all of the biotinylated BAP-H2A was in fact purified. As one can see from Figure 2C (bottom, lane 7), the composition of the purified histone octamers is consistent with the presence of a single BAP-tagged H2A/H2B protein. Whereas, in both input and flowthrough samples (lanes 5,6), all core histones are present in stoichiometric amounts, there is approximately twofold less H2A in the eluate. Confirming that not every BAP-H2A molecule is biotinylated in our experiment, the flowthrough also contains a BAP-H2A band stained by Coomassie (Figure 2B, bottom, lane 6). Note that after elution from streptavidin beads by boiling with Laemmli buffer, significant amounts of streptavidin appear in the sample, overlapping with the H4 histone band (Figure 2C, bottom, lane 7).

Next we wished to verify that PUB-NChIP purified chromatin is enriched in the expected post-translational modifications of histones. As a well-characterized histone PTM, we have chosen γ-H2AX, which forms foci in the nucleus after ionizing irradiation. It was previously shown that Rad18 colocalizes with γ-H2AX under these conditions [24, 25], and we too observed that the chromatin that is proximity biotinylated by BirA-Rad18 often colocalizes with γ-H2AX after ionizing irradiation in 293T cells (Figure S2A, top). The control protein, HP1α, did not show any histone biotinylation that was colocalized with the γ-H2AX foci (bottom). Western blot analyses of histones prepared by PUB-NChIP have shown that, consistent with the observed colocalization between Rad18-induced biotinylation signal and γ-H2AX, the Rad18 proximal chromatin is indeed enriched in γ-H2AX (Figure S2B, bottom).

Next we returned to our principal experimental system, i.e., 293T cells transfected with "BirA-Rad18 + BAP-H2A" and irradiated with UVC. We set out to analyze the acetylation status of the N-terminus of histone H4 in the Rad18-proximal chromatin by mass spectrometry and to compare it with control BirA-GFP biotinylated chromatin. We used SILAC (Stable Isotope Labeling with Aminoacids in Culture [18]) in order to quantitatively compare the presence of specific modifications in the chromatin in proximity to Rad18, as compared to that seen with the GFP-BirA fusion. The cells were grown in medium containing either a light version of lysine and arginine aminoacids or their heavy counterparts (i.e., labeled with C^{13} and N^{15}). The cells were cotransfected with BAP-H2A and either BirA-RAD18 (heavy labeled cells) or BirA-GFP (light labeled cells). Two days after transfection, the cells were UV irradiated, and 6 h later they were pulse-labeled with biotin for 15 min. Then the cells were harvested and mixed, and biotinylated chromatin was purified as described above. The bands corresponding to histone H4 were excised and treated with trypsin; then MRM analysis was performed to quantitatively detect the presence of different acetylated forms of H4 peptide (4-17), comparing the results obtained with experimental (Rad18) and control (GFP) samples.

Figure 2D shows that the H4 histone in close proximity to BirA-Rad18 has a different pattern of acetylation of its N-terminal tail (aa 4-17) as compared to the H4 histone contained in BirA-GFP biotinylated chromatin. Notably, the di- and tetra- acetylated forms are over-represented, as compared to H4 from the control chromatin biotinylated by the GFP-BirA-fusion. Intriguingly, we have also observed under-representation of the tri-acetylated form of H4 in the Rad18 proximal chromatin.

3. The Rad18-specific pattern of H4 acetylation changes after the proximity to Rad18 has been diminished.

The distinctive feature of the PUB-NChIP method is that chromatin in proximity to the BirA-fusion is left with a permanent molecular mark (i.e. biotin), which can persist after the

proximity between BAP- and BirA-fusions has been lost. We decided to take advantage of this property to ask what happens to the Rad18-labeled chromatin at a specified time after the biotinylation.

First, we used immunofluorescence microscopy to monitor the distribution of biotinylated chromatin 6 h after its labeling by BirA-Rad18. Up until the biotin labeling, the cells were treated as in the previous section. Then they were either fixed immediately after the biotin pulse (pulse sample), or else intensively washed to eliminate biotin and then left in fresh medium for 6 h before fixation (chase sample). After biotin removal, the biotinylation levels of histones do not increase with time (Figure 3B, bottom). As before, streptavidin-Cy3 staining of the pulse-labeled cells revealed strong colocalization of labeled BAP-H2A with Rad18 foci (Figure 3A, top). However, the chase sample manifested a different pattern (Figure 3A, middle and bottom). Whereas Rad18 still formed foci, the biotin signal appeared more diffuse, and we generally observed a smaller number of foci per cell, whereas the colocalization of the remaining biotinylated foci with the PCNA or Rad18 foci was considerably less pronounced.

These observations are consistent with the notion that after the BAP-H2A was biotinylated in the proximity of BirA-Rad18 at the moment of the biotin pulse labeling, some Rad18 started to associate with different portions of chromatin, as could be seen by the loss of colocalization of the Rad18 foci with the biotinylated histones. This observation demonstrates that one can monitor the state of proximally-biotinylated chromatin at later times after labeling – e.g., after the proximity with a BirA-fusion may have been lost.

At the next step, we purified the PUB-labeled chromatin after a 6-h chase and analyzed the pattern of H4 acetylation. SILAC labeling and MRM were again used to quantitatively compare the modifications between different samples (Figure 3C). All experimental samples were grown in heavy medium, and as a common reference sample (grown in the light medium), pulse-labeled “BirA-GFP + BAP-H2A” transfected cells were used. Judging by the more modest H/L ratios in the case of the chased sample (Figure 3D), the characteristic under-representation of the tri-acetylated form, which was clearly seen in the RAD18-proximal chromatin at the time of the biotinylation pulse (i.e., 6 h after UV irradiation), became less pronounced and began to resemble the modification pattern seen in the control chromatin (i.e., biotinylated by GFP-BirA).

4. **The pattern of H4 acetylation near Rad18 is different in the case of H2AZ-containing chromatin.**

Another advantage of the PUB-NChIP method resides in the fact that, instead of the canonical core histones (e.g., H3.1, H2A), one can use their alternative forms, which have special roles in chromatin functioning, i.e., have been shown to preferentially associate with different functional states of chromatin.

Immunofluorescence microscopy together with western blot analysis shows that, similar to what is seen with the canonical histone H2A, cotransfection of BirA-Rad18 with BAP-H2AZ, BAP-macroH2A and BAP-H2A.BBD leads to appearance of biotinylated foci in the nuclei that are colocalized with Rad18 and are due to the BAP-histone fusions (Figure 4A and 4B). As expected, BAP-macroH2A foci coincide mostly with DNA-dense areas of the nucleus, whereas BAP-BBD foci associate with less densely stained areas. Given the more complicated distribution of H2AZ in the nucleus [26], no correlation with DNA density was detected for this protein. Overall, the results are consistent with the notion that using alternative histones in

PUB-NChIP should lead to labeling of functional chromatin states associated with these histone variants.

To test the possibility of simultaneously studying histone PTMs both in the vicinity of a protein of interest and in the context of a particular alternative histone, we performed SILAC experiments with the alternative histone H2AZ, which is implicated in many cellular functions, including the cellular response to DNA damage [27-29].

As before, all experimental samples were labeled in heavy medium, and as a reference (light) sample, pulse-labeled "BirA-GFP + BAP-H2A" transfected cells were used. The relative abundance of heavy versus light peptides was analyzed by MRM. As Figure 4C shows, when BAP-H2A is replaced by BAP-H2AZ, levels of the di-acetylated forms of H4 clearly increase in the Rad18-proximal chromatin. The same modification also appears to increase in the control GFP-BirA labeled chromatin containing BAP-H2AZ. The most parsimonious interpretation of this observation is that the two factors (Rad18-proximity vs GFP, and presence of H2AZ vs H2A) affect the di-acetylated form of H4 independently, resulting in an additive effect in the case of 'Rad18 + H2AZ' combination.

Strikingly, however, we have observed a non-additive response to the two factors in the case of the tri- and tetra- acetylated forms of H4. Unlike in the case of 'Rad18 + H2A', there was no decrease in the tri-acetylated form in the case of the 'Rad18 + H2AZ' combination. On the other hand, the tetra-acetylated form was under-represented in the case of 'Rad18 + H2AZ', as compared to 'Rad18 + H2A'.

These results show that the patterns of post-translational histone modifications in proximity to Rad18 protein, as compared to the control GFP protein, depend on the histone composition.

Discussion

We describe here a new methodology to study chromatin in close proximity to a protein of interest in vivo. The well-established approach for such analysis is chromatin immunoprecipitation (ChIP) [1-3]. However, it suffers from the following limitation. Due to the cross-linking of DNA to histones, only the DNA portion of chromatin purified by ChIP can be used, and the protein part is discarded, being irreversibly damaged by the cross-linking procedure. On the other hand, the analysis of histone modifications and replacement histone variants could provide valuable information about the state of chromatin around the particular protein of interest. Until now, the correlations between a protein and histone PTMs could be inferred only indirectly, after comparison of the results of parallel ChIP experiments that use antibodies against the protein and antibodies against particular PTMs. Here, a more direct method to access the chromatin state is proposed.

