Etude des voies de réparation des cassures double brin de l'ADN lors de la recombinaison suicide du locus IgH en physiologie normale et pathologie du lymphocyte B

Hend Boutouil

▶ **To cite this version:**

Hend Boutouil. Etude des voies de réparation des cassures double brin de l'ADN lors de la recombinaison suicide du locus IgH en physiologie normale et pathologie du lymphocyte B. Médecine humaine et pathologie. Université de Limoges, 2018. Français. NNT : 2018LIMO0028 . tel-01893576

HAL Id: tel-01893576
https://theses.hal.science/tel-01893576

Submitted on 11 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse de doctorat

Université de Limoges

UNIVERSITE DE LIMOGES

Ecole Doctorale Sciences Biologiques et Santé (ED n° 615)

UMR CNRS 7276 INSERM U1262

Thèse pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE LIMOGES

Discipline : Biologie Science Santé

Spécialité : Immunogénétique

Présentée et soutenue par

HEND BOUTOUIL

Le 12 septembre 2018

Étude des voies de réparation des cassures double brin de l'ADN lors de la recombinaison suicide du locus *IgH* en physiologie normale et pathologique du lymphocyte B

Thèse dirigée par Sophie PÉRON et Michel COGNÉ

JURY

Rapporteurs

Dr Gaëlle LEGUBE
Pr Pauline SOULAS-SPRAUEL

Examinateurs

Dr Sophie PÉRON
Pr Michel COGNÉ

CRIBL
Contrôle de la Réponse Immune B et Lymphoprolifératives

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

Durant mes trois années de thèse, j'ai bénéficié d'un contrat doctoral avec la faculté de Médecine de Limoges.

Remerciements

Je remercie tout d'abord les membres de mon jury de thèse, mes deux rapporteurs, le Dr LEGUBE et le Pr SOULAS-SPRAUEL qui ont accepté et fait le déplacement de Toulouse et Strasbourg.

Je remercie mes directeurs de thèse, Sophie et Michel. Merci Michel de m'avoir accordé l'opportunité de vivre cette expérience scientifique et humaine au sein de votre équipe. Sophie, je tiens à te remercier pour ton encadrement depuis le stage de M2 jusqu'à la fin de ma thèse. J'ai apprécié de travailler sur ce projet complexe et original à la fois. Je ne m'enfais pas pour la personne qui prendra le relais, elle aura la chance d'être soutenue et motivée en cas de besoin !

Je souhaite remercier mes chers collègues bien évidemment ! Merci pour tous ces moments passés à vos côtés, les bons et les mauvais ! Je pense en premier à mes collègues exclusivement féminines du bureau n°107 : Iman, Zeinab, Audrey, Julie. Grâce à vous, la bonne ambiance était au rendez-vous et le risque de tomber en hypoglycémie réduit à « 0 ». Je n'oublie pas les anciens du bureau, Brice, Nicolas et Mylène, mais aussi Racha et Héloïse, pour votre bonne humeur légendaire !

Je tiens à remercier mes voisins du deuxième étage, à qui je rends souvent visite pendant les temps d'incubation. Je pense surtout au duo Omar/Jean-Marie avec qui on trouve toujours de quoi rigoler.

La liste est très longue alors pour ne pas m'étendre et ne vexer Péronne, je dis un sincère GRAND MERCI à toutes les personnes que j'ai côtoyé au sein de ce laboratoire !

Résumé

La rencontre des lymphocytes B matures avec l'antigène (Ag), au niveau des organes lymphoïdes secondaires, déclenche la maturation terminale, au cours de laquelle deux évènements peuvent avoir lieu : la commutation de classe de l'immunoglobuline (CSR pour Class Switch Recombination) et l'hypermutation somatique (SHM).

Dernièrement, notre laboratoire a décrit pour la première fois la recombinaison suicide du locus *IgH* (LSR pour Locus Suicide Recombination) (Péron et al., 2012). Cette recombinaison engendre une délétion totale de l'ensemble des gènes constants du locus *IgH,* empêchant ainsi l'expression d'Ig, et donc l'absence du récepteur BCR (B Cell Receptor). La cellule B se retrouve privée des signaux de survie délivrés par ce récepteur, et est induite à l'apoptose. La LSR semble opérer par les mêmes étapes que la CSR : 1- la transcription de régions de l'ADN ciblées, 2- la génération de cassures double brin (CDB) à partir de lésions introduites par AID (Activation-induced cytidine deaminase), 3- la réparation de l'ADN lésé majoritairement par le système classique de ligation des extrémités non homologues (C-NHEJ). Cependant, la réparation au cours de la LSR n'a pas été pleinement décrite, et sa détermination constitue l'objectif principale de mon doctorat.

Dans un premier temps, nous avons mis au point un programme bio-informatique « CSReport », afin d'analyser la masse de données générées par le séquençage haut débit de jonctions du locus *IgH* (CSR et LSR) (Boyer et al., 2017). Cet outil nous a permis d'étudier le système de réparation des CDB, à travers la détermination de la structure au point de jonction. De façon inattendue, nos résultats montrent que la réparation de l'ADN dans la LSR est similaire entre la souris et l'Homme et si la CSR fait intervenir le C-NHEJ, la LSR semble faire appel à l'A-EJ (Alternative End Joining) et/ou la HR (Homologous Recombination). Ces observations sont renforcées par les résultats mettant en évidence une différence de l'association de protéines de réparation, ainsi que des marques épigénétiques particulières entre les segments concernés par la CSR et ceux ciblés par la LSR chez la souris.

Nous nous sommes ensuite interrogés sur la LSR dans le lymphome de Hodgkin (HL pour Hodgkin lymphoma), car l'absence de BCR à la surface des cellules de Reed Sternberg pourrait provenir de cet évènement. Les résultats de séquençage

haut débit révèlent une réparation différente au cours de la LSR entre le HL et le contrôle (amygdales saines) ce qui nous laisser stipuler que des altérations intrinsèques aux systèmes de réparation de l'ADN dans les cellules tumorales sont en cause.

Globalement, nous avons développé « CSReport », un outil qui nous permet d'analyser la structure de réparation de l'ADN en partie, et de montrer une réparation similaire des CDB entre la souris et l'Homme et une différence de réparation de l'ADN entre la recombinaison CSR et LSR. De plus, nous avons mis en évidence une altération de la réparation dans des échantillons de lymphomes B (HL et CLL) comparé à des contrôles (amygdales saines).

Mots clés :

Recombinaison suicide / Recombinaison isotypique / Locus *IgH* / Réparation de l'ADN / Lymphomes B

Abstract

Mature B lymphocytes meeting with antigen (Ag) inside secondary lymphoid organs activates their terminal maturation, with occurrence of class switch recombination (CSR) and somatic hyper mutation (SHM).

Recently, our laboratory described for the first time *IgH* locus suicide recombination (LSR) (Péron et al., 2012). This process removes the whole constant genes of the locus, preventing Ig and BCR (B Cell Receptor) expression. The B cell is devoid of survival signals delivered by its receptor and is induced to apoptosis.

LSR seems to operate with same molecular steps as CSR : 1- transcription of targeted DNA regions, 2- generation of double strand breaks (DSB) from DNA lesions induced by AID (Activation-induced cytidine deaminase), 3- DNA repair by classical non homologous end joining (C-NHEJ) pathway. However, DNA repair during LSR was not fully understood, and this is the principal objectif of my PhD studies.

First, we developped a bioinformatic program « CSReport », to analyse high throughput sequencing (HTS) datas of *IgH* locus junctions (CSR and LSR) (Boyer et al., 2017). This tool allowed us to study the DSB repair systems, through determination of the structure at the junction site. Unexpectedly, our results show that DNA repair in DNA during LSR is similar between mice and human, and if CSR implicates C-NHEJ, LSR seems to invlove Alternative end joining (A-EJ) and /or homologous recombination (HR). These observations are consolidated by results showing a difference in the association of repair proteins, and in particular epigenitic marks between DNA segments concerned by CSR and those targeted by LSR in mice.

We asked ourselves about LSR in HL, because BCR absence on its Reed Sternberg cells surface may be a result of this recombination. HTS results reveal a different repair during LSR between HL and the control (healthy tonsils), which let us stipulate that alterations in DNA repair systems of tumoral cells are the cause.

Globaly, we developped « CSReport », a tool which permits us to study DNA repair structure in a part, and to show a similar DSB repair systems between mice and human, and a difference between CSR and LSR repair. Furthermore, we show an alteration in DNA repair of B lymphoma samples (HL and CLL) compared with the control (healthy tonsils).

Key words:

Suicide recombination / Class switch recombination / *IgH* locus / DNA repair / B lymphoma

Liste des abréviations

A

ADN Acide DésoxyriboNucléique

AID Activation-induced cytidine deaminase

Ac Anticorps

Ag Antigène

A-EJ Alternative End Joining

AML Acute myeloid leukemia

ALL Acute Lymphoblastic Leukemia

ARN Pol II ARN Polymérase II

APE1 Endonucléase AP1

ATM Ataxia Telangiectasia Mutated

α alpha

ADNsb ADN simple brin

ADNdb ADN double brin

B

BER Base Excision Repair

B-NHEJ Back-up NHEJ

BCR B Cell Receptor

53BP1 53 binding Protein 1

BRCA1 BReast Cancer1

BSAP B Specific Activator Protein

C

CTCF CCCTC-binding factor

ChIP Chromatin Immunoprecipitation

CSR Class Switch Recombination

CDR Complementarity Determining Region

C Cytidine

CSB Cassure simple brin

CDB Cassure double brin

cHL Classical Hodgkin's Lymphoma

CLL Chronic Lymphocytic Leukemia

CML Chronic myeloid leukemia

C-NHEJ Classical Non-Homologous End Joining

Chk Checkpoint Kinase

Cdc25 Cell division cycle 25

CDK Cyclin-dependent kinase

D

DDR DSB Dependent Response

DNA-PKcs DNA dependent Protein Kinase catalytic subunit

D-loop Displacement loop

dRP Désoxyribose phosphate

δ delta

E

EBV Epstein-Barr Virus

EXO1 Exonucléase1

EOPS Exempte d'organismes pathogènes spécifiques

ε epsilon

F

FEN1 Flap endonuclease 1

G

G Guanine

γ gamma

H

HAT Histones AcétylTransférase

HDAC Histone DésACétylase

HMT Histone MéthylTransférase

HDM Histone Déméthylase

HR Homologous Recombination

hDNA Hétéroduplex d'ADN

HS DNase I Hyper Sensitive

HP1 Heterochromatin protein 1

HJ Jonction Holliday

HL lymphome de Hodgkin

HR Recombinaison homologue

HEK Human embryonic kidney

I

IL-4 Interleukine 4

IFN γ Interféron gamma

Ig Immunoglobuline

K

Kb Kilobase

K Lysine

KAP1 KRAB-associated protein 1

κ Kappa

L

LB Lymphocyte B

L Light

Ligase IV LIG IV

LB Lymphocyte B

LSR Locus suicide Recombination

LS Like Switch

LPS Lipopolysaccharide

LT Lymphocyte T

LTf Lymphocyte T folliculaire

LTc Lymphocyte T cytotoxique

λ Lambda

M

Mb Mégabase

MMR Mismatch Repair

MME Microhomology mediated end joining

MARs Matrix attachment regions

MRE11 Meiotic Recombination 11 homolog

MMSET Multiple myeloma SET domain

MDC1 Mediator of DNA damage checkpoint 1

µ mu

mIg Ig membranaire

N

NBS1 Nijmegen Breakage Syndrome protein 1

NLPHL Nodular Lymphocyte Predominant Hodgkin Lymphoma

NHEJ Non homologous end joining

N Non templated

O

3'OH 3' hydroxyle

P

PTIP Pax transactivation domain-interacting protein

Pol δ Polymérase delta

PIKKs Phosphatidylinositol-3-kinase-like kinase family

PKA Protein Kinase A

PARP-1 Poly ADP Ribose Polymerase 1

PKA Protein Kinase A

Polθ Polymérase theta

PAXX Paralog of XRCC4

PHD Plant Homeodomain

P Palindromique

PL Plasmocyte

R

RSS Recombination Signal Sequence

RPA : Replication Associated Protein A

RIF1 Rap1 interacting factor 1

RAG Recombination Activating Gene

3'RR 3' Regulatory Region

S

SHM Somatic Hyper Mutation

S Switch

sIg Ig secrétée

SCID Severe Combined Immunodeficiency

S Sérine

T

T Thréonine

TdT Terminal désoxynucléotidyl Transférase

TLR Toll Like Receptor

TMEJ Tetha Mediated End Joining

TGF-β Transforming Growth Factor-β

U

U Uridine

UNG Uracile N-Glycosylase

X

XRCC4 X-Ray Repair Cross Complementing Protein 4

XRCC1 X-Ray Repair Cross Complementing Protein 1

XLF XRCC4-Like Factor

Y

Y Tyrosine

Liste des figures

Figure 1 : Structure d'une l'immunoglobuline (IgG)……17
Figure 2 : Les classes d'immunoglobuline……18
Figure 3 : Organisation du locus *IgH*……19
Figure 4 : Organisation des loci de chaîne lourde et chaînes légères d'Ig……20
Figure 5 : Mécanisme de recombinaisons V(D)J……24
Figure 6 : Recombinaisons V_HDJ_H et V_LJ_L des loci *IgH* et *IgL* respectivement….25
Figure 7 : Représentation du mécanisme CSR vers IgG1 (CSRγ1)……26
Figure 8 : Schéma de la transcription germinale lors de la CSR……26
Figure 9 : Représentation schématique de la contraction du locus *IgH* lors de la CSR……27
Figure 10 : Recrutement de AID lors de la CSR……27
Figure 11 : Marche arrière de l'ARN Pol II……27
Figure 12 : Schéma de la réparation par excision de bases (BER)……28
Figure 13 : Schéma de la réparation des mésappariements de bases (MMR)……29
Figure 14 : Régions LS du locus *IgH* murin……30
Figure 15 : Schéma de recombinaison suicide du locus *IgH* (LSR)……30
Figure 16 : Schéma de la voie DDR dans le cas de CDB……31
Figure 17 : Schéma de la réparation des CDB par le C-NHEJ……32
Figure 18 : Schéma de la réparation des CDB par le TMEJ……35
Figure 19 : Schéma de la réparation des CDB par le système de la HR……36
Figure 20 : Structure de la chromatine……37
Figure 21 : Représentation schématique de la fonction de SETD2 dans la HR..38
Figure 22 : Points de contrôles du cycle cellulaire……39
Figure 23 : schéma de la régulation du point de contrôle G2/M par les kinases ATM et ATR……39
Figure 24 : Schémas des différents types de structure pour les jonctions de recombinaison du locus *IgH*……52
Figure 25 : Représentation des jonctions LSR et CSRα dans les LB humains en fonction de leur structure……55
Figure 26 : Représentation des jonctions LSR et CSRγ1 de LB de souris en fonction de leur structure……56

Figure 27 : Comparaison de la fréquence des cellules marquées B220+/IgG1+ par cytométrie en flux……57

Figure 28 : Taux d'enrichissement des protéines impliquées dans des voies de réparation des CDB au niveau des régions Sμ, Sγ1, Sγ3, HS1/2 et HS4……...57

Figure 29 : Taux d'enrichissement de marques d'histones au niveau des régions Sμ, Sγ1, Sγ3, HS1/2 et HS4……58

Figure 30 : Représentation des résultats de l'analyse par CSReport des jonctions LSR obtenues à partir d'échantillons humains en fonction de leur structure….64

Figure 31 : Représentation de la localisation des points de cassure au niveau des régions donneuse Sμ et acceptrice 3'RR2 de la recombinaison LSR dans des librairies d'échantillons humains……65

Figure 32 : Influence du contexte chromatinien sur le recrutement de protéines de réparation lors des recombinaisons du locus *IgH*……73

Liste des tableaux

Tableau 1 : Résultats de l'analyse par CSReport des jonctions LSR issues de librairies d'amygdales saines, de HL et de CLL……..….61/62

Table des matières

Remerciements ... 3
Résumé ... 4
Abstract ... 6
Liste des abréviations ... 8
Liste des figures ... 11
Liste des tableaux ... 12
INTRODUCTION BIBLIOGRAPHIQUE ... 15
I. Les immunoglobulines ... 17
1. Structure ... 17
2. Isotypes ... 18
3. Loci ... 18
4. Éléments cis-régulateurs ... 20
II. Les réarrangements des gènes d'immunoglobulines ... 23
1. Les recombinaisons V(D)J ... 23
1.1. Mécanisme de recombinaison ... 24
1.2. Régulation par exclusion allélique ... 25
2. La recombinaison isotypique des immunoglobulines (CSR) ... 25
1. Mécanisme de recombinaison ... 26
1.1. Transcription germinale ... 26
1.2. Formation de « R-loops » ... 27
1.3. Induction de cassures par AID ... 28
3. La recombinaison suicide du locus *IgH* (LSR) ... 29
3.1. Découverte de la recombinaison ... 29
3.2. Mécanisme moléculaire ... 30
3.3. Potentiel rôle physiologique ... 30
III. La réparation des CDB au locus *IgH* ... 31
1. Détection et prise en charge des cassures ... 31
2. Réparation des cassures ... 32
2.1. Le NHEJ classique (C-NHEJ) ... 32
2.2. Le NHEJ alternatif (A-EJ) ... 34
2.3. Le système de la HR ... 36
3. L'orientation entre les différentes voies de réparation des CDB ... 37
3.1. Le rôle de la structure chromatinienne ... 37
3.2. Le rôle du cycle cellulaire ... 39
3.3. Le rôle de la résection des extrémités ... 40

IIII. Les lymphomes B 41
1. Lymphome de Hodgkin (HL) 41
2. Leucémie lymphoïde chronique (CLL) 42
MATERIEL ET METHODES 43
OBJECTIF DE LA THESE 48
RESULTATS 50
Partie I : Développement d'un protocole d'étude et d'un outil d'analyse des recombinaisons du locus *IgH* 51
Partie II : Signature de réparation des CDB au cours de la LSR et de la CSR : ces deux recombinaisons du locus *IgH* impliquent des voies de réparation différentes ! 55
1. Analyse de la structure de réparation au cours de la LSR 55
2. Recrutement des molécules de réparation aux sites de CSR et LSR 57
3. Analyse du contexte chromatinien des segments mis en jeu lors des recombinaisons CSR et LSR 58
Partie III : La LSR dans le lymphome de Hodgkin 59
DISCUSSION 66
I. Stratégie d'étude de la recombinaison LSR 68
II. Signature de réparation des CDB différente entre LSR et CSR 69
II.1. Recrutement des protéines de réparatation selon le segment de recombinaison concerné 71
II.2. Contexte chromatinien différent entre les segments accepteurs de recombinaison CSR et LSR 72
III. Réparation des CDB altérée lors de la LSR dans le lymphome de Hodgkin 74
CONCLUSION ET PERSPECTIVES 76
BIBLIOGRAPHIE 79
ANNEXES 94

INTRODUCTION BIBLIOGRAPHIQUE

L'organisme met en œuvre deux types de réponses immunitaires afin de se protéger contre les agents pathogènes. Une réponse innée, non spécifique, première barrière protectrice, et une réponse spécifique adaptative, ou acquise. La deuxième correspond au versant cellulaire, c'est-à-dire, à l'action des lymphocytes T cytotoxiques (LTc), et au versant humoral, avec intervention des anticorps (Ac) ou immunoglobulines (Ig) sécrétés par les plasmocytes (PL). Les PL correspondent au stade final de différenciation des lymphocytes B (LB), et sont des cellules spécialisées dans la sécrétion d'Ig. Les LB, à la différence des PL expriment les Ig sous forme membranaire (mIg) très majoritairement et très peu sous la forme sécrétée (sIg). Les mIg forment avec les sous-unités Igα et Igβ le récepteur BCR (pour B Cell Receptor).

Les Ig sont formées de quatre chaînes polypeptidiques liées par des ponts disulfures : deux chaînes lourdes identiques (IgH, H pour Heavy), et deux chaînes légères identiques (IgL, L pour Light), codées par des loci différents. Chaque chaîne de l'Ig est constituée d'une partie constante (C) et d'une partie variable (V). La partie V permet la reconnaissance de l'antigène (Ag). Le tétramère d'environ 150 kDa, est issu de l'assemblage des chaînes IgH avec les chaînes IgL (Porter, 1973).

Chez les mammifères, le développement des LB débute dans la moelle osseuse, organe lymphoïde primaire, et site de formation de LB matures naïfs, à partir de progéniteurs hématopoïétiques. Pendant les stades précoces de développement des LB, ont lieu les réarrangements V(D)J des loci des chaînes d'Ig qui permettent l'expression des Ig. Les LB matures naïfs vont par la suite migrer vers les organes lymphoïdes secondaires comme les ganglions lymphatiques et la rate, en empruntant la circulation sanguine et lymphatique. Dans ces compartiments secondaires, Les LB rencontrent l'Ag qui induit leur maturation terminale.

Des remaniements génétiques majeurs sont vécus par les cellules B pendant leur maturation terminale. Ils comprennent notamment l'hypermutation somatique (SHM) et la commutation de classe (CSR). La SHM porte sur la région variable des gènes d'Ig, et aboutit à la sélection de clones de LB exprimant une Ig de haute affinité vis-à-vis de l'Ag rencontré. La CSR quant à elle, touche les gènes constants d'*IgH* et permet de générer des classes différentes d'Ig tout en maintenant la même affinité pour l'Ag. Ces deux mécanismes visent à assurer une réponse immunitaire adaptée et donc efficace contre l'agent pathogène. La CSR et la SHM font intervenir une enzyme essentielle appelée AID (Activation-induced cytidine deaminase) (Masamichi Muramatsu et al., 2000; Revy et al., 2000).

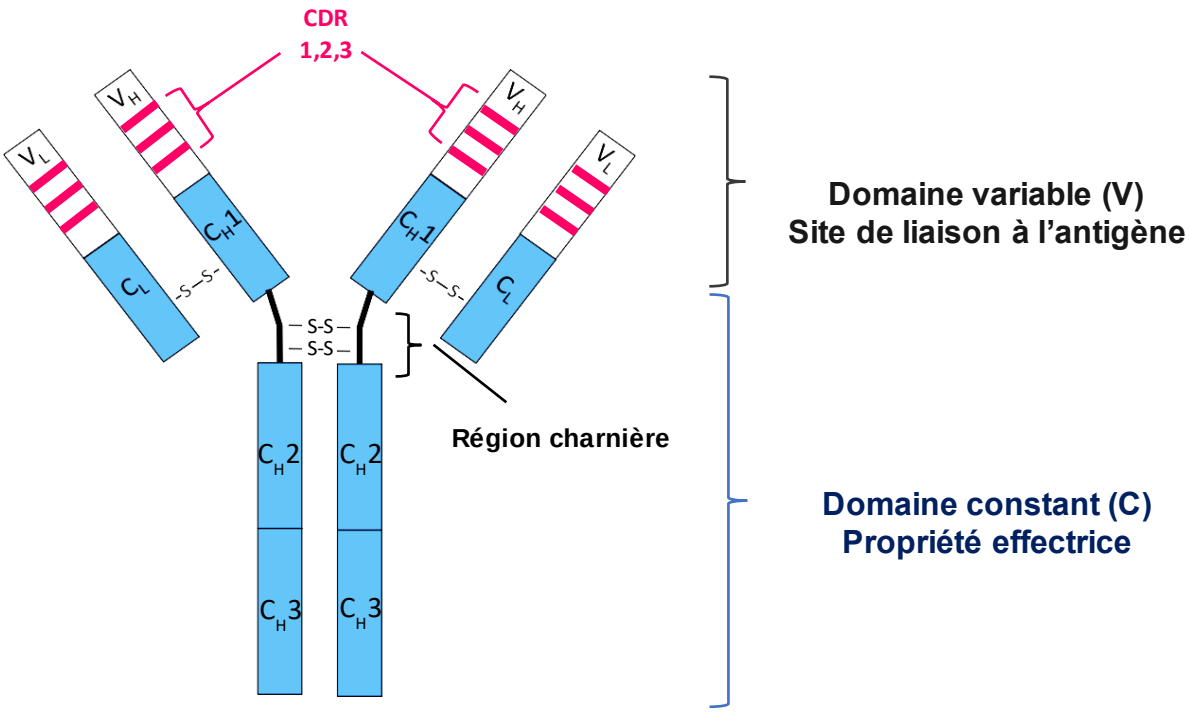


Figure 1 : Structure d'une l'immunoglobuline (IgG). L'immunoglobuline, constituée d'une chaîne légère (L pour light) et d'une chaîne lourde (H pour heavy) présente deux parties principales : le domaine variable (V), site de liaison à l'antigène, et le domaine constant (C) qui confère la propriété effectrice de l'Ig et contient la région charnière située entre les domaines C_H1 et C_H2 de la chaîne lourde.

Dernièrement, un nouveau type de réarrangement génétique du locus *IgH* a été découvert dans les LB, la recombinaison suicide du locus *IgH* (LSR) (Péron et al., 2012). Le processus moléculaire aboutissant à la LSR semble proche de celui de la CSR. En revanche la finalité n'est pas la même, car la LSR empêche l'expression de l'Ig et conduit à la mort de la cellule B affectée.

Le travail que j'ai effectué au cours de ma thèse vise à analyser en détail le mécanisme moléculaire de la réparation des cassures double brin (CDB) de l'ADN générées au cours de la LSR. L'objectif est de déterminer si cette étape fait également appel à la voie de réparation du C-NHEJ qui est impliquée dans la CSR. Ce travail a été mené chez l'Homme et chez la souris. Nous avons aussi initié le travail questionnant l'implication de la LSR dans le HL.

I. Les immunoglobulines

La protection de notre organisme repose sur la reconnaissance d'un large éventail de micro-organismes par le système immunitaire. Parmi les composants de ce système, le BCR exprimé à la surface des LB et formé d'Ig, possède la capacité de fixer de manière spécifique les agents potentiellement pathogènes.

1. Structure

Les Ig sont des glycoprotéines exprimées à la surface des LB ou sécrétées (on parle alors d'Ac) par les PL. Elles sont formées d'une structure commune constituée de deux chaînes polypeptidiques lourdes H identiques associées à deux chaînes polypeptidiques légères L.

Comme le montre la figure 1, les chaînes lourdes et les chaînes légères présentent un domaine variable et un domaine constant. L'union des parties variables des deux types de chaînes porte le site de reconnaissance et de fixation de l'Ag. La partie constante des chaînes lourdes (CH) porte les caractéristiques de fonction propres à chaque classe d'Ig, comme la fixation du complément qui provoque la lyse des cellules, la sécrétion dans la lumière du tractus muqueux digestif ou dans le lait maternel.

Les chaînes légères sont reliées aux chaînes lourdes partenaires par un pont disulfure formé entre leurs résidus cystéines, ainsi que par des interactions non

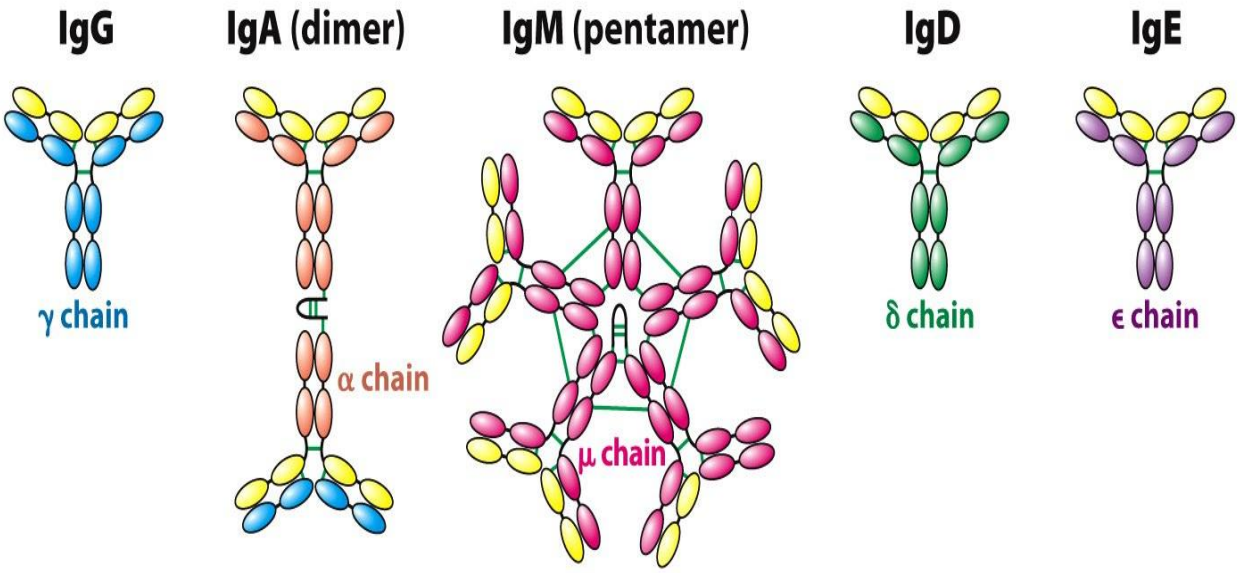


Figure 2 : Les classes d'immunoglobulines. Il existe cinq isotypes d'Ig (IgG, IgA, IgM, IgD et IgE) qui sont déterminés par la classe de la chaîne lourde exprimée (γ,α, μ,δ et ε). *Biochemistry seventh Edition, 2012 W.H. Freeman and company*

covalentes entre les domaines V_H et V_L, et les domaines C_{H1} et C_L. Les deux chaînes lourdes sont également liées entre elles par des ponts disulfures.

2. Isotypes

La majeure diversité des sites de reconnaissance à l'Ag réside dans trois parties hypervariables (Complementarity Determining Region ou CDR) situées à la fois sur les chaînes lourdes et les chaînes légères. Ces régions sont soutenues par quatre parties d'une moindre diversité, nommées « Framework » (Schroeder et al., 2010).

Il existe cinq classes d'Ig déterminées par l'isotype de la chaîne lourde :μ (mu), γ (gamma), α (alpha), δ (delta) et ε (epsilon) qui sont codées par l'Igμ, l'Igγ, l'Igα, l'Igδ et l'Igε respectivement **(Figure 2)**. Parmi les IgG (ou IgA humaines), la région constante des chaînes lourdes peut porter une diversité supplémentaire et déterminer des isotypes appelés « sous-classes ». Les IgG sont réparties en quatre sous-classes : les IgG1, IgG2a, IgG2b ainsi que IgG3 chez la souris. Chez l'Homme, on trouve les IgG1, IgG2, IgG3 et l'IgG4, ainsi que deux sous-classes d'IgA : IgA1 et IgA2. Il existe également deux types de chaînes légères, Igκ (kappa) et Igλ (lambda) qui peuvent s'associer avec chacune des classes de chaîne lourde précédemment citées.

Les cinq classes d'Ig existent sous les deux formes secrétée ou soluble et membranaire. Les Ig membranaires sont exprimées par les LB au cours de leur développement. Les LB au stade immature pré-B, expriment seulement des IgM de surface. Les LB matures qui n'ont pas encore rencontré d'Ag sont caractérisés par l'expression de l'IgD et de l'IgM à leur surface. La stimulation antigénique des cellules B induit l'arrêt de l'expression des IgD, et si le LB subit le processus de CSR alors une Ig autre que l'IgM et l'IgD sera exprimée.

3. Loci

Les gènes d'Ig sont exprimés par trois loci définis. Deux loci codent les chaînes légères, le locus *Igκ* et le locus *Igλ*, et la chaîne lourde est codée par le locus *IgH*. Ces loci portent une série de segments géniques distincts : les segments de variabilité (V), les segments de diversité (D) retrouvés uniquement sur le locus *IgH*, les segments de jonction (J), et les segments constants (C).

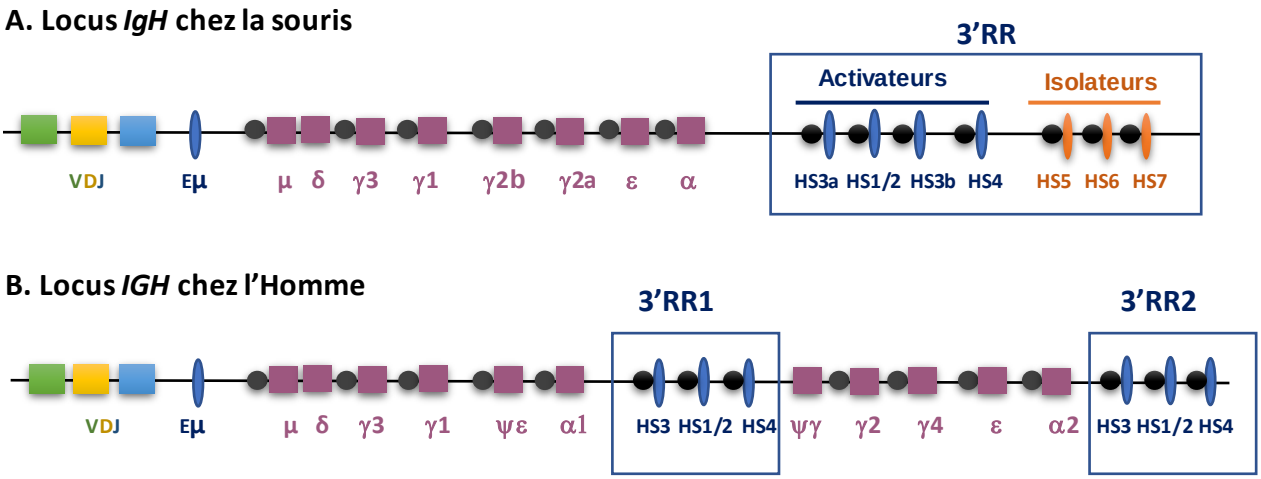


Figure 3 : Organisation du locus *IgH*. A. Le Locus *IgH* chez la souris s'étend sur environ 2,3 Kb, avec les segments V(D)J en position 5' du locus qui codent pour la région variable des Ig et en 3', les gènes constants (C) qui codent pour la classe de l'Ig. Les gènes C sont précédés des régions switch (S) mises en jeu lors de la CSR (sauf le Cδ). En 3' du locus, la région régulatrice 3'RR comprend des activateurs transcriptionnels (HS3a, HS1/2, HS3b et HS4) qui interviennent dans la régulation de la transcription du locus *IgH* et des isolateurs (HS5, HS6 et HS7). **B.** Le Locus *IGH* Chez l'Homme s'étend sur environ 1250 Kb, avec une organisation similaire à celle de la souris, excepté la présence d'une duplication des gènes constants et de la région 3'RR contenant les enhancers HS3, HS1/2 et HS4.

Les segments V, D et J codent pour les régions variables, tandis que les segments C codent pour les régions constantes de l'Ig. Il est à noter que la partie du locus *IgH* codant pour les segments V, D et J affiche une organisation similaire entre la souris et l'Homme **(Figure 3)**. En effet, du côté 5' vers la partie 3' du locus, on trouve les segments V, D, et J et cet ordre est le même chez l'Homme. En revanche, l'organisation de la partie des gènes constants diffère d'une espèce à l'autre. Il existe quatre sites hypersensibles à la DNaseI nommés HS (DNaseI Hyper Sensitive) chez la souris (HS3a, HS1/2, HS3b, et HS4), contre trois seulement chez l'Homme (HS1/2, HS3, et HS4). Chez ce dernier, la région régulatrice en 3' du locus (3'RR pour 3' Regulatory Region) a été dupliquée en même temps que le cluster des gènes constants Cγ-Cε-Cα, avec une « 3'RR1 » en aval de Cα1 et une autre « 3'RR2 » en aval de Cα2 (Birshtein, 2014).

a. Locus des chaînes lourdes d'immunoglobulines (IgH)

Chez la souris, le locus *IgH* se situe sur le chromosome 12 et s'étend sur une longueur de 3 Mb (Brodeur and Riblet, 1984) avec environ 101 gènes V_H fonctionnels, environ 10 segments D qui s'étendent sur 80 Kb en 3' des segments V_H, puis un groupe de 4 segments J_H fonctionnels 0,7 Kb en 3' du segment D distal. Finalement, un intron sépare le dernier J_H du premier exon de la région constante, $C_{H\mu1}$. Il est à noter que la moitié des segments V_H appartient à la famille J558. Du côté 3' du locus, 8 gènes codent pour les 8 classes et sous-classes d'Ig répartis dans un ordre défini : μ, δ, γ3, γ1 γ2b, γ2a, ε et α.

Chez l'Homme, le locus *IGH* est porté par le chromosome 14 avec une taille d'environ 1250 Kb et compte environ 40 segments V_H, 24 segments D et 6 segments J. Il existe deux sous classes d'IgA codées par les gènes Cα1 et Cα2 et 4 sous-classes d'IgG.

b. Loci des chaînes légères (Igκ et Igλ)

Locus *kappa* (*Igκ*)

Le locus *Ig kappa (Igκ)* est localisé sur le chromosome 6 chez la souris et sur le chromosome 2 chez l'Homme (Brekke and Garrard, 2004; Zachau, 2000). Il s'étend

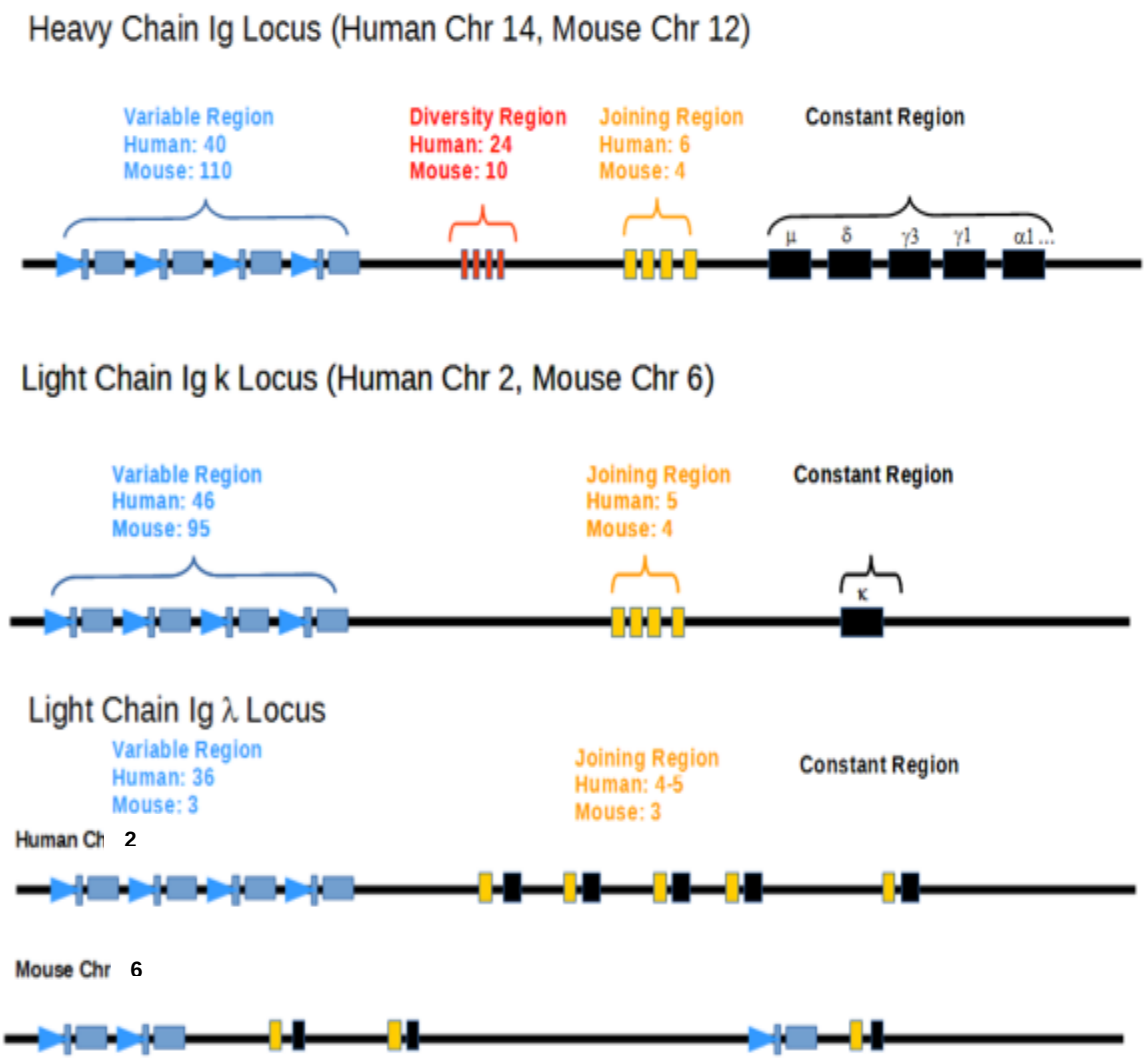


Figure 4 : Organisation des loci de la chaîne lourde et de la chaîne légère d'Ig. Le nombre de segments variable (V), diversité (D) et jonction (J) est indiqué à chaque fois pour la souris et pour l'Homme. Ces gènes en configuration germinale doivent subir les réarrangements V(D)J pour permettre l'expression de l'Ig. (*Stephen Pelsue MBS)*

sur environ 3,3 mégabases (Mb) et comprend des segments Vκ, Jκ et Cκ. Il existe environ 95 gènes Vκ regroupés en 18 familles en fonction de leur homologie de séquence. En effet, si deux séquences affichent une homologie supérieure à 80% elles sont classées dans la même famille de gènes V. En aval de ces gènes Vκ, il existe 4 segments Jκ suivis d'un segment Cκ unique qui code pour la région constante de toutes les chaînes légères κ **(Figure 4)**. Dans le sérum de souris 95% des Ig sont constituées d'une chaîne lourde associée à une chaîne légère Igκ (Kabat et al., 1992).Chez l'Homme, il y a environ 46 segments V, et 5 segments J.

Locus *lambda (Igλ)*

Le locus *lambda* est situé sur le chromosome 6 chez la souris, et sur le chromosome 22 humain (Blomberg et al., 1981; Frippiat et al., 1995). Il couvre une région d'environ 200 kilobases (Kb) et comprend 4 familles de segments Vλ, 3 segments Jλ, ainsi que 4 régions constantes Cλ **(Figure 4)**. De cette façon, les gènes de la chaîne légère λ associent chaque segment J à un segment C. Il est à noter que les segments JCλ4 ne sont pas exprimés à cause d'un site d'épissage défectueux. Seulement 5% des Ig de souris ont recours à une chaîne légère λ, contre 40% chez l'Homme (Kabat et al., 1992). Le locus Igλ humain contient environ 36 gènes V, 5 gènes J et 5 gènes codant pour la partie constante de l'Ig.

4. Éléments cis-régulateurs

a. Éléments cis-régulateurs du locus IgH

Il existe des éléments cis-régulateurs du locus *IgH* dont l'activité de promouvoir l'expression du locus est spécifique des LB. Le premier à être découvert est l'activateur intronique Eμ qui s'étend sur environ 1 Kb, et se situe entre la région J_H et la région Sμ (Banerji et al., 1983). Des travaux sur des lignées cellulaires B ont indiqué que certaines conservaient la capacité de transcrire leurs gènes d'*IgH* malgré la délétion de Eμ suggérant l'existence d'autres activateurs transcriptionnels (Klein S. et al., 1984; Lieberson et al., 1991). En effet, différents éléments de petite taille (core enhancers), ont été successivement identifiés, regroupés au sein de la région 3'RR, qui s'étend sur

environ 30 Kb, et se situe en aval du gène codant pour Cα. (Dariavach Piona et al., 2005; Lieberson et al., 1991). Les rôles respectifs de ces activateurs ont été peu à peu clarifiés. D'un côté, la régulation par Eμ intervient dans les stades précoces de différenciation des LB. Le KO de Eμ a révélé que cet activateur joue un rôle essentiel dans les recombinaisons V(D)J (Ferrier et al., 1990; Perlot et al., 2005).

La région 3'RR chez la souris comprend 4 activateurs transcriptionnels HS (HS3a, HS3b, HS1/2, et HS4) qui sont impliqués lors de l'activation des LB **(Figure 3.A)**. Cette région est suivie en 3' d'éléments isolateurs (HS5, HS6, et HS7) riches en sites de liaison aux facteurs CTCF (CCCTC-binding factor) et PAX5/BSAP (B Specific Activator Protein) (Chatterjee et al., 2011).

PAX5 est un facteur de transcription qui joue un rôle capital dans l'engagement des cellules dans la lignée B, ainsi que dans le système nerveux et les gonades (Adams et al., 1992). Ce facteur de l'identité B se lie au niveau des activateurs de la 3'RR, de même que d'autres facteurs (E2A, Oct1, Oct2, NF-κB, PU1, AP1, YY1...) (Bain et al., 1994; Zhuang et al., 1994; Schubart et al., 1996; Matthias, 1998; Tang and Sharp, 1999; Pinaud et al., 2011). La délétion totale de la 3'RR chez la souris abolie la CSR et la sécrétion d'Ig vers l'ensemble des isotypes (sauf pour Igγ1) (Vincent-Fabert et al., 2010). En revanche, sa délétion dans la lignée murine CH12F3 réduit la CSR vers l'Igα sans pour autant la bloquer, avec détection de transcrits germinaux. Ceci suggère que la 3'RR aurait un rôle dans l'initiation de la CSR autre que l'induction de la transcription des régions S (Kim et al., 2016). La CH12F3 est une lignée de lymphome murin qui possède la capacité à réaliser la CSR vers l'Igα, après stimulation avec de l'IL-4 (Interleukine 4), TGF-β (Transforming Growth Factor β) et CD40L pour 72 heures (Nakamura et al., 1996).

Il existe également des activateurs transcriptionnels du côté 5' du premier segment V (HS1, HS2, HS3a et HS3b) (Pawlitzky et al., 2006), ainsi que dans la région comprise entre les régions V_H et D, avec la présence de 6 sites HS (HS1 jusqu'à HS6) (Featherstone et al., 2010). Nous pouvons aussi citer l'élément régulateur D_{Q52} qui contient l'activateur ainsi que le promoteur. Cette région située du côté 5' du premier segment J est délétée après réarrangement des gènes de la chaîne lourde (Kottmann et al., 1994). La délétion du gène D_{Q52} et d'une partie de son promoteur chez la souris induit une réduction des recombinaisons V(D)J faisant particulièrement usage des

segments J_{H3} et J_{H4}. Il a été admis que la région D_{Q52} contrôle l'accessibilité des éléments J_H et renforce les réarrangements DJ (Nitschke et al., 2001).

b. Éléments cis-régulateurs du locus Igκ

Le locus *Igκ* contient trois activateurs transcriptionnels, un activateur intronique MiEκ (matrix intronic enhancer) situé entre les segments J et la région constante Cμ (Banerji et al., 1983), l'activateur E en 3' (3'Eκ) situé 9 Kb en 3' du gène codant *Igκ* (Meyer and Neuberger, 1989), et l'élément Ed (downstream enhancer) situé plus en aval, à environ 8 Kb de E3' (Liu et al., 2002). La délétion de MiEκ ou de 3'Eκ dans le locus *Igκ* diminue les réarrangements de ce dernier (Gorman et al., 1996; Xu et al., 1996). En revanche, la délétion combinée des deux activateurs bloque complétement les réarrangements Vκ/Jκ indiquant que les régulateurs MiEk et 3'Ek possèdent des rôles redondants mais essentiels pour l'activation des réarrangements au niveau du locus *Igκ* (Inlay et al., 2002).

Des études ont révélé l'implication des deux activateurs transcriptionnels cités précédemment dans la déméthylation du locus *Igκ*. En effet, la déméthylation de l'ADN est requise pour des réarrangements Vκ/Jκ efficaces (Inlay et al., 2002). Additionnellement, la délétion simultanée de E3' et Ed chez la souris induit une diminution de l'expression des gènes *Igκ* réarrangés (Zhou et al., 2010). Dernièrement, l'élément de régulation Dm a été découvert. Il se situe à proximité de l'activateur iEk et participe avec les autres régulateurs à la déméthylation du locus *Igκ*. Sa délétion n'altère pas le statut de méthylation du locus *Igκ* endogène (Levin-Klein et al., 2014). De plus, un site HS10 a été identifié comme régulateur potentiel du locus *Igκ.* Il est situé en 3' de ce locus en aval de Eiκ d'environ 40 Kb. Des études montrent que HS10 et Ed sont requis pour une expression maximale d'*Igκ* au niveau des PL (Zhou et al., 2012).

c. Éléments cis-régulateurs du locus Igλ

Le locus *Igλ* comprend deux activateurs transcriptionnels avec une forte homologie de séquence (90%) entre $E_{\lambda 2\text{-}4}$ et $E_{\lambda 3\text{-}1}$. Le premier est situé à environ 15,5 Kb en 3' du gène Cλ4, et $E_{\lambda 3\text{-}1}$ à environ 35 Kb en 3' de Cλ1 (Hagman et al., 1990).

$E_{\lambda 3\text{-}1}$ s'avère essentiel pour les recombinaisons Vλ/Jλ (Haque et al., 2013). De plus, il existe 4 sites (HS1, HS2, HS3 et HS4) en aval du gène JCλ1 dont HS1 situé du côté 5' de $E_{\lambda 3\text{-}1}$. HS1 est le seul élément qui semble indispensable pour les réarrangements du locus Igλ (Haque et al., 2013).

II. Les réarrangements des gènes d'immunoglobulines

Les gènes qui codent pour les chaînes d'Ig doivent exécuter une série de réarrangements afin de produire une Ig complète et fonctionnelle. Dans la moelle osseuse, pendant les premiers stades de développement du LB, le répertoire primaire des Ig est formé grâce aux recombinaisons V(D)J de la chaîne lourde puis de la chaîne légère. Par la suite, la maturation terminale des cellules B inclue les processus de CSR et de SHM qui vont diversifier le répertoire primaire des Ig et permettre l'optimisation et la « maturation » de la réponse immunitaire. La CSR se produit exclusivement sur le locus *IgH* et permet l'expression d'une classe d'Ig différente de l'IgM initialement exprimée. La SHM quant à elle introduit des mutations aléatoires dans les régions variables des chaînes lourdes et légères, mais secondairement soumises à un processus de sélection cellulaire intense qui ne conserve que les mutations les plus avantageuses pour l'optimisation de l'affinité de leur Ig pour l'Ag.

1. Les recombinaisons V(D)J

Les recombinaisons V(D)J sont au cœur de la capacité des LB à générer un répertoire d'Ig hautement diversifié et spécifique. Dans les précurseurs lymphoïdes communs les segments V_H, D et J_H, pour l'*IgH*, V_L et J_L pour l'*IgL* sont en configuration germinale, et ne permettent pas l'expression d'une Ig complète étant donné leur éloignement physique les uns par rapport aux autres. Les recombinaisons V(D)J sont un ensemble de réarrangements somatiques qui surviennent en absence d'exposition aux Ag lors des étapes les plus précoces du développement lymphoïde dans la moelle osseuse.

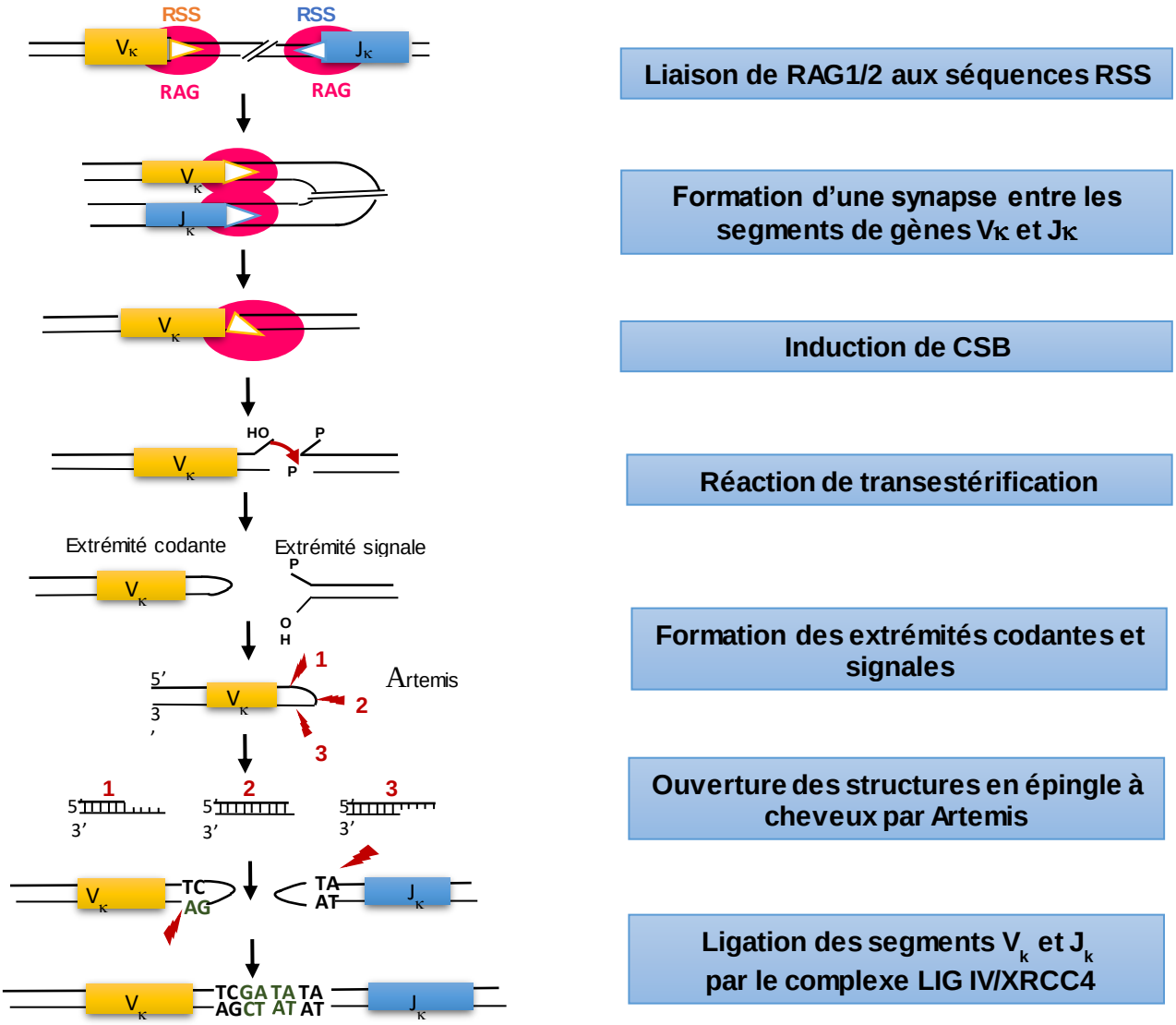


Figure 5 : Mécanisme de recombinaison V(D)J. Les réarrangements V(D)J sont guidés par les séquences RSS qui recrutent le complexe RAG. Ce dernier génère des cassures simple brin puis les extrémités codantes subissent l'action de Artemis et enfin l'étape de réparation de l'ADN fait intervenir la voie du C-NHEJ. Après l'action des RAG, au niveau de la cassure simple brin, le groupement 3' hydroxyle (3'OH) libre à l'extrémité du brin codant du segment Vκ attaque le groupement phosphate sur le brin opposé (réaction de transestérification), créant ainsi une cassure double brin avec des extrémités en structure en épingle à cheveux. Ces extrémités sont dites codantes, tandis que les extrémités de la séquence d'ADN excisée sont appelées signales. Les structures en épingle à cheveux sont clivées par Artemis, un facteur impliqué dans la voie de réparation des cassures double brin de l'ADN du C-NHEJ.

1.1. Mécanisme de recombinaison

Induction de cassures de l'ADN

Le mécanisme de recombinaison V(D)J est similaire entre LB et LT et se déroule suivant plusieurs étapes **(Figure 5)**. Dans un premier temps, la recombinase RAG1/2 forme un complexe avec les séquences RSS flanquant les segments de gènes à réunir. La formation du complexe est initiée par la reconnaissance des nonamères de RSS par RAG1, selon la règle 12/23 évoquée précédemment. Ensuite, les protéines RAG1/2 génèrent deux cassures simple brin (CSB), une sur le brin codant du gène V à la jonction entre la région V et l'heptamère, et l'autre sur le brin non codant au niveau de la jonction heptamère/région J. Au cours de cette étape, le groupement 3' hydroxyle (3'OH) libre à l'extrémité du brin codant du segment V ou J attaque le groupement phosphate sur le brin non codant opposé (réaction de transestérification), créant ainsi une CDB avec des extrémités sous forme d'épingle à cheveux. Ces structures appelées « jonctions codantes » sont clivées par Artemis, un facteur impliqué dans la voie de réparation du C-NHEJ.

Résolution des CDB

Le C-NHEJ est le système qui prend en charge la réparation des CDB générées lors de la recombinaison V(D)J. Les extrémités signales à bouts francs sont liées de façon directe, sans modification de leur séquence nucléotidique. Les extrémités codantes subissent quant à elles une ouverture asymétrique de l'épingle à cheveux par Artemis entrainant l'incorporation aléatoire de nucléotides N (N pour Non templated) par la Terminal désoxynucléotidyl Transférase (TdT) (Alt and Baltimore, 1982; Desiderio et al., 1984). La TdT est une polymérase exprimée pendant les stades précoces du développement des lymphocytes B et T.

Des études ont montré l'absence des nucléotides N au niveau des régions codantes chez des souris déficientes pour *Ku80* , ce qui indique que l'insertion de nucléotides N par la TdT est dépendante du facteur Ku80 (Bogue et al., 1997). Les modifications apportées au niveau de la jonction codante participent à l'augmentation de la diversité du répertoire de reconnaissance des Ig.

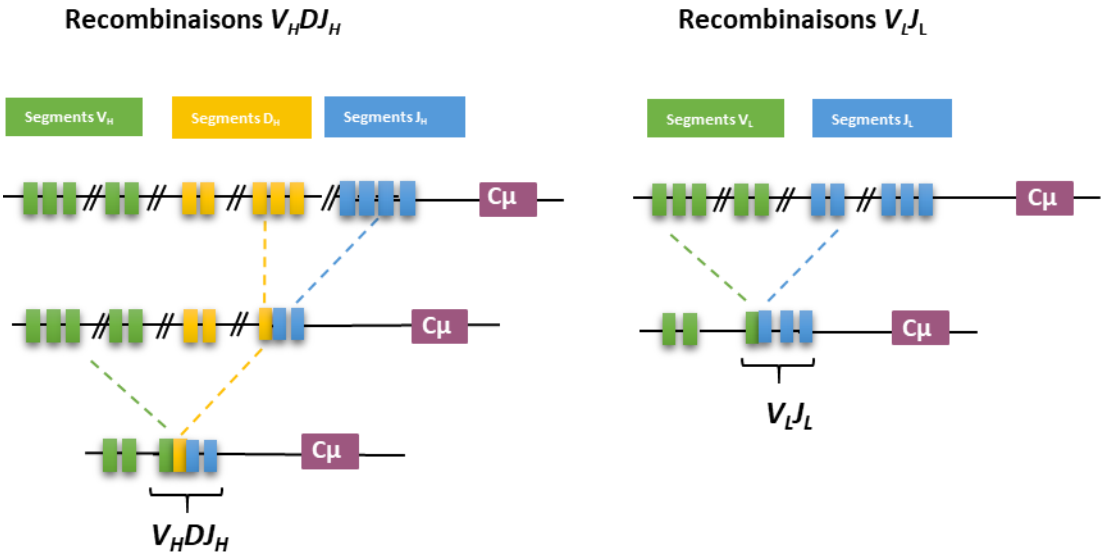


Figure 6 : Recombinaisons V_HDJ_H et V_LJ_L des loci *IgH* et *IgL* respectivement. Les réarrangements V(D)J permettent l'expression d'une Ig fonctionnelle et s'opèrent de façon séquentielle, en recombinant en premier les segments D_H/J_H puis V_H/DJ_H pour le locus de la chaîne lourde (*IgH*) puis V_L/J_L pour le locus de la chaîne légère *Igκ* et *Igλ*. Les réarrangements qui associent un segment V_H au segment DJ_H ou V_L au segment J_L respectivement, ont lieu sur un seul allèle reflétant le mécanisme d'exclusion allélique. Cependant, dans le cas où un allèle présente une recombinaison V(D)J non productive, les segments du second allèle vont être réarrangés. Lorsque les réarrangements sont non productifs, la cellule B est induite à l'apoptose.

1.2. Régulation par exclusion allélique

Les réarrangements V(D)J se produisent de façon aléatoire mais ordonnée. Il se déroulent d'abord au niveau du locus *IgH,* puis aux loci *Igκ* et *Igλ* (**Figure 6**). Dans le premier, les réarrangements D/J_H ont lieu avant les recombinaisons V_H/DJ_H et ce au niveau des deux allèles *IgH*. Tandis que les réarrangements qui associent un segment V_H au segment DJ_H déjà formé, ont lieu sur un seul allèle, reflétant le mécanisme d'exclusion allélique (Alt et al., 1984).

Lorsque les réarrangements V(D)J sont productifs, c'est-à-dire assemblage des différents segments sans codon de terminaison prématurée, le réarrangement sur le second allèle n'a pas lieu. Cependant, dans le cas où un allèle *IgH* présente une recombinaison V(D)J non productive, les segments du second allèle vont se réarranger. Si aucune des deux recombinaisons n'est productive, la cellule sera éliminée par apoptose (Rajewsky, 1996).

Comme pour les chaînes lourdes, l'exclusion allélique concerne également les chaînes légères. L'exon V/J une fois formé sera associé à l'exon constant par épissage du pré-ARNm. La chaîne légère est ensuite associée à la chaîne lourde µ pour produire une IgM complète qui est la classe d'Ig produite et exprimée par défaut à la surface du LB. En effet, les autres classes d'Ig, l'IgD mise à part, ne peuvent être exprimées en absence du processus de CSR, que nous aborderons plus loin.

Il est intéressant de noter que l'étude de la clonalité des LB dans un échantillon biologique par l'analyse des réarrangements V(D)J est couramment utilisée en diagnostique afin d'identifier une lymphoprolifération B anormale ou dans le traitement anti-tumoral pour quantifier la « maladie résiduelle » chez un patient (Libra et al., 2002).

2. La recombinaison isotypique des immunoglobulines (CSR)

La CSR est un réarrangement génétique du locus *IgH* qui se produit après activation des LB par l'Ag, et permet la modification de la classe d'Ig. Elle concerne la partie constante de la chaîne *IgH*, la partie variable quant à elle n'est pas modifiée. Cependant, des études ont montré l'existence d'une recombinaison CSR de moindre fréquence vers l'isotype IgD, et qui repose sur la présence d'une région S atypique en

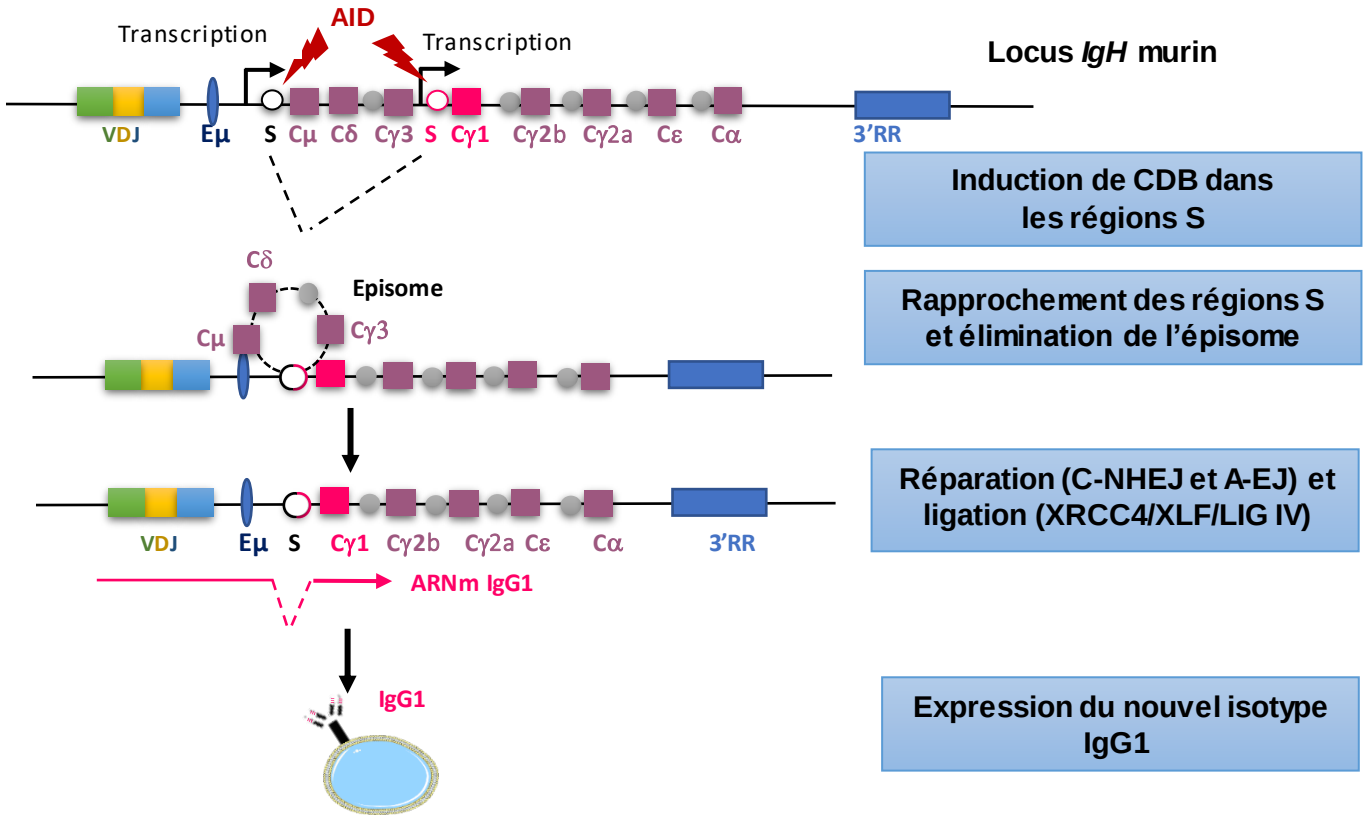


Figure 7 : Représentation du mécanisme CSR vers IgG1 (CSRγ1). Les segments d'ADN donneur et accepteur de la recombinaison CSR (ici Sμ et Sγ1) sont transcrits ce qui permet l'accessibilité à l'ADN à l'action de AID qui génère des lésions de l'ADN aboutissant à la création de CDB. Les régions sont rapprochées et les cassures réparées principalement par le système du NHEJ et permettant l'expression d'une nouvelle classe d'Ig (ici IgG1).

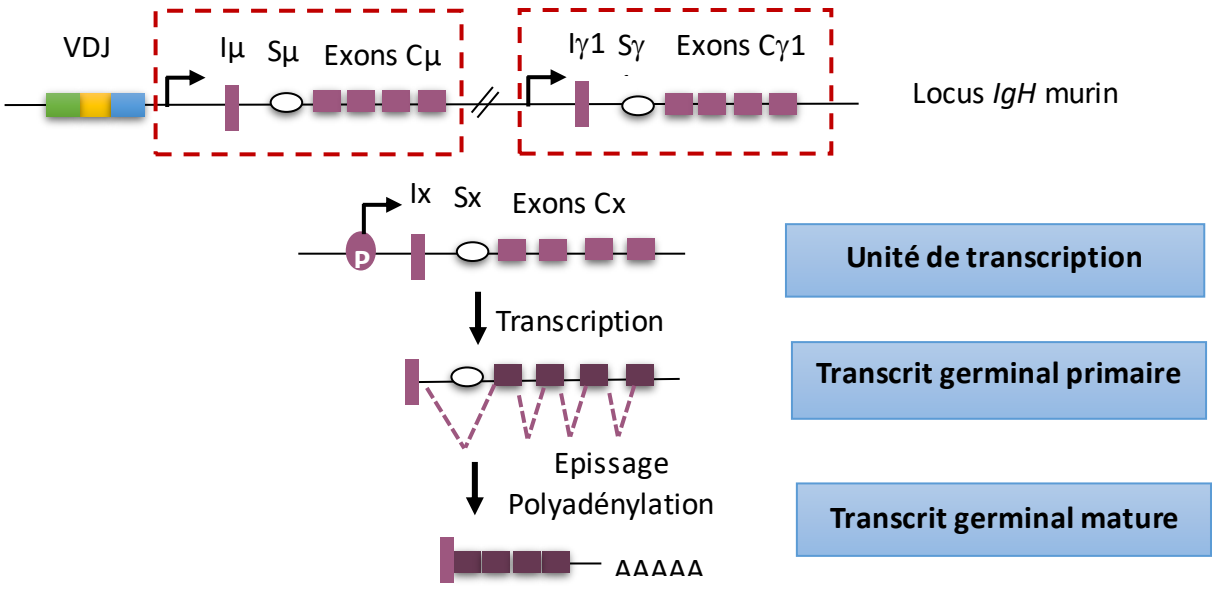


Figure 8 : Schéma de la transcription germinale lors de la CSR. Chaque gène constant possède sa propre unité transcriptionnelle formée d'un promoteur inductible, d'un exon I, d'une région S intronique et d'exons C_H. Il est a noter que le gène Cδ est précédé d'une région pseudo-switch cryptique σδ. Le transcrit primaire subit un épissage alternatif et devient mature non codant.

amont de Cδ (Preud'homme et al., 2000; Chen et al., 2009; Rouaud et al., 2014). L'expression de l'IgD est permise grâce à l'épissage alternatif du transcrit V_HDJ_H/Cμ-Cδ en un transcrit mature V_HDJ_H/Cδ.

1. Mécanisme de recombinaison

Le processus de CSR consiste dans la majorité des cas en une recombinaison intrachromosomique du locus *IgH,* entre une région donneuse Switch μ (Sμ) et une région acceptrice Sx en amont d'un gène constant Cx codant la classe de l'Ig qui va être exprimée **(Figure 7)**. La CSR peut également se produire entre les régions S/S de deux allèles (Laffleur et al., 2014).

1.1. Transcription germinale

La CSR requiert une première étape de transcription des régions S permettant à l'enzyme AID d'induire des lésions de l'ADN qui vont générer des CDB.

Chaque gène constant (sauf Cδ) est constitué par une unité transcriptionnelle indépendante incluant un promoteur et un exon I, une région S intronique, et les exons C_H **(Figure 8)** (Lennon and Perry, 1985). Le gène Cμ est transcrit de façon constitutive dans les LB, même en dehors de toute stimulation, tandis que la transcription des autres gènes constants nécessite une stimulation préalable (Li et al., 1994).

Le choix de la nouvelle classe d'Ig dépend de la combinaison de signaux délivrés par les récepteurs CD40, BCR, TLR (Toll Like Receptor) et des récepteurs cytokiniques. En effet, les cytokines présentes dans l'environnement de la cellule B orientent la CSR vers un isotype donné (Lorenz et al., 1995). Par exemple, l'IL-4 stimule la transcription germinale au niveau des promoteurs Iγ1 et Iε induisant la production d'IgG1 et d'IgE (R L Coffman et al., 1986), l'interféron gamma (IFN γ) oriente la CSR vers IgG2a (Snapper CM et al., 1988), et la TGF-β favorise l'expression de l'IgA (R L Coffman et al., 1986). Au cours de la CSR, les promoteurs travaillent de concert avec les activateurs de la région régulatrice 3'RR (Cogné et al., 1994; Pinaud et al., 2001). Une étude menée chez des souris déficientes pour la région 3'RR a montré une réduction de la CSR vers l'ensemble des isotypes d'Ig, excepté pour IgG1 (Vincent-Fabert et al., 2010). Ces études montrent que les activateurs transcriptionnels de la 3'RR jouent un rôle essentiel pendant la transcription germinale

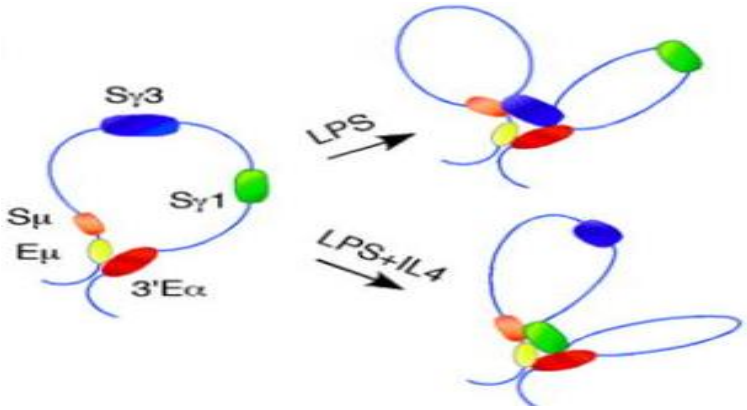


Figure 9 : Représentation schématique de la contraction du locus *IgH* lors de la CSR. Le locus *IgH* des cellules B au repos présente une configuration en boucle avec le rapprochement physique entre les activateurs Eμ et la région 3'RR. Après activation des LB par des cytokines, les segments ciblés par la CSR, (Sγ3 après stimulation par du LPS, ou Sγ1 après stimulation par du LPS+ IL-4), se rapprochent au niveau du complexe Eμ/3'RR déjà formé (Wuerffel et al., 2007).

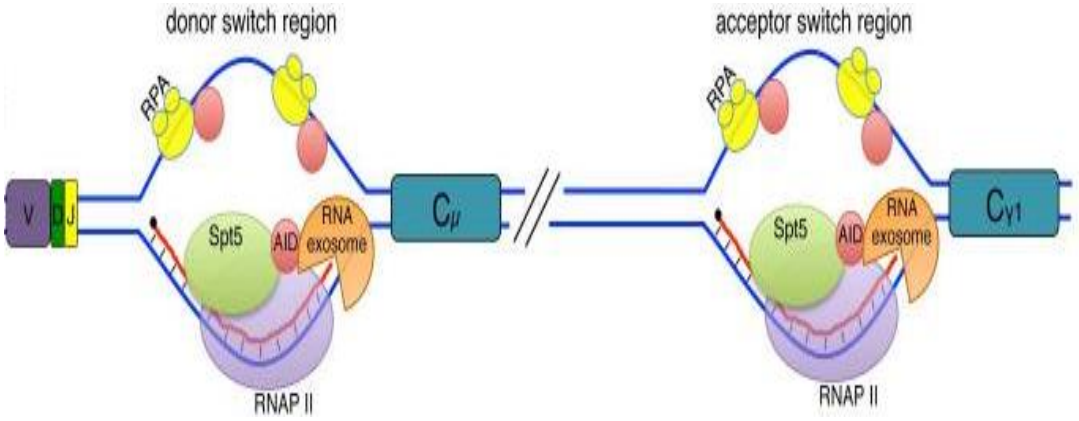


Figure 10 : Recrutement de AID lors de la CSR. La transcription des régions S de la CSR induit la formation de « R-loops » dans lesquelles l'ARN s'hybride avec le brin d'ADN matrice et le brin non matrice est exposé à l'activité de AID (Pefanis and Basu, 2015)

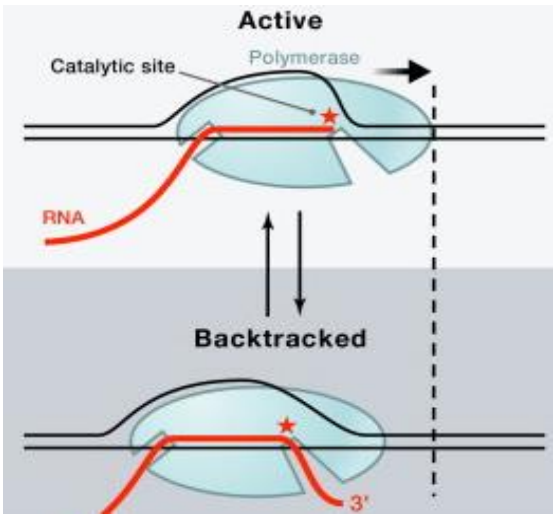


Figure 11 : Marche arrière de l'ARN Pol II. L'ARN Pol II synthétise l'ARN dans le sens 5' vers 3', et peut également faire « marche arrière » en se dirigeant du côté 3' vers 5' de l'Adn matrice.
Adapté de (Nudler, 2012)

des régions S pendant la CSR.

Après stimulation des LB, les promoteurs des transcrits germinaux sont recrutés au niveau du complexe formé par le rapprochement de Eμ et de la 3'RR, formant ainsi une boucle appelée synaptosome **(Figure 9)** (Wuerffel et al., 2007).

1.2. Formation de « R-loops »

La transcription permet de rendre l'ADN des régions S accessible à AID. Les transcrits sont dits « stériles », car ils ne codent pour aucune protéine (Tracy et al., 2000). Durant cette étape, il y a formation d'un hybride ADN/ARN ou « R-loop » (Daniels and Lieber, 1995). La « R-loop » joue un rôle important dans la régulation de la CSR (Zhang et al., 2014). L'enrichissement du brin matrice en G (guanine) favorise la stabilisation des « R-loops » (Zheng et al., 2014). Dans une « R-loop », l'ADN matrice s'hybride avec le brin d'ARN naissant, décalant le brin d'ADN non matrice, qui se retrouve sous forme simple brin (ADNsb) sur lequel s'associe la protéine RPA (Replication Associated Protein A), qui va à son tour recruter AID **(Figure 10)** (Daniels and Lieber, 1995).

In vivo, AID cible les deux brins d'ADN (matrice et non matrice) (Xue et al., 2006). L'action de AID sur le brin matrice est permise grâce à RPA, l'ARN Pol II (ARN polymérase II) et l'ARN exosome. Dans un premier temps, RPA s'associe avec l'ADNsb et recrute AID qui fait appel à l'ARN exosome. Ce dernier composé de neuf protéines « cœur » ayant la capacité de lier l'ARN, ainsi que d'autres sous-unités assurant l'activité ribonucléasique s'accumule au niveau des régions S (Basu et al., 2011). L'ARN exosome a pour fonction la dégradation de l'ARN hybridé ce qui permet d'exposer l'ADNsb à l'action de AID. Pour ce faire, le complexe nécessite un ARN avec une extrémité 3' libre. Cette extrémité est exposée pendant la marche arrière ou « back tracking » de l'ARN Pol II. Ce mouvement se produit souvent lorsque l'enzyme est en « pause » **(Figure 11)** (Adelman et al., 2005)

Pavri a montré que SPT5, facteur associé à l'ARN pol II, interagit avec AID et facilite son recrutement vers l'ARN Pol II amenant de cette manière AID au niveau des régions S (Pavri et al., 2010).

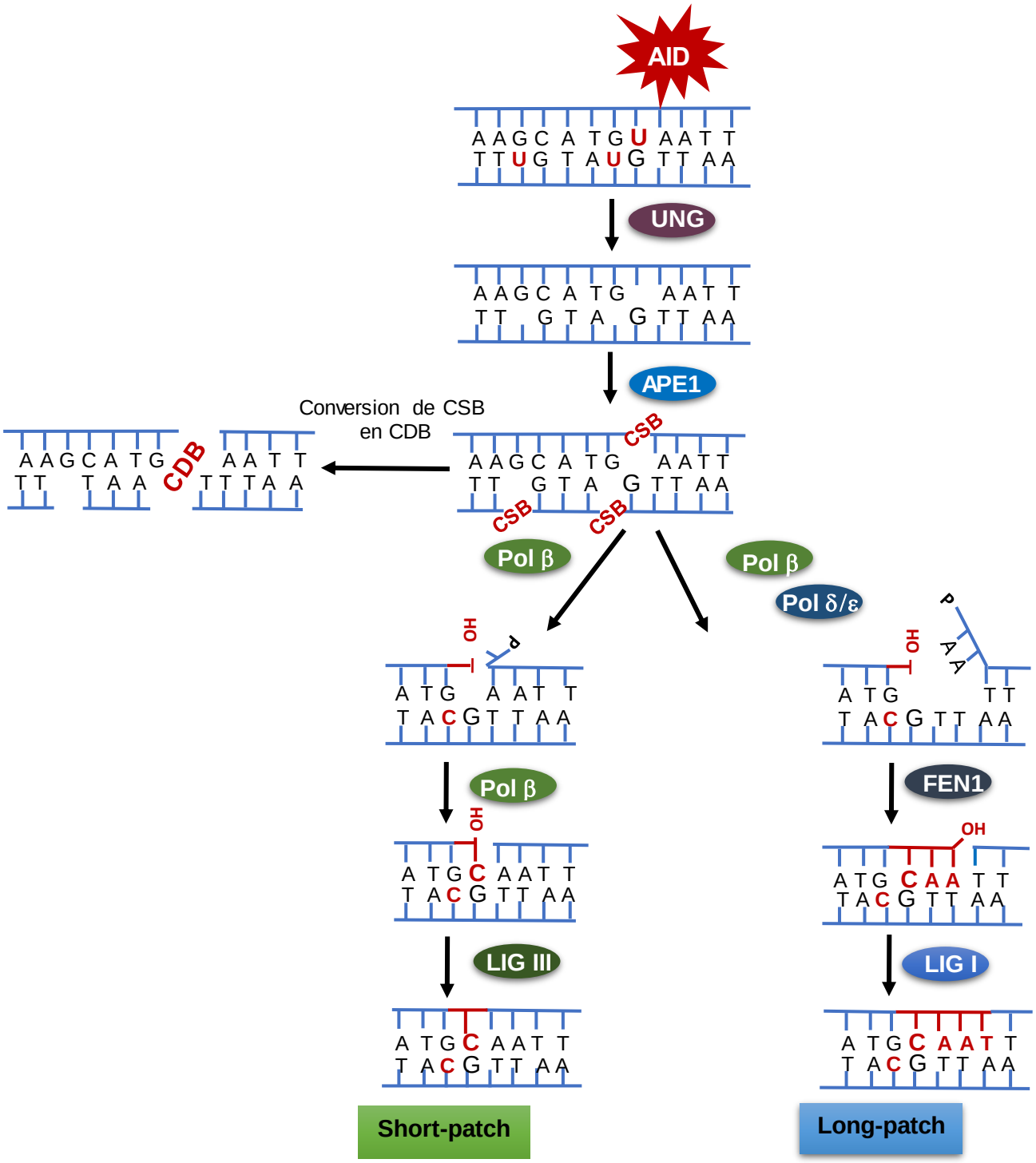


Figure 12 : Schéma de la réparation par excision de bases (BER). AID désamine les cytidines en uridines ce qui crée un mésappariement U/G. Ce dernier est reconnu par l'Uracile N-Glycosylase (UNG) qui clive la liaison N-glycosidique entre le désoxyribose et l'uracile, ce qui génère un site abasique reconnu par l'endonucléase APE1 qui induit une incision aboutissant à une CSB. Par la suite, deux voies peuvent intervenir : la « short-patch » et la « long-patch » lorsque plusieurs mésappariements se retrouvent à proximité. D'après (Robertson et al., 2009)

1.3. Induction de cassures par AID

AID a pour fonction la désamination des cytidines (C) en uridines (U) sur les deux brins d'ADN lors de la transcription des régions S. La majorité des molécules AID sont localisées dans le compartiment cytoplasmique et doivent être importées vers le noyau pour atteindre leurs cibles (Patenaude et al., 2009). Une fois dans le noyau, AID est recruté préférentiellement au niveau de régions spécifiques de l'ADN appelées « hotspots » , ayant pour séquence le motif WRC (W = A/T, R = A/G), et engendre un mésappariement U/G (Betz et al., 1993; Muramatsu et al., 2000).