The principle of PUB-NChIP is somewhat similar to the DamID approach, an alternative to ChIP way to map the DNA binding sites of nuclear proteins [30, 31]. This methodology is based on expression of a fusion of bacterial Dam-methylase to a protein of interest that leads to labeling the protein's genomic binding site by adenine methylation. The label on DNA can be detected by various methods. Our approach is conceptually similar, but instead of Dam-methylase, we fuse the biotin-ligase BirA to the protein of interest and, instead of DNA, we label the histone part of the chromatin. The scopes of these approaches are also different, as PUB-NChIP is mostly aimed at analyzing the histone PTMs. However, we cannot exclude that PUB-NChIP could be used for DNA analysis as well. Although we have not explored this possibility in the present work, one can point to several potential advantages of PUB-NChIP in this respect. There are a wide variety of commercially available resources for detection and/or purification of biotinylated macromolecules, and NChIP with biotinylated chromatin has been shown to work [9]. In addition, by substituting the canonical core histone by its replacement variants, the analysis can be made dependent on the context of particular functional states of chromatin.

As far as the histone PTM analysis is concerned, we present here a proof of principle that it can indeed be done with the PUB-NChIP purified chromatin, focusing mostly on the acetylation status of H4 N-termini. Our approach to analyzing histone PTMs is based on SILAC, which allows us to quantitatively analyze the histone modifications in the Rad18-proximal chromatin, their dynamics after labeling, and their dependence on histone variants. This particular approach is well suited for the type of mass spectrometer (ion-trap) we used. However, because it requires trypsin digestion, the information about correlations between different histone modifications on individual histone molecules is not accessible. On the other hand, the mass-spectrometry instruments with higher resolution (such as FTICR and Orbitrap) can be used in the so called top-down approach that targets histones without tryptic digestion [32, 33]. Using these instruments for downstream analysis of the PUB-NChIP prepared chromatin, one will be able to detect correlations between PTMs on individual histone molecules. It should be noted that such correlation analysis is impossible with the traditional ChIP technique.

In addition to histones, some other proteins associate with DNA sufficiently strongly and thus are expected to be purified in PUB-NChIP [34, 35]. Thus, instead of the histone component of chromatin, one can also use PUB-NChIP for analysis of the non-histone fraction of chromatin as well, which is complicated in the case of regular ChIP due to the cross-linking.

In respect to the biological significance of our findings, their simplest interpretation is that the chromatin in the vicinity of Rad18 is locally destabilized to facilitate the access to DNA of the damage repair machinery. This is consistent with the notion that acetylation of the N-terminal tails of H4 loosens up histone-DNA interactions and hence creates a relaxed nucleosome. Our observation of a less-pronounced pattern of histone modifications 6 h after labeling with biotin (pulse-chase experiment) is consistent with the notion that, after the repair is completed, the chromatin returns to its original state.

The acetylation pattern of the H2AZ-containing chromatin in proximity to Rad18 is more challenging to interpret. It is possible that the chromatin containing H2AZ histone in proximity to Rad18 corresponds to a specific stage of DNA-lesion processing, which could be reflected in the different histone PTM signature. Regardless of the specific mechanistic interpretation of these observations, our data strongly demonstrate that one can analyze how changes in specific histone modifications in the proximity of a nuclear protein of interest depends on the particular chromatin type, here represented by a specific histone variant.

Acknowledgments

We thank Dr. L.L. Pritchard for the critical reading of the manuscript. This work was supported by grants from “La Ligue Contre le Cancer” (9ADO1217/1B1-BIOCE), the “Institut National du Cancer” (247343/1B1-BIOCE) and the Centre National de la Recherche Scientifique [CNRS-INCA-MSHE Franco-Pologne #3037987) to VO, and by NCB Kazakhstan (0103_00404) to AK. AK thanks Mr. Guillaume Giraudet for his help in organizing a visit from NCB to IGR.

Figure legends:

Figure 1. Analysis of post-translational modifications of a specific protein fraction using Proximity Utilizing Biotinylation (PUB).

A. Experimental scheme. 293T cells are transfected with the plasmids expressing BAP-PCNA and BirA-PolH. Two days later, the cells are treated with UVC (20 J/m^2), and 6 h afterwards, the cells are pulse-labeled (5 min) with biotin. The cells are then harvested, and the biotinylated BAP-PCNA is separated from the non-biotinylated BAP-PCNA using streptavidin-sepharose pulldown.

B. Ubiquitination status of PolH-proximal PCNA. Western blot with α-PCNA antibodies. (1) untrasfected cells. (2, 4) BAP-PCNA cotransfected with BirA-GFP fusion. (3, 5) BAP-PCNA cotransfected with Bir-PolH fusion. (2, 3) Flowthrough fraction. (4, 5) Eluate. Note that the BAP-PCNA from the flowthrough fraction was further purified via Ni-NTA chromatography, in order to decrease the signal from endogenous PCNA. The endogenous PCNA, the BAP-PCNA fusion and the ubiquitinated BAP-PCNA are indicated by arrows.

C. Ub-BAP-PCNA is preferentially biotinylated due to its specific interaction with PolH.
Western blot analysis of nuclear extracts from transfected 293T cells. Middle – strepatividin-HRP, detecting two forms of biotinylated BAP-PCNA in the presence of different Bir-PolH fusions, treated or untreated with UV. Top left: Expression of different forms of BirA-PolH, detected by α-His western blot. Bottom left: PCNA in the transfected 293 cells, showing that BAP-PCNA constitutes an undetectable fraction of the total PCNA in transfected cells.

Figure 2. PUB-NChIP reveals a specific pattern of H4 acetylation in the Rad18-proximal chromatin.

A. Chromatin is biotinylated at the sites of RAD18 foci. Confocal microscopy analysis of colocalization of (top) Rad18 and biotinylated chromatin or (bottom) H2A histone and biotinylated chromatin.

B. Biotinylation signal is due to BAP-H2A in proximity to Rad18-BirA.
Left: Western blot analysis with streptavidin-HRP, showing that a specific signal appears only after BirA-Rad18 and BAP-H2A are cotransfected. (1) control untransfected sample. (2) BirA-Rad18. (3) BirA-Rad18 + BAP-H2A. (4) BAP-H2A. The 2 forms of H2A (non-ubiquitinated and ubiquitinated), both biotinylated in this experiment, are indicated by asterisks. Top, α–His antibodies; bottom, streptavidin-HRP.
Right: Western blot analysis of the biotinylation status of BAP-H2A after different times of biotin-labeling. Top, α–His antibodies; bottom, streptavidin-HRP.

C. Purification of biotinylated chromatin. Bottom: SDS-PAGE gel stained with Coomassie blue, with the identities of the histones indicated. The H4 band in the elution also contains streptavidin, and thus is thicker than the other histone bands. Middle: Western blot analysis of the same samples with Streptavidin-HRP, showing that all biotinylated BAP-H2A was pulled down with immobilized streptavidin. Top: Agarose gel electrophoresis and ethidium bromide staining for the same samples, showing that mainly the mononucleosome fractions were purified.

D. Relative acetylation status of H4 N-terminal tails (aa 4-17). The presence of a particular peptide was monitored by MRM. For the mono-, di- and tri- acetylated peptides, it was difficult to distinguish between modifications at different positions, hence we present results without reference to the position. Due to the variations in ratio between the heavy and light isotope labeled chromatin in each experiment, the H/L ratio for every peptide has been normalized by dividing it by the H/L_{um} ratio for the unmodified peptide (UM). The graph present the average normalized H/M ratios and standard deviations for all modified peptides from 2 independent experiments. The standard deviation for the unmodified peptide (UM) represents the standard deviation for the H/L value for this peptide (as H/L_{um}) in two different experiments.

Figure 3. The Rad18-specific pattern changes after the proximity with Rad18 has been diminished.

A. Decrease in colocalisation between Rad18 and biotinylated chromatin after a 6-h chase. Top, confocal microscopy showing strong colocalization between Rad18 and biotinylated chromatin in the nucleus of 293T cells immediately after pulse labeling. Middle, confocal microscopy showing different localization of Rad18 and biotinylated chromatin 6 h after pulse labeling. Bottom, zoomed area from the chase sample showing an example of biotinylated foci that do not colocalize with the Rad18 foci, and vice versa (indicated by arrows). Right, scatterplot with the coefficient of correlation (top left corner) between the intensities of the red and green signals for every pixel (above a background threshold), showing stronger colocalization of the biotin and Rad18 signals in the pulse sample.

B. No increase in biotinylation signal after chase. Western analysis showing that the level of biotinylation of BAP-H3 and BAP-H2A is not increased 6 h after the cells have been washed free of biotin. Top, α-His antibody. Bottom, strepavidin-HRP. P, pulse sample; Ch, chase sample.

C. Scheme of the pulse-chase experiment. Cells grown in light SILAC medium are transfected by BirA-GFP and BAP-H2A (reference sample). Cells grown in heavy SILAC medium are transfected with BirA-Rad18 (or BirA-GFP) and BAP-H2A. The heavy labeled cells are either harvested immediately after biotin labeling or washed to remove biotin and left for 6 h before harvest. The heavy and light labeled cells are then mixed, and chromatin is prepared as in Fig.2C.

D. Acetylation status of H4 N-terminal tails. The analysis was performed, and is presented, as described in the legend to Figure 2D.

Figure 4. The pattern of H4 acetylation near Rad18 is different in the case of H2AZ-containing chromatin.

A. Alternative histones are biotinylated at the locations of Rad18. Confocal microscopy of 293T cells cotransfected by BirA-Rad18 and BAP-H2ABBD (Top), BAP-macroH2A (Middle), and BAP-H2AZ (Bottom). Shown are staining with Topro-3 (left), α–Rad18 (middle) and streptavidin (right).
BAP-histone fusions. Top, Middle: Western blot analysis showing that a specific biotinylation signal appears only after BirA-Rad18 and BAP-macroH2A or BAP-H2ABBD fusions are cotransfected. (1, 4) BAP-histone + BirA-Rad18; (2, 5) BAP-histone; (3, 6) control untransfected samples. Top, α–His antibodies; Middle, streptavidin. Bottom: Western blot

analysis for the BAP-H2AZ fusion. The 2 forms of H2AZ (non-ubiquitinated and ubiquitinated), both biotinylated in this experiment, are indicated by arrows.

C. Acetylation status of H4 N-terminal tails. The analysis was performed, and is presented, as described in the legend to Figure 2D.

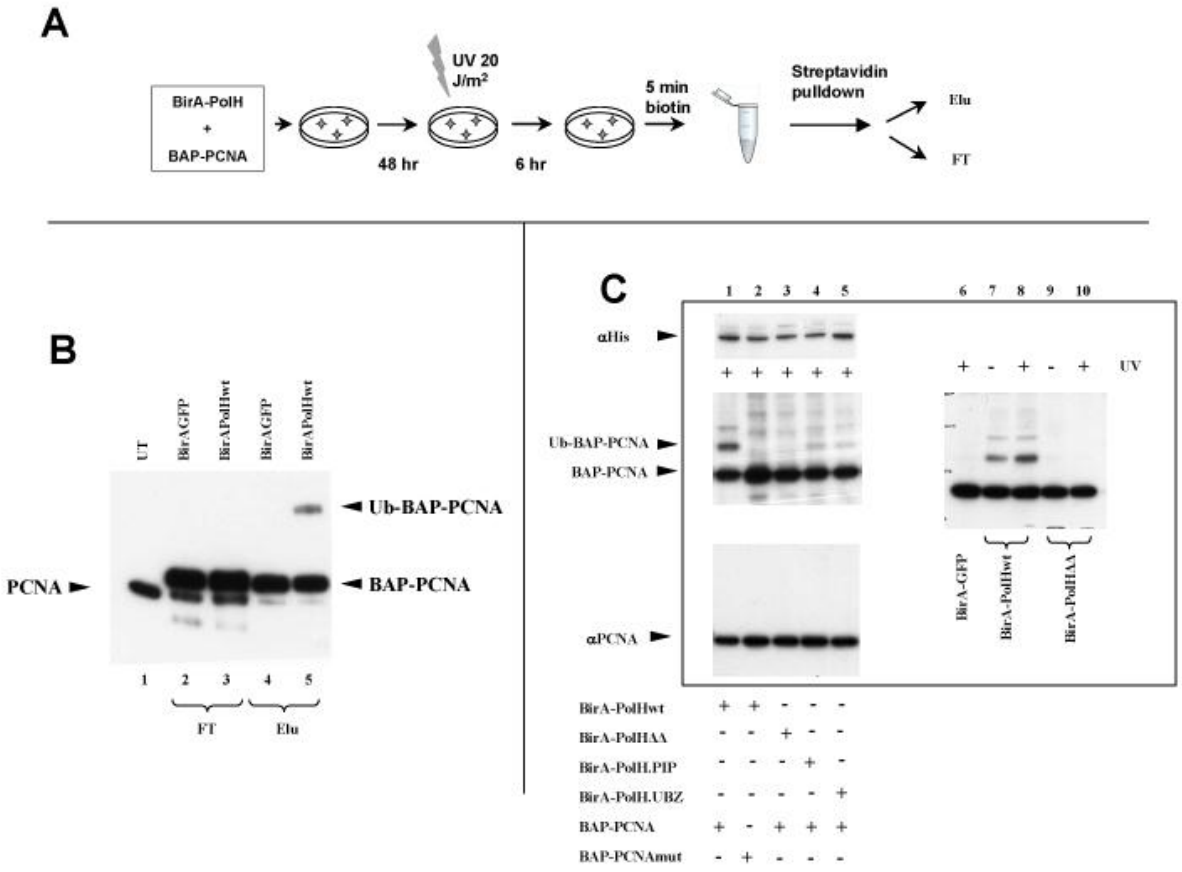

A
BirA-PolH
+
BAP-PCNA
UV 20
J/m2
48 hr
6 hr
5 min
biotin
Streptavidin
pulldown
Elu
FT
B
UT
BirAGFP
BirAPolHwt
BirAGFP
BirAPolHwt
Ub-BAP-PCNA
PCNA
BAP-PCNA
1 2 3 4 5
FT
Elu
C
1 2 3 4 5 6 7 8 9 10
αHis
UV
Ub-BAP-PCNA
BAP-PCNA
αPCNA
BirA-GFP
BirA-PolHwt
BirA-PolHAA
BirA-PolHwt
BirA-PolHAA
BirA-PolH.PIP
BirA-PolH.UBZ
BAP-PCNA
BAP-PCNAmut

Figure 1.

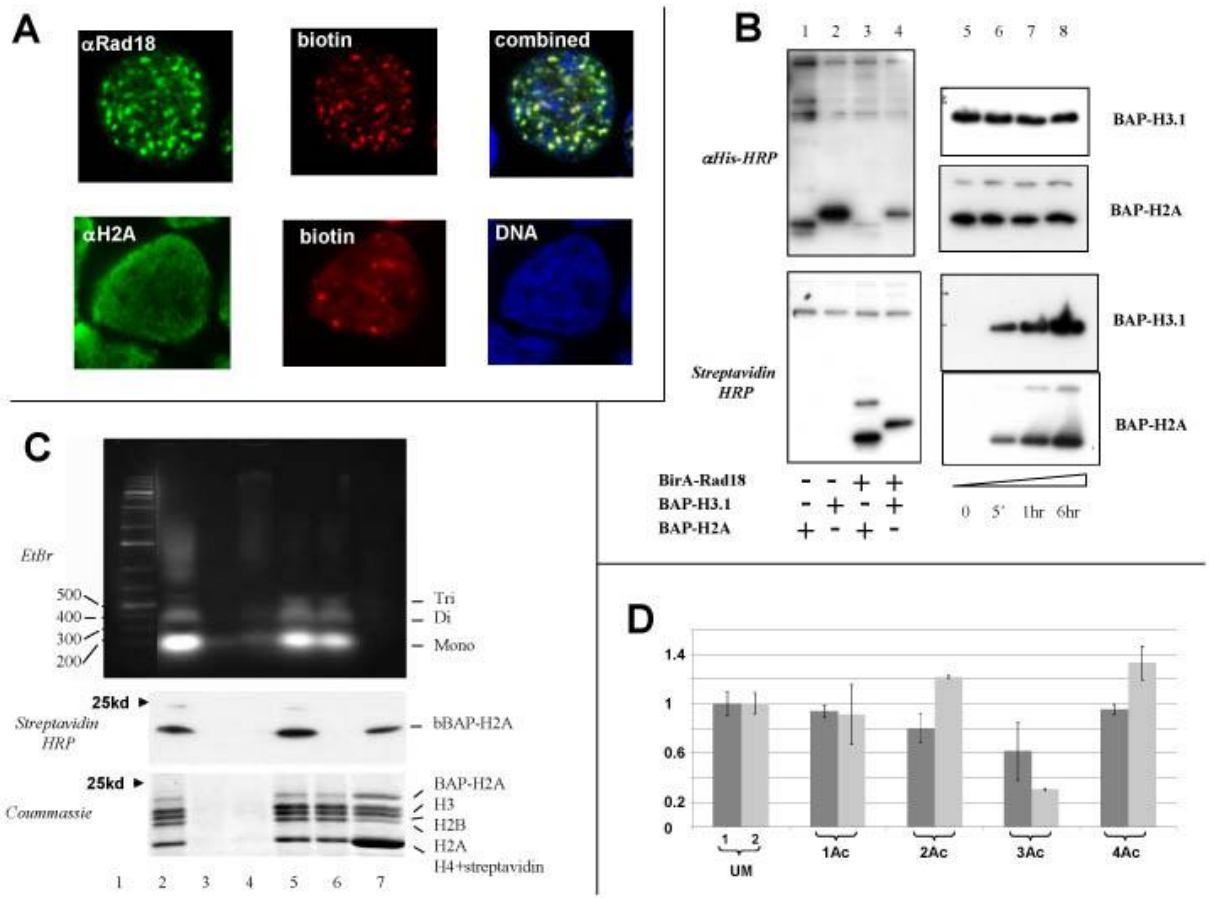

A
αRad18
biotin
combined
αH2A
biotin
DNA
B
1 2 3 4
5 6 7 8
αHis-HRP
BAP-H3.1
BAP-H2A
Streptavidin HRP
BAP-H3.1
BAP-H2A
BirA-Rad18 - - + +
BAP-H3.1 - + - +
BAP-H2A + - + -
0 5' 1hr 6hr
C
EtBr
500
400
300
200
Tri
Di
Mono
25kd
Streptavidin HRP
bBAP-H2A
25kd
Coummassie
BAP-H2A
H3
H2B
H2A
H4+streptavidin
1 2 3 4 5 6 7
D
1.4
1
0.6
0.2
0
1 2
UM
1Ac
2Ac
3Ac
4Ac