Deux principaux mécanismes vont être déployés afin de réparer ce mésappariement : la réparation par excision de base (Base Excision Repair ou BER) et la réparation des mésappariements de bases (Mismatch Repair ou MMR).

1- L'uracile obtenue par l'action de AID peut être éliminée grâce à la voie BER **(Figure 12)** avec dans un premier temps reconnaissance de la base endommagée par l'Uracile N-Glycosylase (UNG), qui va créer un site abasique, en clivant la liaison glycosidique entre l'uracile et le désoxyribose (Rada et al., 2002). Ensuite, il y a recrutement de l'endonucléase apurinique/apyrimidique (APE1) qui effectue une incision au niveau de ce site et crée des extrémités 3'OH et 5' désoxyribose phosphate (dRP) (Doetsch and Cunningham, 1990). La résolution des cassures par la voie BER peut s'opérer selon deux voies qui sont « short-patch » et « long-patch » **(Figure 12)** (Muramatsu et al., 2000; Krokan and Bjørås, 2013). La voie short-patch fait appel à plusieurs acteurs protéiques dont les composants cœurs sont UNG, APE1, Pol β et la LIG III (Krokan and Bjørås, 2013). Le clivage par APE1 génère un résidu 5' abasique qui sera enlevé et remplacé par Pol β. La voie long-patch est initiée de façon similaire à la voie précédente, mais elle concerne le remplacement de deux nucléotides et plus (Frosina et al., 1996). Elle fait intervenir Pol β, Pol δ/ε, la Flap endonucléase 1 (FEN1) et la LIG I. L'incision par APE1 libère une extrémité 3' OH qui sert de substrat directe pour Pol β. Cette dernière remplace le nucléotide enlevé ainsi que plusieurs bases en aval, déplaçant le brin du côté 3' du site de clivage (Robertson et al., 2009; Kim and Wilson, 2012; Krokan and Bjørås, 2013). Le brin déplacé est ensuite enlevé par FEN1 suite à sa stimulation par PARP1 (Prasad et al., 2001) puis intervient l'étape de ligation qui est réalisée par la LIG I.

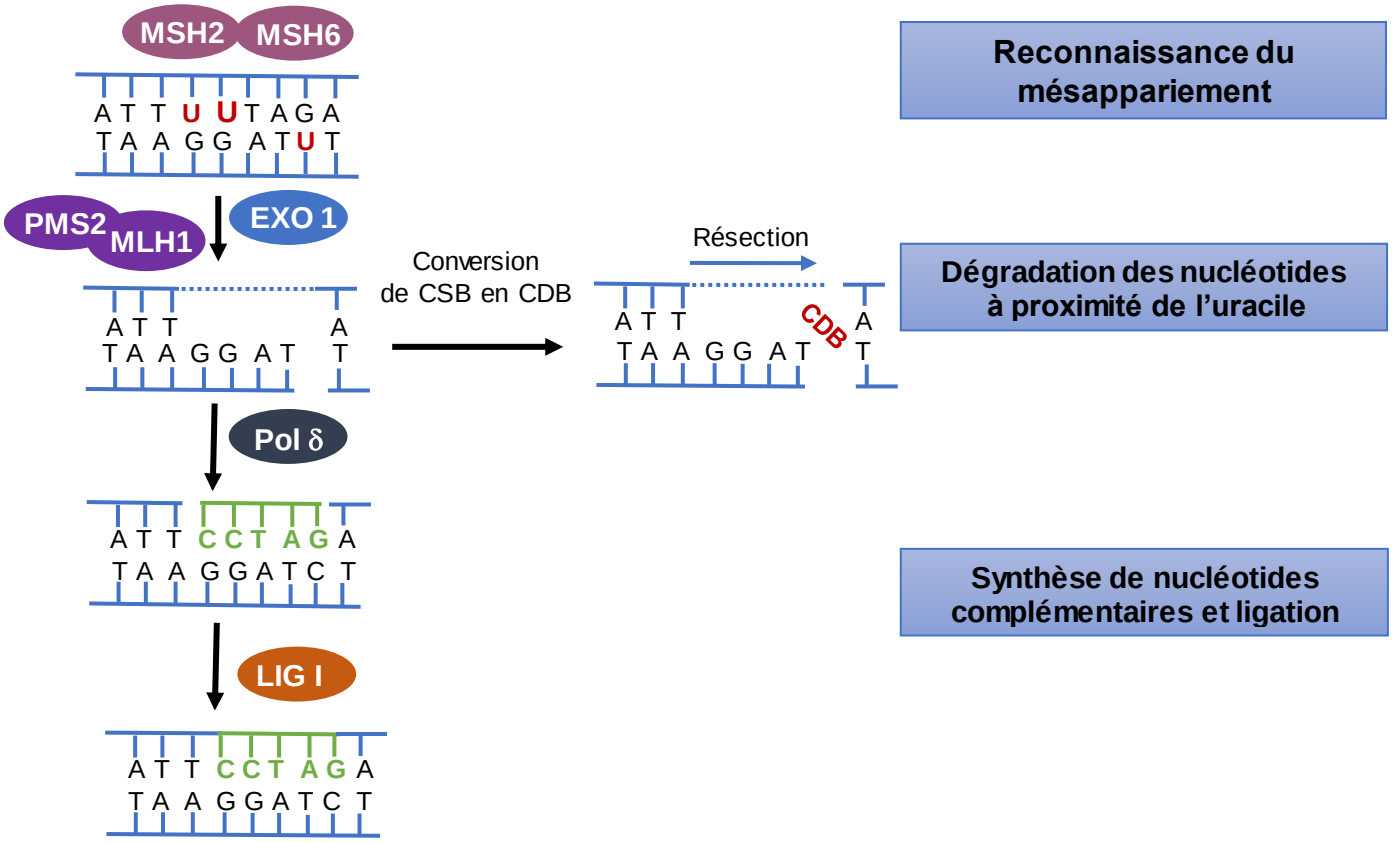


Figure 13 : Schéma de la réparation des mésappariements de bases (MMR). Le mésappariement U/G est reconnu par l'hétérodimère MSH2/MSH6 qui recrute PMS2/MLH1 et l'exonucléase (EXO1) qui dégrade les nucléotides autours de l'uracile dans le sens 5' vers 3' jusqu'à atteindre la CSB du brin d'ADN opposé, enfin l'ADN Pol δ reconstruit le brin excisé et la ligase referme le brin.

2- Le complexe MSH2/MSH6 peut également reconnaitre les mésappariements d'une base ou MSH2/MSH3 lorsque le nombre de bases mésappariées est plus élevé (Rada et al., 2004) **(Figure 13)**. Après cette reconnaissance, il y a recrutement d'autres facteurs comme l'hétérodimère endonucléasique PMS2/MLH1 et l'exonucléase 1 (EXO1) qui dégradent les nucléotides autour de l'uracile, et enfin l'ADN Pol δ reconstruit le brin excisé et la ligase referme le brin. C'est la voie de réparation des mésappariements de bases (MMR). Chez la souris, des mutations dans les gènes codant des protéines impliquées dans la voie MMR (MSH2,MSH6,MLH1,PMS2 ou EXO1) altèrent de façon drastique le processus de CSR (Ehrenstein and Neuberger, 1999; Schrader et al., 2003; Chaudhuri J and Alt FW, 2004; Bardwell et al., 2004; Péron et al., 2008).

La recombinaison CSR ferait intervenir de façon plus importante la voie du BER. Le déficit en *Ung* réduit de manière significative la CSR jusqu'à moins de 10% (Di Noia and Neuberger, 2002). Toutefois, Stavnezer et son équipe ont montré que lorsque les CSB générées par le BER se retrouvent à proximité les unes des autres, elles peuvent générer spontanément des CDB **(Figure 12)**. En revanche, quand les CSB sont éloignées, l'hétérodimère MSH2/MSH6 peut être recruté au niveau des mésappariements U/G qui n'ont pas été pris en charge par UNG (Stavnezer et al., 2008) **(Figure 13)**. Les voies BER et MMR ne semblent donc pas en concurrence et peuvent survenir simultanément et de façon synergique, afin de favoriser l'excision de mésappariements et/ou de sites abasiques au niveau de lésions créées par AID (Girelli Zubani et al., 2017).

3. La recombinaison suicide du locus *IgH* (LSR)

3.1. Découverte de la recombinaison

La recombinaison LSR a été découverte dernièrement par notre laboratoire (Péron et al., 2012). Au sein de la région 3'RR, des régions Like Switch (LS) ont été identifiées comme nouvelles cibles de l'enzyme AID. Comme pour les régions S ciblées par AID lors de la CSR, les régions LS sont caractérisées par des séquences d'ADN répétées (Chauveau and Cogné, 1996; Pinaud et al., 1997).

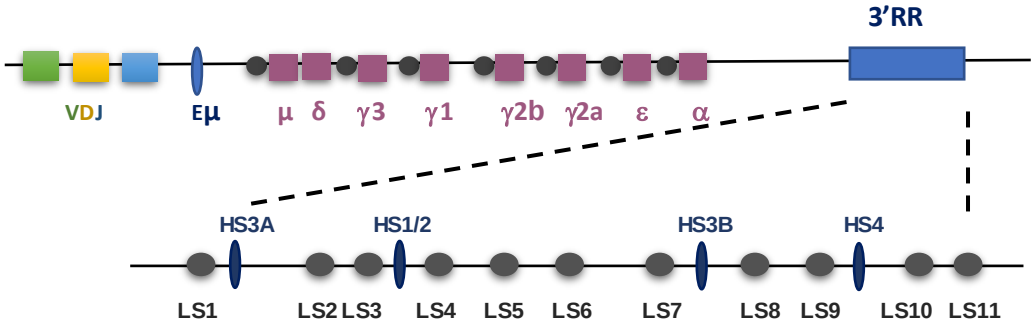


Figure 14 : Régions LS du locus *IgH* murin. Les régions LS (LS1 jusqu'à LS11) présentent des motifs d'ADN répétés de façon similaire aux régions S et leurs tailles varient de 0,5 à 1 kb. Les régions LS sont localisées de part et d'autre des activateurs transcriptionnels (HS3A, HS1/2, HS3B et HS4) au sein de la région régulatrice 3'RR.

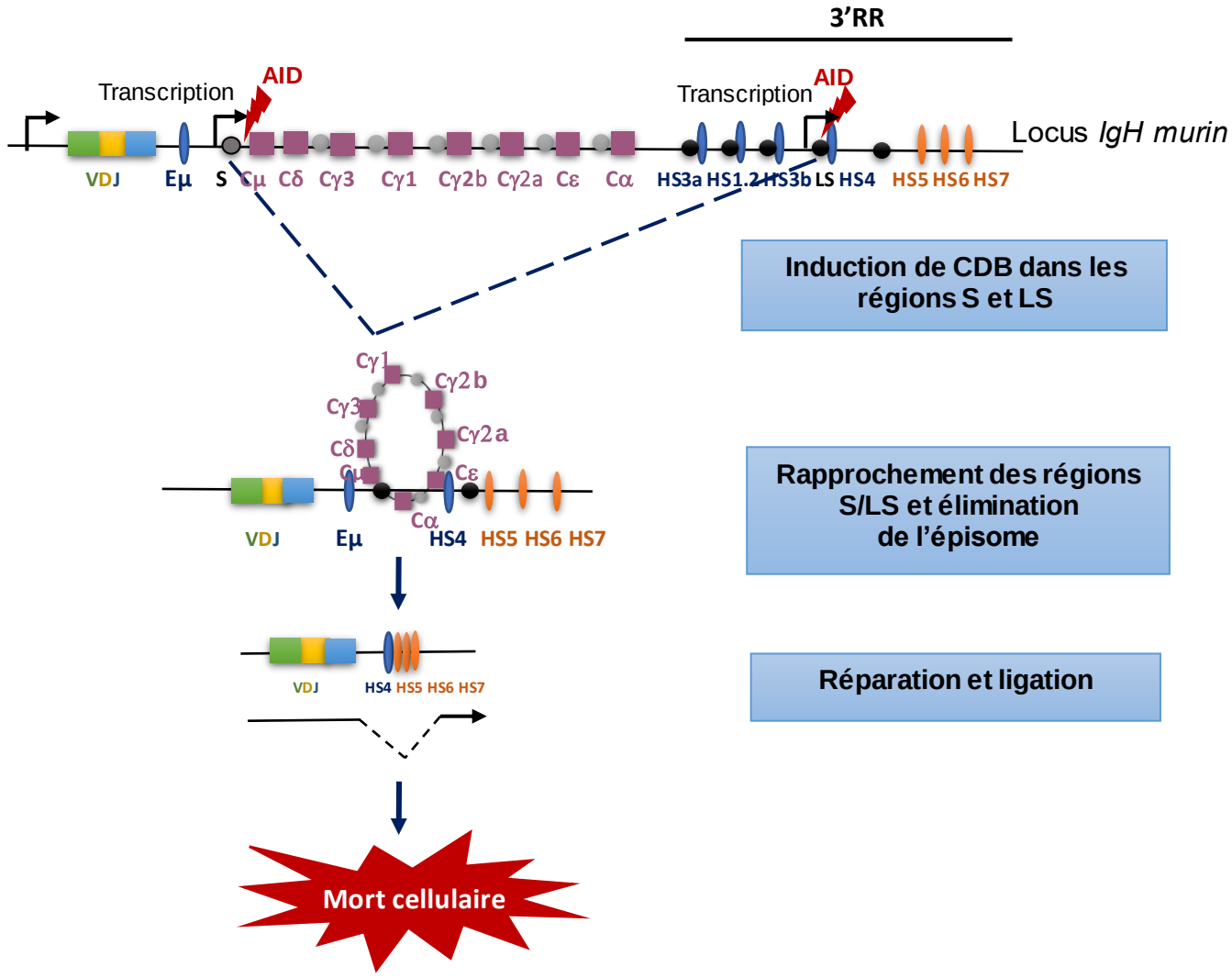


Figure 15 : Schéma de recombinaison suicide du locus *IgH* (LSR). Les segments donneur et accepteur de la recombinaison LSR (Sμ et LS) sont transcrits puis ciblés par AID qui induit des CDB. Les deux régions sont rapprochées et la totalité des gènes constants est éliminée sous forme d'épisome empêchant l'expression d'Ig ce qui induit la mort de la cellule B concernée.

L'étude a identifié 11 régions LS chez la souris **(Figure 14)**, avec des séquences similaires à celles des segments Sγ. Les séquences LS sont conservées entre espèces chez lesquelles la séquence complète du locus *IgH* est disponible (souris, Homme, lapin et chien). Chez l'Homme, elles sont particulièrement définies avec une homologie de séquence supérieure à 80%. Cette conservation inter-espèce soulève l'hypothèse d'un rôle fonctionnel de ces régions.

Les auteurs ont également mis en évidence une transcription germinale de la 3'RR après stimulation des cellules B, ainsi que des mutations, avec un taux qui se rapproche de celui de la région Sμ (0.42%) (Péron et al., 2012). Ces résultats, en plus de la détection de CDB au voisinage des séquences LS suggèrent un mécanisme moléculaire similaire entre les deux recombinaisons dépendantes de AID.

3.2. Mécanisme moléculaire

La recombinaison LSR est un réarrangement du locus *IgH* ciblant la région donneuse Sμ et une région acceptrice LS située dans la 3'RR **(Figure 15)**. Cette recombinaison engendre la délétion du fragment d'ADN intermédiaire sous forme d'épisome. Ce dernier contient la totalité des gènes codant les parties constantes des différentes classes d'Ig. Lorsque la LSR a lieu sur l'allèle fonctionnel, les chaînes lourdes des Ig ne sont plus exprimées empêchant la constitution d'un BCR. En absence de ce récepteur, et du signal de survie qu'il délivre, les cellules sont induites à l'apoptose.

3.3. Potentiel rôle physiologique

La LSR pourrait être le côté sombre de la CSR, avec un rôle important dans le maintien de l'homéostasie cellulaire B. D'après notre hypothèse, la LSR ciblerait des populations de LB présentant un potentiel danger pour l'organisme, comme les cellules auto-immunes ou les cellules tumorales. Ce processus pourrait provoquer la mort de cellules B ayant subi une stimulation suboptimale ou sous-optimale, alors qu'une stimulation adéquate activerait la transcription de régions S pour réaliser la CSR et promouvoir la survie de LB exprimant une nouvelle classe d'Ig.

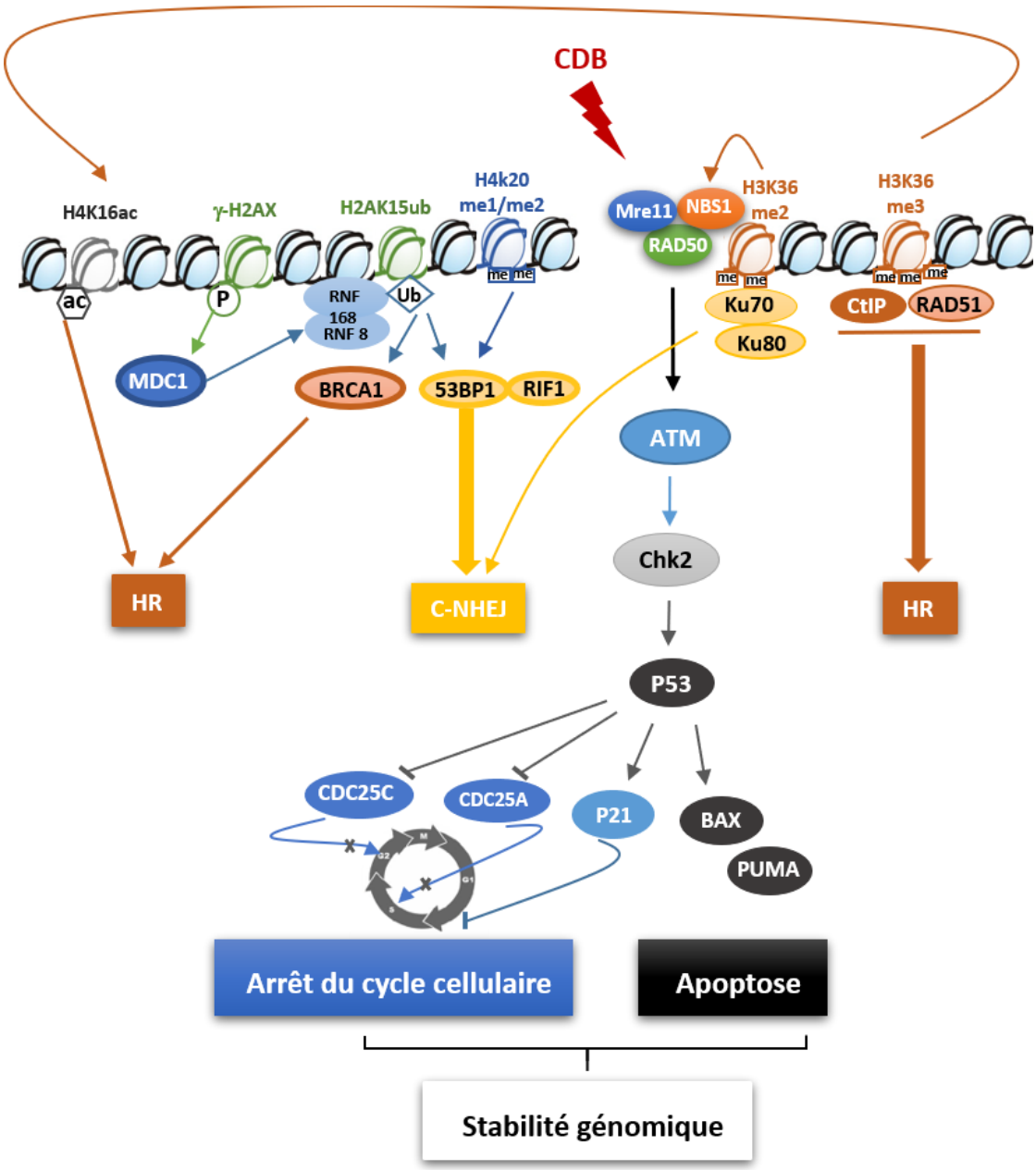


Figure 16 : Schéma de la voie DDR dans le cas de CDB. Cette voie fait intervenir des molécules senseurs des lésions de l'ADN comme le complexe MRN qui détecte et active des protéines kinases dont ATM. ATM phosphoryle H2AX (γ-H2AX) ainsi que d'autres acteurs comme CtIP qui favorise la résection des extrémités de l'ADN, 53BP1 qui l'inhibe, ou encore P53 qui induit l'arrêt du cycle cellulaire et l'apoptose. Les différentes marques d'histones recrutent les protéines qui interviennent dans la réparation des CDB appartenant aux voies distinctes du C-NHEJ ou de la HR.

III. La réparation des CDB au locus *IgH*

Les CDB générées en partie au cours de la CSR menacent l'intégrité du génome. En réponse à cette menace, la cellule active la voie DDR qui constitue un réseau de facteurs assurant la détection, la signalisation et la réparation de ces cassures. Dans le cas où la réparation des CDB ne s'effectue pas, la cellule s'engage en mort programmée.

1. Détection et prise en charge des cassures

Dans un premier temps, le complexe MRN composé de MRE11 (Meiotic Recombination 11 homolog), RAD50 et NBS1 (Nijmegen Breakage Syndrome protein 1) détecte la CDB et prévient la dissociation des extrémités d'ADN lésées **(Figure 16)**. Ce complexe est requis pour l'activation de différentes kinases de la famille PIKKs (Phosphatidylinositol 3-kinase-like serine/theronine kinases) dont ATM et la DNA-PKc s (Lee and Paull, 2005).

ATM existe sous forme dimérique inactive qui se dissocie et s'autophosphoryle une fois recrutée au site de cassure (Bakkenist and Kastan, 2003). Cette kinase déclenche une cascade d'activation à travers la phosphorylation de plusieurs protéines dont H2AX sur son résidu Ser139 pour former « γ-H2AX » des deux côtés de la CDB (Rogakou et al., 1998), MDC1 (Mediator of DNA damage checkpoint 1) qui est recruté au niveau de γH2AX participe à l'amplification du signal sur des milliers de Kb autour de la cassure (Rogakou et al., 1998; Clouaire and Legube, 2015) active les facteurs RNF8/RNF168 qui ubiquitinylent la lysine 13 et /ou lysine 15 de l'histone H2A/H2AX et former l'H2AK13ub et l'H2AK15ub respectivement (Mattiroli et al., 2012). Ces dernières peuvent recruter des molécules antagonistes qui sont le facteur de susceptibilité aux cancers du sein BRCA1 qui favorise la résection et la réparation par la HR, et 53BP1 qui l'inhibe et oriente la réparation vers le système du C-NHEJ (Rogakou et al., 1998; Goldberg et al., 2003; Stewart et al., 2003).

D'autres marques épigénétiques jouent un rôle important dans le recrutement de facteurs intervenant dans les différents systèmes de réparation. L'H3K36me3 et l'H4K20me2/1 sont des marques épigénétiques qui sont démasquées après l'apparition de CDB. La première recrute RAD51 et active la réparation par la HR, alors

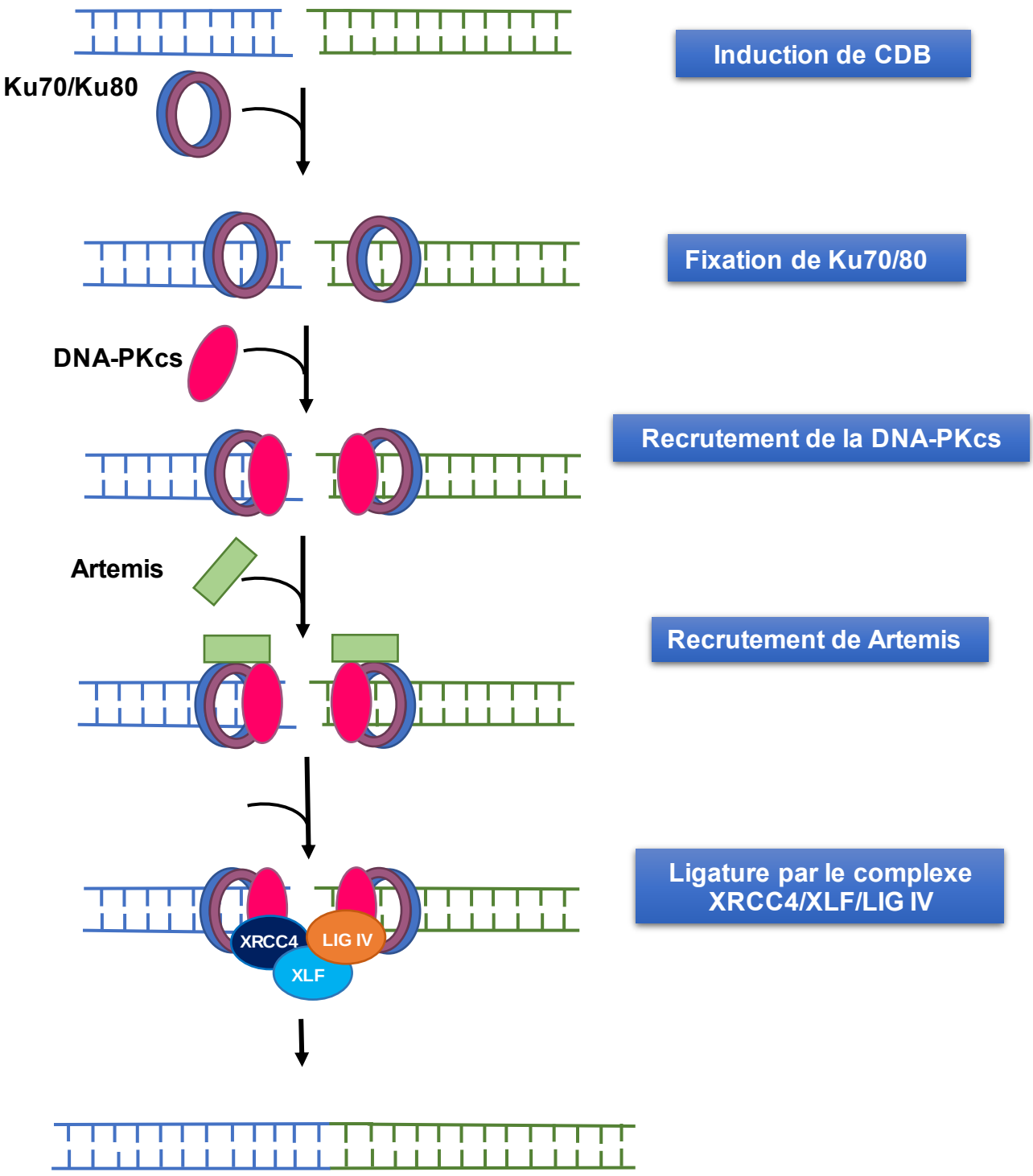


Figure 17 : Schéma de la réparation des CDB par le C-NHEJ. Le système C-NHEJ est une des voies principales de réparation des CDB de l'ADN. Les extrémités des CDB sont détectées par l'hétérodimère Ku70/Ku80 qui va recruter et activer la DNA-PKcs pour former le complexe DNA-PK capable de recruter Artemis qui clive les structures en épingle à cheveux. Cette étape est nécessaire pour que le complexe XRCC4/XLF/LIG IV intervienne dans l'étape de ligation.

que l'H4K20me1/2 recrutent 53BP1 et les acteurs du C-NHEJ. Il est intéressant de noter que l'H3K36me3 augmente la forme acétylée de l'H4K16 dans des cellules humaines (Li and Wang, 2017).

ATM active également les points de contrôle du cycle cellulaire comprenant les points G1/S, S/G2 et G2/M, ce qui met en pause le cycle afin de permettre à la cellule d'effectuer les réparations nécessaires au maintien de son intégrité génomique. À ce stade, soit la cellule parvient à réparer les lésions, soit elle ne parvient pas à le faire et déclenche le processus d'apoptose par l'intermédiaire de P53 qui active la transcription de facteurs pro-apoptotiques comme BAX et PUMA (Roos and Kaina, 2006).

2. Réparation des cassures

Différents systèmes peuvent intervenir afin de réparer les CDB dont deux principaux. D'un côté, il y a la HR, processus qui nécessite des séquences d'ADN homologues, et connu pour sa fidélité. De l'autre, le NHEJ, qui ne requiert pas d'homologie de séquence et induit de fréquentes délétions ou insertions de nucléotides « indels » au site de jonction. Le NHEJ à son tour est divisé en deux voies : Le C-NHEJ qui ne nécessite pas de résection des extrémités d'ADN et produit une jonction directe dépourvue de microhomologie de séquence ou avec peu de microhomologie, et l'A-EJ (Alternative end joining) qui génère des microhomologies plus longues et nécessite une résection des extrémités lésées.

2.1. Le NHEJ classique (C-NHEJ)

2.1.1. Mécanisme

Dans un premier temps, les protéines Ku70 et Ku80 s'assemblent entre elles et forment un anneau autour des extrémités d'ADNdb clivées **(Figure 17)**. Il a été suggéré qu'un hétérodimère est recruté à chaque extrémité de la cassure et qu'il ne se dissociera de l'ADN qu'une fois les extrémités jointes (Blier et al., 1993). Ku70/Ku80 recrute et active la sous unité catalytique de la protéine kinase dépendante de l'ADN (ADN-PKcs) via la partie C-terminale de Ku80 (Singleton et al., 1999). La kinase s'active en s'autophosphorylant et active Artemis, enzyme indispensable dans le

clivage des structures en épingle à cheveux générées durant les réarrangements V(D)J (Ma et al., 2002). En effet, les cellules déficientes pour *Artemis* présentent un défaut au niveau des jonctions codantes sans affecter les jonctions signales (Moshous et al., 2001; Rooney et al., 2002). Finalement, l'ADN-PKcs recrute le complexe XRCC4/XLF/LIG IV qui assure l'étape de ligation (Ahnesorg et al., 2006; Kanaar et al., 1998).

2.1.2. Altérations dans les composants du C-NHEJ

De façon générale, les cellules déficientes en *Ku70*, *Ku80*, *Xrcc4, Xlf* et *Lig IV*, affichent des recombinaisons V(D)J altérées ainsi qu'une sensibilité pour les agents induisant les CDB (Nussenzweig et al., 1996; Frank et al., 1998; Li et al., 1995; IJspeert et al., 2016). L'absence de la DNA-PKcs ne semble pas engendrer d'effets aussi sévères que ceux de la délétion d'un composant cœur du C-NHEJ comme Ku70 ou Ku80, ce qui laisse penser que ces derniers exercent leurs rôles de façon indépendante de la DNA-PKcs (Barnes et al., 1998; Gao et al., 1998; Yan et al., 2007). Il a été montré que le déficit en *Xrcc4* ou *Lig IV* dans les cellules B activées induit une réduction de la CSR, ce qui souligne l'importance du C-NHEJ dans ce processus (Soulas-Sprauel et al., 2007; Yan et al., 2007; Han and Yu, 2008). D'un autre côté, la CSR résiduelle qui s'élève à environ 50% du taux de CSR normal indique clairement l'implication d'une voie alternative au C-NHEJ. Par ailleurs, il a été montré que la délétion combinée de *Ku70* ainsi que la *LIG IV* favorise les translocations *IgH*/*c-Myc* (Boboila et al., 2010a). L'augmentation de la fréquence de ces translocations a été observée également chez des souris déficientes pour *Ku80* (Difilippantonio et al., 2000). Ces études indiquent l'implication d'une voie de réparation alternative dans la génération de translocations. En revanche, et de façon inattendue chez l'Homme, c'est le C-NHEJ qui semble favoriser la génération de translocations, car l'altération de la LIG IV entraine une diminution de la fréquence d'apparition des translocations (Ghezraoui et al., 2014).

2.2. Le NHEJ alternatif (A-EJ)

2.2.1. Mécanisme

L'A-EJ connu également sous le nom de B-NHEJ (Back-up NHEJ) est un système de réparation qui fonctionne quand le C-NHEJ est altéré, ou en concomitance mais de façon mineure (Boboila et al., 2010a). La voie A-EJ se distingue du C-NHEJ par le fait qu'elle est indépendante de Ku (McVey and Lee, 2008).

Il existerait plusieurs systèmes impliqués dans l'A-EJ dont le MMEJ (Microhomology Mediated End Joining) et le TMEJ (Tetha Mediated End Joining). La voie alternative à l'opposé du C-NHEJ produit de façon fréquente des séquences avec des microhomologies au point de réparation. Comme cité précédemment, la réparation par l'A-EJ est plus susceptible aux délétions et aux translocations chromosomiques que le C-NHEJ (Boboila et al., 2010a; Yan et al., 2007). Parmi les acteurs identifiés dans le MMEJ, il y a PARP-1 (Poly ADP Ribose Polymerase 1), CtIP, le complexe MRN, XRCC1 et la Ligase III (LIG III). PARP-1 reconnait les CDB et recrute le complexe MRN, qui est pour rappel, un acteur clé de la voie DDR (Haince et al., 2008). L'étape de résection des extrémités d'ADN est assurée par les protéines CtIP et le complexe MRN permettant par la suite l'étape de ligation par le complexe XRCC1/LIG III (El-Khamisy et al., 2003).

L'ADN polymérase theta (Pol θ ou PolQ) intervient de façon spécifique dans l'A-EJ à travers la voie TMEJ, et semble agir de façon Ku-indépendante. En effet, le KO de *Polθ* n'a pas d'effets sur la réparation par le C-NHEJ (Mateos-Gomez et al., 2015), et ne semble pas affecter la fréquence de CSR dans la lignée CH12F3, ainsi que chez la souris (Martomo et al., 2008; Li et al., 2011). Cependant, la réparation par Pol θ semble occuper une place importante dans certains cancers où la voie HR est altérée, comme dans le cancer du sein. Dans ce cas, Polθ est fréquemment sur-exprimée, et ce statut est généralement associé à un mauvais pronostic (Ceccaldi et al., 2015; Higgins et al., 2010).

Pol θ possède la capacité de bloquer la voie HR en interagissant avec RAD51 pour limiter l'assemblage du nucléofilament, et peut aussi réduire l'efficacité de la formation de la « D-loop », favorisant de cette façon la réparation en faveur de l'A-EJ (Ceccaldi et al., 2015).

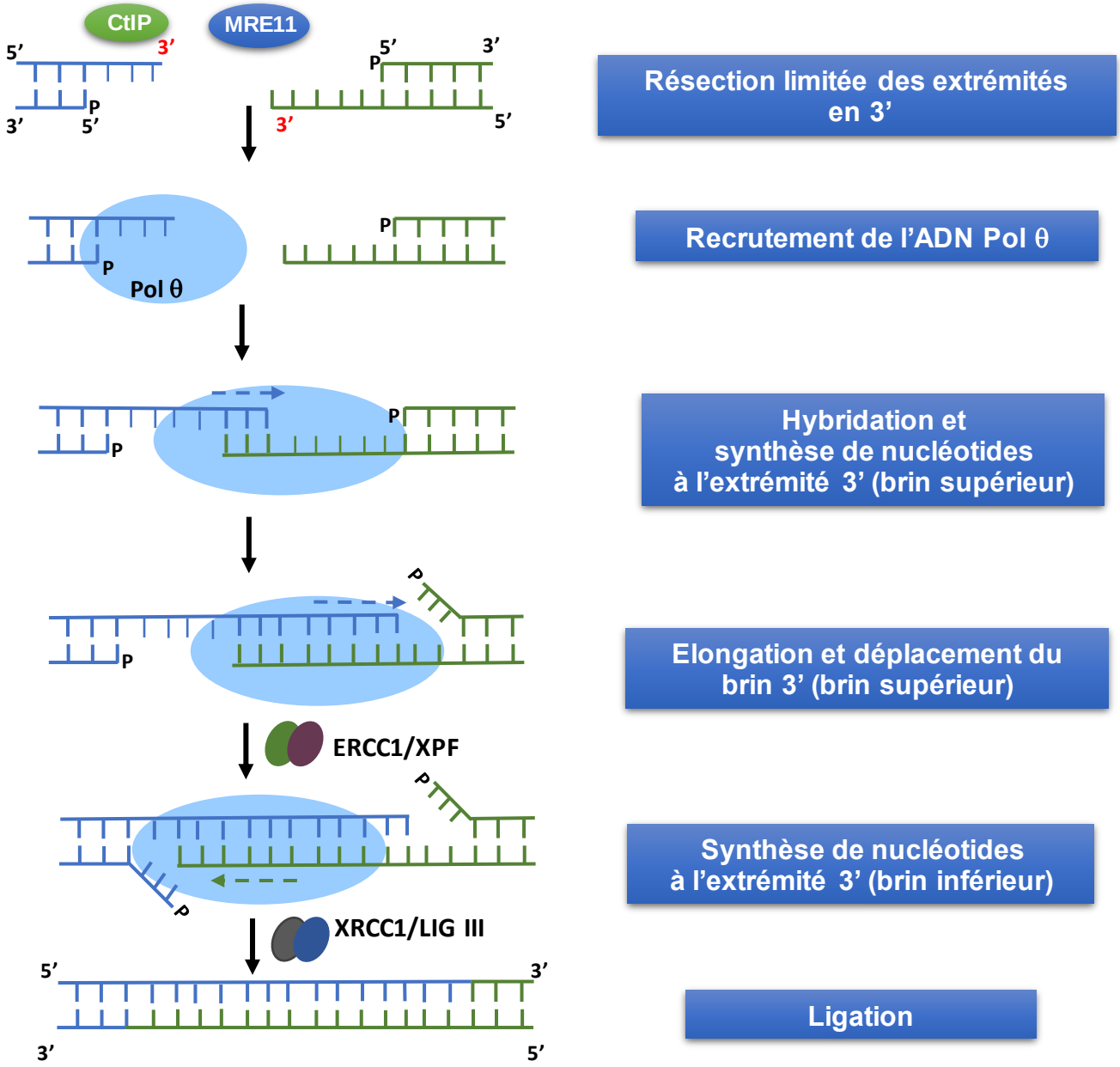


Figure 18 : Schéma de la réparation des CDB par le TMEJ. Au niveau des CDB les extrémités 5' subissent une résection limitée par l'action de CtIP et MRE11, puis Pol θ intervient au niveau d'une séquence de complémentarité entre les deux ADNsb pour initier la synthèse de nucléotides au niveau des extrémités 3' (brin supérieur et brin inférieur). Finalement l'ADN simple brin sortant est dégradé par ERCC1/XPF et la ligation réalisée par le complexe XRCC1/LIG III.

La réparation des CDB médiée par la Pol θ passe par plusieurs étapes **(Figure 18)**. Dans un premier temps, les CDB sont détectées par PARP-1, qui recrute MRE11, acteur clé de la voie DDR (Haince et al., 2008; Robert et al., 2009), qui assure avec CtIP l'étape initiale de résection des extrémités 5' lésées. Cette résection permet l'exposition de courtes régions homologues entre les deux molécules d'ADNsb. La protéine PARP-1 recrute également Pol θ au site de la cassure (Wang et al., 2006). La polymérase se sert des extrémités 3' débordantes générées au cours de la résection comme amorce de synthèse. Après hybridation des séquences homologues, elle assure l'incorporation de nucléotides dans le sens 5' vers 3', puis l'élongation qui induit le déplacement du brin d'ADN (Yousefzadeh et al., 2014; Kent et al., 2015). Les queues 3' qui ne sont pas complémentaires sont alors enlevées par le complexe endonucléasique ERCC1/XPF, puis intervient l'étape finale de ligation par le complexe XRCC1/LIG III et ou la LIG I (El-Khamisy et al., 2003).

La réparation par Polθ laisse une signature spécifique, avec utilisation de microhomologies et d'insertions au point de réparation (Chan et al., 2010; Koole et al., 2014). Comme il a été décrit dans la littérature, l'A-EJ semble occuper une fonction primordiale quand les mécanismes « standards » de réparation sont défectueux. Cependant, elle peut intervenir même quand la HR est opérationnelle (Truong et al., 2013).

2.2.2. Altérations des composants de l'A-EJ

Les ligases I et III possèdent des fonctions redondantes, car en absence de la Lig VI, la Lig I ou la Lig III peut intervenir seule dans la réparation médiée par l'A-EJ au cours de la CSR (Masani et al., 2016). En l' absence d'un autre composant ,XRCC1, les jonctions CSR présentent des séquences de microhomologie plus courtes, suggérant son implication dans l'A-EJ (Saribasak et al., 2011). Cependant, dans les CH12F3 avec double délétion de *Xrcc1* et de la *Lig IV,* le processus de CSR est réalisé de manière comparable aux cellules déficientes uniquement pour la *Lig IV*. Ces résultats indiquent que XRCC1 intervient dans la réparation, mais ne semble pas indispensable (Han et al., 2012). La réduction dans l'usage de microhomologie de séquence a été observée également en absence de *Parp-1,* ce qui montre que cette protéine est également impliquée sans être essentielle dans l'AEJ au cours de la CSR (Robert et al., 2009).

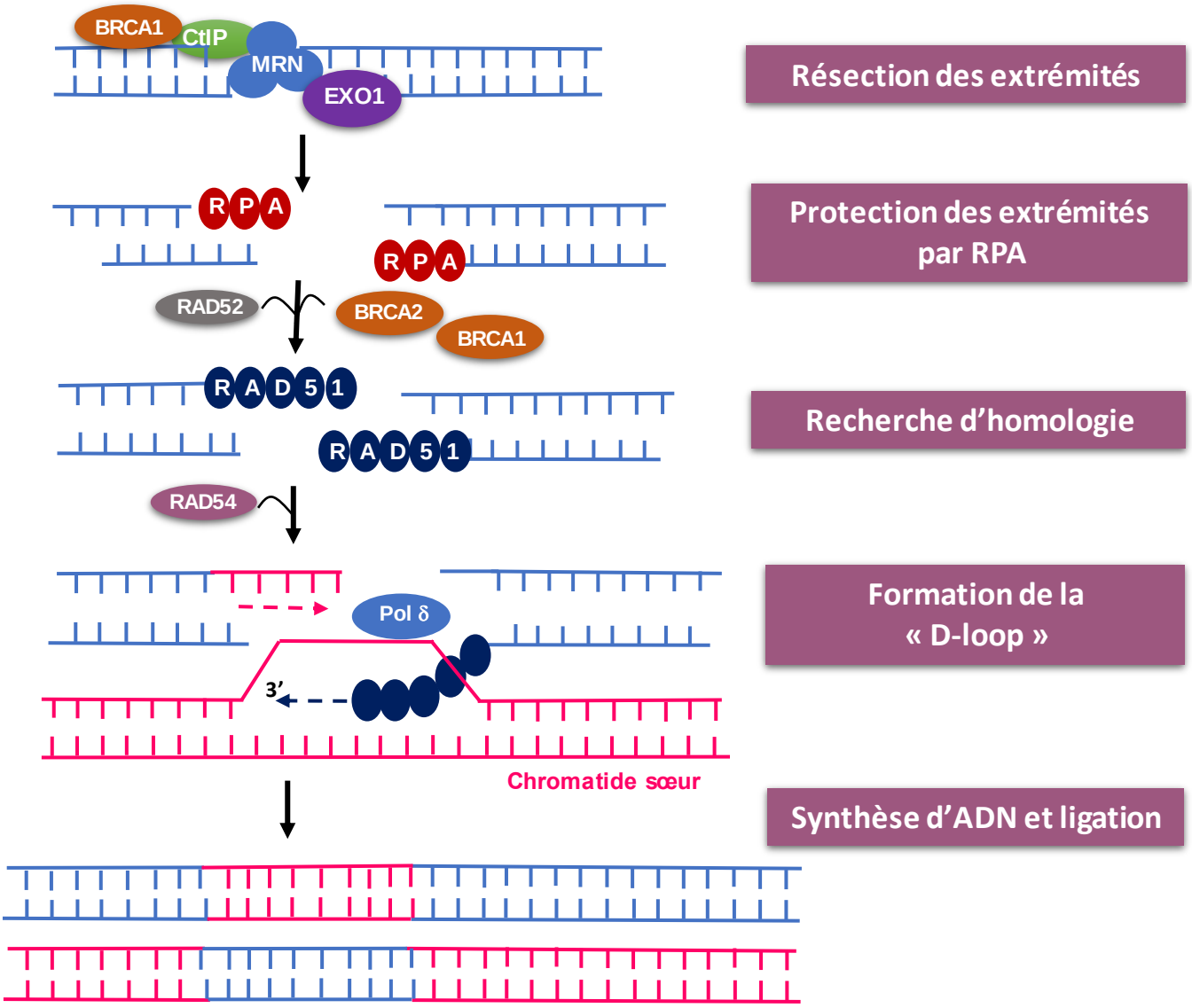


Figure 19 : Schéma de la réparation des CDB par le système de la HR. La CDB est détectée et les extrémités 5' subissent la résection par un ensemble de facteurs (BRCA1, CtIP, MRN et EXO1), les extrémités 3' sortantes simple brin sont protégées par RPA. RAD52, BRCA1/2 participent à la formation du nucléofilament composé de RAD51, puis la recherche d'homologie de séquence est médiée par RAD54 permettant l'invasion de la chromatide sœur et la création de la « D-loop ». Par la suite il y a synthèse des nucléotides manquants et ligation du fragment nouvellement synthétisé.

2.3. Le système de la HR

2.3.1. Mécanisme

Parmi les mécanismes déployés lors de la réparation des CDB, la HR est celui qui permet une réparation fidèle. Cette voie requiert la présence d'un brin d'ADN qui possède une séquence homologue d'environ 100 pb pour diriger la réparation (Dudás and Chovanec, 2004), et s'effectue durant les phases S et G2 du cycle cellulaire, après la réplication de l'ADN.

La HR se déroule en trois grandes étapes **(Figure 19)**. La première appelée phase « pré-synaptique » est initiée par la résection des extrémités d'ADN dans le sens 5' vers 3', engendrant des extrémités 3' simple brin débordantes (Truong et al., 2013). Cette opération est réalisée par BRCA1, CtIP, MRE11 et EXO1 (Lavin, 2004; Sartori et al., 2007; Prakash et al., 2015). Pendant cette étape, l'ADNsb est recouvert et protégé par la RPA qui empêche la formation de structures secondaires. La recherche d'une séquence d'ADN homologue nécessite le recrutement de la recombinase RAD51, puis la cellule fait appel à RAD52 et BRCA2 pour favoriser la formation d'un filament de RAD51 sur l'ADNsb préalablement couvert par RPA.

Durant la phase « synaptique », le filament de RAD51 rejoint l'ADNdb de la chromatide sœur issue de la réplication. Il envahit la séquence homologue trouvée créant ainsi une boucle de déplacement « D-loop » (Displacement loop). Durant ce processus, la protéine RAD54 est requise à la fois pour la stabilisation du filament, la formation de la « D-loop », ainsi que la transition de l'étape d'invasion de l'ADN bicaténaire par le nucléofilament vers l'étape de synthèse de l'ADN et donc la dissociation de RAD51 de l'hétéroduplex d'ADN (Heyer et al., 2006).

Finalement, une fois le brin endommagé associé à son brin complémentaire intact, intervient la phase « post-synaptique ». À ce stade, la synthèse du nouveau brin d'ADN s'effectue de façon principale par la polymérase delta (Pol δ) et RAD54 stimule l'extension de manière ATP dépendante, puis intervient l'étape de ligation (Solinger and Heyer, 2001).

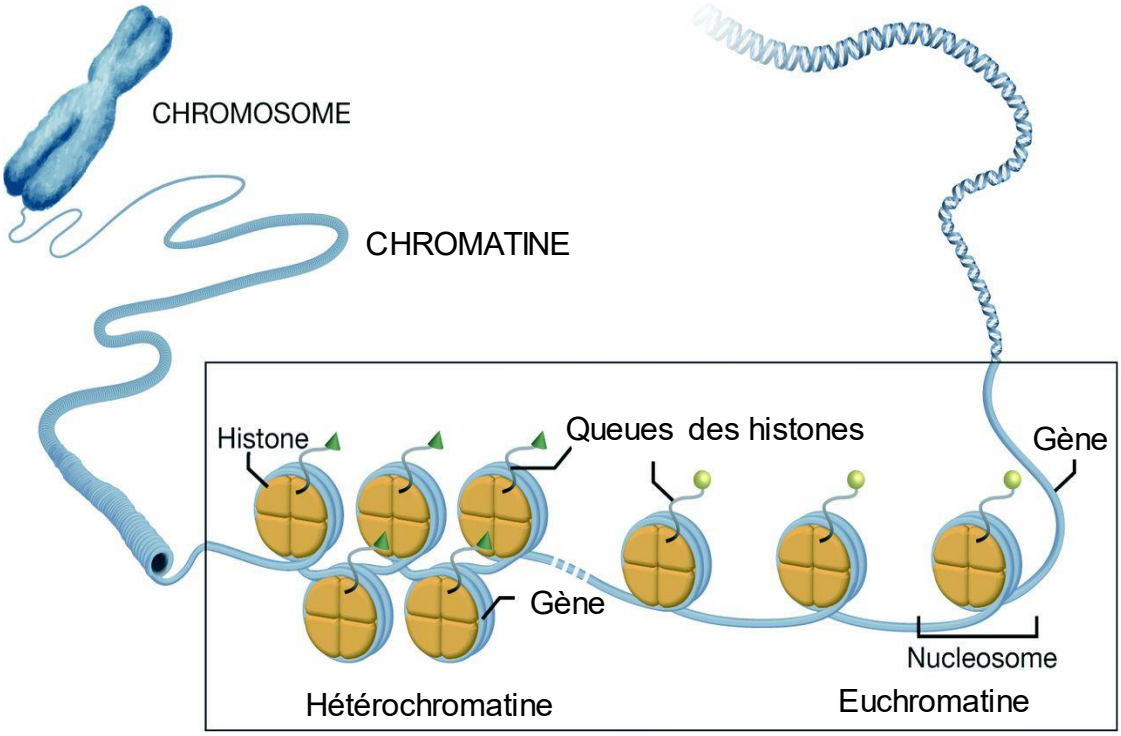


Figure 20 : Structure de la chromatine. L'ADN est enroulé autour de plusieurs unités de base appelées « nucléosomes ». Chaque nucléosome est composé de deux paires d'histones H2A, H2B, H3 et H4. Les nucléosomes sont liés les uns aux autres par l'histone H1. La chromatine peut être sous forme condensée « hétérochromatine » ou sous forme relâchée « euchromatine » selon son enrichissement en marques épigénétiques particulières. Adapté d'après (Pieterman et al., 2014)

2.3.2. Altérations des composants de la HR

Comme nous l'avons précédemment mentionné, les protéines RAD51, RAD52 et RAD54 occupent des fonctions importantes au cours du processus de HR. L'altération du gène *Rad51* chez la souris induit une létalité au stade embryonnaire (Lim and Hasty, 1996; Tsuzuki et al., 1996), alors que celle de *Rad52* réduit le processus de HR sans générer de sensibilité accrue aux dommages de l'ADN (Rijkers et al., 1998). D'autres études ont montré que le déficit en *Rad52* induit une augmentation de la CSR, avec une diminution de l'usage de jonctions comprenant des séquences de microhomologie (Zan et al., 2017).

3. L'orientation entre les différentes voies de réparation des CDB

La question qui peut se poser est comment s'oriente la réparation des CDB vers la voie HR ou celles du NHEJ (C-NHEJ et A-EJ). Afin de répondre à cette interrogation, des études menées *in vivo* ont montré que la délétion de *Ku* induit une augmentation de la réparation par le système HR, suggérant une compétition entre les deux grand systèmes (Adachi et al., 2001; Pierce et al., 2001). De plus, chez l'Homme, la coopération de Ku avec DNA-PKcs régule le choix entre le C-NHEJ et l'A-EJ tout en inhibant la voie HR (Fattah et al., 2010).

3.1. Le rôle de la structure chromatinienne

La chromatine est composée d'une unité fondamentale qui est le nucléosome, constitué d'un filament d'ADN enroulé régulièrement autour de complexes protéiques formés chacun par deux copies d'histones (H) H2A, H2B, H3 et H4. Les nucléosomes s'enroulent sur eux-mêmes de façon plus ou moins « serrée », formant des fibres de chromatine plus ou moins denses. Le niveau de condensation le plus élevé correspond à l'hétérochromatine, définie comme fermée et inaccessible aux protéines se liant de façon directe sur l'ADN **(Figure 20)**.

Il s'est avéré que cette structure chromatinienne joue un rôle important dans le choix de recrutement des facteurs de réparation de l'ADN. Tout d'abord, les CDB qui se produisent dans une région contenant de l'hétérochromatine en phase G2 du cycle

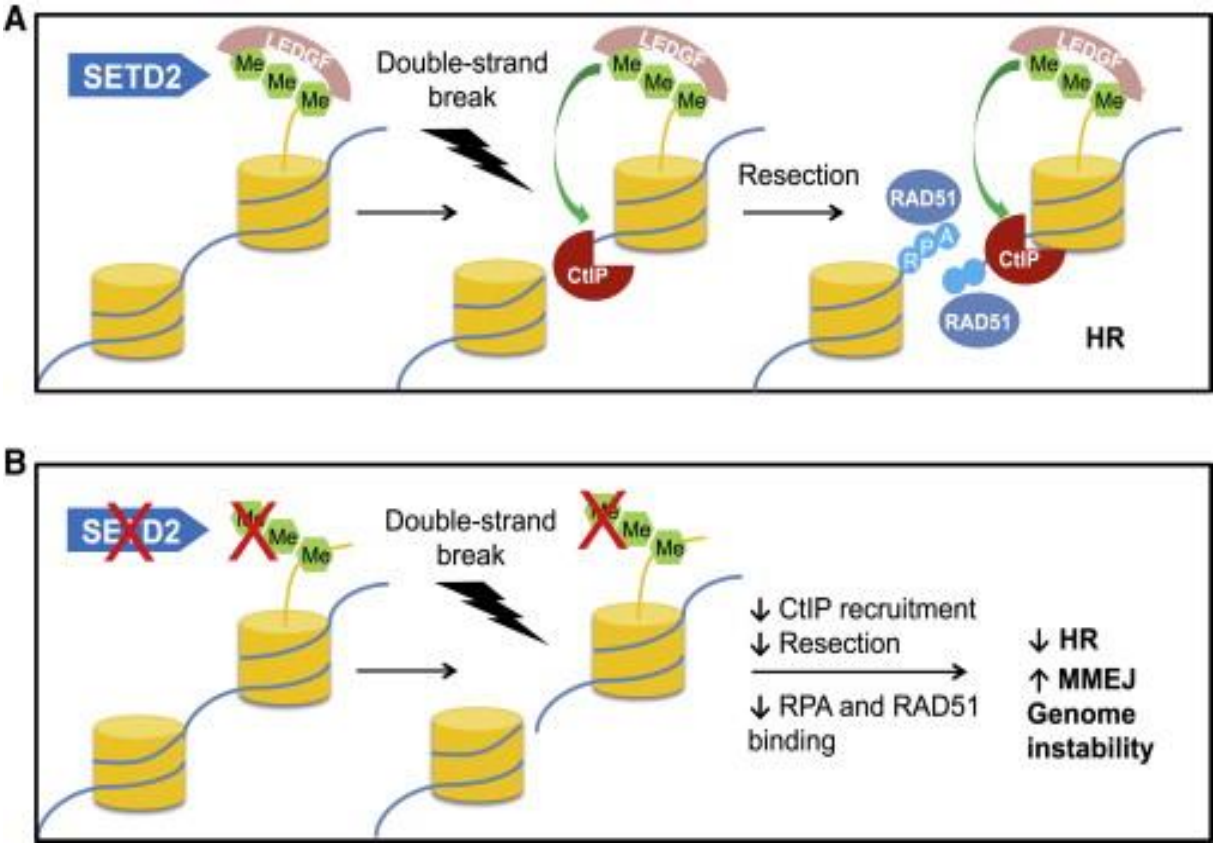


Figure 21 : Représentation schématique de la fonction de SETD2 dans la HR. A. La tri-méthylation de l'H3K36 recrute le facteur LEDGF (Lens epithelium-derived growth factor p75). Après l'apparition de CDB, ce dernier recrute CtIP qui active la résection et déclenche la voie HR en facilitant l'association de RPA et de RAD51. **B.** En absence de l'H3K36me3 (absence de SETD2), LEDGF ne peut plus s'associer à la chromatine ce qui réduit la résection de l'ADN médiée par CtIP favorisant la voie de l'A-EJ plutôt que celle de la HR. D'après (Pfister et al., 2014)

cellulaire sont résolues préférentiellement par la HR (Beucher et al., 2009). Il a été montré que Ku70 est présent au niveau des CDB situées dans l'euchromatine ainsi que dans l'hétérochromatine, ce qui indique une réparation possible par le C-NHEJ dans les deux configurations (Lorat et al., 2012).

Il est intéressant de noter que l'efficacité et la vitesse de réparation est régulée par l'état de condensation de la chromatine (Lorat et al., 2012). La chromatine condensée rend difficile l'accès et donc le recrutement des différents acteurs de réparation au niveau de l'ADN lésé. Pour y remédier, les histones subissent des modifications dont la phosphorylation de l'histone H2AX sur son résidu Serine 139 situé du côté C-terminal. Cette phosphorylation est réalisée par les kinases ATM, ATR et la DNA-PKcs appartenant à la famille des PIKK (Stucki and Jackson, 2006).

Les CDB qui se produisent au niveau de gènes en état de transcription active sont réparées de façon générale par la HR. En effet, la tri-méthylation de l'histone H3 lysine 36 (H3K36me3) qui est une marque épigénétique associée à la transcription, permet le recrutement de RAD51 et des autres acteurs de la HR (Aymard et al., 2014).

Des études ont montré que l'altération de SETD2, la principale tri-méthyltransférase de l'H3K36, induit une réduction importante de la résection et du recrutement d'acteurs de la HR, ce qui semble profiter à la voie A-EJ **(Figure 21)** (Pfister et al., 2014; Clouaire and Legube, 2015). De façon étonnante, chez la levure, Set2, qui semble aussi jouer un rôle déterminant dans le choix de la voie de réparation à tendance à favoriser l'emploi du NHEJ par rapport à la HR (Jha and Strahl, 2014).

Il a été montré que la forme di-méthylée de l'H3K36 (H3K36me2) localisée au niveau de la lésion recrute NBS1 et Ku70, ce qui favorise la prise en charge des cassures par la voie du C-NHEJ (Fnu et al., 2011).

Après l'apparition de CDB, l'H4K20me1 et l'H4K20me2 recrutent 53BP1 à travers son domaine Tudor (Botuyan et al., 2006). Comme cité précedemment, 53BP1 empêche la résection des extrémités d'ADN lésées nécessaire à la réparation par la HR, et favorise de cette façon l'intervention du C-NHEJ (Yu and Chen, 2004; Bothmer et al., 2010).

L'H4K16ac est une autre marque épigénétique située généralement au niveau de gènes actifs (Taylor et al., 2013). En cas d'absence de lésions de l'ADN, l'H4K16ac empêche l'interaction entre H4K20me2/1 et 53BP1, inhibant l'activation de la voie C-NHEJ. L'apparition de CDB induit une désacétylation transitoire et localisée de l'H4K16 levant l'inhibition de 53BP1 (Hsiao and Mizzen, 2013).

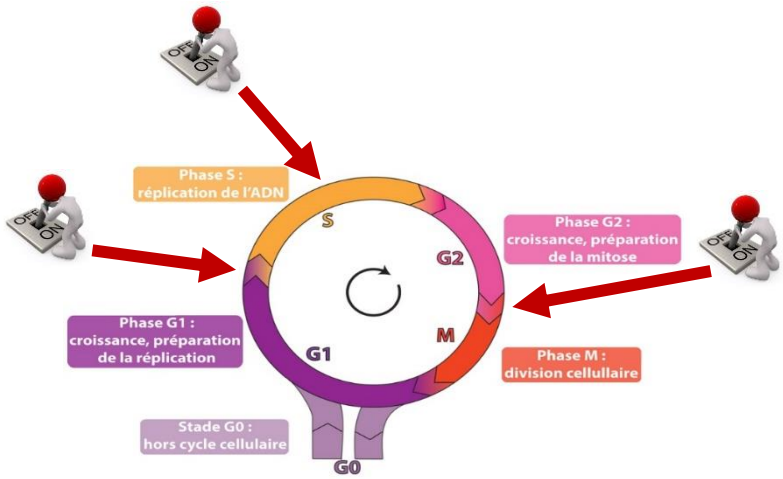


Figure 22 : Points de contrôles du cycle cellulaire. Les points de contrôle du cycle cellulaire qui sont G1/S, S/G2 et G2/M sont activés entre autres par la kinase ATM. Cette activation est réalisée lors de la réponse aux lésions de l'ADN permettant de mettre en pause le cycle cellulaire afin que la cellule puisse effectuer les réparations nécessaires et préserver ainsi l'intégrité de son génome.

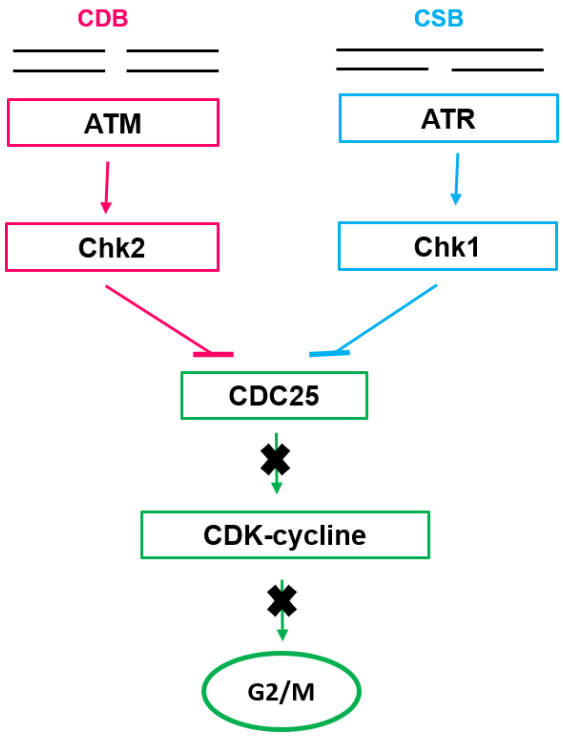


Figure 23 : Schéma de la régulation du cycle cellulaire par les kinases ATM et ATR (ici point de contrôle G2/M). Les lésions de l'ADN sont détectées par les senseurs, ATR ou ATM sont activées et activent à leur tour des effecteurs dont les protéines ChK1 et Chk2 respectivement. Ces dernières sont responsables du contrôle du cycle cellulaire. En effet, Chk1 et Chk2 activées inhibent l'activité des phosphatases de la famille CDC25 ce qui empêche la déphosphorylation des CDK. Les CDK ne peuvent plus former de complexes avec les cyclines. Les complexes CDK-cyclines sont nécessaires pour la transition entre les phases du cycle cellulaire, et leur absence induit l'arrêt du cycle cellulaire au point de contrôle (G2/M dans cet exemple).

En plus de la structure chromatinienne, des équipes se sont intéressées aux différents compartiments du noyau. Il s'est avéré que cette compartimentalisation possède un impact sur le choix de la voie de réparation des CDB. En effet, la chromatine située à l'intérieur de la lamina nucléaire à tendance à inhiber la HR (Lemaître et al., 2014). Ceci est en accord avec des études qui montrent que la HR est activée au niveau de lésions localisées dans les gènes actifs de l'euchromatine (Aymard et al., 2014; Pfister et al., 2014).

3.2. Le rôle du cycle cellulaire

Le cycle cellulaire comprend des points de contrôle « checkpoints » qui sont G1/S, intra-S et G2/M **(Figure 22)**. Ces derniers sont finement régulés par la cellule en prolifération afin d'empêcher sa progression vers la phase suivante en cas de présence de lésions importantes de l'ADN (Langerak and Russell, 2011).

La voie DDR est impliquée dans l'arrêt du cycle cellulaire à travers l'activation de deux voies de signalisation majeures **(Figure 23)**. La première voie est assurée par la kinase Chk1 qui est activée par ATR et répond majoritairement à la présence de CSB (Latif et al., 2004). La deuxième médiée par ATM qui active la kinase Chk2 est déployée en cas de CDB (Zannini et al., 2014). De façon générale, ces deux voies inhibent l'activité des phosphatases de la famille Cdc25 (Cell division cycle 25), empêchant ainsi l'activation des kinases régulatrices du cycle cellulaire dont les CDK (cyclin-dependent kinases). Ces CDK ont pour rôle principal l'activation de l'entrée de la cellule en phase de mitose (phase M).

Le C-NHEJ intervient majoritairement lors de CDB générées au cours des recombinaisons V(D)J et de la CSR (Rooney et al., 2004). Ce système est opérationnel durant l'ensemble du cycle cellulaire. En revanche, la HR est quasiment inactive lors de la phase G1, son activité est augmentée à partir de la phase S jusqu'à la fin de la phase G2. Ce qui coïncide avec la disponibilité de chromatides sœur, et le pic de protéines CtIP activées (Yu and Baer, 2000). Récemment, il a été montré que pendant ces deux phases d'activité de la HR, l'usage du C-NHEJ est réduit par l'effet de CYREN (Cell cycle regulator of NHEJ) qui s'associe à l'hétérodimère Ku70/Ku80 (Arnoult et al., 2017).

L'A-EJ quant à lui semble opérer comme système de sauvegarde lorsque le C-NHEJ est indisponible, et particulièrement en phase G1 avec une HR inactive. En effet,

des cellules déficientes en C-NHEJ et irradiées en phase G1 présentent une fréquence augmentée de translocations indiquant l'intervention potentielle de l'A-EJ (El-Khamisy et al., 2003). L'A-EJ intervient également au cours de la phase G2, lorsque la résection prolongée inhibe le recrutement des facteurs du C-NHEJ (Frit et al., 2014).

3.3. Le rôle de la résection des extrémités

Le C-NHEJ ne nécessite pas de résection de l'ADN et produit des jonctions de réparation directes ou avec de petites microhomologies (Lieber, 2008). En revanche, l'A-EJ utilise des séquences de microhomologie et nécessite donc une résection des extrémités lésées (Haber, 2008)

Il semble exister une compétition entre le recrutement des systèmes du NHEJ (classique et alternatif) et celui de la HR pour réparer les CDB. Cette compétition concerne plus particulièrement la machinerie de résection des extrémités lésées, dont le complexe MRN, BRCA-1 et CtIP d'une part, et 53BP1 et l'hétérodimère Ku70/Ku80 qui au contraire préservent ces extrémités. Les cellules B déficientes pour 53BP1 affichent une augmentation ATM dépendante de la résection générant des extrémités d'ADN qui seront prises en charge par l'A-EJ et/ou la HR (Bothmer et al., 2010; Tang et al., 2013). De la même façon, le déficit de BRCA1 altère la résection bloquant la HR et facilitant l'intervention du C-NHEJ (Bunting et al., 2010).

Ainsi, l'étape de résection des extrémités 5' est indispensable pour que la HR fonctionne. Une fois cette opération réalisée, un retour en arrière est inexécutable, à ce stade, une réparation par le C-NHEJ est quasiment impossible. Cependant, lorsque la résection est lancée, alors que la HR n'est pas opérationnelle, l'A-EJ peut prendre le relais. En effet, l'inhibition de l'A-EJ (inhibition de PARP-1 et ou la LIG III), en absence de HR, engendre une diminution importante des translocations (Soni et al., 2014).

Il est à noter qu'une dérégulation des systèmes de réparation des lésions de l'ADN peut générer un cancer, et dans le cas de cellules B à des lymphomes B.

IIII. Les lymphomes B

Le lymphome est par définition un cancer qui touche le système lymphatique. Je présenterai de manière succincte deux lymphomes B pour lesquels j'ai initié le travail d'analyse de la LSR durant ma thèse, le HL et la CLL.

1. Lymphome de Hodgkin (HL)

Le HL est caractérisé par la présence de cellules atypiques nommées « cellules de Reed-Sternberg ». Ces dernières sont de grande taille et présentent un noyau particulièrement volumineux à contours irréguliers. Ce type de cellules permet de distinguer le HL des autres lymphomes dits "non hodgkiniens".

Ce lymphome est divisé principalement en deux grandes classes : le HL classique (cHL pour classical HL), qui est le plus fréquent avec environ 95% des cas de HL, et le lymphome à prédominance lymphocytaire (NLPHL pour Nodular Lymphocyte Predominant HL) qui représente les 5% restants (Smith, 2010). Cependant, des différences subsistent au niveau de la composition cellulaire de ces lymphomes. En effet, le cHL est subdivisé en plusieurs types : lymphome de forme scléro-nodulaire, comprenant un mélange de cellules de Reed-Sternberg et de globules blancs (environ 70% des HL), le lymphome à cellularité mixte, avec beaucoup plus de cellules de Reed-Sternberg (environ 15% des HL), le lymphome riche en lymphocytes (environ 10% des cas de HL) qui ne présente pas de cellules de Reed-Sternberg, mais d'autres cellules anormales sont retrouvées, raison pour laquelle il est classé dans les lymphomes hodgkiniens, et pour finir, il y a le lymphome à déplétion lymphocytaire (moins de 5% des cas) qui se caractérise par un nombre extrêmement important de cellules de Reed-Sternberg dans les ganglions lymphatiques cancéreux.

Le HL est d'origine cellulaire B monoclonale et se caractérise par l'absence d'expression de BCR sur les cellules tumorales (Marafioti et al., 1997). La perte d'expression des Ig dans les cellules de Reed-Sternberg est le résultat de mutations non-sens dans les gènes codant pour la partie variable de l'IgH(Marafioti et al., 1997). Les cellules de Reed-Sternberg sont généralement des LB transformés. En effet, elles perdent l'expression de la majorité des marqueurs de la lignée B à un degré qui les rend uniques parmi les lymphomes B (Kanzler et al., 1996; Marafioti et al., 1997; Schwering et al., 2003).

L'échappement cellulaire à l'apoptose joue un rôle primordial dans la pathogénicité de la cellule tumorale, et ceci s'applique spécialement aux cellules de Reed-Sternberg. Plusieurs des voies de signalisation sont activées de façon aberrante et contribuent à la survie de ces cellules. Un autre facteur qui empêche les cellules de mourir est l'infection de ces dernières par le virus d'Epstein-Barr (EBV pour Epstein-Barr Virus). Effectivement, dans environ 40% des cas de HL, les cellules de Reed-Sternberg sont infectées par l'EBV et expriment la protéine LMP2A qui se comporte comme un substitut du BCR et/ou la protéine LMP1, substitut de CD40, ces protéines virales sont responsables d'une signalisation constitutive oncogénique (Glaser et al., 1997).

2. Leucémie lymphoïde chronique (CLL)

La leucémie est un cancer qui touche les leucocytes dans la moelle osseuse. Il peut être divisé en leucémies aigues et chroniques. Elles peuvent également être classées en quatre grand types qui sont la leucémie myéloïde aigue (AML pour Acute myeloid leukemia), leucémie lymphoïde aigue (ALL), leucémie myéloïde chronique (CML) et la CLL (Rodriguez-Abreu et al., 2007).

La CLL est une forme de leucémie dans laquelle il existe un nombre excessif de LB matures, mais peu fonctionnels dans le sang et la moelle osseuse. En effet, une partie des cellules B matures présentent un défaut d'apoptose qui induit leur accumulation dans le sang, les ganglions, la rate et la moelle osseuse. Cette accumulation engendre une hyperplasie, autrement dit, une augmentation du volume des ganglions et de la rate.

D'un point de vue moléculaire, le statut mutationnel de l'*IGHV* permet de diviser la CLL en deux catégories : les CLL avec taux de mutations élevé des gènes *IGHV*, et celles sans (ou avec peu de) mutations. Cette deuxième catégorie est la forme la plus agressive de la leucémie, car les patients atteints de ce type de CLL affichent un taux de survie plus faible (Hamblin et al., 1999).

Dans ces deux types de lymphomes B, J'ai initié le travail de caractérisation de la fréquence et de la structure des jonctions LSR.

MATERIEL ET METHODES

Stimulation de lymphocytes B murins

Des souris C57BL/6J WT ont été maintenues dans une animalerie exempte d'organismes pathogènes spécifiques (EOPS), puis sacrifiées à 6 ou 8 semaines. Les splénocytes ont été extraits et stimulés *in vitro* pendant 48 heures à 72 heures avec du LPS B4 (Invivogen, Hrl-3pelps) (1µg/ml) et de l'IL-4 (Peprotech, 214-14) (20ng/ml) pour induire la CSR vers l'Igγ1 (CSRγ1) et la LSR. Les expériences ont été réalisées en accord avec le comité d'éthique de l'Université de Limoges et le CHU de Limoges (Comité Régional d'Ethique sur l'Expérimentation Animale du Limousin, CHU Dupuytren Limoges, France).

Vérification de la stimulation par cytométrie en flux

48 heures après la stimulation des splénocytes, les cellules ont été marquées avec 5µl des anticorps anti-B220 BV510 (Biolegend, 103247) et anti-IgG1 APC (BD, 553440) pendant 20 minutes dans l'obscurité et à 4°C. Sur les cellules vivantes sélectionnées par les caractéristiques FSC/SSC, le pourcentage de la population cellulaire IgG1+ a été analysé à partir des cellules B220+ en utilisant le logiciel BD Diva version 6.1.3.

Amplification des jonctions CSR et LSR chez la souris

Les jonctions CSRγ1 ont été amplifiées à partir de 100ng d'ADN génomique extrait de splénocytes de souris par PCR nichée comme décrit dans notre article précédent (Boyer et al., 2017). Les jonctions LSR ont également été amplifiées à partir de 100ng d'ADN génomique de splénocytes par PCR nichée avec la polymérase Phusion High Fidelity (New England Biolabs, M0530S). Les premières amorces employées sont : Smum1 (CAG TTG AGG CCA GCA GGT) et LS43R (CTA TAG CCA TGT GGG GCT GT) selon le programme suivant : 94°C/5min, 35 cycles de 94°C/1min, 60°C/1min, 72°C/1 min30, et une étape finale à 72°C/7min. La seconde paire d'amorces est la suivante : Smum2 (CAG GTC GGC TGG ACT AAC TC) et LS42R (TTC AAC AAT CTC CCC CAC TG) selon les conditions suivantes : 94°C/5min, 35

cycles de 94°C/1min, 62°C/1min, 72°C/1min et une dernière étape à 72°C/7min. La PCR a été réalisée en triplicate pour chaque échantillon.

Immunoprécipitation de la chromatine

Les expériences de ChIP ont été réalisées selon le protocole décrit dans l'article (Iacovoni et al., 2010). Les cellules ont été traitées avec 1% de formaldéhyde 15 minutes à température ambiante. Après une série de lavages, les culots ont été repris dans du tampon de lyse et incubés pendant 10 minutes à 4°C, et les noyaux séparés et lysés. La chromatine a été soniquée pour générer des fragments d'environ 200 pb à 1000 pb et incubée avec des billes d'agarose. Environ 150ng de chromatine a été utilisée pour chaque immunoprécipitation avec les anticorps suivants : anti-H3K36me3 (Abcam ab9050), anti-H3K36me2 (Merck Millipore 07-274), anti-H4K20me1 (Abcam ab9051), anti-H4K16ac (Merck Millipore 07-329), anti-RAD51 (Santa Cruz Biotechnology 398587), anti-XRCC4 (Santa Cruz Biotechnology 8285), anti-polθ (Abcam 80906) et anti-PARP1 (Cell Signaling 9542S). La chromatine a été incubée avec environ 2 µg de chaque anticorps toute une nuit à 4°C, puis des billes ont été ajoutées et incubées pendant 2 heures à 4°C. La chromatine et les inputs ont été resuspendus dans du TE10.1. L'ADN a été incubé avec de la RNAse (0,05mg/mL) pendant 20 minutes à 37°C, puis du SDS 10% et incubation toute une nuit à 70°C. De la protéinase K a été rajoutée et l'ADN extrait par la méthode classique au phénol/chloroforme/isoamino-alcool. Les ADN purifiés à l'issue des expériences de ChIP ont été soumis à la PCR en temp réel en utilisant le kit TB Green Premix Ex Taq II (Takara) et l'appareil StepOnePlus (ThermoFisher), afin d'analyser leur enrichissement au niveau de séquences d'ADN d'intérêt. Plusieurs couples d'amorces spécifiques ont été utilisées : Sµ F (GCTAAACTGAGGTGATTACTCTGAGGTAAG), Sµ R(GTTTAGCTTAGCGGCCCAGCTCATTCCAGT), Sγ1 F (ATAAGTAGTAGTTGGGGATTC), Sγ1 R (CTCAGCCTGGTACCTTATACA), Sγ3 F (AATCTACAGAGAGCCAGGTGG), Sγ3 R (TGGTTTTCCATGTTCCCACTT) (Zan et al., 2017), ainsi que les couples d'amorces : HS1/2 F (ATCAGTACCAGAAACAAGGC), HS1/2 R (TTGGGGTGAACCTGCAGC), HS4 F (TTTAGTCTCAGCAAGACCC), HS4 R (AATGGGGCTTTCCACGCC) (Péron et al., 2012).

Amplification des jonctions CSR et LSR à partir d'échantillons humains

L'ADNg a été extrait selon la méthode classique au phénol/chloroforme/isoamino-alcool à partir d'échantillons humains obtenus par le Centre de Ressources Biologiques de Limoges (CRBioLim, CHU de Limoges). A partir de l'ADNg d'échantillons humains (amygdales pédiatriques saines, amygdales de patients atteints de lymphome hodgkinien, cellules mononucléées du sang périphérique (PBMC) de patients atteints de CLL. Les jonctions CSRα ont été amplifiées par PCR à partir de 200ng comme cité précédemment (Pan et al., 2001), ainsi que les jonctions LSR en utilisant la polymérase Herculase II Fusion (Agilent 600677). Une première PCR a été réalisée avec les amorces : Eμh1 (AGG CTG ACC GAA ACT GAA AA) et 3'FarHS4 Rev (CAA GCG TCA AGG TGT GGA C) et les conditions suivantes: 98°C/30sec, 2 cycles de 98°C/10sec, 64°C/30sec, 68°C/4min, 3 cycles de 98°C/10sec, 62°C/30sec, 68°C/4min, 25 cycles de 98°C/10sec, 60°C/30sec, 68°C/4min er 68°C/ 5min. la PCR nichée utilise le deuxième couple d'amorces Sμ1b (CAG GGA ACT GGG GTA TCA AG) et Probe 3'Far HS4 Rev (GGA CGC GGT TTG CTT TTA T) selon le programme suivant: 98°C/30sec, 30 cycles de 98°C/10sec, 58°C/30sec, 68°C/4 min et 68°C/ 5min. Les PCR ont été réalisées en triplicate pour chaque échantillon. L'étude a été approuvée par le comité éthique (N° : CPP12-012/2012-A00630-43, Comité de Protection des Personnes du Sud-Ouest et Outre-Mer IV).

Préparation des librairies de séquençage haut débit (Ion Proton)

Des librairies de séquençage haut débit de fragments d'environ 200 pb ont été préparées à partir de 100ng de produits de trois PCR de jonctions de LSR ou de CSR préparées à partir de l'ADN des échantillons murins et humains mentionnés précédemment. Les librairies ont été préparées en utilisant le kit « Ion Xpress Library Kit » (Life Technologies, 4471269). La concentration de chaque librairie a été déterminée par dosage au Qubit à l'aide du kit DsDNA high sensitivity (Thermofisher Scientific, Q32854), et un mélange équimolaire des librairies a été déposé et séquencé sur des puces Ion PI v3 du séquenceur haut débit Ion Proton (Life Technologies). Le nombre de puces dépend du nombre de librairies à charger (60 librairies

maximum/puce). Les données de séquençage ont été analysées avec le programme CSReport.