Figure 2

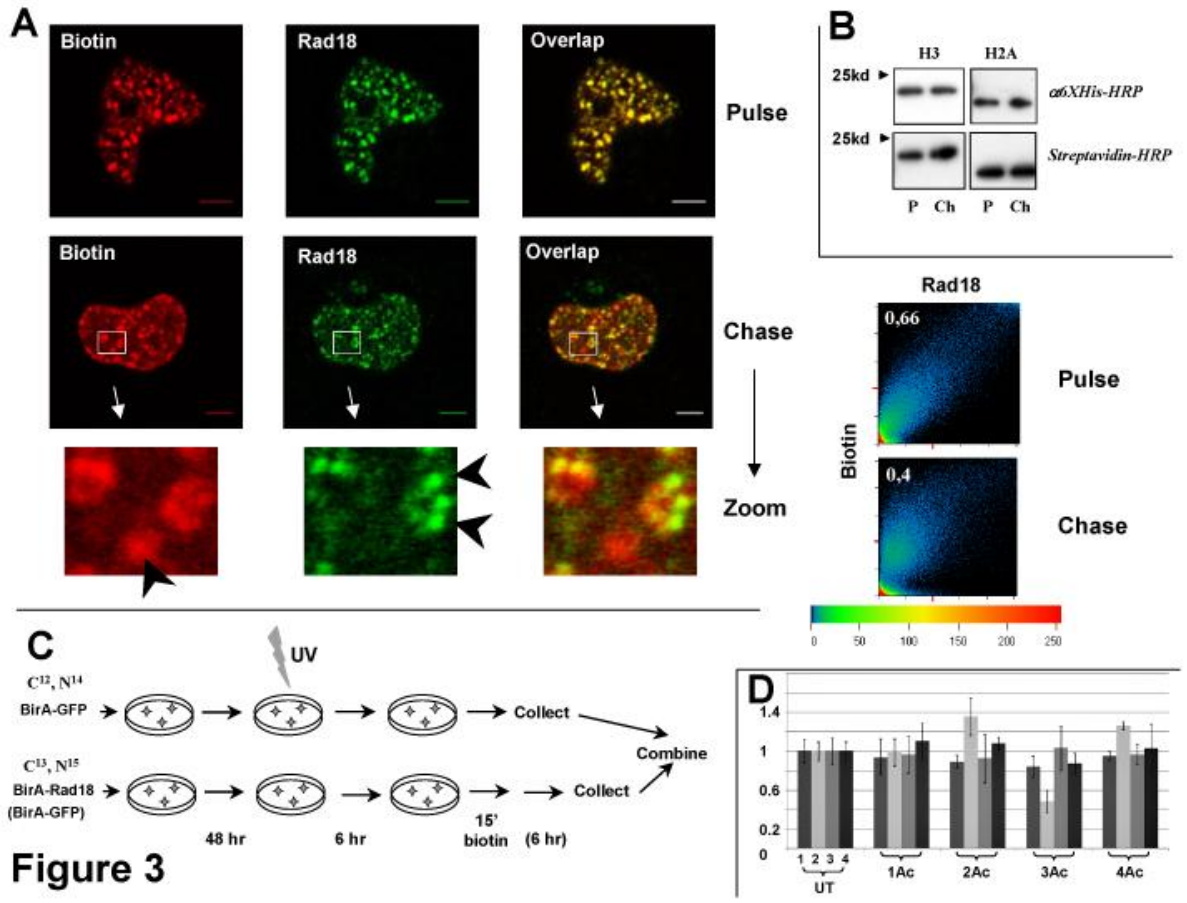

A
Biotin
Rad18
Overlap
Pulse
Chase
Zoom
B
H3
H2A
25kd
25kd
α6XHis-HRP
Streptavidin-HRP
P Ch P Ch
Rad18
Biotin
0,66
0,4
Pulse
Chase
0 50 100 150 200 250
C
UV
C12, N14
BirA-GFP
Collect
Combine
C13, N15
BirA-Rad18
(BirA-GFP)
Collect
48 hr
6 hr
15' biotin
(6 hr)
D
1.4
1
0.6
0.2
0
1 2 3 4
UT
1Ac
2Ac
3Ac
4Ac

Figure 3

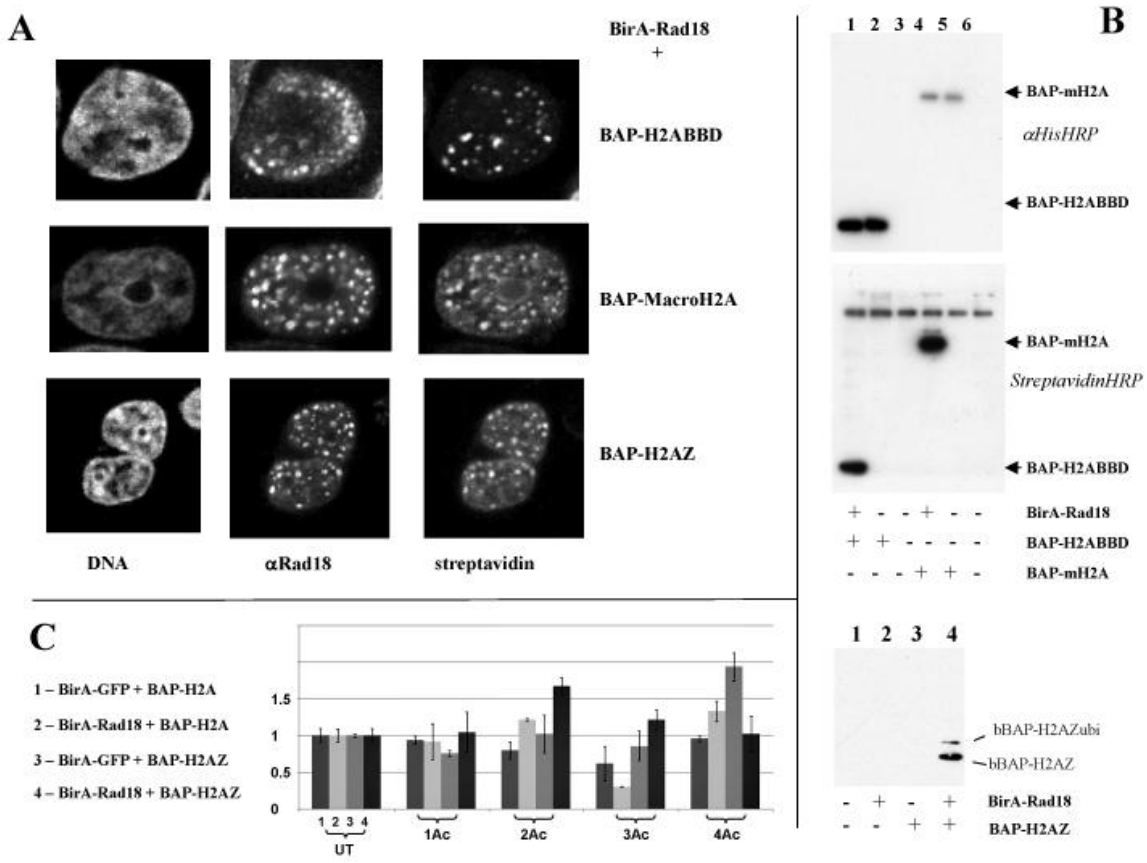

A
BirA-Rad18
+
BAP-H2ABBD
BAP-MacroH2A
BAP-H2AZ
DNA
αRad18
streptavidin
B
1 2 3 4 5 6
BAP-mH2A
αHisHRP
BAP-H2ABBD
BAP-mH2A
StreptavidinHRP
BAP-H2ABBD
+ - - + - - BirA-Rad18
+ + - - - - BAP-H2ABBD
- - - + + - BAP-mH2A
C
1 – BirA-GFP + BAP-H2A
2 – BirA-Rad18 + BAP-H2A
3 – BirA-GFP + BAP-H2AZ
4 – BirA-Rad18 + BAP-H2AZ
1.5
1
0.5
0
1 2 3 4
UT
1Ac
2Ac
3Ac
4Ac
1 2 3 4
bBAP-H2AZubi
bBAP-H2AZ
- + - + BirA-Rad18
- - + + BAP-H2AZ