OBJECTIF DE LA THESE

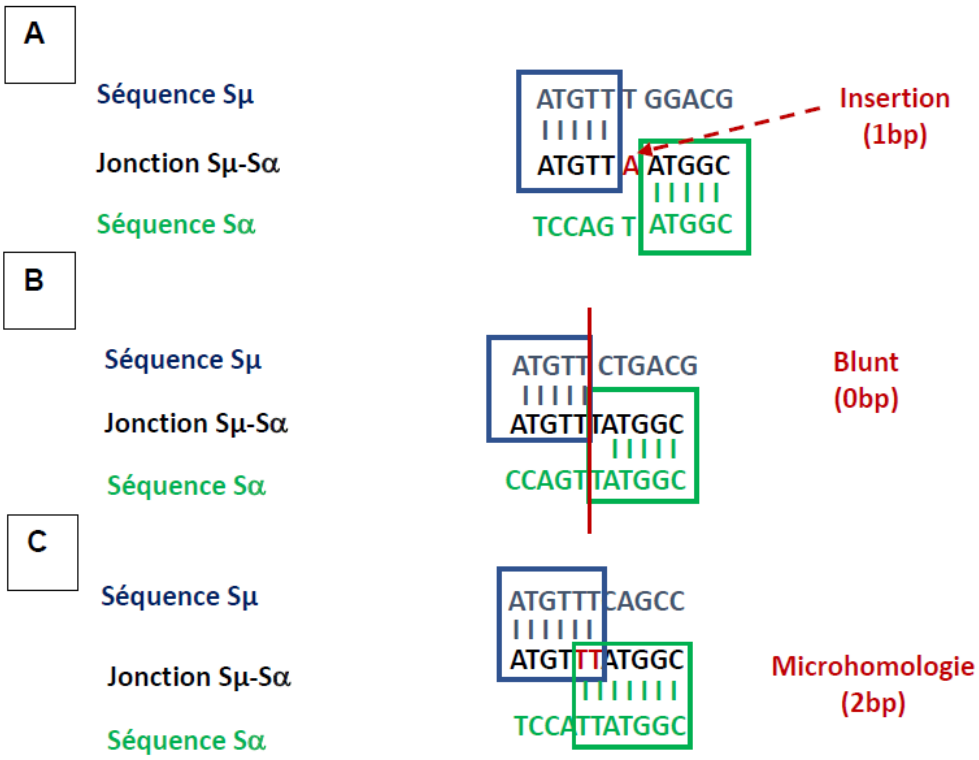


Figure 24 : Schémas des différents types de structure des jonctions de recombinaison du locus *IgH* (ici jonctions CSRα). A. Alignement de la séquence de la jonction avec les séquences donneuse et acceptrice révélant une jonction de recombinaison avec insertion (1 pb). **B.** Alignement donnant lieu à une jonction franche ou blunt (0 pb). **C.** Alignement présentant l'identification d'une jonction avec une séquence de microhomologie (2 pb).

Au cours de ma thèse j'ai eu pour objectif de participer à définir le mécanisme moléculaire de la recombinaison LSR dans les LB, et plus exactement, la caractérisation des systèmes de réparation des CDB intervenant dans ce processus. La LSR est un évènement fugace comparé à la CSR, puisqu'il induit la mort de la cellule B qui l'exécute. Pour cette raison, son étude est difficile, pour y remédier mon travail de doctorat a consisté dans un premier temps à adapter le séquençage haut débit pour l'analyse des jonctions de recombinaison du locus *IgH*. Ceci a nécessité l'élaboration d'un logiciel informatique automatisant l'analyse. Nous nous sommes aussi interrogés sur la potentielle implication de la LSR dans des cas de lymphomes B humains, et plus particulièrement, dans le HL, car les cellules tumorales ne présentent pas de BCR à leur surface. Nous avons également travaillé à partir d'échantillons de CLL, lymphome non hodgkinien comme contrôle.

RESULTATS

Partie I : Développement d'un protocole d'étude et d'un outil d'analyse des recombinaisons du locus *IgH*

Face aux difficultés rencontrées lors du séquençage des jonctions LSR par la méthode classiquement employée (clonage et séquençage Sanger), nous avons choisi d'utiliser un séquenceur haut débit Ion Proton (Life technologies), afin de séquencer des librairies issues de PCR nichées amplifiant les jonctions LSR et CSR pour chaque échantillon. Le séquençage haut débit possède une haute profondeur de lecture qui permet de séquencer de façon massive de nombreuses jonctions de recombinaison. Cependant, la quantité de séquences (plusieurs millions de séquences par librairie) est telle que l'analyse des jonctions ne peut être effectuée manuellement et doit être automatisée.

Dans un premier temps, l'analyse des résultats de séquençage des jonctions de recombinaison LSR et CSR a été réalisée avec la version en ligne de Blast. Cette dernière a permis l'analyse d'un petit nombre de jonctions, car l'étape de vérification visuelle prend beaucoup de temps pour chaque résultat d'alignement. Nous avons donc développé « CSReport », un programme adapté à l'analyse de jonctions de recombinaison qui fonctionne avec une série d'algorithmes et prend en charge le traitement des données brutes dès leur sortie du séquenceur haut débit. Les résultats d'analyse des jonctions de recombinaison par CSReport peuvent être représentés de façon synthétique dans un tableau Excel et sous forme de graphe. Cet outil nous permet de détecter, quantifier et de révéler la structure des recombinaisons du locus *IgH* (CSR et LSR), à partir des données issues du séquenceur haut débit. Cependant, son utilisation peut être élargie au traitement de fichiers obtenus par d'autres séquenceurs haut débit. Il peut également rendre possible l'analyse de diverses recombinaisons génétiques, comme les translocations *IgH*/*c-Myc*, et ce chez plusieurs espèces, à condition que les séquences de référence soient disponibles.

Les différentes étapes de traitement de CSReport ont été mises au point selon la logique d'analyse manuelle des jonctions de recombinaison, avec une attention particulière portée sur la détermination de la structure de ces jonctions. En effet, la structure d'une jonction apporte des informations précieuses sur le système impliqué dans la réparation de l'ADN.

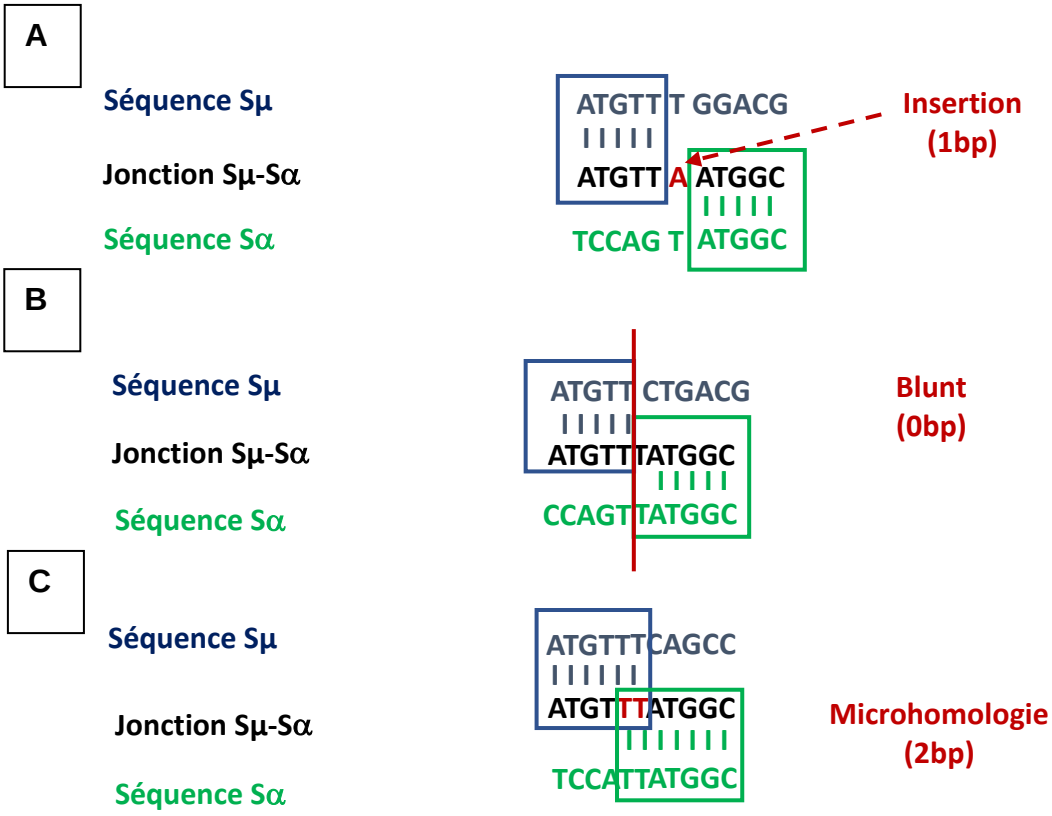


Figure 24 : Schémas des différents types de structure pour les jonctions de recombinaison du locus *IgH* (ici jonctions CSRα). A. Alignement révélant une jonction de recombinaison avec insertion (1 pb). B. Alignement donnant lieu à une jonction franche (0 pb). C. Alignement aboutissant à l'identification d'une jonction avec microhomologie (2 pb).

- Dans un premier temps, CSReport assure la détection des recombinaisons à travers l'alignement des séquences contre le locus *IgH*, en faisant appel à Blastn.
- Un minimum d'alignement de 30 pb est requis pour chacune des deux séquences donneuse et acceptrice de la recombinaison.
- Une première évaluation de la structure est obtenue par la segmentation de Blastn :

- La structure d'une jonction de recombinaison est déterminée par le nombre de bases nucléotidiques qui chevauchent l'extrémité 3' du premier segment et l'extrémité 5' du second, comme le montre la figure 14. Si ce nombre est positif, la jonction est dite « avec microhomologie », c'est-à-dire, une séquence d'ADN identique aux deux segments. En cas d'absence de chevauchement, la jonction est appelée « blunt », autrement dit, avec des extrémités franches, ou alors, il peut y avoir des nucléotides non identifiés par l'alignement entre les deux séquences, la jonction est alors nommée « avec insertion » **(Figure 24)**.
- Lorsqu'une longue insertion est trouvée (>=10 pb), un deuxième alignement est réalisé par Blast-short, ce qui évite une perte de séquences pouvant provenir de multiples cassures au niveau des régions S, créant des jonctions complexes, observées dans des études précédentes (Pan-Hammarström et al., 2006; Reina-San-Martin et al., 2007).
- Les jonctions avec des microhomologies (>2 pb) sont alignées une nouvelle fois en utilisant le module de Biopython (pairwise2), pour s'assurer de l'absence de décalage dans l'alignement pouvant survenir avec Blastn.

- La dernière mission de CSReport est le regroupement des résultats de chaque librairie de façon individuelle dans des fichiers sous format Excel.

La mise au point de notre programme a été effectuée à partir de librairies issues de produits d'amplification de jonctions CSR. Les caractéristiques de structure de ces jonctions ont largement été décrites dans la littérature. De plus, cette analyse structurale est réalisée de façon routinière au sein de notre laboratoire, permettant la confrontation des résultats déjà obtenus de façon manuelle à ceux générés de manière automatisée par CSReport.

Nous avons donc effectué cette comparaison pour une vingtaine de jonctions CSR sélectionnées de façon aléatoire, et les deux types d'analyse ont abouti à des

résultats similaires. Ensuite, nous avons procédé à l'analyse de 5000 jonctions créées de façon informatique, par concaténation de segments Sμ et Sα simulant des jonctions CSRα. CSReport a identifié 3278 jonctions sur les 5000 analysées. L'analyse manuelle détaillée des jonctions non identifiées a révélé la présence d'homologie de séquence entre Sμ et Sα. Cette homologie est responsable de la perte de sensibilité dans l'étape initiale d'alignement avec Blastn. En effet, ce dernier n'arrive pas à statuer sur l'identité d'un segment donné, s'il provient de la région Sμ ou alors s'il appartient à Sα. La perte de certaines jonctions de recombinaison en raison de l'homologie existante entre les segments donneur et accepteur est un point faible de CSReport. Toutefois, une analyse manuelle utilisant blast aboutira à ce même manque de résultat. De plus, des mutations au point de jonction peuvent induire un biais d'analyse des jonctions vers celles contenant des insertions, mais cet effet est aussi retrouvé dans d'autres méthodes d'analyse.

Finalement, CSReport donne le nombre de séquences lues pour chaque jonction identifiée et le nombre de jonctions uniques. Les jonctions uniques sont définies comme affichant des points de cassure identiques ou très proches d'un à deux nucléotides, et sont regroupées en « clusters », afin de permettre leur quantification.

Une augmentation de la tolérance dans la localisation des points de cassure altère la diversité des échantillons évaluée par l'inverse de diversité de Simpson. Cet indice nous donne une estimation de la diversité des jonctions dans un échantillon donné. Plus il est augmenté, plus il y a de jonctions de structures différentes dans la librairie de séquençage analysée, et inversement. Nous avons défini un seuil de tolérance maximum de cette localisation des cassures de 10 pb. CSReport est également capable de résumer les motifs dans lesquels les cassures sont retrouvés de façon fréquente. Par exemple, les motifs « AGCT » et « GGG » sont souvent séquencées au site de jonction dans les recombinaisons Sμ/Sα de la lignée CH12F3 (Boyer et al., 2017).

CSReport apporte une solution sans précédent concernant l'analyse d'une quantité importante de jonctions de recombinaison à partir de librairies de séquençage haut débit. Cet outil libre d'accès est à la portée de chaque utilisateur en possession d'un ordinateur classique. Notre programme facilite et accélère l'étude des recombinaisons du locus *IgH* et harmonise l'analyse et l'interprétation des résultats.

Nous avons utilisé CSReport pour l'analyse des jonctions de CSRα et CSRγ1 chez la souris ainsi que des jonctions CSRα chez l'Homme et publié les résultats (Boyer et al., 2017)

The Journal of Immunology

This information is current as of April 10, 2018.

CSReport: A New Computational Tool Designed for Automatic Analysis of Class Switch Recombination Junctions Sequenced by High-Throughput Sequencing

François Boyer, Hend Boutouil, Iman Dalloul, Zeinab Dalloul, Jeanne Cook-Moreau, Jean-Claude Aldigier, Claire Carrion, Bastien Herve, Erwan Scaon, Michel Cogné and Sophie Péron

J Immunol 2017; 198:4148-4155; Prepublished online 17 April 2017;
doi: 10.4049/jimmunol.1601924
http://www.jimmunol.org/content/198/10/4148

Supplementary Material **http://www.jimmunol.org/content/suppl/2017/04/16/jimmunol.1601924.DCSupplemental**

References This article **cites 16 articles**, 6 of which you can access for free at:
http://www.jimmunol.org/content/198/10/4148.full#ref-list-1

Why *The JI*? Submit online.

- **Rapid Reviews! 30 days*** from submission to initial decision
- **No Triage!** Every submission reviewed by practicing scientists
- **Fast Publication!** 4 weeks from acceptance to publication

**average*

Subscription Information about subscribing to *The Journal of Immunology* is online at:
http://jimmunol.org/subscription

Permissions Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts

The Journal of Immunology is published twice each month by
The American Association of Immunologists, Inc.,
1451 Rockville Pike, Suite 650, Rockville, MD 20852
Copyright © 2017 by The American Association of Immunologists, Inc. All rights reserved.
Print ISSN: 0022-1767 Online ISSN: 1550-6606.

Downloaded from http://www.jimmunol.org/ by guest on April 10, 2018

CSReport: A New Computational Tool Designed for Automatic Analysis of Class Switch Recombination Junctions Sequenced by High-Throughput Sequencing

François Boyer,*,†,[1] Hend Boutouil,*,†,[1] Iman Dalloul,*,† Zeinab Dalloul,*,† Jeanne Cook-Moreau,*,† Jean-Claude Aldigier,*,† Claire Carrion,*,† Bastien Herve,‡ Erwan Scaon,‡ Michel Cogné,*,†,§ and Sophie Péron*,†

B cells ensure humoral immune responses due to the production of Ag-specific memory B cells and Ab-secreting plasma cells. In secondary lymphoid organs, Ag-driven B cell activation induces terminal maturation and Ig isotype class switch (class switch recombination [CSR]). CSR creates a virtually unique *IgH* locus in every B cell clone by intrachromosomal recombination between two switch (S) regions upstream of each C region gene. Amount and structural features of CSR junctions reveal valuable information about the CSR mechanism, and analysis of CSR junctions is useful in basic and clinical research studies of B cell functions. To provide an automated tool able to analyze large data sets of CSR junction sequences produced by high-throughput sequencing (HTS), we designed CSReport, a software program dedicated to support analysis of CSR recombination junctions sequenced with a HTS-based protocol (Ion Torrent technology). CSReport was assessed using simulated data sets of CSR junctions and then used for analysis of Sμ-Sα and Sμ-Sγ1 junctions from CH12F3 cells and primary murine B cells, respectively. CSReport identifies junction segment breakpoints on reference sequences and junction structure (blunt-ended junctions or junctions with insertions or microhomology). Besides the ability to analyze unprecedentedly large libraries of junction sequences, CSReport will provide a unified framework for CSR junction studies. Our results show that CSReport is an accurate tool for analysis of sequences from our HTS-based protocol for CSR junctions, thereby facilitating and accelerating their study. ***The Journal of Immunology*****, 2017, 198: 4148–4155.**

B cells ensure humoral immune responses due to their capacity to recognize a wide variety of Ags via their BCR for Ag. BCR is the association of an Ig, composed of two identical IgH chains covalently bound to two identical IgL chains, which are associated with an Igα/Igβ heterodimer at the B cell membrane. In mature B cells, in secondary lymphoid organs, Ag binding by the BCR with T cell–dependent coactivation signaling induces Ag-driven activation of B cells, which then undergo terminal maturation, ensuring production of Ag-specific memory B cells and Ab-secreting plasma cells. Two molecular events absolutely necessary for efficient humoral responses against Ag occur, as follows: Ig isotype class switch (class switch recombination [CSR]) and somatic hypermutation (SHM) (1, 2). CSR is an intrachromosomal recombination that occurs in the *IgH* locus between two switch (S) regions located upstream of each C segment encoding the C region of the Ig. SHM consists in the introduction of mutations in the V regions of *IgH* ($V_HD_HJ_H$) and *IgL* (V_LJ_L) loci, which recognize the Ag. CSR enables synthesis of different Ig isotypes, each characterized by distinct effector functions, without affecting the antigenic specificity of Ig. SHM permits the selection of B cell clones with high affinity for Ag. CSR is divided into three steps, as follows: 1) germline transcription of targeted S regions; 2) DNA lesions induced by the enzyme activation-induced cytidine deaminase (1, 3) leading to double-strand breaks (DSB) in DNA; and finally 3) DNA repair mainly by the nonhomologous end joining (NHEJ) pathway.

Analysis of CSR junctions has been used in basic and clinical research (for review, see Ref. 4) because it provides insight into the amount of class-switched B cells in samples and documents the predominant CSR molecular pathway. In response to Ag, B cells expand and CSR creates a virtually unique *IgH* locus in every B cell clone, and CSR junctions can be amplified by PCR in a given sample. This is especially needed to evaluate the potential alterations of CSR in various immune defects. Structure of the repair

*Université de Limoges, Contrôle de la Réponse Immune B et Lymphoprolifération, UMR 7276, F-87000 Limoges, France; †CNRS, Contrôle de la Réponse Immune B et Lymphoprolifération, UMR 7276, F-87000 Limoges, France; ‡Université de Limoges, plateforme Biologie Intégrative Santé Chimie Environnement, F-87000 Limoges, France; and §Institut Universitaire de France, F-75000 Paris, France

[1]F.B. and H.B. contributed equally to this work.

ORCIDs: 0000-0002-0351-685X (F.B.); 0000-0003-3590-1602 (J.C.-M.); 0000-0003-3221-6319 (C.C.); 0000-0002-8519-4427 (M.C.); 0000-0002-0056-3271 (S.P.).

Received for publication November 10, 2016. Accepted for publication February 19, 2017.

This work was supported by grants from Conseil Régional du Limousin, Association pour la Recherche sur le Cancer, and Ligue contre le Cancer (Comité de la Haute-Vienne, Comité de la Corrèze). F.B. is supported by Fondation Partenariale de l'Université de Limoges and Association Limousine pour l'Utilisation du Rein Artificiel à Domicile. H.B. is supported by Région Limousin and Université de Limoges. M.C. is supported by Institut Universitaire de France and Fondation pour la Recherche Médicale.

The sequences presented in this article have been submitted to the European Molecular Biology Laboratory–European Bioinformatics Institute European Nucleotide Archive (http://www.ebi.ac.uk/ena) under the accession number ERP019632.

See related articles in this issue: Langerak et al. (*J. Immunol.* 198, 3765; DOI: https://doi.org/10.4049/jimmunol.1602050) and IJspeert et al. (*J. Immunol.* 198, 4156; DOI: https://doi.org/10.4049/jimmunol.1601921).

Address correspondence and reprint requests to Dr. Sophie Péron, Université de Limoges, CNRS UMR 7276, Centre de Biologie et de Recherche en Santé, Faculté de Médecine, 2 rue du Docteur Marcland, Limoges 87000, France. E-mail address: sophie.peron@inserm.fr

The online version of this article contains supplemental material.

Abbreviations used in this article: CSR, class switch recombination; DSB, double strand break; HTS, high-throughput sequencing; NHEJ, nonhomologous end joining; SHM, somatic hypermutation; S region, switch region.

Copyright © 2017 by The American Association of Immunologists, Inc. 0022-1767/17/$30.00

Downloaded from http://www.jimmunol.org/ by guest on April 10, 2018

junction between donor and acceptor segments indeed reflects the DNA DSB repair pathway used during CSR: it has been shown that a recombination junction between two S regions may be repaired by alternative end joining DNA (5, 6) in the case of defective NHEJ. CSR recombination junction analysis protocols (6, 7) until now relied on cloning and subsequent sequencing by Sanger's method of individual products obtained by a single PCR or a nested PCR designed to amplify DNA fragments from switch junctions. Alignment of sequences with germline reference sequences assigns breakpoints on the *IgH* locus for each donor and acceptor segments involved, characterizes individual junction structure, and identifies mutations around the CSR recombination sites.

To overcome the laborious currently used manual processes and increase in-depth switch junction analysis, we applied a protocol using high-throughput sequencing (HTS). As a proof of concept, this was done with data generated in our laboratory using the Ion Torrent technology (Life Technologies, Thermo Fisher Scientific) based on simple library preparation utilizing DNA segments from amplification of switch junctions by nested PCR. For processing of large data sets from HTS, available Blastn and BWA algorithms for sequences analysis are inappropriate, because they are unable to automatically summarize results and structural patterns of junctions. The massive amount of HTS data thus requires a specifically designed tool to process sequencing data into insightful information. A big challenge was to develop a new computational tool that automatically analyzes HTS data, which identifies CSR junctions between two regions of the *IgH* locus and summarizes breakpoints and structure (blunt junctions, junctions with microhomology, or insertions inserted into the junction). During this study, we developed CSReport, a program processing HTS data for automatic analysis of CSR recombination junctions. CSReport is a Python code using Blast$^+$ and Biopython module. It runs in ~10 min on a laptop computer for a typical 1 million read library. Our results validate CSReport as a successful and accurate computational tool for supporting analysis of extremely large numbers of sequences (as from our HTS-based protocol for CSR junctions), gaining in-depth assessment, and facilitating manipulation and extraction of valuable information.

Materials and Methods

CH12F3 cell culture and CSR induction in vitro

CH12F3 cells were maintained in RPMI 1640 supplemented with Ultra Glutamine, 10% FCS (Lonza), sodium pyruvate (Lonza), penicillin/streptomycin (Lonza), nonessential amino acids (Lonza), and 2-ME. Cells were stimulated to undergo CSR toward IgA with murine IL-4 (5 ng/ml; PeproTech), human TGF-β1 (1 ng/ml; R&D Systems), and murine anti-CD40 Ab (200 ng/ml; eBioscience) (8). After 72-h activation, cells were labeled with anti-mouse IgA (IgA-FITC; BD Biosciences-Pharmingen) and anti-mouse IgM-PE-cyanine7 (eBioscience) Abs. IgA-bearing cells, representing the population that had undergone CSR, were sorted by a BD FACS Aria III cell sorter for DNA extraction.

Mice and CSR induction in vitro

C57BL/6J wild-type mice (6–8 wk old) were maintained in a specific opportunistic pathogen-free facility and immunized i.p. with 100 μg nuclear protein–keyhole limpet hemocyanin (Biosearch Technologies) or 50 μg nuclear protein–AECM–Ficoll (Biosearch Technologies) supplemented with CFA (Sigma-Aldrich). A boost was performed 3 wk later. After 7 d, animals were sacrificed and splenocytes were recovered. Splenocytes were stimulated in vitro to undergo CSR toward Igγ1 with LPS B4 (1 μg/ml; InvivoGen) and murine IL-4 (20 ng/ml; PeproTech) for 72 h. Experimental procedures were performed in accordance with the ethics committee review board of Limoges University (Limoges, France) and Hospital (Comité Régional d'Ethique sur l'Expérimentation Animale du Limousin, Centre Hospitalier Universitaire Dupuytren, Limoges, France).

Amplification of Sμ/Sα junctions from CH12F3 cells

Genomic DNA from sorted CH12F3 cells was extracted by the phenol/chloroform/isoamyl alcohol method. Sμ/Sα junctions were amplified by nested PCR using Phusion HF polymerase from BioLabs. The first PCR used the following primers: Smum1 (5′-CAG TTG AGG CCA GCA GGT-3′) and mCαCH1 reverse 2 (5′-TTC TTG GAC GGC GTT AGA GT-3′), and the following protocol: 94°C for 5 min; 35 cycles of 94°C for 1 min, 60°C for 1 min, and 72°C for 1:30 min; and 72°C for 7 min. The second PCR was performed with Smum2 (5′-CAG GTC GGC TGG ACT AAC TC-3′) and mCαCH1 reverse 1 (5′-ATG TTG CAC GGA ACA TTT CA-3′) under these conditions: 94°C for 5 min; 35 cycles of 94°C for 1 min, 62°C for 1 min, and 72°C for 1 min; and 72°C for 7 min. PCRs were done in triplicate.

Amplification of Sμ/Sγ1 junctions from mouse splenocytes

CSR junctions were amplified from murine splenocyte genomic DNA by nested PCR. First PCR primer pairs were as follows: Sμnest1 forward (5′-AGA GAC CTG CAG TTG AGG CC-3′) and sg1D reverse (5′-AGG ATG TCC ACC CTC ACC CAG GC-3′), and touchdown conditions as follows: 98°C for 30 s; 2 cycles of 98°C for 10 s, 69°C for 30 s, and 72°C for 90 s; 2 cycles of 98°C for 10 s, 67°C for 30 s, and 72°C for 90 s; 2 cycles of 98°C for 10 s, 65°C for 30 s, and 72°C for 90 s; and 25 cycles of 98°C for 10 s, 60°C for 30 s, 72°C for 90 s, and a final step at 72°C for 7 min. Primers for the second PCR were as follows: SμNest forward 2 (5′-CCA GCC ACA GTA ATG ACC CAG-3′) and DownSg1 reverse (5′-TTG TTA TCC CCC ATC CTG TCA CCT-3′), according to the following protocol: 98°C for 30 s, and 35 cycles of 98°C for 10 s, 65°C for 30 s, 72°C for 90 s, and 72°C for 7 min. PCRs were done in triplicate for each sample.

Amplification of Sμ/Sα junctions from human PBMCs

CSR junctions were amplified as previously (9). Amplified CSR junctions from DNA of PBMCs from four healthy individuals were equally mixed to prepare sequencing library. Study was approved by the ethics committee (*N°*: CPP12-012/2012-A00630-43, Comité de Protection des Personnes du Sud-Ouest et Outre-Mer IV).

Library preparation for ion proton sequencing

Libraries of 200-bp read length were prepared from 100 ng equally mixed triplicate PCR products as a starting material. A first enzymatic fragmentation was performed with Ion Xpress Plus Fragment Library kit (Life Technologies). Fragmented products were then linked to barcodes to identify each library. Ligated DNA fragments were migrated on E-Gel SizeSelect 2% agarose gels, and fragments of desired length were collected and amplified with cycler parameters, as follows: 95°C for 5 min, and 5 cycles of 95°C for 15 s, 58°C for 15 s, and 70°C for 1 min. Library concentrations were determined by Qubit DsDNA high sensitivity (Invitrogen), and an equimolar mixture was prepared. Libraries were run on an Ion PI v3 chip on the Ion Proton sequencer (Life Technologies).

Sequences are available on European Molecular Biology Laboratory–European Bioinformatics Institute European Nucleotide Archive (http://www.ebi.ac.uk/ena; study accession number: ERP019632).

Pipeline HTS data analysis

Raw sequencing data consisted of fastq files containing adapter-trimmed reads and were obtained from the sequencer's processing pipeline. The algorithm of CSReport aims at interpreting each read as an assembly of a segment from the donor S region Sμ and segment from an acceptor S region. First, these reads were used as query sequences in a Blast search (Blastn) (10) with ad hoc options (-dust no-, parameter adapted for repeated regions, and -culling_limit-1- restricting the junction mapping to segments with highest score). Reference sequences were slices from the murine C57BL/6 chromosome 12 sequence (gi|372099098 GRCm38.p3) and from the *IgH* locus sequence from *Homo sapiens* chromosome 14 reference (NC_000014.9). Only reads that map both regions, that is, a first hit in donor S region Sμ and a second hit in the target S region (user defined), with a minimum of 30 bp for each mapping, were kept for further analysis. A first junction structure evaluation was performed based on segmentation given by Blastn algorithm. The junction structure evaluation consists in determining the number of bases in the overlap between the 3′ end of the first segment and the 5′ end of the second segment. If this count is strictly positive, the junction is identified as microhomology. If no overlap is found, the junction is identified as either blunt or insertion if some bases are not mapped between the segments' ends. When a long (>10-bp) segment in the middle of a read was not mapped, a second Blast search was performed for the inserted segment (Blastn-short). This rescued some reads that may have originated from multiple breaks in S regions (complex junctions), as already reported (11, 12). When a microhomology (>2 bp) was found, realignment was performed to ensure that reported microhomologies did not contain any mismatches (as the first Blastn algorithm may have introduced mismatches to extend alignment): this

Downloaded from http://www.jimmunol.org/ by guest on April 10, 2018

alignment against each germline segment was performed using pairwise 2 from Biopython module (Python implementation of Needleman–Wunsch algorithm). Results of segmentation and junction structure for each filtered read were then reported in a tabular text file. In a subsequent step, these results were processed by statistical summarization procedures that may be selected by user at time of execution as they depend on the validated definition of an individual unique junction: by default, a junction is uniquely defined by one structure/size and two breakpoint positions, and all reads sharing these features are clustered into a single junction's identification. Clustering based on breakpoint proximity can also be used to cluster junctions with very close breakpoints. These steps resulted in a consolidated set of identified junctions described by structure and breakpoint positions. This set was then used to report the structural profile and breakpoint distributions. CSReport workflow is described on the diagram in Supplemental Fig. 1. CSReport as well as other postprocessing scripts are available upon request as a Python code. To use CSReport on other species, one just has to provide an *IgH* reference sequence and an annotation file for S region definition.

Results

Assessment of CSReport using a simulated data set of switch junctions

Use of simulated data sets is a common method to test sensitivity and resolution of computational tools (13) as they compare resulting outputs with expected outcomes. First, CSReport was used to analyze a computer-generated sequence file (Fasta) of 5000 randomly chosen concatenations of germline segments of Sμ and Sα (each 50–100 bp in length). By generation principle, all sequences are thus blunt junctions and each of their breakpoints is known. Fig. 1A shows structural profiles of such junctions in terms of either junctions with insertions (more simply insertions), blunt junctions, or junctions with microhomologies (more simply microhomologies). As expected, blunt junctions are the most frequent reported structure (~50% of sequences). Microhomologies are also found at significant frequencies. As already emphasized by previous studies (4), junction analysis is inherently biased to microhomologies as such structures can arise by chance, and this statistical effect (which can be analytically characterized for random sequences) is further increased when recombined regions share extensive homology. If breakpoints are located in short homologous regions, each alignment will extend further than the actual breaks and a microhomology will be found independently of a microhomology-mediated repair mechanism. Finally, it has to be stressed that about only two thirds of junctions (3278 of 5000) were reported by CSReport: a manual curation of some unidentified junctions showed that the homology between Sμ and Sα was again responsible for this loss of sensitivity as the initial Blastn search failed to properly address each segment to its correct S region.

Accuracy of junction analysis can also be evaluated using a simulated data set of Sμ-Sα sequences. As shown in Fig. 1B and 1C, segment assignment using CSReport was highly accurate as 90 and 97% of break positions were correctly identified within 10 bp in Sμ and Sα, respectively. Insets showed a slight dispersion of break positions toward longer alignments for the same reasons

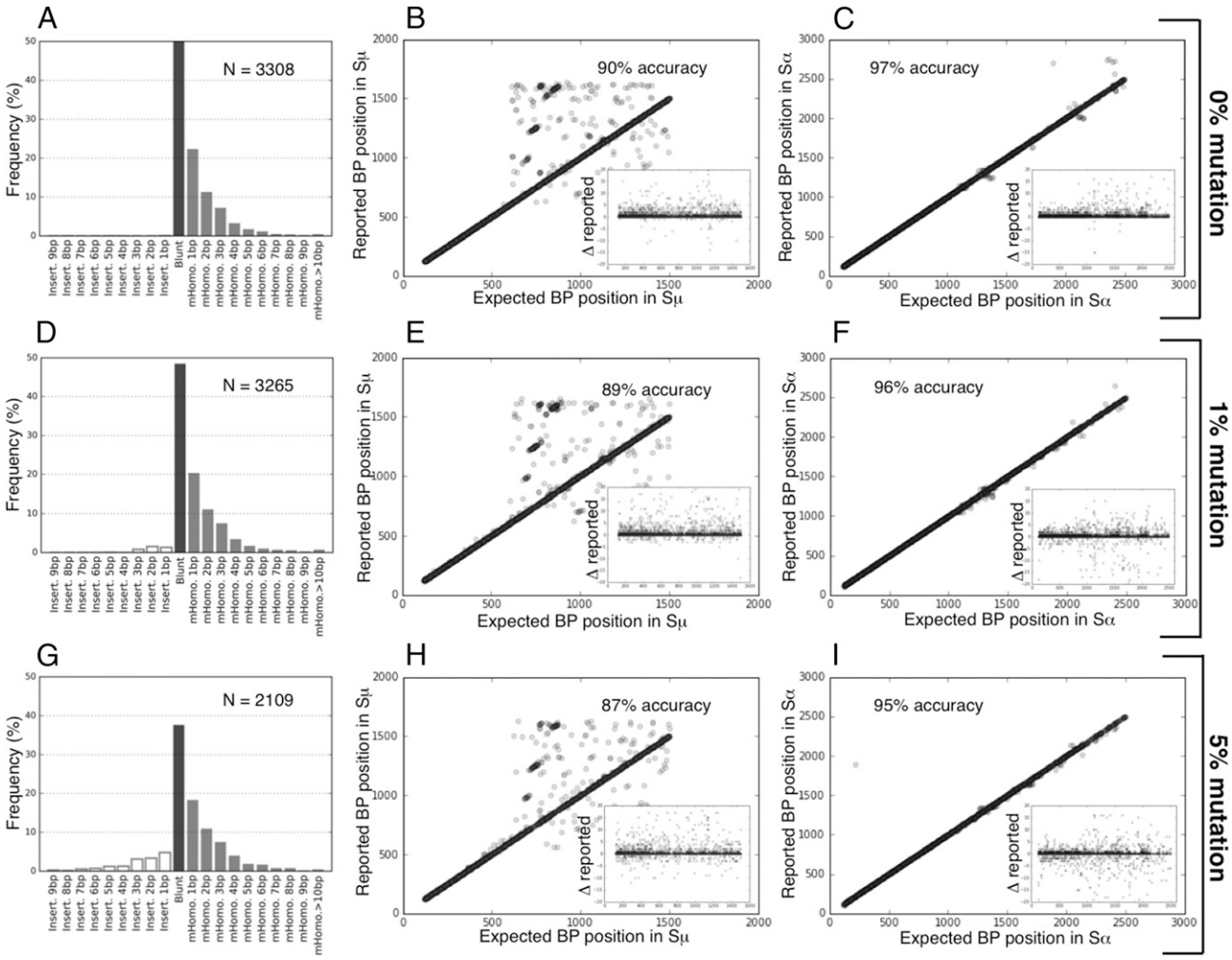


FIGURE 1. CSReport accurately identifies switch junctions in simulated Sμ-Sα sequences. Structural profiles and reported breakpoint positions are obtained from simulated Sμ-Sα blunt junctions, randomly generated from concatenation of Sμ and Sα segments: (**A**–**C**) without mutations, (**D**–**F**) with 1% random mutations, and (**G**–**I**) with 5% random mutations.

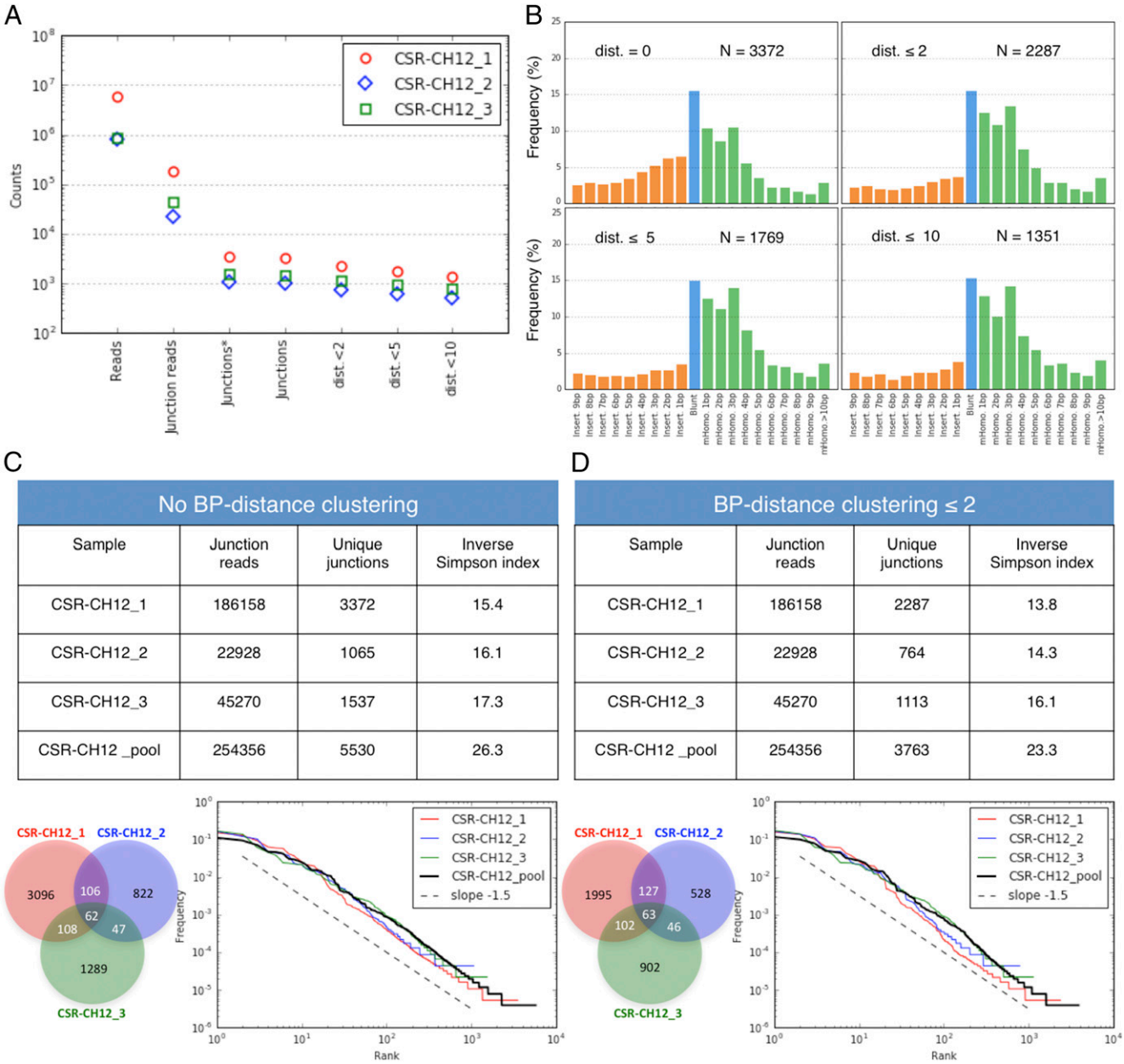
Downloaded from http://www.jimmunol.org/ by guest on April 10, 2018

as the observed levels of statistical microhomologies. Therefore, the accuracy of junction assignment by CSReport has to be evaluated with some tolerance in break positions (in this study, 10 bp) as this bias is only related to the extent of homology between both S regions and will be inherent to any alignment-based method.

As mutations frequently occur close to breakpoints of switch junctions, CSReport had to be tested in mutated sequences. Simulated mutated data sets were generated with a uniform substitution rate of 1 and 5% and analyzed with CSReport. Fig. 1 shows sensitivity and accuracy results for mutated junctions. As expected, the strongest effect is obtained with 5% mutations. Compared with unmutated sequences, fewer junctions were identified (2109 versus 3278). The structural profile (Fig. 1G) not only displays the previously discussed bias to microhomologies but also shows insertions at low frequencies. Breakpoint assignment accuracy was also slightly reduced but remained at high levels so that segments were correctly identified for most junctions (Fig. 1H, 1I). Nevertheless, 1% mutation does not seem to modify results compared with no mutations (Fig. 1D–F).

Statistics of Sμ-Sα CSR junctions in CH12F3 cells

CSReport was used to analyze switch junctions from in vitro CSR-induced CH12F3 cells. Results from three independent samples are reported in Fig. 2. To document bottlenecks in the pipeline and influence of filter parameters, counts of reads/junctions are shown in Fig. 2A. For each sample, filtration of reads that map to a valid



Sample	Junction reads	Unique junctions	Inverse Simpson index
CSR-CH12_1	186158	3372	15.4
CSR-CH12_2	22928	1065	16.1
CSR-CH12_3	45270	1537	17.3
CSR-CH12 _pool	254356	5530	26.3

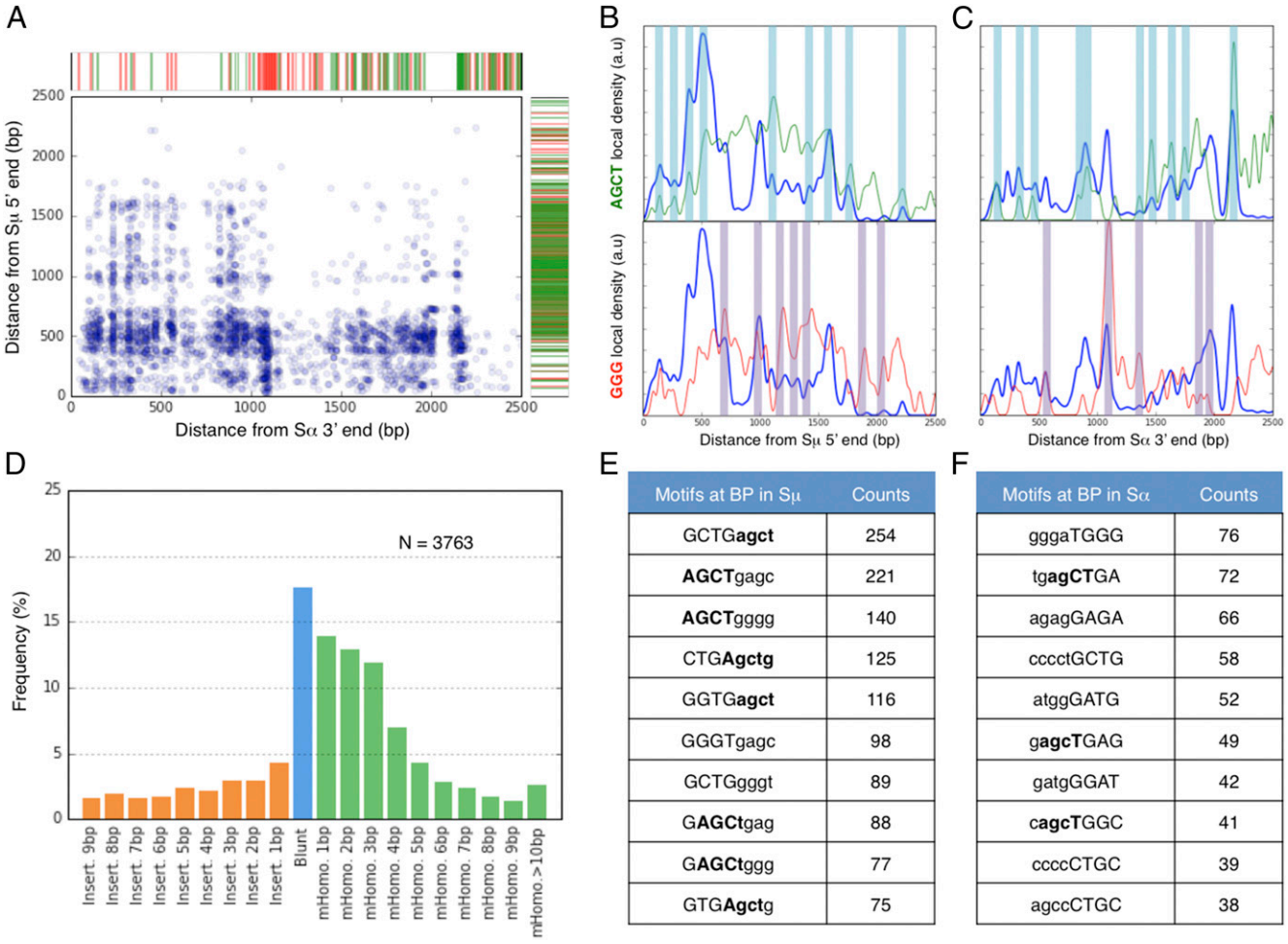
Sample	Junction reads	Unique junctions	Inverse Simpson index
CSR-CH12_1	186158	2287	13.8
CSR-CH12_2	22928	764	14.3
CSR-CH12_3	45270	1113	16.1
CSR-CH12 _pool	254356	3763	23.3

FIGURE 2. Quantitative assessment of switch junctions in CSR-induced CH12F3 cells. (**A**) Overview of junction counts for three independent samples of CH12F3 cells. Reads: sequences in raw fastq file (Ion Proton output). Junction reads: sequences that can be segmented in a strand-aware consistent switch junction. Junctions*: uniquely defined by two breakpoint positions and one structure. Junctions: same as before, and >10-bp insertions are excluded. dist < *x*: junctions with close breakpoint positions are counted only once. (**B**) Structural profile of sample CSR-CH12_1 for four different values of the breakpoint-distance clusterization parameter (similar trends were observed for the other samples). (**C** and **D**) Statistics of junction diversity across samples for junctions obtained without breakpoint-distance clusterization (C) or with clusterization of breakpoint distance ≤2. Venn diagram comparing the three independent samples. Rank–frequency curves of unique junctions are compared with a generalized Zipf law $f \approx k^{-1.5}$.

Downloaded from http://www.jimmunol.org/ by guest on April 10, 2018

CSR junction was highly selective as only 3–5% of input sequences were kept at this step. CSR quantification in a sample relied on determining the number of independent switch recombination events and therefore depended on the definition of a unique junction. A logical choice was to define a junction by its structure (insertion, blunt, or microhomology/size) and the positions of both breakpoints. When counting junctions, 2–5% of junction reads were found to be independent, indicating that most junctions were sequenced multiple times. A reasonable choice was to exclude junctions with large inserts (>10 bp) as they are only partially assembled and represent a small fraction of identified junctions. A further step of summarization was done by clustering identified junctions with very close breakpoints on both S regions. This was parameterized by the maximal distance in base pairs between break positions for two junctions to be clustered. From Fig. 2B, counts of independent junctions after clustering were seen to significantly depend on the distance parameter, whereas the structural profile did not depend on this parameter for non-null values. It thus indicates that distance clustering is efficient in identifying junctions that differ by only a few base pairs in breakpoint positions. A manual curation of these examples suggests this arises when an insertion is present right at the junction site. It is likely that this insertion is an unfortunate sequencing error (indels are by far the most frequent errors in the Ion Proton system), and these reads should not be considered as single junctions. Increasing the distance clustering parameter further than 2 bp would only reduce the number of independent junctions and seems an overcorrection of this effect as the structure profiles remain unchanged. Detailed investigation of clustering effects is shown in Fig. 2C and 2D in terms of junction counts and diversity. For all three samples, counts were indeed reduced after clustering, but it has to be noted from the Venn diagrams that only nonredundant junctions were significantly removed. Also, sample diversity was slightly reduced after clustering. All these results on three independent samples clearly demonstrated homogeneity (as shown on rank–frequency graphs) so that the pooled data set provided a deeper insight into CSR junctions under these conditions, with greater statistical power.

Summarization of 3763 independent junctions provided descriptions of Sμ-Sα CSR sequences at an unprecedented level. The distribution of breakpoints along S regions thus gave indications of motif targeting during the generation of DSB. Fig. 3 indicates colocalization of breakpoints, both in Sμ-Sα, with motifs AGCT or GGG. The density of breakpoints shares local maxima with the density of one of these two typically abundant switch motifs. The clear correlation observed in this study indicates that breaks will



E

Motifs at BP in Sμ	Counts
GCTG**agct**	254
AGCTgagc	221
AGCTgggg	140
CTG**Agctg**	125
GGTG**agct**	116
GGGTgagc	98
GCTGgggt	89
G**AGCT**gag	88
G**AGCT**ggg	77
GTG**Agct**g	75

F

Motifs at BP in Sα	Counts
gggaTGGG	76
tg**agCT**GA	72
agagGAGA	66
cccctGCTG	58
atggGATG	52
g**agcT**GAG	49
gatgGGAT	42
c**agcT**GGC	41
ccccCTGC	39
agccCTGC	38

FIGURE 3. CSReport provides valuable DNA break/repair information on large data sets. Structural and positional results obtained from analysis of Sμ-Sα junctions in CSR-induced CH12F3 cells (pool of three independent samples). (**A**) Positions of junction breakpoint expressed in terms of distance from S region ends. Each point in the graph is a unique junction defined by 2 bp positions (in Sμ and in Sα). Positions of palindromic AGCT motifs (green) and GGG motifs (red) are reported on right (Sμ) and top sides (Sα). (**B** and **C**) Estimation of local density of AGCT motifs (green), GGG motifs (red), and breakpoint along Sμ (B) and Sα (C) regions. (**D**) Structural profile of Sμ-Sα junctions. (**E** and **F**) Most frequent 8-nt motifs at breakpoint in Sμ (E) and Sα (F). Excised DNA at junctions is shown in lowercase; AGCT motifs are emphasized in bold.

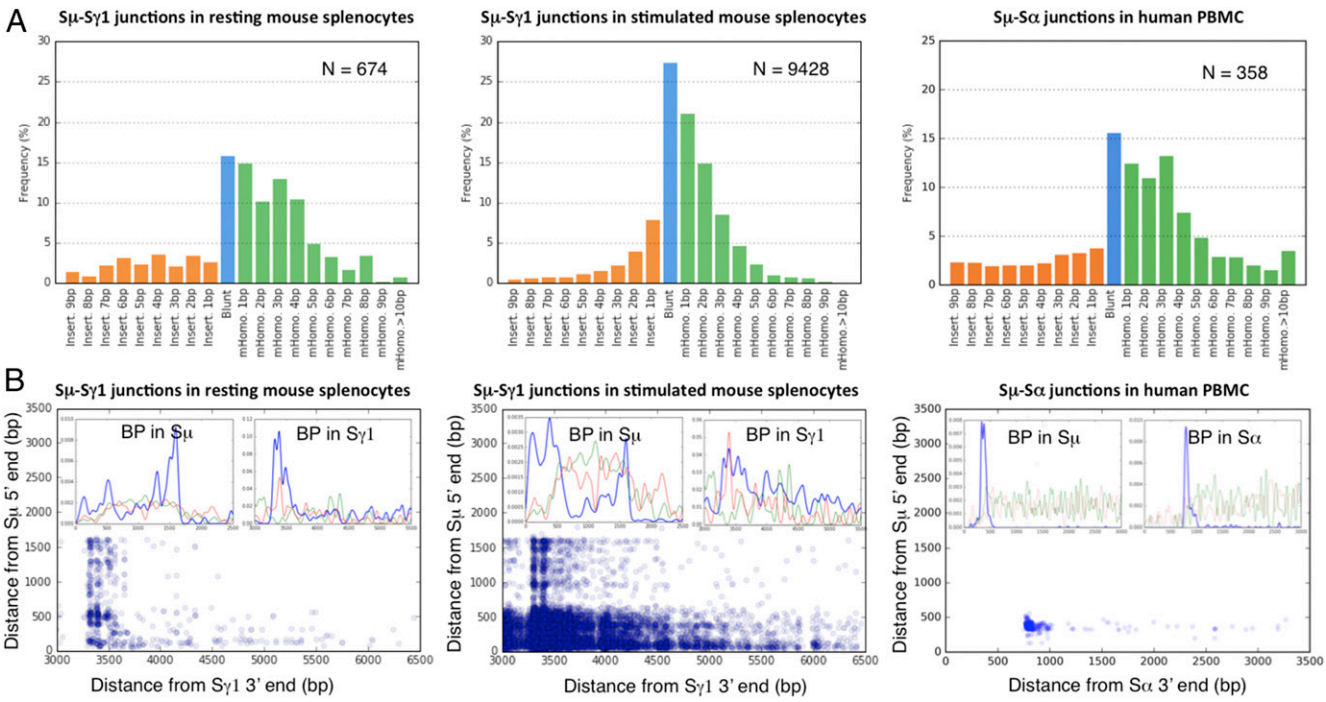
Downloaded from http://www.jimmunol.org/ by guest on April 10, 2018

be preferentially generated at local maxima in motif density, regardless of the average motif density of the subregion. As CSReport ensures 1-bp resolution, a more precise characterization of the break mechanism specificity can be obtained by looking at the most frequent nucleotide motif at breakpoints from both regions. Again, both AGCT and GGG were found to account for a large proportion of junction motifs, in accordance with the commonly accepted activation induced-cytidine deaminase–induced break mechanism. Finally, statistics about junction structures (Fig. 3D) can be obtained on a large number of independent junctions. Despite previously discussed bias toward insertions (due to mutations) and microhomologies (by chance, due to region homology), blunt junctions were found to be the most frequent structures. Microhomologies of 1–3 bp were also found at significant levels as the results of shared motifs between Sμ and Sα. This profile is therefore characteristic of NHEJ-resolved DSB. As already emphasized in a thorough review of CSR junction studies (4), detailed structural information is critical to infer DNA repair pathways. Sμ-Sα switch junctions are known to be noticeably biased to statistical microhomologies, but studies have been limited to a few manually curated junction sequences. In this work, our experimental HTS-based protocol and our computational analysis with CSReport give valuable detailed and accurate information on a large set of switch junctions for better elucidation of recombination mechanisms.

Statistics of CSR Sμ-Sγ1 junctions in resting versus stimulated mouse splenocytes

The comparative analysis of CSR Sμ-Sγ1 junctions in resting versus stimulated mouse splenocytes was performed using CSReport, and results are shown in Fig. 4. As expected, in vitro LPS plus IL-4 stimulation of splenocytes efficiently induced CSR as a greater number of independent Sμ-Sγ1 junctions was found (n = 9428) compared with those before stimulation (n = 674). Interestingly, the structural profile was also modified as blunt junctions were reported in a higher proportion of stimulated B cells (27.5%) than in resting B cells (15.9%). Whereas microhomology frequencies exhibited a smooth decrease as a function of microhomology size in stimulated cells, microhomologies of 1–4 bp were found approximately at the same amounts as blunt junctions in resting cells.

Quantitative and qualitative differences can also be documented in terms of breakpoint positions and distinctive motifs at junctions. Fig. 4B shows how possible breakpoint positions were used upon stimulation: breaks resulting in valid junctions span over 3 kb in Sγ1 in stimulated cells, whereas they appear more localized to the 3′ end of this region in resting splenocytes. In contrast, distribution of break positions in Sμ was similar in both conditions (spanning over the first 1.6 kb). Again, local density maxima of breakpoints in the S regions colocalized with local density maxima of palindromic AGCT and GGG motifs, as highlighted by the prevalence of such



C

Sμ-Sγ1 junctions in resting mouse splenocytes

BP-motifs in Sμ		BP-motifs in Sγ1	
G**AGCT**gag	43	tgcaGAGA	38
CTG**Agct**g	32	gagc**AGCT**	38
AGCTgagc	32	gtatGAAG	36
GCTG**agct**	23	g**agcT**GGA	32
GTG**Agctg**	20	accaGGCT	18

Sμ-Sγ1 junctions in stimulated mouse splenocytes

BP-Motifs in Sμ		BP-Motifs in Sγ1	
GCTG**agct**	275	gagc**AGCT**	109
AGCTgagc	194	**tagcT**GGT	72
GGTG**agct**	186	cagcCAGG	68
CTG**Agct**g	171	c**agcT**TAG	65
GGGTgagc	153	gctt**AGCT**	53

Sμ-Sα junctions in human PBMC

BP-Motifs in Sμ		BP-Motifs in Sα	
AGCTgagc	27	ctggGCTG	35
ACTGacct	17	tgggCTGG	26
GGGCaaaa	16	tggcCTGG	18
ACTGggct	15	ctg**aGCTG**	12
AGCTgacc	14	ggaTGGG	12

FIGURE 4. CSReport identifies switch junctions in mouse splenocytes and human PBMCs. (**A**) Structural profiles, (**B**) distribution of breakpoints, and (**C**) most frequent 8-nt motifs at breakpoint. Left and middle figures are obtained from Sμ-Sγ1 junctions, respectively, in resting and stimulated mouse splenocytes. Right figures are obtained from Sμ-Sα junctions in human PBMCs.

Downloaded from http://www.jimmunol.org/ by guest on April 10, 2018

patterns in the distinctive break sequences (Fig. 4C). Again, Sμ breaks exhibited more similarity in motif usage when stimulated cells were compared with resting cells.

Discussion

Sequence analysis of switch junctions from B cells reveals valuable information on the mechanism and effectiveness of CSR, and this is done almost routinely during studies of B cell functions in a variety of mouse models and clinical research. We aimed to study CSR junctions using a protocol based on HTS (Ion Torrent technology and Ion Proton sequencer). This technique challenges currently used methods already used due to the extremely large amount of collected data. To address the lack of automated tools for the analysis of large data sets of switch junction sequences, we designed CSReport, a software that ensures assessment of switch recombination junctions from Fastq or Fasta files. Typically, CSReport analyzes millions of reads that represent a few thousand single CSR junctions (Fig. 2A). In commonly used methods for CSR junction study, bias in junction mapping and junction structural characterization can arise from the chosen alignment algorithm as well as from the author interpretation. Blastn remains the more appropriate mapping software in regard to the BWA aligner, mostly used for HTS data analysis that we excluded because results obtained with BWA were very far from manual analysis. Blastn parameters were set up, but can be modified by the user, and we used the -no dust- option appropriate to the analysis of repeated sequences as S regions and the -culling limit- 1- returning the highest quality mapping result. This implies that the correct reference sequence is used, which is particularly important for different murine genetic backgrounds. To evaluate errors in junction analysis induced by CSReport, we subjected it to analysis of artificially designed blunt-ended Sμ-Sα junctions with known breakpoints (Fig. 1). Obtained results showed that CSReport interpretation induced a shift of blunt-ended junctions toward junctions using microhomologies (Fig. 1A) in the same way as a manual analysis, whatever the used aligner. In addition, mutations in proximity to the junction site induced false detection of junctions with insertions (Fig. 1D–G). These effects are most likely to be present in previous methods of junction analysis as well. As HTS data are more prone to sequencing errors, it appeared that clustering of unique junctions can lower the rate of abnormal detection of junctions with insertions (Fig. 2A, 2B). Clustering is controlled by the breakpoint-distance parameter (maximal distance of breakpoint of two individual junctions for them to be clustered). Based on our results, we recommend using CSReport with a breakpoint distance $\leq$2 bp. Increasing further this parameter will only result in fewer reported junctions. It is of note that this 2-bp clustering does not significantly affect diversity of the samples (Fig. 2C, 2D), as evaluated by Inverse Simpson Index, which can be only determined with large amounts of data. More detailed evaluation of sample clonality is provided by rank–frequency curves that also require substantial depth. After accuracy assessment of CSR junction identification by CSReport, we submitted results from HTS of five libraries made using amplified CSR junctions from DNA of CSR-stimulated CH12F3 cells (Fig. 3) and normal mouse splenocytes before and after 3 d of IL-4 plus LPS in vitro stimulation (Fig. 4). CSReport permits the automatic determination of breakpoint positions along the reference sequences, structural features of CSR junctions, and the most frequent DNA motifs at the junction breakpoints. As expected, Sμ-Sγ1 junction count was greatly increased after splenocyte stimulation, and CSReport revealed structural features for all tested samples with 25–50% of blunt-end junctions, whereas junctions with microhomologies longer than 4 bp were quite rare, according to results in the literature (4).

Finally, we propose CSReport as an automated computational tool able to support analysis of large data sets from CSR junction HTS. Widespread use of CSReport will normalize studies analyzing switch junctions as it provides a single pipeline providing breakpoint positions and junction structure (blunt-end junctions or junctions with insertion or microhomology). CSReport conforms to the definition of microhomology as perfectly matched nucleotides shared by donor and acceptor junction segments without any imperfect homology or mismatch, as previously described (14). CSReport also includes specific identification of complex sequential CSR (Sμ-Sx-Sy junctions) by realigning long insertions ($>$10 bp) that may not have been mapped during the first alignment step. In addition, the use of CSReport will reduce the bias in the delineation of CSR junction characteristics due to operator interpretation. CSReport is able to support Fastq and Fasta files and is meant to process data sets regardless of the sequencing technology used to produce them. It is of note that read size is not a limitation criterion, as in Ig or TCR repertoire analyses (15, 16), which need ~500-bp length sequencing, because CSReport only keeps for analysis those reads that map on donor and acceptor regions for $>$30 bp. Nevertheless, large fragments could certainly improve capture of junctions containing reads.

Another interesting point is that CSReport permits CSR junction study regardless of the species, as demonstrated by Sμ-Sα1 junction analysis done on human PBMCs (Fig. 4). Finally, in addition to the count and characterization of CSR junctions, CSReport presents results in the form of clear and easily manipulated tabular files of valuable information. CSReport facilitates and accelerates the study of CSR junctions. It also harmonizes analysis and interpretation.

Acknowledgments

We thank the GénoLim platform and the Cytométrie, Imagerie et Microscopie platform (Génomique Environnement Immunité Santé et Thérapeutique Institut, Université de Limoges, Centre de Biologie et de Recherche en Santé, Centre Hospitalier Universitaire de Limoges) for sequencing and cell sorting, and S. Desforges, B. Remerand, and D. Champagnol for help with animal care.

Disclosures

The authors have no financial conflicts of interest.

References

1. Muramatsu, M., K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai, and T. Honjo. 2000. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 102: 553–563.
2. Teng, G., and F. N. Papavasiliou. 2007. Immunoglobulin somatic hypermutation. *Annu. Rev. Genet.* 41: 107–120.
3. Revy, P., T. Muto, Y. Levy, F. Geissmann, A. Plebani, O. Sanal, N. Catalan, M. Forveille, R. Dufourcq-Labelouse, A. Gennery, et al. 2000. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the hyper-IgM syndrome (HIGM2). *Cell* 102: 565–575.
4. Stavnezer, J., A. Björkman, L. Du, A. Cagigi, and Q. Pan-Hammarström. 2010. Mapping of switch recombination junctions, a tool for studying DNA repair pathways during immunoglobulin class switching. *Adv. Immunol.* 108: 45–109.
5. Soulas-Sprauel, P., G. Le Guyader, P. Rivera-Munoz, V. Abramowski, C. Olivier-Martin, C. Goujet-Zalc, P. Charneau, and J.-P. de Villartay. 2007. Role for DNA repair factor XRCC4 in immunoglobulin class switch recombination. *J. Exp. Med.* 204: 1717–1727.
6. Ehrenstein, M. R., and M. S. Neuberger. 1999. Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. *EMBO J.* 18: 3484–3490.
7. Pan, Q., H. Rabbani, F. C. Mills, E. Severinson, and L. Hammarström. 1997. Allotype-associated variation in the human γ3 switch region as a basis for differences in IgG3 production. *J. Immunol.* 158: 5849–5859.
8. Nakamura, M., S. Kondo, M. Sugai, M. Nazarea, S. Imamura, and T. Honjo. 1996. High frequency class switching of an IgM+ B lymphoma clone CH12F3 to IgA+ cells. *Int. Immunol.* 8: 193–201.

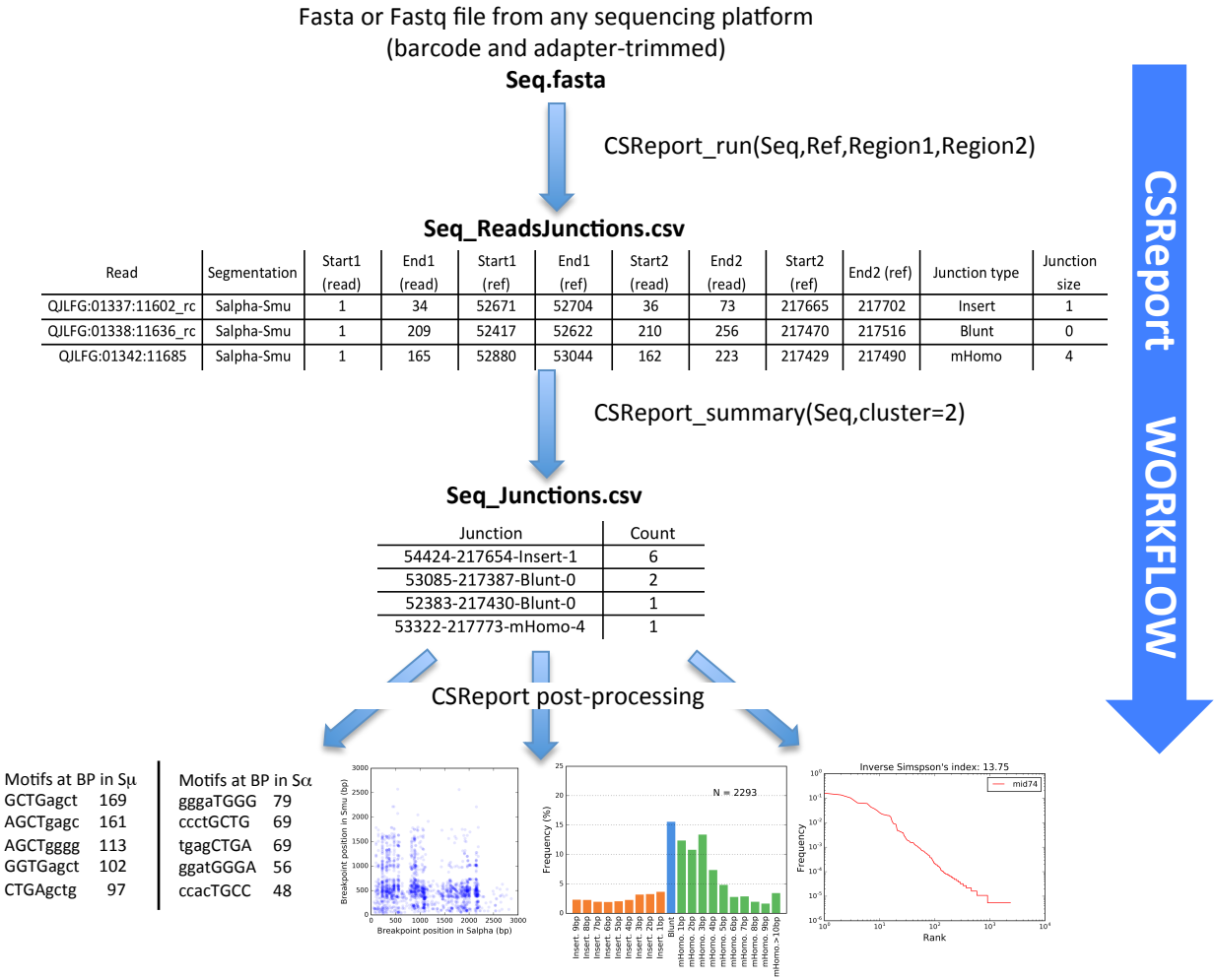
Downloaded from http://www.jimmunol.org/ by guest on April 10, 2018

9. Pan, Q., C. Petit-Frére, S. Dai, P. Huang, H. C. Morton, P. Brandtzaeg, and L. Hammarström. 2001. Regulation of switching and production of IgA in human B cells in donors with duplicated alpha1 genes. *Eur. J. Immunol.* 31: 3622–3630.
10. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215: 403–410.
11. Pan-Hammarström, Q., A. Lähdesmäki, Y. Zhao, L. Du, Z. Zhao, S. Wen, V. L. Ruiz-Perez, D. K. Dunn-Walters, J. A. Goodship, and L. Hammarström. 2006. Disparate roles of ATR and ATM in immunoglobulin class switch recombination and somatic hypermutation. [Published erratum appears in 2006 *J. Exp. Med.* 203: 251.] *J. Exp. Med.* 203: 99–110.
12. Reina-San-Martin, B., J. Chen, A. Nussenzweig, and M. C. Nussenzweig. 2007. Enhanced intra-switch region recombination during immunoglobulin class switch recombination in 53BP1$^{-/-}$ B cells. *Eur. J. Immunol.* 37: 235–239.
13. Safonova, Y., A. Lapidus, and J. Lill. 2015. IgSimulator: a versatile immunosequencing simulator. *Bioinformatics* 31: 3213–3215.
14. Pan, Q., C. Petit-Frére, A. Lähdesmäki, H. Gregorek, K. H. Chrzanowska, and L. Hammarström. 2002. Alternative end joining during switch recombination in patients with ataxia-telangiectasia. *Eur. J. Immunol.* 32: 1300–1308.
15. IJspeert, H., P. A. van Schouwenburg, D. van Zessen, I. Pico-Knijnenburg, A. P. Stubbs, and M. van der Burg. 2017. Antigen Receptor Galaxy: a user-friendly, web-based tool for analysis and visualization of T and B cell receptor repertoire data. *J. Immunol.* 198: 4156–4165.
16. Langerak, A. W., M. Brüggemann, F. Davi, N. Darzentas, D. Gonzalez, G. Cazzaniga, V. Giudicelli, M.-P. Lefranc, M. Giraud, E. A. Macintyre, et al. 2017. High-throughput immunogenetics for clinical and research applications in immunohematology: potential and challenges. *J. Immunol.* 198: 3765–3774.

Downloaded from http://www.jimmunol.org/ by guest on April 10, 2018

Figure S1

A



Read	Segmentation	Start1 (read)	End1 (read)	Start1 (ref)	End1 (ref)	Start2 (read)	End2 (read)	Start2 (ref)	End2 (ref)	Junction type	Junction size
QJLFG:01337:11602_rc	Salpha-Smu	1	34	52671	52704	36	73	217665	217702	Insert	1
QJLFG:01338:11636_rc	Salpha-Smu	1	209	52417	52622	210	256	217470	217516	Blunt	0
QJLFG:01342:11685	Salpha-Smu	1	165	52880	53044	162	223	217429	217490	mHomo	4

Junction	Count
54424-217654-Insert-1	6
53085-217387-Blunt-0	2
52383-217430-Blunt-0	1
53322-217773-mHomo-4	1

B

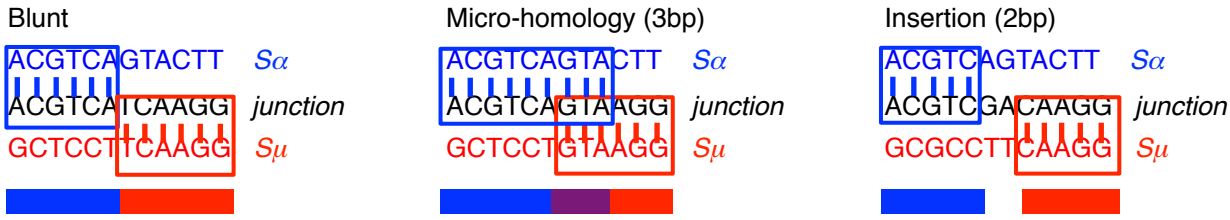


Figure S1. CSReport identifies swicth junctions from raw sequencing reads and provides graphical results for direct interpretation. (A) Workflow diagram of CSReport. Input files are sequencing reads in fasta or fastq format. The first command (CSReport_run) calls Blastn and assembles junctions from realignment of segments'ends (individual structure and BP). The second command (CSReport_summary) identifies single junctions from the junction reads, accordingly to the user-defined parameter of distance clustering. Other CSReport functions produces graphical output to represent structural profile, BP distribution, junctional diversity or involved nucleotidic motifs. (B) Definition of blunt junction (left), junction with micro-homology (middle), or junction with insertion (right). CSReport conforms to those definitions

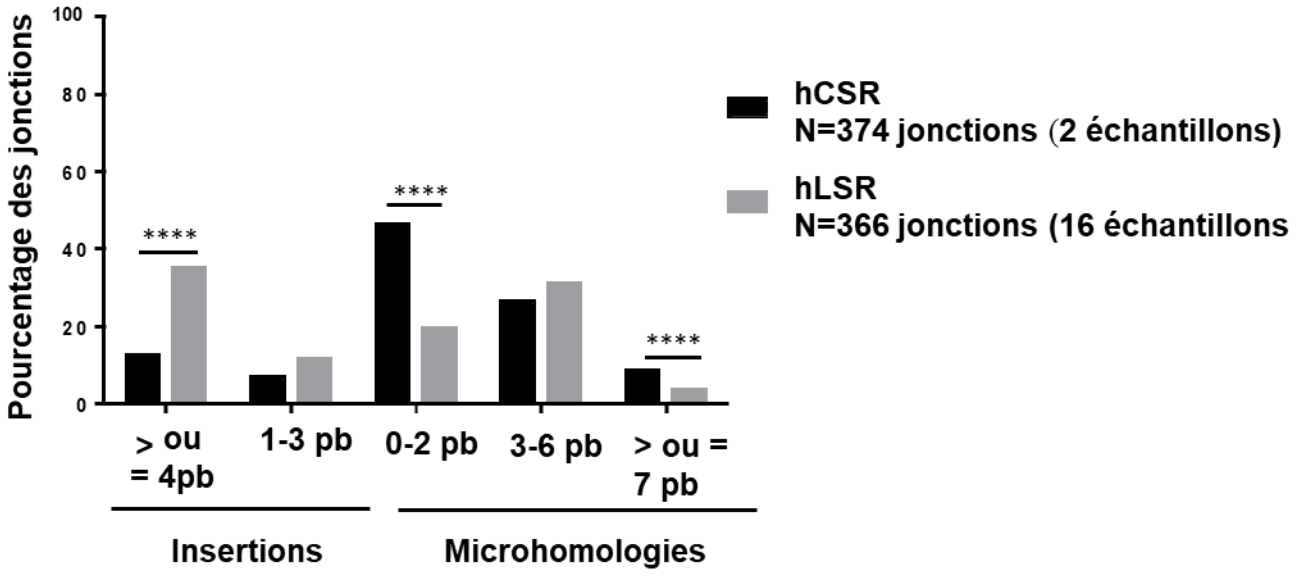


Figure 25 : Signature de réparation des CDB au cours de la LSR et la CSR chez l'Homme. La signature de réparation est différente entre les jonctions de recombinaison LSR et CSRα. L'histogramme représente le pourcentage de jonctions en fonction de leur structure (jonctions avec insertions, blunt ou de 0 à 2 pb de microhomologie, et jonctions avec des microhomologies de taille plus importante). Chi Square Test : **** P<0,001.

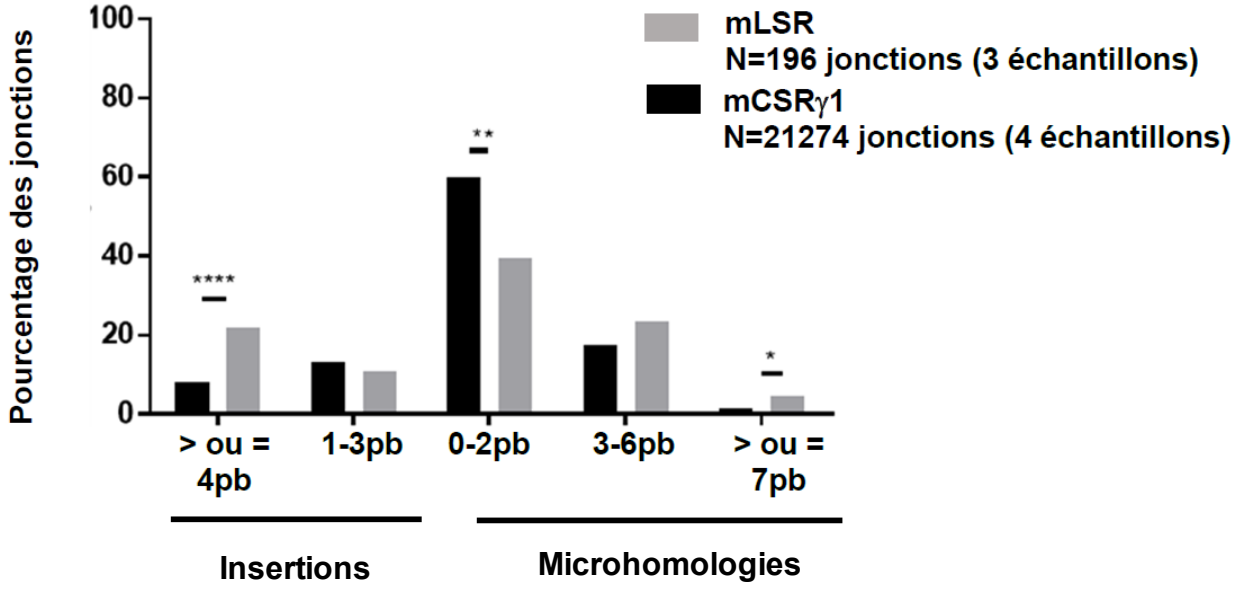


Figure 26 : Signature de réparation des CDB au cours de la LSR et la CSR chez la souris. La signature de réparation est différente entre les jonctions LSR et CSRγ1. L'histogramme représente le pourcentage de jonctions en fonction de leur structure (jonctions avec insertions, blunt ou de 0 à 2 pb de microhomologie, et jonctions avec des microhomologies de taille plus importante). Chi Square Test : * P<0.5, ** P<0.01, **** P <0,001.

Partie II : Signature de réparation des CDB au cours de la LSR et de la CSR : ces deux recombinaisons du locus *IgH* impliquent des voies de réparation différentes !

Depuis la découverte de la recombinaison LSR par notre laboratoire (Péron et al., 2012), il n'y a pas eu d'autres travaux publiés sur ce processus. En effet, cet évènement furtif qui induit la mort de la cellule B concernée rend l'étude complexe. Toutefois, la mise en place de notre protocole d'étude des jonctions CSR et LSR par séquençage haut débit, puis leur analyse par CSReport rend l'étude moins complexe.

Comme mentionné précédemment, la LSR est une recombinaison intrachromosomique du locus *IgH* qui s'opère dans les LB activés et qui est dépendante de l'activité de AID. La LSR a lieu entre une région Sμ et la 3'RR, et empreinte le même schéma d'étapes que la recombinaison CSR : 1) transcription des régions cibles, 2) lésion de l'ADN par AID générant des CDB, 3) réparation de l'ADN. La dernière étape de réparation de l'ADN au cours de la CSR est bien documentée dans la littérature. Il a été montré que le C-NHEJ est la voie de réparation impliquée de façon majoritaire pendant la CSR, et que l'A-EJ peut intervenir de façon moins importante particulièrement lorsque le C-NHEJ est défaillant (Soulas-Sprauel et al., 2007; Yan et al., 2007; Han and Yu, 2008)

Pour savoir si la réparation des CDB de la LSR est similaire à celle de la CSR, nous avons analysé la structure des jonctions LSR et CSR. Effectivement, l'analyse de la signature de réparation d'une recombinaison est indicative du système de réparation employé, et pour ce faire nous avons analysé ces jonctions par CSReport.

1. Analyse de la structure de réparation au cours de la LSR

Nous nous sommes intéressés à la réparation des CDB pendant la LSR chez la souris et chez l'Homme, afin de savoir si les deux ont recours aux mêmes voies ou si les cassures sont réparées différemment. Nous avons donc procédé à l'analyse de jonctions LSR en gris et CSRα en noir en fonction de leur structure **(Figure 25)**.

Nous observons un profil de structure qui diffère entre les jonctions LSR et CSRα humaines. Concernant la LSR, le pourcentage de jonctions franches et avec de courtes microhomologies (0 à 2 pb) est réduit de moitié par rapport à la CSR (de 46,3%

à 19,3%). Une diminution importante est également notée pour les recombinaisons LSR avec des microhomologies supérieures ou égales à 7 pb de 8,3% à 3,4%. A l'inverse, l'usage des insertions supérieures ou égales à 4 pb semble très augmenté dans la LSR de 12,3% à 35%. Les résultats chez l'Homme montrent que la signature de réparation de la LSR diffère de celle de la CSR. L'usage de microhomologie de taille modérée dans les jonctions LSR est inchangée par rapport à la CSR, et ce type de jonctions avec les jonctions comportant des insertions supérieures ou égales à 4 pb sont les jonctions les plus fréquentes dans la LSR. En revanche, ce sont les jonctions franches et avec de courtes microhomologies qui ressortent dans la CSR.

Nous avons également réalisé l'analyse par CSReport des jonctions CSRγ1 et LSR issues de LB de souris. Les résultats sont représentés dans la figure 26.

Dans un premier temps, nous retrouvons une répartition des recombinaisons CSRγ1 en accord avec la littérature, c'est-à-dire, une utilisation importante des jonctions franches et avec de courtes microhomologies (0 à 2 pb) qui atteint 59,84%, et qui indique une réparation majoritaire par la voie du C-NHEJ. Concernant la LSR, il y a un pourcentage moins important (39,29%) des jonctions franches et avec de courtes microhomologies de 0 à 2 pb. En revanche, il y a une augmentation de l'usage de longues microhomologies supérieures ou égales à 7 pb, de 1,57% à 4,59% dans le cas des jonctions LSR, ainsi que les insertions supérieures ou égales à 4 pb, de 8,18% à 21,94%. Ces résultats montrent des profils de structure différents entre les jonctions LSR et CSRγ1 murines. Cette variation de structure chez la souris est très similaire à celle observée précédemment chez l'Homme, et nous renseigne sur le fait que l'Homme et la souris semblent faire appel aux mêmes voies de réparation des CDB au cours des évènements de LSR.