Figure 4

References

1. **Orlando V: Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem Sci* 2000, 25:99-104.**
2. **Ren B, Dynlacht BD: Use of chromatin immunoprecipitation assays in genome-wide location analysis of mammalian transcription factors. *Methods Enzymol* 2004, 376:304-315.**
3. **Weinmann AS, Farnham PJ: Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. *Methods* 2002, 26:37-47.**
4. **Buck MJ, Lieb JD: ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* 2004, 83:349-360.**
5. **Negre N, Lavrov S, Hennetin J, Bellis M, Cavalli G: Mapping the distribution of chromatin proteins by ChIP on chip. *Methods Enzymol* 2006, 410:316-341.**
6. **Park PJ: ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* 2009, 10:669-680.**
7. **Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, et al: Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* 2007, 4:651-657.**
8. **O'Neill LP, Turner BM: Immunoprecipitation of native chromatin: NChIP. *Methods* 2003, 31:76-82.**
9. **Ooi SL, Henikoff JG, Henikoff S: A native chromatin purification system for epigenomic profiling in Caenorhabditis elegans. *Nucleic Acids Res* 2010, 38:e26.**
10. **Chadwick BP, Willard HF: Histone H2A variants and the inactive X chromosome: identification of a second macroH2A variant. *Hum Mol Genet* 2001, 10:1101-1113.**
11. **Chadwick BP, Willard HF: A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. *J Cell Biol* 2001, 152:375-384.**
12. **Costanzi C, Pehrson JR: Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature* 1998, 393:599-601.**
13. **Kulyyassov A, Shoaib M, Pichugin A, Kannouche P, Ramanculov E, Lipinski M, Ogryzko V: PUB-MS – a mass-spectrometry–based method to monitor protein-protein proximity in vivo. *Journal of Proteome Research* 2011, in press.**
14. **Saade E, Mechold U, Kulyyassov A, Vertut D, Lipinski M, Ogryzko V: Analysis of interaction partners of H4 histone by a new proteomics approach. *Proteomics* 2009, 9:4934-4943.**
15. **Garcia BA, Mollah S, Ueberheide BM, Busby SA, Muratore TL, Shabanowitz J, Hunt DF: Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat Protoc* 2007, 2:933-938.**
16. **Shevchenko A, Wilm M, Vorm O, Mann M: Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* 1996, 68:850-858.**
17. **Fernandez-Suarez M, Chen TS, Ting AY: Protein-protein interaction detection in vitro and in cells by proximity biotinylation. *J Am Chem Soc* 2008, 130:9251-9253.**
18. **Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M: Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 2002, 1:376-386.**

19. **Moldovan GL, Pfander B, Jentsch S: PCNA, the maestro of the replication fork. *Cell* 2007, 129:665-679.**
20. **Haracska L, Johnson RE, Unk I, Phillips B, Hurwitz J, Prakash L, Prakash S: Physical and functional interactions of human DNA polymerase eta with PCNA. *Mol Cell Biol* 2001, 21:7199-7206.**
21. **Haracska L, Johnson RE, Unk I, Phillips BB, Hurwitz J, Prakash L, Prakash S: Targeting of human DNA polymerase iota to the replication machinery via interaction with PCNA. *Proc Natl Acad Sci U S A* 2001, 98:14256-14261.**
22. **Kannouche PL, Wing J, Lehmann AR: Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage. *Mol Cell* 2004, 14:491-500.**
23. **Prakash S, Johnson RE, Prakash L: Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. *Annu Rev Biochem* 2005, 74:317-353.**
24. **Huang J, Huen MS, Kim H, Leung CC, Glover JN, Yu X, Chen J: RAD18 transmits DNA damage signalling to elicit homologous recombination repair. *Nat Cell Biol* 2009, 11:592-603.**
25. **Watanabe K, Iwabuchi K, Sun J, Tsuji Y, Tani T, Tokunaga K, Date T, Hashimoto M, Yamaizumi M, Tateishi S: RAD18 promotes DNA double-strand break repair during G1 phase through chromatin retention of 53BP1. *Nucleic Acids Res* 2009, 37:2176-2193.**
26. **Viens A, Mechold U, Brouillard F, Gilbert C, Leclerc P, Ogryzko V: Analysis of human histone H2AZ deposition in vivo argues against its direct role in epigenetic templating mechanisms. *Mol Cell Biol* 2006, 26:5325-5335.**
27. **Ahmed S, Dul B, Qiu X, Walworth NC: Msc1 acts through histone H2A.Z to promote chromosome stability in Schizosaccharomyces pombe. *Genetics* 2007, 177:1487-1497.**
28. **Kalocsay M, Hiller NJ, Jentsch S: Chromosome-wide Rad51 spreading and SUMO-H2A.Z-dependent chromosome fixation in response to a persistent DNA double-strand break. *Mol Cell* 2009, 33:335-343.**
29. **Morillo-Huesca M, Clemente-Ruiz M, Andujar E, Prado F: The SWR1 histone replacement complex causes genetic instability and genome-wide transcription misregulation in the absence of H2A.Z. *PLoS One*, 5:e12143.**
30. **Greil F, Moorman C, van Steensel B: DamID: mapping of in vivo protein-genome interactions using tethered DNA adenine methyltransferase. *Methods Enzymol* 2006, 410:342-359.**
31. **van Steensel B, Henikoff S: Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat Biotechnol* 2000, 18:424-428.**
32. **Pesavento JJ, Kim YB, Taylor GK, Kelleher NL: Shotgun annotation of histone modifications: a new approach for streamlined characterization of proteins by top down mass spectrometry. *J Am Chem Soc* 2004, 126:3386-3387.**
33. **Zhang L, Freitas MA, Wickham J, Parthun MR, Klisovic MI, Marcucci G, Byrd JC: Differential expression of histone post-translational modifications in acute myeloid and chronic lymphocytic leukemia determined by high-pressure liquid chromatography and mass spectrometry. *J Am Soc Mass Spectrom* 2004, 15:77-86.**
34. **Lambert JP, Mitchell L, Rudner A, Baetz K, Figeys D: A novel proteomics approach for the discovery of chromatin-associated protein networks. *Mol Cell Proteomics* 2009, 8:870-882.**

35. Shen Z, Sathyan KM, Geng Y, Zheng R, Chakraborty A, Freeman B, Wang F, Prasanth KV, Prasanth SG: A WD-repeat protein stabilizes ORC binding to chromatin. *Mol Cell* 2010, 40:99-111.

Supplementary Figures

Legends to supplementary figures

Supplementary Figure S1.

A. BirA-GFP cotransfected with BAP-H2A produces homogeneous Biotinylation throughout the nucleus. Confocal microscopy analysis of the localization of GFP (left) and biotinylated chromatin (right).

B. BAP-H3.1 is biotinylated at the sites of RAD18 foci. Confocal microscopy analysis of colocalization of Rad18 (left) and biotinylated chromatin (right) after cotransfection of 293T cells with "BirA-Rad18 + BAP-H3.1" followed by UVC irradiation.

Supplementary Figure S2. Rad18 proximal chromatin is enriched in γ-H2AX after IR.

A. Colocalization of γ-H2AX and chromatin biotinylated in proximity to BirA-Rad-18.
Confocal microscopy analysis of the colocalization (right) of γ-H2AX (left) and biotinylated chromatin (center) after cotransfection of 293T cells with "BirA-Rad18 + BAP-H2A" (top) or control "BirA-HP1α + BAP-H2A" (bottom), ionizing irradiation 48 h later, and biotin pulse labeling 3 h after irradiation.

B. Western blot analysis of biotinylated chromatin.
The 293T cells were transfected with "BirA-Rad18 + BAP-H2A" (left) or control "BirA-HP1α + BAP-H2A" cells (right), treated with ionizing irradiation 48 h afterwards, and pulse labeled with biotin 3 h after that. After NChIP purification, the amounts of biotinylated chromatin were normalized by western blot with α-H3 antibodies (top). The normalized amounts were analyzed by Western with γ-H2AX antibodies.

Supplementary Figure S3(a-e). MS spectra.

Shown are MS spectra of peptides corresponding to the light and heavy versions of non- (a), mono- (b), di- (c), tri- (d) and tetra- (e) acetylated H4 N-terminal tail peptide (aa 4-17). The ions that match the predicted y- and b-series ions are indicated.