Il est bien documenté que La CSR fait intervenir la voie du C-NHEJ de façon majoritaire, l'A-EJ participe également à un degré moindre, et surtout en cas de défaillance du C-NHEJ (Soulas-Sprauel et al., 2007; Yan et al., 2007; Han and Yu, 2008). L'A-EJ génère des jonctions contenant des microhomologies de séquence ce qui est évocatif des résultats de séquençage haut débit des jonctions LSR.

Pour déterminer si la différence de signature de réparation de la LSR par rapport à celle de la CSR est dû à un recrutement différent des voies de réparation de l'ADN, nous avons cherché à analyser le recrutement de XRCC4 (C-NHEJ), PARP1 et Pol θ

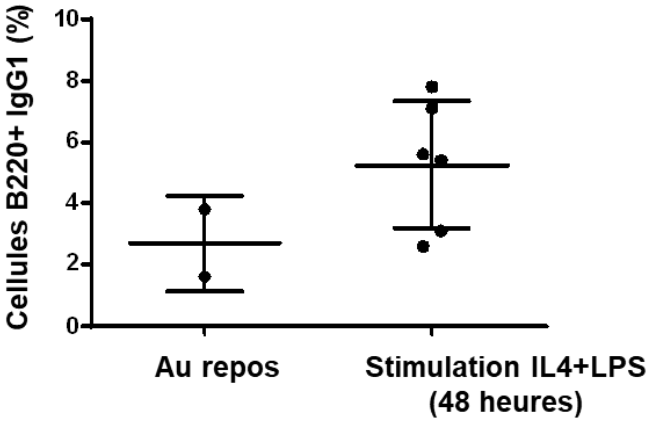


Figure 27 : Analyse du taux de CSR vers l'IgG1. Les LB B220+/IgG1+ ont été quantifiés par cytométrie en flux dans les LB matures au repos (N=2), et dans les LB activés *in vitro* pendant 48h (N=6) avec du LPS et IL4. La stimulation induit une augmentation de l'expression de l'IgG1 à la surface des LB.

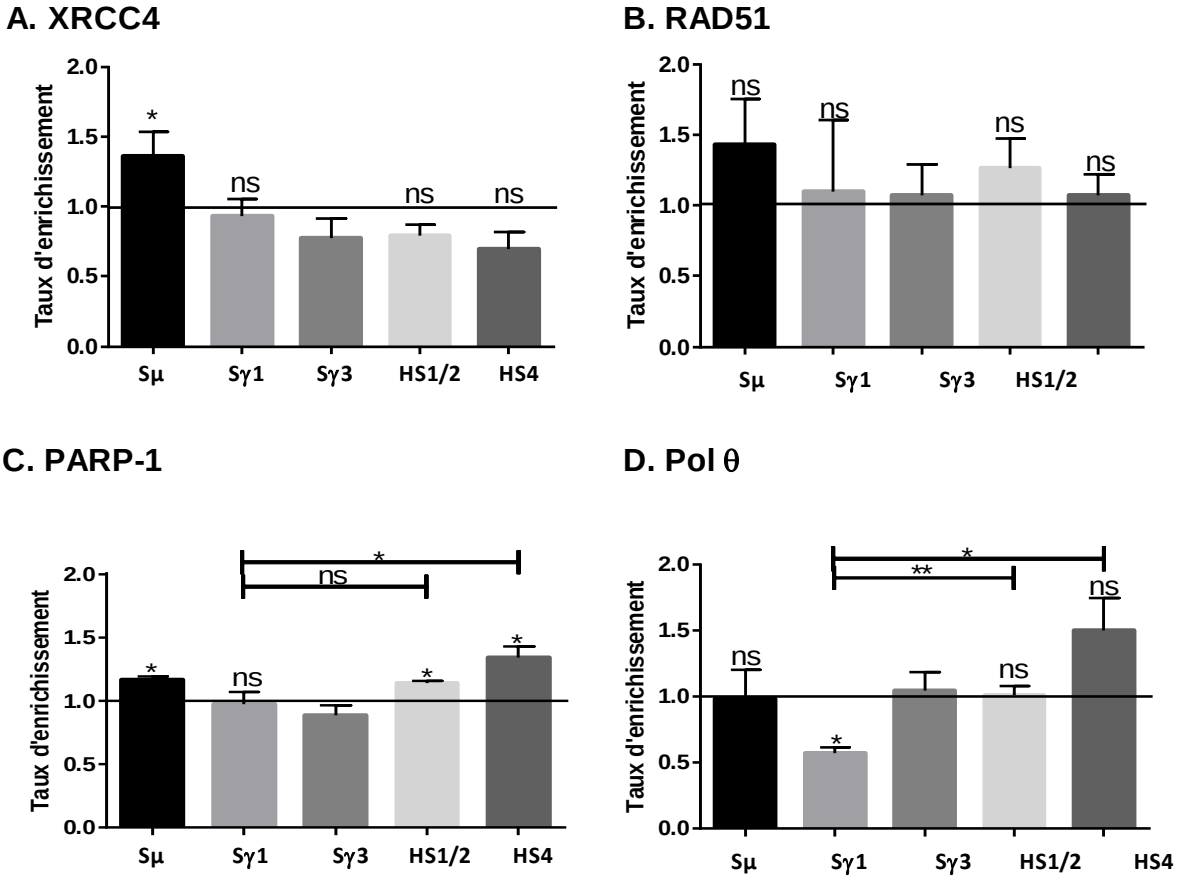


Figure 28 : Recrutement du C-NHEJ, de la HR et de l'A-EJ au locus *IgH* de LB activés. Taux d'enrichissement des protéines XRCC4 (N ≥5) (A), RAD51 (N ≥5) (B), PARP-1 (N≥3) (C) ou Polθ (N≥3) (D) au niveau des régions Sμ, Sγ1, Sγ3, HS1/2 et HS4. Les histogrammes représentent la moyenne des expériences +/- SEM. Les taux d'enrichissement sont comparés à celui du contrôle négatif Sγ3, et comparés entre les différentes régions analysées. XRCC4 et PARP1 sont recrutés au niveau de Sμ. PARP-1 est également recruté au niveau de HS1/2 et HS4. Nous ne détectons aucun enrichissement de RAD51 ou Polθ sur le locus *IgH.* Test statistique unpaired T test, ns : non significatif, * : P<0.05, ** : P<0.01.

(A-EJ), RAD51 (HR) sur les segments du locus *IgH* ciblés par la CSR et la LSR. Etant donné que les résultats de structure sont similaires entre l'Homme et la souris, les expériences de ChIP ont été réalisées uniquement chez la souris.

2. Recrutement des molécules de réparation aux sites de CSR et LSR

Nous nous sommes intéressés au recrutement des protéines impliquées dans les différentes voies de réparation des CDB. Des expériences de ChIP des protéines citées précédemment ont été réalisées à partir de la chromatine de LB murins issus de 3 souris WT poolés et stimulés *in vitro* pendant 48 heures par du LPS et de l'IL-4 orientant la CSR vers l'isotype Igγ1. L'efficacité de la stimulation des LB à 48 heures de culture a été vérifiée par le marquage des cellules B $IgG1^+$ **(Figure 27)**.

Nous observons une augmentation du pourcentage de cellules B exprimant l'IgG1 d'environ 3% au repos à 6% après stimulation.

L'analyse par ChIP a été réalisée au niveau de régions ciblées par la CSRγ1 qui sont Sμ et Sγ1, et les régions concernées par la LSR, HS1/2 et HS4 **(Figure 28)**. La région Sγ3 qui n'est pas concernée par les recombinaisons CSR et LSR est utilisée comme contrôle négatif.

Les valeurs d'enrichissement sont rapportées aux valeurs d'input obtenues sans anticorps (Mock), et comparées au contrôle Sγ3, qui n'est pas ciblé par des CDB dans les conditions utilisées.

De façon attendue, aucune protéine ne semble enrichie au niveau du segment Sγ3. En revanche, nous observons un taux d'enrichissement augmenté des protéines XRCC4 et PARP-1 au niveau de la région Sμ par rapport à la région Sγ3. De façon étonnante nous n'avons pas noté d'enrichissement de XRCC4 au niveau de Sγ1. PARP-1 est retrouvée aussi associée à HS1/2 et HS4. Le recrutement de Polθ au niveau de HS4 par rapport à Sγ3 n'est pas statistiquement significatif, mais nous observons que son association aux segments HS1/2 et HS4 est augmentée comparé à Sγ1.

Finalement, les résultats indiquent que dans les LB activés 48 heures (LPS+ IL-4), les protéines XRCC4 et PARP-1 sont recrutées au niveau de la région Sμ qui est le segment donneur des recombinaisons LSR et CSR, et que PARP-1, et peut-être

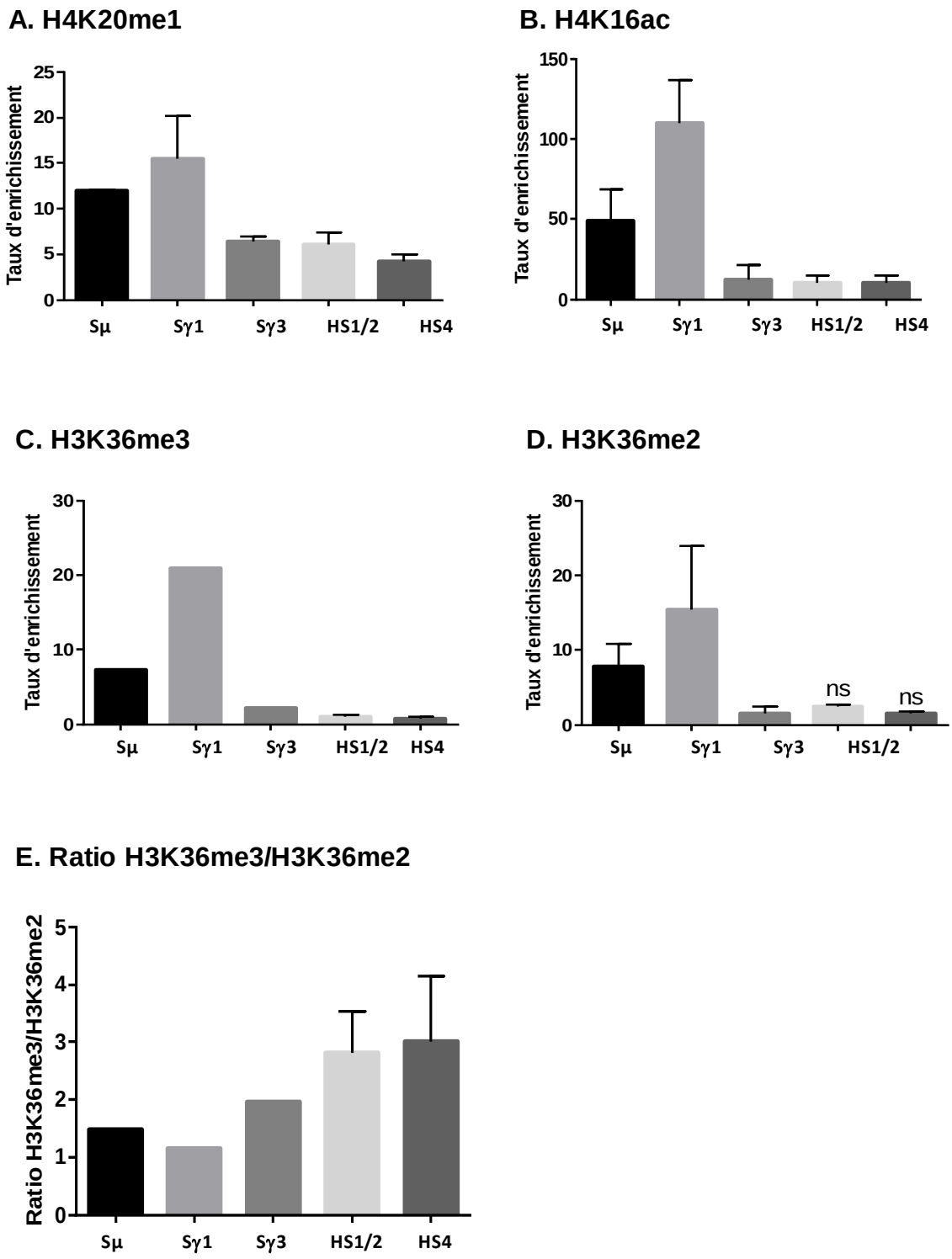


Figure 29 : Enrichissement des marques d'histones H4K20me1, H4K16ac, H3K36me2 et H3K36me3 au locus *IgH* de LB activés. Taux d'enrichissement des histones H4K20me1 (N=2) (A), H4K16ac (N=2) (B), H3K36me3 (N=2) (C), H3K36me2 (N=2) (D), et le ratio H3K36me3/me2 (E) au niveau des régions Sμ, Sγ1, Sγ3, HS1/2 et HS4. Les histogrammes représentent la moyenne des expériences +/- SEM. Les taux d'enrichissement sont comparés à celui du contrôle négatif Sγ3. Les marques épigénétiques sont enrichies au niveau de Sμ et Sγ1, et le ratio H3K26me3/me2 est augmenté au niveau des régions HS1/2 et HS4. Test statistique unpaired T test, ns : non significatif.

Polθ, sont recrutées au niveau de la 3'RR acceptrice de la LSR. Ces résultats sont en accord avec la littérature selon laquelle XRCC4 est impliquée dans le C-NHEJ au cours de la CSR (Soulas-Sprauel et al., 2007). L'absence de XRCC4 et la présence de PARP-1, et peut-être de Polθ, au niveau de la 3'RR suggère que la réparation des CDB dans cette région peut être prise en charge par l'A-EJ plutôt que par le C-NHEJ. Ceci explique les résultats de séquençage haut débit que nous avons obtenu au sujet de la structure de réparation des jonctions LSR.

Afin de mieux comprendre nos résultats, nous nous sommes interrogés sur l'étude du contexte chromatinien qui est connu pour influencer la réparation des CDB, au niveau des segments concernés par les recombinaisons CSR et LSR.

3. Analyse du contexte chromatinien des segments mis en jeu lors des recombinaisons CSR et LSR

Les modifications des histones participent à l'orientation du recrutement des acteurs de la réparation de l'ADN. En effet, les marques épigénétiques influencent le recrutement d'un système de réparation de l'ADN en particulier plutôt qu'un autre (Clouaire and Legube, 2015). Pour déterminer si la différence de signature de réparation observée entre la LSR et la CSR peut être associée à des différences de chromatine au niveau des régions concernées par ces recombinaisons, nous avons choisi d'analyser par ChIP des marques d'histones qui ont été impliquées dans l'orientation de la réparation des CDB.

Nous avons analysé le taux d'enrichissement de marques épigénétiques particulières qui sont l'H4K20me1 et l'H4K16ac, l'H3K36me3, l'H3K36me2, ainsi que le ratio H3K36me3/H3K36me2, dans des régions concernées par la LSR Sμ, HS1/2 et HS4, et celles ciblées par la CSR Sμ et Sγ1 **(Figure 29)**.

La région Sγ3 n'étant pas concernée par la recombinaison est utilisée comme contrôle négatif. Les résultats présentés ci-dessous sont préliminaires et doivent être renforcés par des expériences indépendantes supplémentaires. Cependant, nous pensons que certaines tendances qui se dégagent de ses résultats sont d'intérêt dans le questionnement de l'orientation vers des voies de réparation des CDB différentes au cours de la CSR et de la LSR.

Les résultats de ChIP montrent un enrichissement important de l'H4K20me1 **(Figure 29.A)** et de l'H4K16ac **(Figure 29.B)** au niveau des régions Sμ et Sγ1 dans les LB activés. Comme cité précédemment, H4K20me1 est une marque qui participe au recrutement de 53BP1 au niveau des CDB, et peut ainsi favoriser le C-NHEJ dans la réparation de la jonction Sμ/Sγ1, ce qui est en accord avec la littérature (Han and Yu, 2008; Soulas-Sprauel et al., 2007; Yan et al., 2007). L'H4K16ac est associée aux segments d'ADN transcriptionnellement actifs et est décrite comme inhibitrice de la voie du C-NHEJ. Les associations de l'H4K20me1 et de l'H4K16ac au niveau des régions HS1/2 et HS4 de la 3'RR sont bien inférieures à celles observées au niveau des régions Sμ et Sγ1. L'immunoprécipitation des marques tri-méthylées et di-méthylées de l'H3K36 sont elles aussi particulièrement détectées au niveau de Sμ et Sγ1 par rapport à HS1/2 et HS4 **(Figure 29.C et D)**. De façon intéressante, le ratio H3K36me3/H3K36me2 est plus augmenté dans les régions acceptrices de la LSR **(Figure 29.E)**, ce qui indique une dominance de la forme H3K36me3 qui est connue pour favoriser le recrutement de la voie HR, par rapport à la forme H3K36me2, qui favorise le C-NHEJ.

Les résultats préliminaires de ChIP des marques d'histone étudiées suggèrent que le contexte chromatinien diffère entre les régions du locus *IgH* impliquées dans la CSR et la LSR.

Partie III : La LSR dans le lymphome de Hodgkin

Nous avons focalisé nos efforts sur le HL, afin de savoir si la LSR est associée à la perte d'expression du BCR à la surface des cellules de Reed Sternberg. Il a été montré que ces cellules tumorales ont perdu l'expression des Ig suite à des mutations non-sens dans les gènes qui codent pour la partie variable de l'IgH (Marafioti et al., 1997). Cependant, nous avons souhaité savoir si la LSR peut aussi y participer. Nous avons donc choisi des échantillons d'amygdales saines (absence de lymphome) et de CLL (lymphome non hodgkinien) comme contrôles. Ce travail a été réalisé en collaboration avec le Pr. Jean Feuillard et son équipe.

1.1. Quantification des jonctions LSR par séquençage haut débit

Nous avons obtenu à l'aide du Centre de Ressources Biologiques du CHU de Limoges (CRBioLim) des coupes de tissus congelés de témoins sains (amygdales, N= 16), et de ganglions de patients atteints de HL (N= 19), ainsi que de l'ADNg de PBMC de patients atteints de CLL (N= 19). Nous avons extrait l'ADNg des tissus et pour chaque échantillon, nous avons amplifié les jonctions LSR, pour ensuite procéder à la préparation des librairies de séquençage haut débit. Les librairies ont été séquencées sur plusieurs puces afin de ne pas les surcharger, et les séquences ont été analysées par CSReport. Le tableau 1 résume pour chaque librairie, le nombre de séquences totales lues par le séquenceur, le nombre de séquences identifiant une jonction LSR, le nombre de jonctions LSR, et l'indice inverse de diversité de Simpson quand son calcul est possible. Comme cité précédemment, cet indice est positivement corrélé à la diversité des jonctions.

Amygdales saines	N séquences au total	N séquences identifiant une jonction LSR	N jonctions LSR	Indice inverse de Simpson	**CLL**	N Séquences au total	N séquences identifiant une jonction LSR	N jonctions LSR	Indice inverse de Simpson	**HL**	N séquences au total	N Séquences identifiant une jonction LSR	N Jonctions LSR	Indice inverse de Simpson
Amg 89	1 291 383	356	14	1,16	CLL 39	1 310 306	204	46	ND	HL 17	2 338 720	0	0	ND
Amg 90	851 551	972	61	3,21	CLL 40	9 910 975	1	1	ND	HL 18	1 868 135	0	0	ND
Amg 91	856 566	15 565	104	1,29	CLL 49	8 842 705	744	117	5,35	HL 19	1 878 718	55	1	ND
Amg 92	1 066 800	93	23	3,99	CLL 50	1 103 324	29	11	3,32	HL 20	2 535 788	35	1	ND
Amg 93	1 422 363	34 469	143	1,06	CLL 51	1 619 561	15	7	ND	HL 21	2 321 055	82	2	ND
Amg 94	914 714	105	21	2,59	CLL 52	1 064 705	89	9	ND	HL 55	1 313 772	536	11	1,64
Amg 55	1 437 581	146	11	1,17	CLL 53	1 190 872	47	10	ND	HL 56	1 530 167	0	0	ND
Amg 56	1 555 364	1 001	56	2,99	CLL 54	1 813 061	637	102	ND	HL 57	1 034 302	13	7	3,76
Amg 57	1 529 032	370	26	1,3	CLL 49b	3 153 014	42	25	12,08	HL 58	1 093 980	55	10	5,19
Amg 58	1 851 620	295	30	1,57	CLL 50b	1 771 320	0	0	ND	HL 59	929 006	0	0	ND

Amg 59	1 068 325	84	17	1,76	CLL 51b	3 060 515	54	22	7,4	HL 60	897 373	55	2	1,04
Amg 60	1 943 209	20 213	79	1,96	CLL 52b	2 744 078	36	6	1,43	HL 61	1 063 986	0	0	ND
Amg 61	2 129 842	443	95	ND	CLL 53b	2 015 854	206	74	32,2	HL 62	874 090	13	2	1,17
Amg 62	1 645 180	38	12	ND	CLL 54b	2 672 939	32	3	1,81	HL 63	1 252 928	24	1	ND
Amg 63	2 031 500	168	25	ND	CLL 55	2 848 050	0	0	ND	HL 64	994 123	3	1	ND
Amg 64	1 887 932	76 819	129	ND	CLL 56	2 504 767	2	2	2	HL 65	855 581	429	47	5,22
					CLL 57	3 304 148	9	9	9	HL 66	664 726	13	6	2,96
					CLL 58	3 446 821	6	4	3,6	HL 67	806 398	175	8	1,08
					CLL 59	3 309 678	82	25	4,93	HL 68	1 311 291	5 423	118	2,55
Moyenne	1 467 685	9 446	53	2	Moyenne	3 036 142	118	25	8	Moyenne	1 345 481	364	11	3
HEK 293T	2 321 673	2	0	1,8										

Amg : Amygdales ; **CLL :** Chronic Lymphocytic Leukemia ; **HL :** Hodgkin's Lymphoma ; **HEK :** Human Embryonic Kidney ; **ND :** Non déterminé

Tableau 1 : Résultats de l'analyse par CSReport des jonctions LSR issues de librairies d'amygdales saines, de HL et de CLL. Ce tableau comparatif résume les résultats obtenus avec le nombre de séquences correspondant à des jonctions de recombinaison LSR, le nombre de jonctions LSR et l'inverse de l'indice de diversité de Simpson, pour les échantillons d'amygdales saines, les CLL et HL.

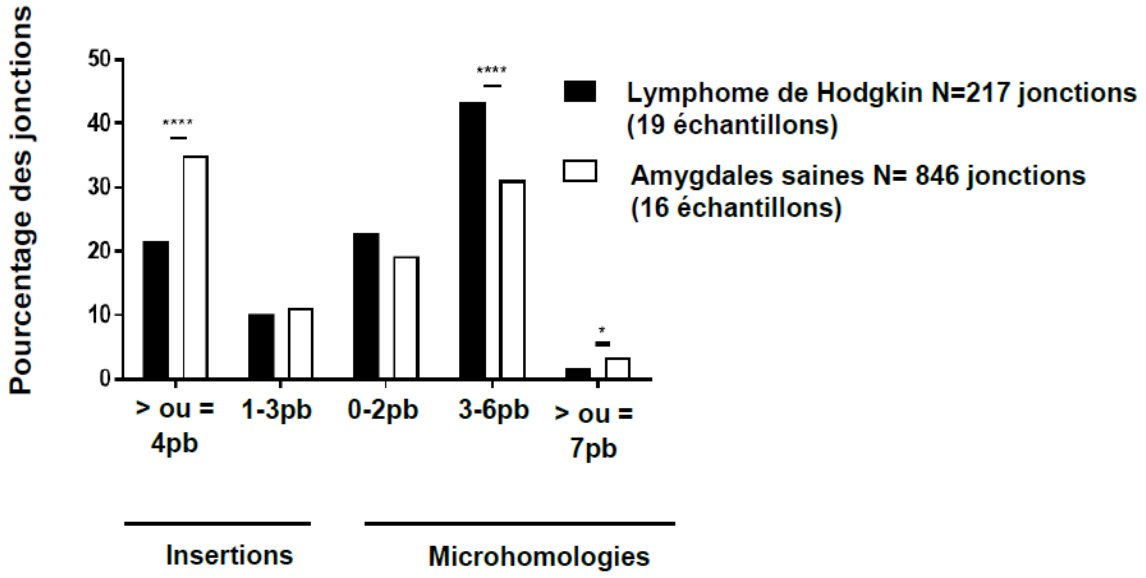
Nous observons en moyenne 1 467 685 et 1 345 481 séquences lues par le séquenceur concernant les librairies LSR issues d'échantillons d'amygdales saines et de PBMC de CLL respectivement. Ce nombre double dans le cas des librairies de HL, avec en moyenne 3 036 142 séquences obtenues. Parmi la totalité de ces séquences, CSReport a identifié 9 446 séquences correspondant à des recombinaisons LSR dans les échantillons sains, 118 et 364 dans les librairies de CLL et HL respectivement. Ces dernières comprennent des séquences de jonctions amplifiées par PCR, ainsi que celles qui ont été lues à plusieurs reprises, par la machine lors du séquençage. Afin d'avoir une estimation du nombre de séquences qui correspondent à des jonctions LSR non redondantes (jonctions différentes ou uniques), nous avons défini certains critères de sélection sur notre programme. En effet, CSReport regroupe les séquences de recombinaison LSR qui affichent une structure de réparation (blunt, insertion ou microhomologie/taille en pb) et des points de cassures très proches (< ou = 2 pb) au niveau des deux régions donneuse et acceptrice de la recombinaison. Après application de ces réglages, nous obtenons en moyenne 53 jonctions LSR différentes par librairie d'amygdales, 25 et 11 par échantillon de CLL et HL respectivement. Il est à noter que l'indice inverse de diversité de Simpson est calculé de façon automatique par le programme, et va de 2 en moyenne pour les amygdales, 3 dans le cas du HL jusqu'à 8 pour les CLL. Ceci indiquerait que la diversité de jonctions LSR semble moins importante dans les échantillons d'amygdales par rapport aux échantillons de HL et de CLL. Nous avons également préparé une librairie à partir d'ADNg de cellules rénales humaines embryonnaires (HEK 293T) comme contrôle négatif, et comme attendu CSReport n'a pas identifié de jonction LSR parmi les 3 669 440 séquences lues par le séquenceur haut débit.

Nous avons réussi à identifier des jonctions LSR à partir des librairies d'amygdales, de LLC et de HL. Toutefois, il faut rester vigilant quant à l'interprétation. Une quantification de ces jonctions n'est pas faisable actuellement, car le nombre de cellules B par échantillon n'a pas encore été déterminé.

1.2. Analyse structurale des jonctions LSR dans le HL et la CLL

L'analyse par CSReport des séquences de LSR pour chaque échantillon permet d'obtenir des informations sur la structure du point de jonction et d'obtenir une

A.



B.

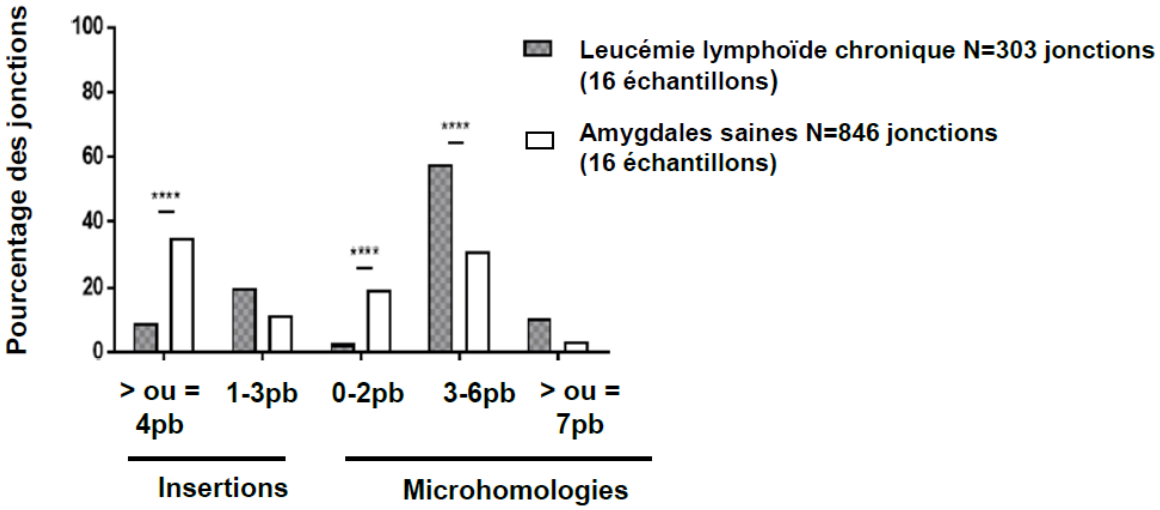


Figure 30 : Signature de réparation des CDB au cours de la LSR dans des échantillons de lymphomes B humains. A. Comparaison du pourcentage de jonctions LSR en fonction de leur structure (jonctions avec insertions, jonctions franches (0 à 2pb), et jonctions avec microhomologies) entre les échantillons de lymphome de Hodgkin et les amygdales saines (contrôles). **B.** Comparaison du pourcentage de jonctions LSR en fonction de leur structure (jonctions avec insertions, jonctions franches (0 à 2pb), et jonctions avec microhomologies) entre les échantillons de leucémie lymphoïde chronique et les amygdales saines (contrôles). La signature de réparation est différente entre les prélèvements de lymphomes B (ici lymphome de Hodgkin et leucémie lymphoïde chronique) et les prélèvements sains (amygdales). Chi Square Test : * P<0.5, **** P <0,001

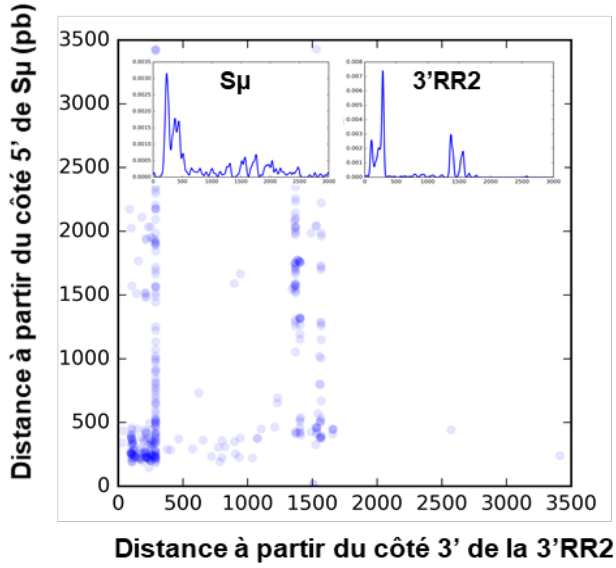
information sur le système de réparation de l'ADN impliqué au cours de la recombinaison. Nous avons choisi d'analyser cette structure dans des échantillons de HL en comparaison avec des amygdales et des échantillons de CLL comme contrôles (absence de lymphome et lymphome non hodgkinien), comme le montre la figure 30.

Les histogrammes représentent le pourcentage de jonctions LSR en fonction de leur structure, dans le cas du HL **(Figure 30.A)** et de la CLL **(Figure 30.B)**. Nous observons une augmentation des jonctions de recombinaisons contenant des microhomologies de 3 à 6 pb dans les échantillons de HL (43,32% versus 31,09% pour les contrôles) et dans le cas des CLL (57,8%). Nous observons également une réduction dans l'usage des jonctions avec insertions supérieures ou égales à 4 pb, de 34,99% pour les amygdales contrôles à 21,66% et 9,2% dans le HL et la CLL respectivement. Dans les échantillons de CLL, l'usage de jonctions franches et avec de courtes microhomologies (0 à 2 pb) est diminué à 2,6%. Les résultats obtenus indiquent que la réparation de l'ADN au cours de la LSR semble différente dans les échantillons pathologiques (HL et CLL), par comparaison aux échantillons sains (amygdales, absence de lymphome). Cette différence peut s'expliquer par des systèmes de réparation des CDB dérégulés dans les cellules tumorales, ou par un contexte chromatinien altéré. Un ciblage modifié des CDB dans la 3'RR2 des échantillons pathologiques pourrait aussi induire un biais de structure, lorsque ces cassures sont générées au niveau de segments affichant une forte homologie avec la région donneuse Sμ. Nous avons donc déterminé par CSReport la localisation des cassures au niveau de Sμ et de la 3'RR2 dans les trois types d'échantillons étudiés (amygdales, HL et CLL) **(Figure 30.A.B et C)**.

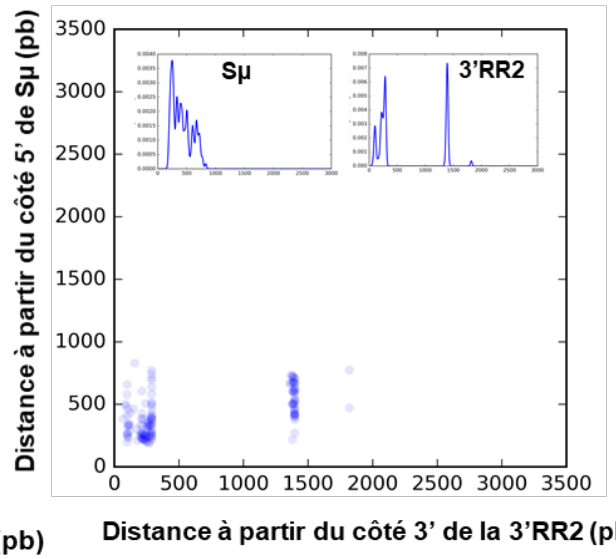
Les points de cassures des jonctions LSR provenant des amygdales saines **(Figure 31.A)** sont localisés principalement au niveau de deux régions, dans un segment d'ADN situé 500 pb en amont de l'extrémité 3' de la 3'RR2, et à environ 1500 pb. La majorité des points de rupture sont localisés dans les 500 pb en 3' de la 3'RR2. Ceci est certainement dû à un biais d'amplification des jonctions de LSR puisque cette région est à proximité des amorces que nous utilisons. Cependant, la région localisée à 1500 bp semble être un point de ciblage préférentiel. Nous détectons moins de jonctions avec un point de rupture aux environs de 1500 bp qu'aux environs de 500 bp, mais ceci doit être dû à l'amplification préférentielle des plus courtes jonctions de

LSR pendant la PCR. De façon générale, les CDB sont distribuées de façon similaire entre les échantillons pathologiques de lymphome B (HL et CLL) et les amygdales saines.

A. Amygdales (N= 846 jonctions LSR)



B. HL (N= 217 jonctions LSR)



C. CLL (N= 303 jonctions LSR)

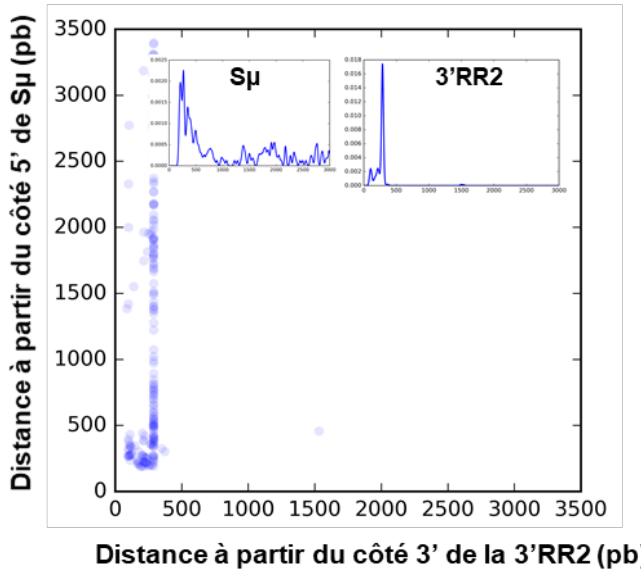


Figure 31 : Représentation de la localisation des points de cassure au niveau des régions donneuse Sμ et acceptrice 3'RR2 de la recombinaison LSR dans des librairies d'échantillons humains. A. Jonctions LSR issues d'amygdales saines.**B.** Jonctions LSR issues de HL.**C.** Jonctions LSR issues de CLL.

DISCUSSION

Notre laboratoire s'intéresse aux mécanismes moléculaires régissant les recombinaisons génétiques du locus *IgH* des LB, et a permis la mise en évidence de la recombinaison LSR chez la souris (Péron et al., 2012). La LSR entre la région Sμ et la 3'RR délète le locus *IgH* de tous les gènes constants, et empêche de cette façon l'expression d'une Ig et du BCR à la surface du LB. Cette absence de récepteur prive la cellule des signaux nécessaires à sa survie. Ainsi, la recombinaison LSR, en induisant la mort du LB qui la subit pourrait occuper une fonction importante dans la physiologie de la cellule B. Cette hypothèse est renforcée par la conservation des séquences LS de la 3'RR ciblées par la LSR, entre les espèces pour lesquelles les séquences complètes des loci *IgH* sont disponibles (Homme, souris, chien et lapin) (Péron et al., 2012). La LSR semble être la contrepartie délétère de la CSR. En effet, la CSR permet au LB d'exprimer l'Ig la plus adaptée pour la réponse immune contre un Ag, tandis que la LSR induit la mort de la cellule B concernée.

Les mécanismes moléculaires des recombinaisons LSR et CSR semblent similaires. Les deux types de recombinaison passent par trois étapes principales : une étape essentielle de transcription des segments d'ADN donneur et accepteur de la recombinaison, la génération de lésions au niveau de leur ADN par AID aboutissant à des CDB, et enfin, la réparation des cassures, majoritairement par le C-NHEJ dans la CSR (Soulas-Sprauel et al., 2007; Yan et al., 2007; Han and Yu, 2008). La première description de la LSR chez la souris avait mis en évidence le recrutement de l'ARN Pol II au niveau de la 3'RR et sa transcription, l'insertion de mutations induites par AID dans l'ADN de cette région, et la dépendance de la recombinaison LSR à l'enzyme AID (Péron et al., 2012). La réparation des CDB au cours de la LSR a été déduite de la structure des jonctions de réparation qui semblent comparables à la structure des jonctions CSR. Cependant, le manque de preuve directe de l'activité d'une voie particulière de réparation des CDB au cours de la LSR nous a poussé à approfondir la compréhension de ce point.

de séquences lues par la machine, le nombre de jonctions, ainsi que la structure de chaque jonction individuellement, ce qui renseigne sur le système de réparation de l'ADN impliqué.

II. Signature de réparation des CDB différente entre LSR et CSR

L'utilisation de CSReport qui a été mise au point à partir de jonctions CSR a été étendue à l'analyse de jonctions LSR chez la souris et chez l'Homme. L'analyse par CSReport des données de séquençage haut débit des jonctions de recombinaisons LSR et CSR nous a permis d'aller plus en avant dans l'étude de la réparation de l'ADN au cours de la LSR. Les résultats que nous avons obtenus montrent clairement que la structure au point de réparation est différente entre les processus de LSR et de CSR de manière similaire chez l'Homme et chez la souris. La CSR fait intervenir de façon majoritaire la voie du C-NHEJ, et les résultats obtenus sur la structure de réparation est en accord avec ce qui est attendu (Han and Yu, 2008; Soulas-Sprauel et al., 2007; Yan et al., 2007). Cependant, la signature de réparation de la LSR est plus proche de celle observée dans la CSR lors du défaut de C-NHEJ (Boboila et al., 2010b). Dans ce dernier cas, la réparation des CDB au niveau des régions S fait appel à l'A-EJ. Ainsi la structure des jonctions LSR semble résulter de l'implication d'un système de réparation autre que le C-NHEJ. Toutefois, certaines particularités de la 3'RR pourraient aussi être à l'origine de la présence de microhomologies de séquence au niveau de la jonction de réparation :

- la structure des extrémités de la CDB peut favoriser la microhomologie au point de jonction lorsque l'extrémité est cohésive avec un simple brin sortant de taille importante (Schrader et al., 2007). L'hybridation directe de brins débordants entre les séquences donneuse et acceptrice de la recombinaison présentant une homologie de séquence est susceptible également de générer l'augmentation de la microhomologie à la jonction de réparation.

- au niveau des extrémités lésées de l'ADN, la structure cohésive à long segment d'ADNsb sortant peut-être le produit de la résection d'un brin d'ADN à l'extrémité de la CDB dans la région 3'RR. Il serait donc intéressant d'évaluer si CtIP, MRE11 ou EXO1, acteurs principaux de la résection de l'ADN aux CDB, sont recrutés

à la 3'RR des LB activés. Si la résection est augmentée à la 3'RR, la réparation sera spécifique de ce type de CDB et différente de celle employée dans la CSR.