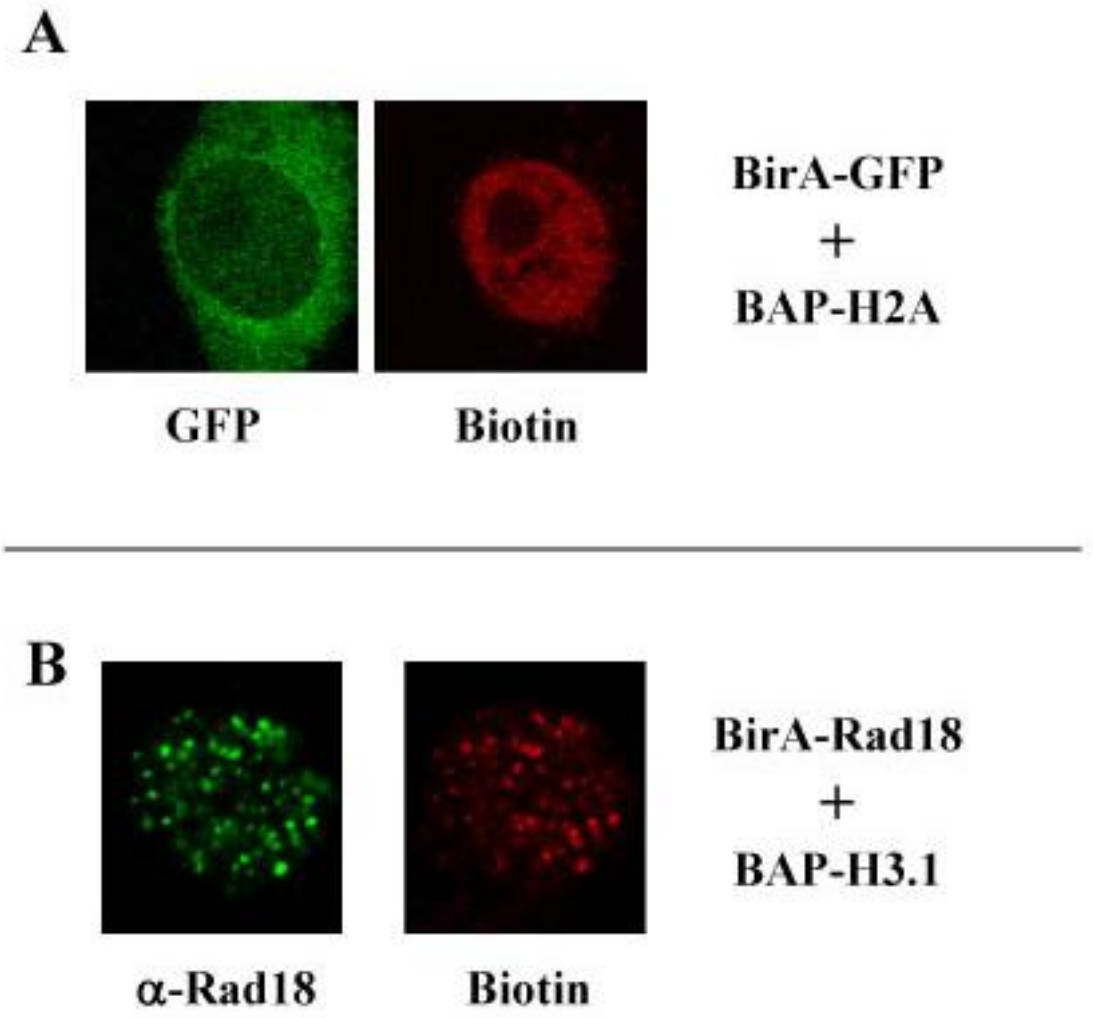


Figure S1

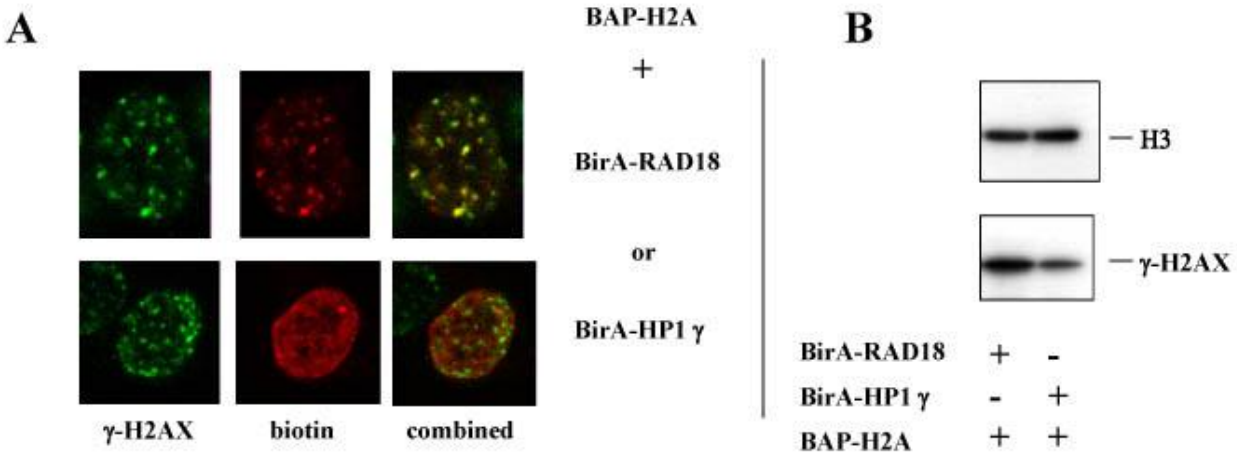


Figure S2

Histone H4 residues 4-17 (GKGGKGLGKGGAKR) UM
747.4 (Light)
764.4 (Heavy)

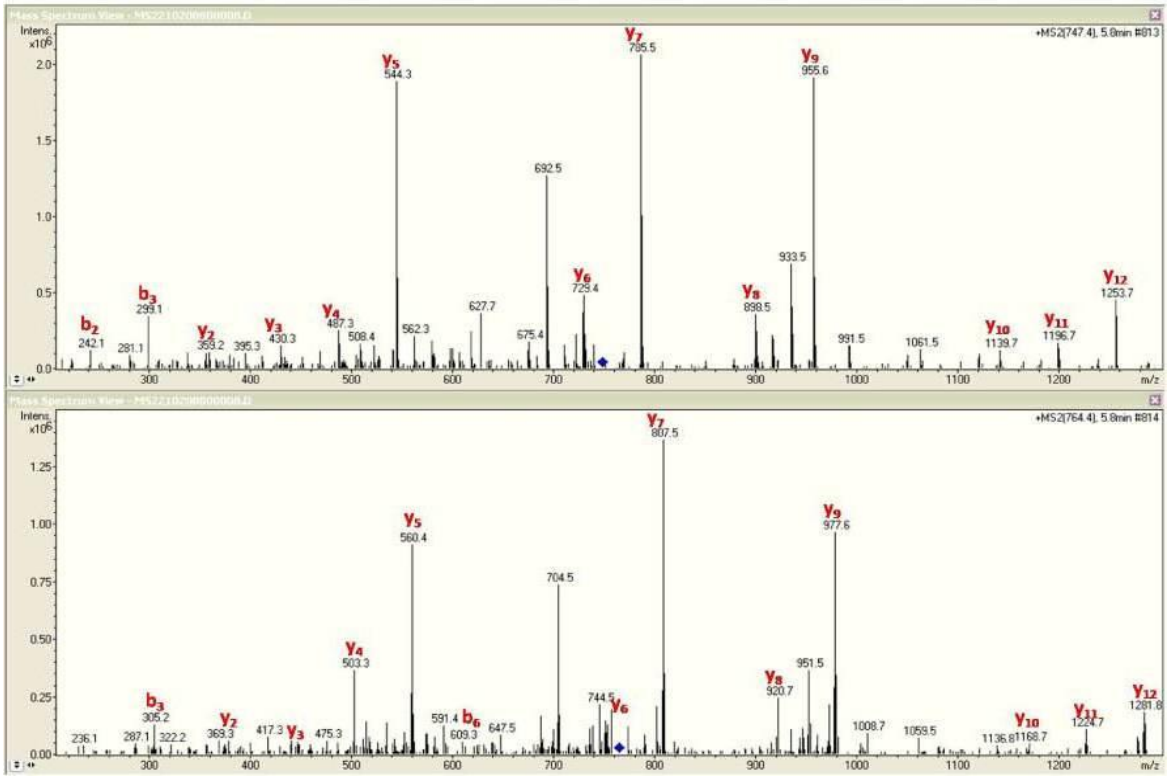


Figure S3a

Histone H4 residues 4-17 (GKGGKGLGKGGAKR) 1ac
740.4 (Light)
757.4 (Heavy)

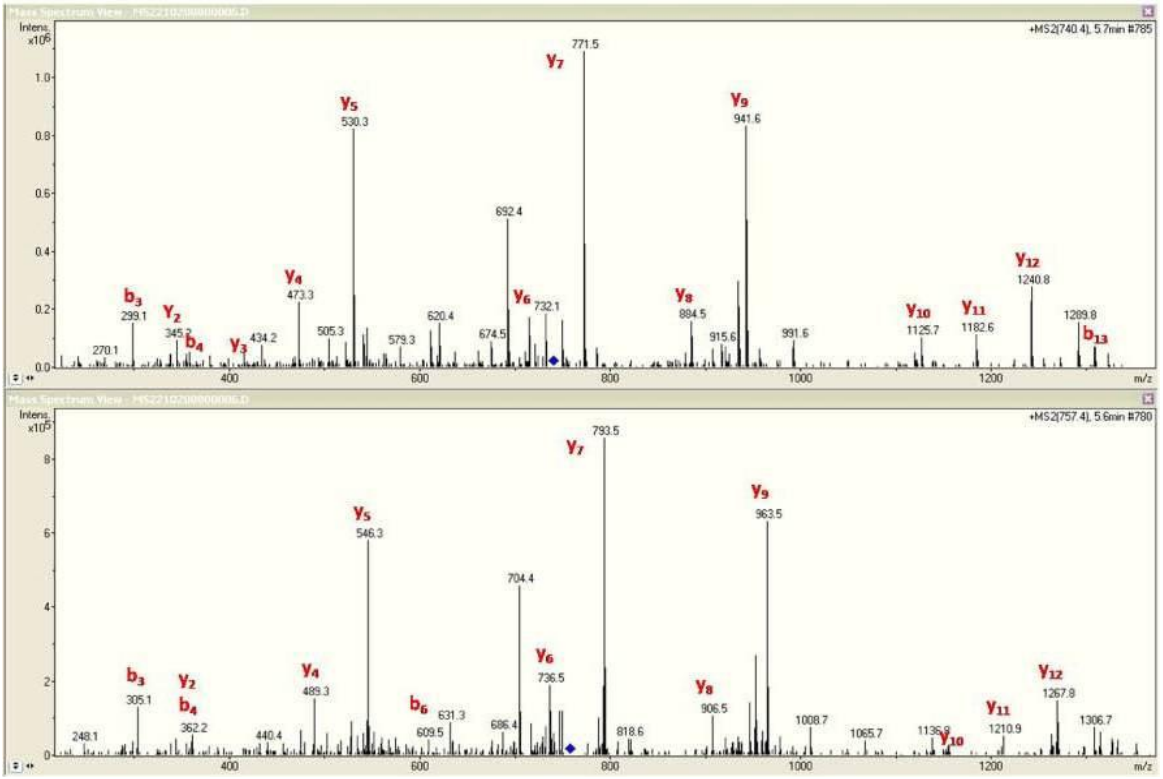


Figure S3b

Histone H4 residues 4-17 (GKGGKGLGKGGAKR) 2ac
733.3 (Light)
750.3 (Heavy)

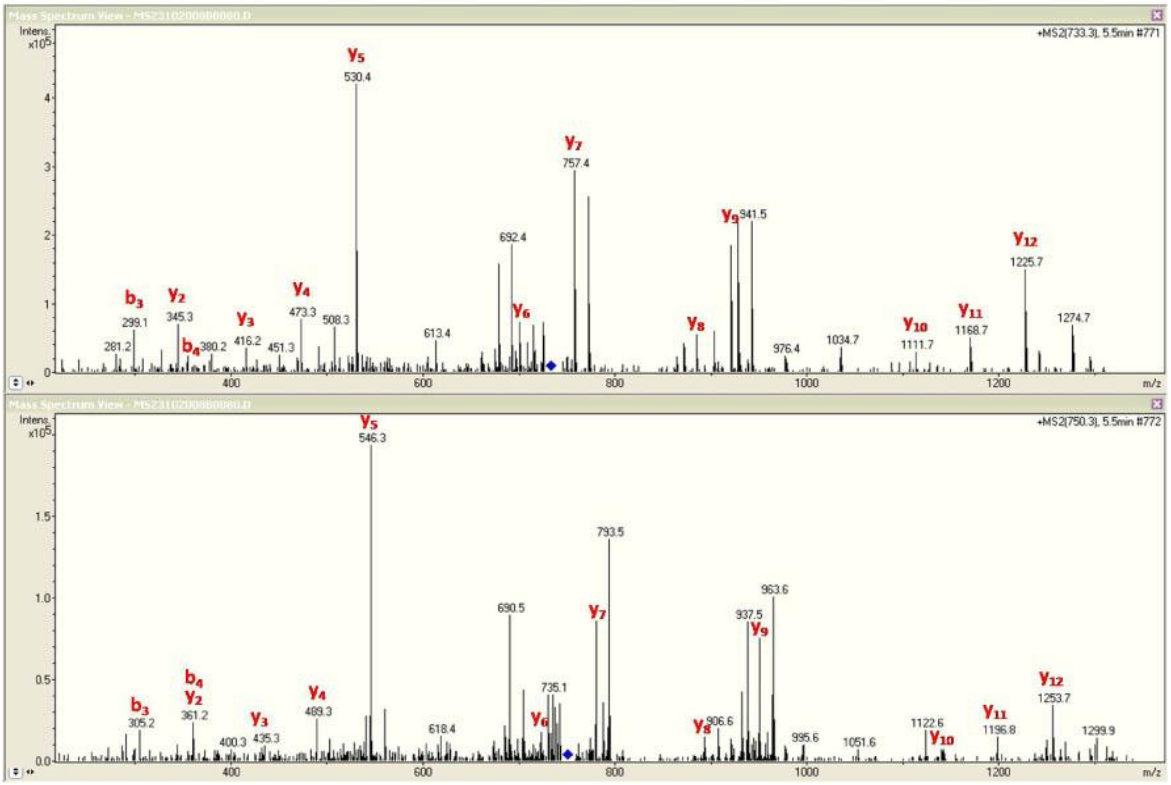


Figure S3c

Histone H4 residues 4-17 (GKGGKGLGKGGAKR) 3ac
726.3 (Light)
740.3 (Heavy)

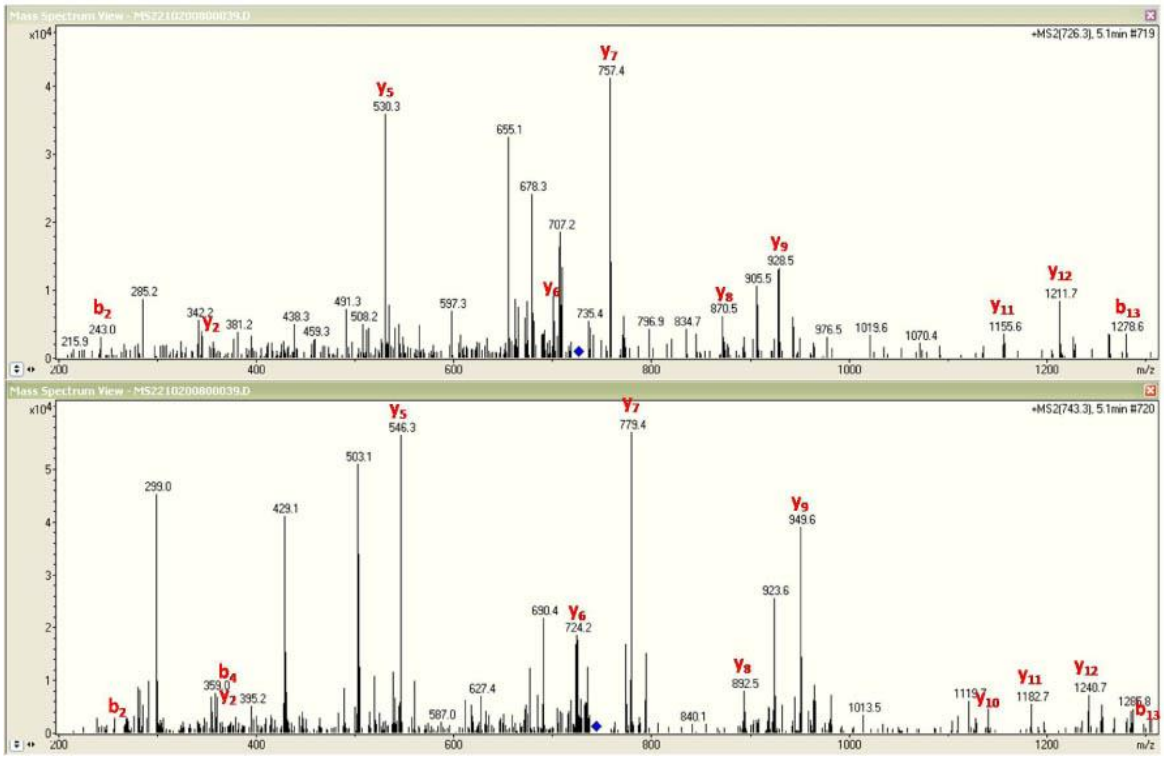


Figure S3d

Histone H4 residues 4-17 (GKGGKGLGKGGAKR) 4ac
719.3 (Light)
736.3 (Heavy)

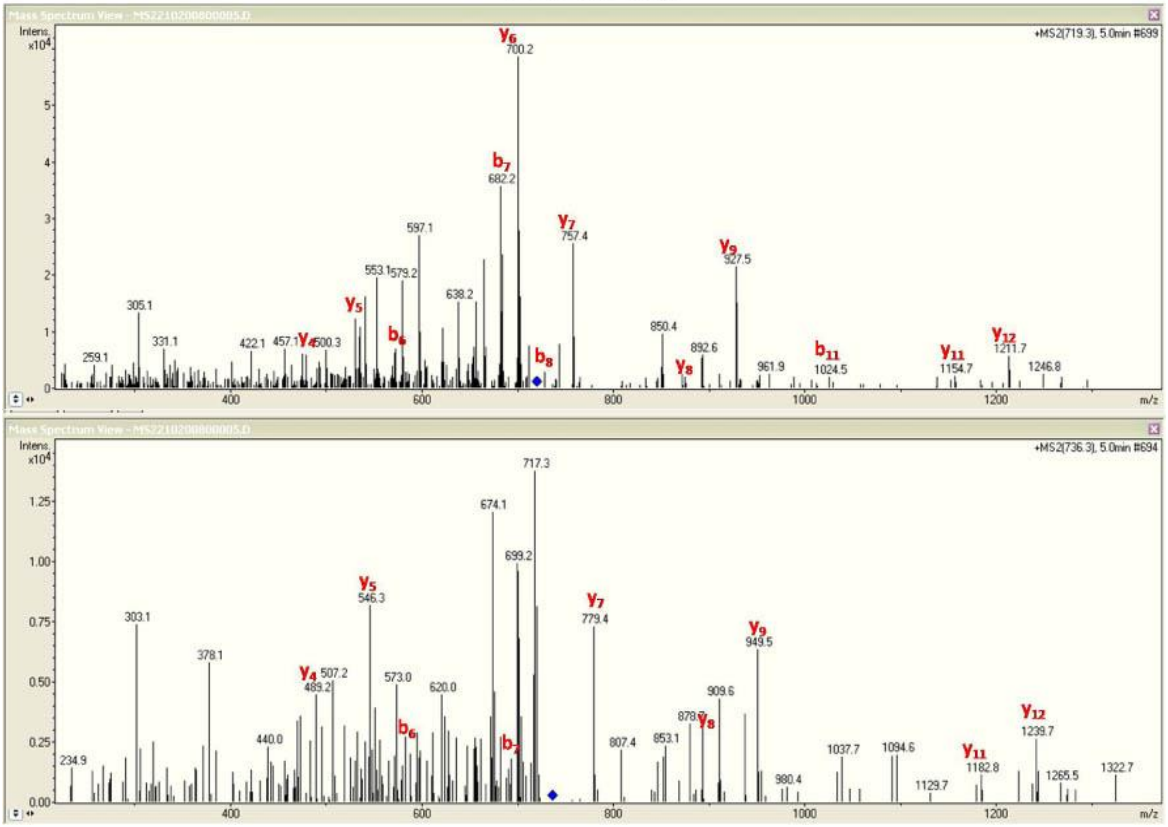


Figure S3e

Development of New Approaches to Study the Role of Chromatin in DNA Damage Response

Abstract

In eukaryotic cells, the genome is packed into chromatin, a hierarchically organized complex composed of DNA and histone and non-histone proteins. In this thesis we have addressed the role of chromatin in cellular response to DNA damage (DDR) using various methodologies encompassing functional genomics and proteomics. First, we analyzed histone post-translational modifications (PTM) in the context of specific kind of DNA lesions (ICL-Interstrand Crosslinks) in Fanconi anemia using quantitative proteomics methodology, SILAC (Stable Isotope Labeling of Amino acids during Cell Culture). Using mass spectrometry (MS), we have successfully identified and quantified a number of histone PTM marks in histone H3 and H4, mainly acetylations and methylations, which have shown dependence upon functional FA-pathway. As a next step, we applied a functional genomics approach to study DDR in FA cells. In this analysis we first monitored the expression profile of histone modifying enzymes related to histone acetylations and methylations. Our results suggest some correlations between histone PTMs and gene expression of histone modifying enzymes, although conclusive evidence warrants further investigations. Next, we analyzed the total transcriptome after DNA damage induction in FA mutant and wild type cells. We also included in this analysis IR irradiation, in an attempt to dissociate more generic DDR from more specific changes that are associated with the role of FA pathway to the DNA ICLs. By performing a factorial interaction analysis, we were able to isolate the part of transcriptional response to DNA damage that was requiring functional FA pathway, as well as the genes that were sensitized to DNA damage by the inactivation of FA pathway. In the final part of the thesis, we attempted to solve one of the limitations that we encountered in the histone PTM analysis. The current approaches used to study histone PTMs from particular loci involves classical chromatin immunoprecipitation, which due to involvement of formaldehyde crosslinking render the protein part mostly unavailable for MS-based proteomics. We have proposed a novel methodology, which is based upon the biotin tagging of histones proximal to a protein of interest and subsequent purification of nucleosomes carrying the tagged histone. This methodology does not involve any crosslinking, enabling us to purify histones from specific loci, and subject them to large scale MS-based histone PTM analysis. A time dimension can also be added to our approach, as we can follow the modification status of particular fraction of histones once they get biotinylated. Another advantage is the use of alternate variant histones, which allows us to study the PTM profile of different functional states of chromatin. This methodology certainly has an edge on current techniques to study histone PTMs pattern associated with a particular protein of interest or with particular chromatin state.

KEYWORDS: chromatin, histone PTM, DNA damage response, transcriptome analysis, SILAC, biotinylation, native chromatin immunoprecipitation.

Développement de Nouvelles Approches pour Etudier le Rôle de la Chromatine en Réponse aux Dommages de l'ADN

Résumé

Le génôme des cellules eucaryotes est condensé au sein d'une structure complexe hiérarchiquement organisée : la chromatine. La chromatine est composée d'ADN, de protéines histone et non-histone. Cette thèse a pour but d'étudier le rôle de la chromatine dans la réponse cellulaire aux dommages de l'ADN (DDR) par les méthodologies de génomique fonctionnelle et de protéomique. Nous avons tout d'abord analysé les modifications post-traductionnelles (PTM) des histones dans le cadre des pontages inter-brins (ou "Interstrand Crosslinks", ICL), type particulier de lésions de l'ADN, en choisissant le modèle de l'Anémie de Fanconi (FA). Ceci a été réalisé grâce aux techniques de protéomique quantitative SILAC (Stable Isotope Labeling of Amino acid during Cell culture) et de spectrométrie de masse (MS). Nous avons ainsi réussi à identifier et à quantifier de nombreuses PTMs dans les histones H3 et H4, et à démontrer que certaines de ces PTM sont dépendantes d'une voie fonctionnelle de la signalisation de FA. Nous avons également approfondi l'étude des DDR dans les cellules de FA par une approche de génomique fonctionnelle. Pour cela, nous avons analysé le profil d'expression d'enzymes associées à l'acétylation et à la méthylation des histones. Nos résultats suggèrent l'existence de corrélations entre le profil d'expression de ces enzymes et les PTMs des histones. Des études complémentaires sont nécessaires en vue de confirmer ces corrélations. Nous avons également comparé le transcriptome de deux lignées cellulaires de FA (mutée en FANCC et corrigée en FANCC) après induction de dommages à l'ADN. Afin de différencier les changements spécifiquement associés à la voie de signalisation de FA en réponse aux ICL de l'ADN des réponses plus générales aux dommages de l'ADN, nous avons inclus des cellules traitées par rayonnement ionisant. En réalisant une analyse d'interactions factorielles, nous avons pu identifier une réponse transcriptionnelle aux dommages de l'ADN nécessitant une voie fonctionnelle de la signalisation de FA. Nous avons également tenté de pallier aux limitations rencontrées dans l'analyse des PTMs des histones. En effet, les PTMs des histones que nous avons identifiées représentent l'ensemble des modifications, c'est-à-dire les PTMs concernant les histones se trouvant immédiatement à proximité du site du dommage et en relation directe avec celui-ci, et les PTMs se trouvant à distance du dommage et pouvant ne pas être en relation directe avec celui-ci. Les approches courantes pour identifier les PTMs se trouvant à des loci particuliers sont basées sur l'immunoprécipitation classique de la chromatine où l'utilisation de formaldéhyde altère les protéines, ce qui en rend impossible l'analyse par MS. Nous avons proposé une nouvelle méthodologie basée sur la biotinylation expérimentale d'histones situées à proximité d'une protéine particulière, suivie de la purification des nucléosomes contenant ces histones biotinylées. Contrairement à l'immunoprécipiatation classique de la chromatine, cette méthode n'induit pas d'altération des protéines, permettant ainsi de purifier les histones à partir d'un locus spécifique et d'analyser à grande échelle leurs PTMs par MS. Cette approche permet aussi de suivre dans le temps les PTMs d'une fraction des histones juste après leur biotinylation. Enfin, elle présente l'avantage de pouvoir étudier le profil des PTMs de différents états fonctionnels de la chromatine grâce à l'utilisation de variants d'histones.

MOTS CLÉS : chromatine, modification post-traductionnelle des histones (PTM), réponse aux dommages de l'ADN (DDR), analyse du transcriptome, SILAC, biotinylation, immunoprécipitation de chromatine native.

CNRS UMR8126, Molecular Interactions and Cancer
Institut Gustave Roussy, PR1, 114 rue Edouard Vaillant, 94805, Villejuif, France

Frais de reprographie pris en charge par la taxe d'apprentissage collectée par l'IGR.