- la structure cohésive des extrémités de CDB dans la région 3'RR peut également être générée par une réduction de l'activité de AID au niveau de cette région. En effet, au cours du processus de CSR, les CDB sont issues de la conversion de CSB sur les deux brins d'ADN générées lors de la prise en charge de mésappariements U/G induits par AID. Il est facilement imaginable que la diminution de l'activité de AID sur la 3'RR, à travers la création de CSB plus éloignées que dans les régions S, favoriserait la formation de CDB avec de long segments d'ADN sortants. Cette hypothèse est probable, et des données de ChIP-seq qui ont montré dans la 3'RR le recrutement de AID mais l'absence de RPA vont dans ce sens (Yamane et al., 2011). Cependant, les mutations dépendantes de AID au niveau de la 3'RR et leur taux, bien qu'inférieur à celui observé dans la région Sμ (0.28% dans les séquences LS versus 0.42% dans Sμ), démontrent une activité soutenue de AID dans cette région (Péron et al 2012).

- une autre explication à l'augmentation de la microhomologie au niveau du site de réparation de la LSR par rapport à la CSR peut être une homologie de séquence entre les régions Sμ et la 3'RR. En effet, il est connu que le taux d'homologie entre les séquences donneuse et acceptrice de la CSR influence la structure de réparation et l'usage de la microhomologie. Effectivement, les régions S sont caractérisées par une répétition de séquences cœur. Les régions Sμ (4 Kb) (Nikaido et al., 1981), Sε (1 Kb) (Nikaido et al., 1982) et Sα (4,2 Kb) (Arakawa et al., 1993) sont homologues entre elles et sont particulièrement riches en motifs « GGGGT » et « GAGCT ». Les régions cœur de Sγ3 (2,5 Kb), Sγ1 (8 Kb), Sγ2b (5 Kb), et Sγ2a (2,5 Kb) contiennent une séquence de répétition de 49 pb qui présente plusieurs fois le motif « AGCT » trouvé également dans les séquences Sμ, Sε et Sα (Zarrin et al., 2005). Il a été montré que les segment Sα et Sμ possèdent une forte homologie de séquence supérieure à celle constatée entre Sμ et Sγ. Ainsi les jonctions de recombinaison Sμ/Sα affichent de plus longues séquences de microhomologie comparé aux jonctions Sμ/Sγ, (Pan-Hammarström et al., 2007; Stavnezer et al., 2010, 2010). Cependant, les régions LS ont été identifiées par la recherche de segments similaires aux régions S, et les régions LS contiennent de longs segments d'ADN (0,5 à 1 Kb) contenant la répétition d'un motif de 49 pb avec une homologie de 60% chez la souris et 80% chez l'Homme aux régions Sγ (Péron et

al., 2012). De ce fait, la structure de réparation de la LSR au site de jonction Sμ/LS attendue se rapprocherait de celle des jonctions Sμ/Sγ1.

Finalement, ni la structure cohésive des extrémités de la CDB ni une forte homologie entre les séquences donneuse et acceptrice de la recombinaison n'expliquent l'augmentation des insertions observées au niveau des jonctions LSR. En revanche, ceci est une des caractéristiques, avec la microhomologie de petite taille, à la jonction de réparation impliquant la voie A-EJ et particulièrement Polθ (Chan et al., 2010; Mateos-Gomez et al., 2015).

II.1. Recrutement des protéines de réparatation selon le segment de recombinaison concerné

Les arguments cités précédemment nous ont poussé à explorer la possibilité que la LSR fasse appel à un système de réparation autre que celui du C-NHEJ. Nous avons donc analysé le recrutement de protéines impliquées dans les différentes voies de réparation des CDB dont XRCC4 (C-NHEJ), RAD51 (HR), PARP-1 et Pol θ (A-EJ) au niveau des régions ciblées par la LSR et la CSR, afin de déterminer si ce recrutement diffère en fonction des régions du locus *IgH*. Ceci a été réalisé par ChIP et bien que les taux d'enrichissement soient faibles en valeur, nous observons des différences statistiquement significatives. L'association de XRCC4 et PARP-1 au niveau de la région Sμ donneuse de la CSR est en accord avec l'implication du C-NHEJ et éventuellement de l'A-EJ (Soulas-Sprauel et al., 2007; Yan et al., 2007; Han and Yu, 2008). L'absence de détection de XRCC4 au niveau de HS1/2 et HS4 et le recrutement de PARP-1 (et Polθ) est en accord avec une orientation de la réparation des CDB de l'ADN vers le système A-EJ au niveau de la 3'RR. L'absence du recrutement de la protéine XRCC4 au niveau de Sγ1 est étonnante mais la conformation en boucle du locus *IgH* dans les LB, et la synapse des régions S lors de l'activation des LB (Wuerffel et al., 2007) peut permettre la réparation des CDB par les molécules présentes au niveau de Sμ. L'absence de recrutement préférentiel de RAD51 sur les segments étudiés du locus *IgH* est en accord avec son absence d'implication dans le CSR et est en défaveur d'un recours à la HR dans la LSR. La mise en évidence, pour la première fois, du recrutement de PARP-1 et potentiellement de Polθ sur la 3'RR est en faveur de l'implication de la l'A-EJ au cours de la réparation des CDB dans la 3'RR pendant la LSR.

II.2. Contexte chromatinien différent entre les segments accepteurs de recombinaison CSR et LSR

Il est déjà connu que les marques épigénétiques orientent l'intervention d'un système de réparation particulier (Clouaire and Legube, 2015). Notre choix s'est porté sur l'analyse de l'enrichissement en H4K20me1, H4K16ac, H3K36me3 et H3K36me2 dans les régions citées précédemment (Sμ, Sγ1, HS1/2, HS4 et Sγ3). Nos résultats, bien que préliminaires, permettent d'avoir une idée sur les marques de chromatine associées au locus *IgH* dans les LB activés. Nous avons observé que H4K20me1 est associée aux régions Sμ et Sγ1 dans les LB activés. H4K20me1 est impliquée dans le recrutement et la stabilisation de 53BP1 au niveau d'une CDB, et peut ainsi favoriser le C-NHEJ (Yu and Chen, 2004; Bothmer et al., 2010).

Ces résultats sont donc en accord avec l'implication du C-NHEJ au cours de la CSR. Nous n'avons pas détecté d'association de la forme monométhylée de l'histone H4K20 (H4K20me1) au sein de la 3'RR. Ceci est cohérent avec nos résultats de ChIP qui ne mettent pas en évidence de recrutement de XRCC4 au niveau de la 3'RR dans les LB activés. La forme acétylée H4K16ac est rerouvée elle aussi associée aux régions Sμ et Sγ1. Cette marque décrite pour son enrichissement au niveau de la chromatine transcriptionnellement active (Taylor et al., 2013) a été montrée comme bloquant le recrutement de 53BP1 au niveau des CDB, empêchant ainsi une réparation par le C-NHEJ et favorisant la résection et l'intervention de l'A-EJ et/ou la HR (Hsiao and Mizzen, 2013).

Dans la recombinaison CSR, il semblerait donc que les marques épigénétiques étudiées orientant la réparation des CDB vers les voies du C-NHEJ et de l'AEJ et/ ou HR coexistent au niveau des segments donneur et accepteur. Il est probable que la marque H4K20me1 permet le recrutement de 53BP1 et du C-NHEJ à Sμ et Sγ1 au cours du CSR. L'enrichissement de l'H4K16ac au niveau des régions Sμ et Sγ1 dans des LB activés par LPS et IL-4 n'est pas étonnante. En effet, dans cette condition, les deux régions sont activement transcrites. En revanche, elle est absente dans la 3'RR de LB activés. Cette absence est inattendue puisque que selon la littérature, l'H4K16ac est retrouvée dans des régions transcriptionnellement actives (Taylor et al., 2013), et qu'il a été montré que la région 3'RR est transcrite dans les LB activés (Péron et al., 2012; Saintamand et al., 2015). Selon nos résultats, Sγ3 n'est pas une région du locus *IgH* enrichie pour les marques H4K20me1 et H4K16ac. Ceci est cohérent avec la

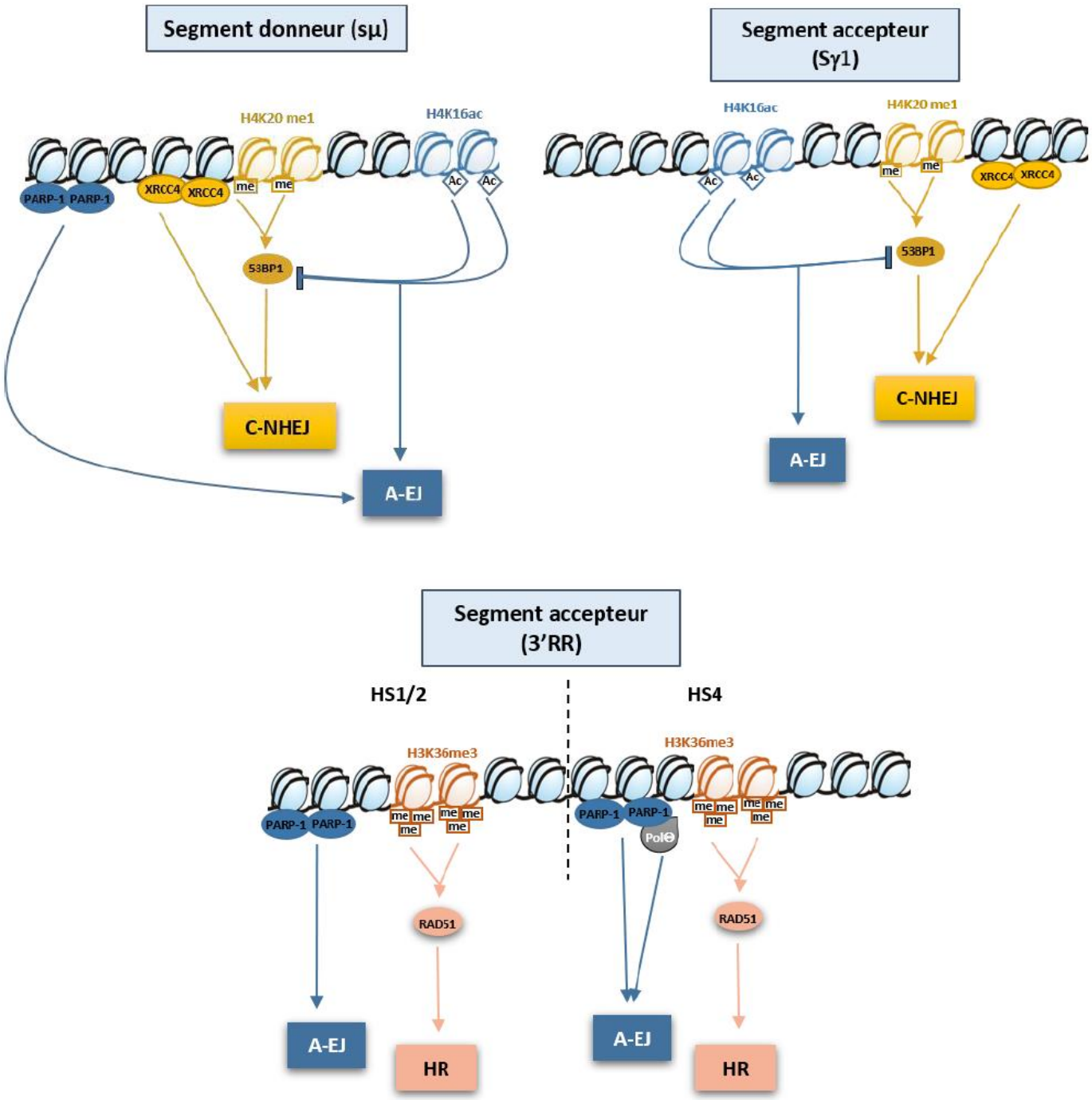


Figure 32 : Influence du contexte chromatinien sur le recrutement de protéines de réparation lors des recombinaisons du locus *IgH*. Des marques d'histones différentes au niveau du segment accepteur de la CSR (ici Sγ1) et du segment accepteur de la LSR (3'RR) recrutent des protéines qui sont impliquées dans des voies distinctes de réparation des CDB de l'ADN comme le C-NHEJ et l'A-EJ pour Sγ1 et l'A-EJ et éventuellement la HR pour la région 3'RR.

condition d'activation des LB (LPS et IL-4) qui n'entraine ni la transcription ni la lésion de l'ADN au niveau de cette région.

Finalement, l'étude des marques d'histones H4K20me1 et H4K16ac au locus *IgH* des LB activés est informative pour les régions concernées par la CSR. Nous avons détecté l'H3K36me3 et l'H3K36me2 au niveau de toutes les régions du locus *IgH* analysées. Cependant, nous avons constaté que le ratio H3K36me3/H3K36me2 est augmenté au niveau des segments HS1/2 et HS4, par rapport aux régions Sμ et Sγ1. L'enrichissement de la forme H3K36me3 a été impliqué dans l'orientation de la réparation des CDB vers la voie HR (Aymard et al., 2014), alors que H3K36me2 semble plutôt associée à la réparation des CDB par le C-NHEJ (Fnu et al., 2011). Le ratio H3K36me3/H3K36me2 augmenté dans la 3'RR serait en faveur d'une réparation des CDB par la voie HR à ce niveau.

De façon générale, nos données de ChIP concernant les molécules de la réparation de l'ADN, et les marques épigénétiques étudiées semblent en faveur d'une absence de recrutement de la voie C-NHEJ au niveau de la 3'RR dans les LB activés. Ainsi, des CDB générées au niveau de cette région sont susceptibles de ne pas être prises en charge par la voie du C-NHEJ mais plutôt par l'A-EJ et/ou la HR **(Figure 32)**.

L'implication de l'un de ces deux systèmes expliquerait la signature de réparation que nous avons démontré dans la LSR. En revanche, le fait que nous n'ayons pas détecté RAD51 au locus *IgH* de LB activés, et les microhomologies de séquence de taille modérée retrouvées au point de jonction suggèrent que l'A-EJ serait la meilleure candidate pour la réparation majeure des CDB au cours de la LSR.

Il serait intéressant d'approfondir l'étude de la LSR, car elle pourrait jouer un rôle primordial dans le maintien de l'homéostasie cellulaire B. Les LB pourraient « se sacrifier » afin de protéger notre organisme, lors de stimulations non adéquates du BCR, dans le cas de cellules B auto-immunes ou cellules tumorales. Quels sont les facteurs qui prennent part dans le choix délicat entre CSR et survie ou LSR et apoptose ? Comment ce phénomène est-il régulé ? Des investigations méritent d'être menées pour mieux cerner cette recombinaison fatale.

III. Réparation des CDB altérée lors de la LSR dans le lymphome de Hodgkin

Le travail initié dans les lymphomes B, et en particulier le HL ne permet pas encore de répondre à la question initialement posée, c'est-à-dire, la possibilité que la LSR puisse être à l'origine de la perte de BCR à la surface des cellules de Reed-Sternerg. En effet, dans ce lymphome, les cellules tumorales ne présentent pas de BCR à leur surface. Il a été montré que l'absence de ce récepteur peut être attribuée à des mutations non-sens dans les gènes qui codent pour les Ig (Kanzler et al., 1996; Schwering et al., 2003). Nous avons donc émis l'hypothèse que cette absence de récepteur pourrait également être le résultat du mécanisme de LSR. De plus, les cellules tumorales possèdent un cycle cellulaire et des processus pro-apoptotiques dérégulés (Bai et al., 2005; Pettigrew and Cotter, 2009), ce qui pousserait les cellules à persister au lieu de succomber à la recombinaison fatale.

Nous avons réussi à détecter des jonctions LSR dans les échantillons sains (amygdales) et pathologiques (HL et CLL), et nous avons constaté que le nombre moyen de jonctions LSR est diminué dans les échantillons de lymphomes. La détection de LSR dans les échantillons de ganglions de lymphome de HL et d'amygdales saines (contrôles) s'explique par la réaction LSR en cours. Effectivement, au sein de ces organes lymphoïdes secondaires, des LB matures sont activés, prolifèrent et réalisent les évènements de maturation terminale (CSR et SHM) et LSR.

Dans le HL, les prélèvements ayant été choisis sur la base d'une richesse d'infiltration de cellules anormales, qui reste néanmoins faible en pourcentage, les jonctions LSR pourraient être attribuées aux cellules de Reed-Sternberg selon notre hypothèse de travail. Dans ce cas, la diminution du nombre de jonctions LSR pourrait être la conséquence de l'enrichissement des cellules tumorales dans les prélèvements. Cependant, ces résultats ne peuvent être comparés que si nous connaissons pour chaque prélèvement l'enrichissement en LB normaux et cancéreux. Ce travail, en cours, consiste à mettre au point l'analyse des réarrangements V(D)J par séquençage haut débit à partir de produits d'amplification. Ainsi, la clonalité des LB présents dans l'échantillon pourra être analysée et utilisée pour comparer la « clonalité » des jonctions LSR. Toutefois, pour conclure définitivement, l'idéal serait de travailler sur de l'ADNg obtenu à partir de cellules de Reed-Sternberg microdisséquées, afin d'avoir un matériel non contaminé.

Dans notre étude, les prélèvements de CLL ont été utilisés comme contrôles dans lesquels aucune anomalie du nombre de jonctions LSR n'est attendue. Nos résultats préliminaires montrent une diminution du nombre moyen de jonctions LSR dans ce type de prélèvement, par rapport aux échantillons d'amygdales saines contrôles. L'ADNg des cas de CLL est issu des cellules mononucléées du sang, et les jonctions LSR détectées doivent être portées par l'allèle *IgH* non productif de LB normaux et tumoraux circulants exprimant un BCR. Il serait nécessaire de monitorer la LSR dans les cellules mononucléées de sang de témoins sains, et de quantifier la clonalité dans ces échantillons afin de conclure si les échantillons de CLL présentent des jonctions LSR de façon similaire aux témoins.

De façon étonnante, l'ensemble des résultats a permis de mettre en évidence une altération de la signature de réparation de la LSR, dans les prélèvements issus de HL et de CLL. Puisque les points de cassure situés sur le segment 3'RR2 des jonctions LSR semblent être distribués de façon similaire quel que soit le type d'échantillon, la différence de signature de réparation observée ne peut pas être attribuée à l'utilisation de différents segments donneurs et accepteurs de la recombinaison. Nous stipulons donc l'hypothèse que des altérations intrinsèques aux systèmes de réparation de l'ADN dans les LB des échantillons tumoraux peuvent être en cause. Pour en savoir plus, il serait intéressant de quantifier la transcription et la traduction des acteurs de la réparation des CDB de l'ADN comme XRCC4, PARP-1 et Polθ, à partir des ARN et extraits protéiques de cellules issues des échantillons analysés ci-dessus.

CONCLUSION ET PERSPECTIVES

Finalement, les travaux de mon doctorat ont permis de mettre au point un programme libre de droits permettant de faciliter et d'harmoniser l'analyse des jonctions de réparation issues des recombinaisons du locus *IgH* (CSR et LSR). L'élaboration et la mise au point de cet outil a fait l'objet d'une publication dans Journal of Immunology, dans laquelle je suis co-premier auteur (Boyer et al., 2017).

Le développement de CSReport a rendu l'analyse de la LSR plus aisée, que ce soit dans des LB purifiés ou non, cultivés *in vitro,* ou à partir de tissus humains et murins. Ainsi, nous avons montré pour la première fois une différence de signature de réparation de l'ADN entre la LSR et la CSR, et ce chez l'Homme et chez la souris. Le travail, principalement mené chez la souris, semble indiquer que le recrutement des systèmes de réparation diffère entre les régions Sμ, Sγ1 et la 3'RR dans les LB activés. Les faibles valeurs du taux d'enrichissement pour les immunoprécipitations peuvent être liées à la fugacité de la liaison des molécules de réparation, ou bien au fait que 48 heures après le début de la stimulation n'est pas la bonne fenêtre temporelle, et il faudrait que nous réitérions ces expériences au cours d'une cinétique après stimulation jusqu'à 72 heures où l'on détecte classiquement le pourcentage maximum de cellules vivantes ayant réalisé la CSR. Il serait aussi intéressant de mener ce travail dans des LB humains activés *in vivo* et/ou *in vitro*. De plus, nous avons observé que les régions Sμ, Sγ1 et la 3'RR présentent un contexte chromatinien différent les unes des autres, en accord avec les résultats de recrutement des molécules de réparation.

Globalement l'ensemble de nos résultats tendent à montrer que la LSR et la CSR impliquent différents systèmes de réparation des CDB, le C-NHEJ pour la CSR, et selon nos données l'A-EJ pour la LSR. Ceci se traduit par des signatures de réparation différentes aux points de jonction, et le contexte chromatinien de chaque segment d'ADN impliqué pourrait influencer l'orientation de la réparation des CDB vers une voie donnée.

Pour finaliser cette étude, il faudrait réaliser une analyse de la LSR, dans des LB déficients pour les molécules de réparation présumées impliquées au cours de la LSR, comme PARP-1 et Polθ qui font partie de la voie A-EJ. Il existe peu de données concernant l'orientation de la réparation des CDB vers l'A-EJ. En effet, ce système est souvent perçu comme utilisé par défaut, alors qu'il pourrait être la voie majoritaire de certains processus comme potentiellement la recombinaison LSR. Il est possible que

les résultats que nous avons obtenus puissent participer à densifier ce champ de recherche.

Mon travail de thèse a ouvert plusieurs perspectives de travail à court et à moyen terme. Parmi ces pistes, il y a la quantification relative des jonctions LSR par rapport à la clonalité des LB dans les échantillons humains, la compréhension de l'altération de la signature de réparation au cours de la LSR, mais aussi la détermination de la façon dont le segment accepteur de recombinaison (au locus *IgH* ou en dehors) influence le recrutement du système de réparation. Ce dernier point pourrait apporter des éléments de compréhension des translocations du locus *IgH*.

BIBLIOGRAPHIE

Adachi, N., Ishino, T., Ishii, Y., Takeda, S., and Koyama, H. (2001). DNA ligase IV-deficient cells are more resistant to ionizing radiation in the absence of Ku70: Implications for DNA double-strand break repair. Proc. Natl. Acad. Sci. U. S. A. *98*, 12109–12113.

Adams, B., Dörfler, P., Aguzzi, A., Kozmik, Z., Urbánek, P., Maurer-Fogy, I., and Busslinger, M. (1992). Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. Genes Dev. *6*, 1589–1607.

Adelman, K., Marr, M.T., Werner, J., Saunders, A., Ni, Z., Andrulis, E.D., and Lis, J.T. (2005). Efficient Release from Promoter-Proximal Stall Sites Requires Transcript Cleavage Factor TFIIS. Mol. Cell *17*, 103–112.

Ahnesorg, P., Smith, P., and Jackson, S.P. (2006). XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. Cell *124*, 301–313.

Alt, F.W., and Baltimore, D. (1982). Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions. Proc. Natl. Acad. Sci. U. S. A. *79*, 4118–4122.

Alt, F.W., Yancopoulos, G.D., Blackwell, T.K., Wood, C., Thomas, E., Boss, M., Coffman, R., Rosenberg, N., Tonegawa, S., and Baltimore, D. (1984). Ordered rearrangement of immunoglobulin heavy chain variable region segments. EMBO J. *3*, 1209–1219.

Arakawa, H., Iwasato, T., Hayashida, H., Shimizu, A., Honjo, T., and Yamagishi, H. (1993). The complete murine immunoglobulin class switch region of the alpha heavy chain gene-hierarchic repetitive structure and recombination breakpoints. J. Biol. Chem. *268*, 4651–4655.

Arnoult, N., Correia, A., Ma, J., Merlo, A., Garcia-Gomez, S., Maric, M., Tognetti, M., Benner, C.W., Boulton, S.J., Saghatelian, A., et al. (2017). Regulation of DNA repair pathway choice in S and G2 phases by the NHEJ inhibitor CYREN. Nature *549*, 548–552.

Aymard, F., Bugler, B., Schmidt, C.K., Guillou, E., Caron, P., Briois, S., Iacovoni, J.., Daburon, V., Miller, K.M., Jackson, S.P., et al. (2014). Transcriptionally active chromatin recruits homologous recombination at DNA double strand breaks. Nat. Struct. Mol. Biol. *21*, 366–374.

Bai, M., Papoudou-Bai, A., Kitsoulis, P., Horianopoulos, N., Kamina, S., Agnantis, N.J., and Kanavaros, P. (2005). Cell cycle and apoptosis deregulation in classical Hodgkin lymphomas. Vivo Athens Greece *19*, 439–453.

Bain, G., Maandag, E.C.R., Izon, D.J., Amsen, D., Kruisbeek, A.M., Weintraub, B.C., Krop, I., Schlissel, M.S., Feeney, A.J., Roon, M. van, et al. (1994). E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. Cell *79*, 885–892.

Bakkenist, C.J., and Kastan, M.B. (2003). DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. Nature *421*, 499–506.

Banerji, J., Olson, L., and Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell *33*, 729–740.

Bardwell, P.D., Woo, C.J., Wei, K., Li, Z., Martin, A., Sack, S.Z., Parris, T., Edelmann, W., and Scharff, M.D. (2004). Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice. Nat. Immunol. *5*, 224–229.

Barnes, D.E., Stamp, G., Rosewell, I., Denzel, A., and Lindahl, T. (1998). Targeted disruption of the gene encoding DNA ligase IV leads to lethality in embryonic mice. Curr. Biol. CB *8*, 1395–1398.

Basu, U., Meng, F.-L., Keim, C., Grinstein, V., Pefanis, E., Eccleston, J., Zhang, T., Myers, D., Wesemann, D.R., Januszyk, K., et al. (2011). The RNA Exosome Targets the AID Cytidine Deaminase to Both Strands of Transcribed Duplex DNA Substrates. Cell *144*, 353–363.

Betz, A.G., Rada, C., Pannell, R., Milstein, C., and Neuberger, M.S. (1993). Passenger transgenes reveal intrinsic specificity of the antibody hypermutation mechanism: clustering, polarity, and specific hot spots. Proc. Natl. Acad. Sci. U. S. A. *90*, 2385–2388.

Beucher, A., Birraux, J., Tchouandong, L., Barton, O., Shibata, A., Conrad, S., Goodarzi, A.A., Krempler, A., Jeggo, P.A., and Löbrich, M. (2009). ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2. EMBO J. *28*, 3413–3427.

Birshtein, B.K. (2014). Epigenetic Regulation of Individual Modules of the immunoglobulin heavy chain locus 3' Regulatory Region. Front. Immunol. *5*.

Blier, P.R., Griffith, A.J., Craft, J., and Hardin, J.A. (1993). Binding of Ku protein to DNA. Measurement of affinity for ends and demonstration of binding to nicks. J. Biol. Chem. *268*, 7594–7601.

Blomberg, B., Traunecker, A., Eisen, H., and Tonegawa, S. (1981). Organization of four mouse lambda light chain immunoglobulin genes. Proc. Natl. Acad. Sci. U. S. A. *78*, 3765–3769.

Boboila, C., Jankovic, M., Yan, C.T., Wang, J.H., Wesemann, D.R., Zhang, T., Fazeli, A., Feldman, L., Nussenzweig, A., Nussenzweig, M., et al. (2010a). Alternative end-joining catalyzes robust IgH locus deletions and translocations in the combined absence of ligase 4 and Ku70. Proc. Natl. Acad. Sci. U. S. A. *107*, 3034–3039.

Boboila, C., Yan, C., Wesemann, D.R., Jankovic, M., Wang, J.H., Manis, J., Nussenzweig, A., Nussenzweig, M., and Alt, F.W. (2010b). Alternative end-joining catalyzes class switch recombination in the absence of both Ku70 and DNA ligase 4. J. Exp. Med. *207*, 417–427.

Bogue, M.A., Wang, C., Zhu, C., and Roth, D.B. (1997). V(D)J Recombination in Ku86-Deficient Mice: Distinct Effects on Coding, Signal, and Hybrid Joint Formation. Immunity *7*, 37–47.

Bothmer, A., Robbiani, D.F., Feldhahn, N., Gazumyan, A., Nussenzweig, A., and Nussenzweig, M.C. (2010). 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. J. Exp. Med. *207*, 855–865.

Boyer, F., Boutouil, H., Dalloul, I., Dalloul, Z., Cook-Moreau, J., Aldigier, J.-C., Carrion, C., Herve, B., Scaon, E., Cogné, M., et al. (2017). CSReport: A New Computational Tool Designed for Automatic Analysis of Class Switch Recombination Junctions Sequenced by High-Throughput Sequencing. J. Immunol. Baltim. Md 1950 *198*, 4148–4155.

Brekke, K.M., and Garrard, W.T. (2004). Assembly and analysis of the mouse immunoglobulin kappa gene sequence. Immunogenetics *56*, 490–505.

Brodeur, P.H., and Riblet, R. (1984). The immunoglobulin heavy chain variable region (Igh-V) locus in the mouse. I. One hundred Igh-V genes comprise seven families of homologous genes. Eur. J. Immunol. *14*, 922–930.

Bunting, S.F., Callén, E., Wong, N., Chen, H.-T., Polato, F., Gunn, A., Bothmer, A., Feldhahn, N., Fernandez-Capetillo, O., Cao, L., et al. (2010). 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. Cell *141*, 243–254.

Ceccaldi, R., Liu, J.C., Amunugama, R., Hajdu, I., Primack, B., Petalcorin, M.I.R., O'Connor, K.W., Konstantinopoulos, P.A., Elledge, S.J., Boulton, S.J., et al. (2015). Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. Nature *518*, 258–262.

Chan, S.H., Yu, A.M., and McVey, M. (2010). Dual Roles for DNA Polymerase Theta in Alternative End-Joining Repair of Double-Strand Breaks in Drosophila. PLOS Genet. *6*, e1001005.

Chatterjee, S., Ju, Z., Hassan, R., Volpi, S.A., Emelyanov, A.V., and Birshtein, B.K. (2011). Dynamic changes in binding of immunoglobulin heavy chain 3' regulatory region to protein factors during class switching. J. Biol. Chem. *286*, 29303–29312.

Chaudhuri J, and Alt FW (2004). Class-switch recombination: interplay of transcription, DNA deamination and DNA repair.

Chauveau, C., and Cogné, M. (1996). Palindromic structure of the IgH 3'locus control region. Nat. Genet. *14*, 15–16.

Chen, K., Xu, W., Wilson, M., He, B., Miller, N.W., Bengten, E., Edholm, E.-S., Santini, P.A., Rath, P., Chiu, A., et al. (2009). Immunoglobulin D enhances immune surveillance by activating antimicrobial, pro-inflammatory and B cell-stimulating programs in basophils. Nat. Immunol. *10*, 889–898.

Clouaire, T., and Legube, G. (2015). DNA double strand break repair pathway choice: a chromatin based decision? Nucl. Austin Tex *6*, 107–113.

Cogné, M., Lansford, R., Bottaro, A., Zhang, J., Gorman, J., Young, F., Cheng, H.L., and Alt, F.W. (1994). A class switch control region at the 3' end of the immunoglobulin heavy chain locus. Cell *77*, 737–747.

Daniels, G.A., and Lieber, M.R. (1995). RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. Nucleic Acids Res. *23*, 5006–5011.

Dariavach Piona, Williams Gareth T., Campbell Kathryn, Pettersson Sven, and Neuberger Michael S. (2005). The mouse IgH 3'-enhancer. Eur. J. Immunol. *21*, 1499–1504.

Desiderio, S.V., Yancopoulos, G.D., Paskind, M., Thomas, E., Boss, M.A., Landau, N., Alt, F.W., and Baltimore, D. (1984). Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells. Nature *311*, 752–755.

Di Noia, J., and Neuberger, M.S. (2002). Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. Nature *419*, 43–48.

Difilippantonio, M.J., Zhu, J., Chen, H.T., Meffre, E., Nussenzweig, M.C., Max, E.E., Ried, T., and Nussenzweig, A. (2000). DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation. Nature *404*, 510–514.

Doetsch, P.W., and Cunningham, R.P. (1990). The enzymology of apurinic/apyrimidinic endonucleases. Mutat. Res. *236*, 173–201.

Dudás, A., and Chovanec, M. (2004). DNA double-strand break repair by homologous recombination. Mutat. Res. *566*, 131–167.

Ehrenstein, M.R., and Neuberger, M.S. (1999). Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. EMBO J. *18*, 3484–3490.

El-Khamisy, S.F., Masutani, M., Suzuki, H., and Caldecott, K.W. (2003). A requirement for PARP-1 for the assembly or stability of XRCC1 nuclear foci at sites of oxidative DNA damage. Nucleic Acids Res. *31*, 5526–5533.

Fattah, F., Lee, E.H., Weisensel, N., Wang, Y., Lichter, N., and Hendrickson, E.A. (2010). Ku regulates the non-homologous end joining pathway choice of DNA double-strand break repair in human somatic cells. PLoS Genet. *6*, e1000855.

Featherstone, K., Wood, A.L., Bowen, A.J., and Corcoran, A.E. (2010). The Mouse Immunoglobulin Heavy Chain V-D Intergenic Sequence Contains Insulators That May Regulate Ordered V(D)J Recombination. J. Biol. Chem. *285*, 9327–9338.

Ferrier, P., Krippl, B., Blackwell, T.K., Furley, A.J., Suh, H., Winoto, A., Cook, W.D., Hood, L., Costantini, F., and Alt, F.W. (1990). Separate elements control DJ and VDJ rearrangement in a transgenic recombination substrate. EMBO J. *9*, 117–125.

Fnu, S., Williamson, E.A., De Haro, L.P., Brenneman, M., Wray, J., Shaheen, M., Radhakrishnan, K., Lee, S.-H., Nickoloff, J.A., and Hromas, R. (2011). Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. Proc. Natl. Acad. Sci. U. S. A. *108*, 540–545.

Frank, K.M., Sekiguchi, J.M., Seidl, K.J., Swat, W., Rathbun, G.A., Cheng, H.L., Davidson, L., Kangaloo, L., and Alt, F.W. (1998). Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. Nature *396*, 173–177.

Frippiat, J.P., Williams, S.C., Tomlinson, I.M., Cook, G.P., Cherif, D., Le Paslier, D., Collins, J.E., Dunham, I., Winter, G., and Lefranc, M.P. (1995). Organization of the human immunoglobulin lambda light-chain locus on chromosome 22q11.2. Hum. Mol. Genet. *4*, 983–991.

Frit, P., Barboule, N., Yuan, Y., Gomez, D., and Calsou, P. (2014). Alternative end-joining pathway(s): bricolage at DNA breaks. DNA Repair *17*, 81–97.

Frosina, G., Fortini, P., Rossi, O., Carrozzino, F., Raspaglio, G., Cox, L.S., Lane, D.P., Abbondandolo, A., and Dogliotti, E. (1996). Two pathways for base excision repair in mammalian cells. J. Biol. Chem. *271*, 9573–9578.

Gao, Y., Sun, Y., Frank, K.M., Dikkes, P., Fujiwara, Y., Seidl, K.J., Sekiguchi, J.M., Rathbun, G.A., Swat, W., Wang, J., et al. (1998). A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. Cell *95*, 891–902.

Ghezraoui, H., Piganeau, M., Renouf, B., Renaud, J.-B., Sallmyr, A., Ruis, B., Oh, S., Tomkinson, A.E., Hendrickson, E.A., Giovannangeli, C., et al. (2014). Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. Mol. Cell *55*, 829–842.

Girelli Zubani, G., Zivojnovic, M., De Smet, A., Albagli-Curiel, O., Huetz, F., Weill, J.-C., Reynaud, C.-A., and Storck, S. (2017). Pms2 and uracil-DNA glycosylases act jointly in the mismatch repair pathway to generate Ig gene mutations at A-T base pairs. J. Exp. Med. *214*, 1169–1180.

Glaser, S.L., Lin, R.J., Stewart, S.L., Ambinder, R.F., Jarrett, R.F., Brousset, P., Pallesen, G., Gulley, M.L., Khan, G., O'Grady, J., et al. (1997). Epstein-Barr virus-associated Hodgkin's disease: epidemiologic characteristics in international data. Int. J. Cancer *70*, 375–382.

Goldberg, M., Stucki, M., Falck, J., D'Amours, D., Rahman, D., Pappin, D., Bartek, J., and Jackson, S.P. (2003). MDC1 is required for the intra-S-phase DNA damage checkpoint. Nature *421*, 952–956.

Gorman, J.R., van der Stoep, N., Monroe, R., Cogne, M., Davidson, L., and Alt, F.W. (1996). The Ig(kappa) enhancer influences the ratio of Ig(kappa) versus Ig(lambda) B lymphocytes. Immunity *5*, 241–252.

Haber, J.E. (2008). Alternative endings. Proc. Natl. Acad. Sci. *105*, 405–406.

Hagman, J., Rudin, C.M., Haasch, D., Chaplin, D., and Storb, U. (1990). A novel enhancer in the immunoglobulin lambda locus is duplicated and functionally independent of NF kappa B. Genes Dev. *4*, 978–992.

Haince, J.-F., McDonald, D., Rodrigue, A., Déry, U., Masson, J.-Y., Hendzel, M.J., and Poirier, G.G. (2008). PARP1-dependent kinetics of recruitment of MRE11 and NBS1 proteins to multiple DNA damage sites. J. Biol. Chem. *283*, 1197–1208.

Hamblin, T.J., Davis, Z., Gardiner, A., Oscier, D.G., and Stevenson, F.K. (1999). Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood *94*, 1848–1854.

Han, L., and Yu, K. (2008). Altered kinetics of nonhomologous end joining and class switch recombination in ligase IV–deficient B cells. J. Exp. Med. *205*, 2745–2753.

Han, L., Mao, W., and Yu, K. (2012). X-ray repair cross-complementing protein 1 (XRCC1) deficiency enhances class switch recombination and is permissive for alternative end joining. Proc. Natl. Acad. Sci. U. S. A. *109*, 4604–4608.

Haque, S.F.Y., Bevington, S.L., and Boyes, J. (2013). The Eλ3–1 enhancer is essential for V(D)J recombination of the murine immunoglobulin lambda light chain locus. Biochem. Biophys. Res. Commun. *441*, 482–487.

Heyer, W.-D., Li, X., Rolfsmeier, M., and Zhang, X.-P. (2006). Rad54: the Swiss Army knife of homologous recombination? Nucleic Acids Res. *34*, 4115–4125.

Higgins, G.S., Harris, A.L., Prevo, R., Helleday, T., McKenna, W.G., and Buffa, F.M. (2010). Overexpression of POLQ confers a poor prognosis in early breast cancer patients. Oncotarget *1*, 175–184.

Hsiao, K.-Y., and Mizzen, C.A. (2013). Histone H4 deacetylation facilitates 53BP1 DNA damage signaling and double-strand break repair. J. Mol. Cell Biol. *5*, 157–165.

Iacovoni, J.S., Caron, P., Lassadi, I., Nicolas, E., Massip, L., Trouche, D., and Legube, G. (2010). High-resolution profiling of γH2AX around DNA double strand breaks in the mammalian genome. EMBO J. *29*, 1446–1457.

IJspeert, H., Rozmus, J., Schwarz, K., Warren, R.L., van Zessen, D., Holt, R.A., Pico-Knijnenburg, I., Simons, E., Jerchel, I., Wawer, A., et al. (2016). XLF deficiency results in reduced N-nucleotide addition during V(D)J recombination. Blood *128*, 650–659.

Inlay, M., Alt, F.W., Baltimore, D., and Xu, Y. (2002). Essential roles of the kappa light chain intronic enhancer and 3' enhancer in kappa rearrangement and demethylation. Nat. Immunol. *3*, 463–468.

Jha, D.K., and Strahl, B.D. (2014). An RNA polymerase II-coupled function for histone H3K36 methylation in checkpoint activation and DSB repair. Nat. Commun. *5*, 3965.

Kabat, E.A., Wu, T.T., Foeller, C., Perry, H.M., and Gottesman, K.S. (1992). Sequences of Proteins of Immunological Interest (DIANE Publishing).

Kanaar, R., Hoeijmakers, J.H., and van Gent, D.C. (1998). Molecular mechanisms of DNA double strand break repair. Trends Cell Biol. *8*, 483–489.

Kanzler, H., Küppers, R., Hansmann, M.L., and Rajewsky, K. (1996). Hodgkin and Reed-Sternberg cells in Hodgkin's disease represent the outgrowth of a dominant tumor clone derived from (crippled) germinal center B cells. J. Exp. Med. *184*, 1495–1505.

Kent, T., Chandramouly, G., McDevitt, S.M., Ozdemir, A.Y., and Pomerantz, R.T. (2015). Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase θ. Nat. Struct. Mol. Biol. *22*, 230–237.

Kim, Y.-J., and Wilson, D.M. (2012). Overview of base excision repair biochemistry. Curr. Mol. Pharmacol. *5*, 3–13.

Kim, A., Han, L., Santiago, G.E., Verdun, R.E., and Yu, K. (2016). Class-Switch Recombination in the Absence of the IgH 3' Regulatory Region. J. Immunol. *197*, 2930–2935.

Klein S., Sablitzky F., and Radbruch A. (1984). Deletion of the IgH enhancer does not reduce immunoglobulin heavy chain production of a hybridoma IgD class switch variant. EMBO J. *3*, 2473–2476.

Koole, W., Schendel, R. van, Karambelas, A.E., Heteren, J.T. van, Okihara, K.L., and Tijsterman, M. (2014). A Polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. Nat. Commun. *5*, 3216.

Kottmann, A.H., Zevnik, B., Welte, M., Nielsen, P.J., and Köhler, G. (1994). A second promoter and enhancer element within the immunoglobulin heavy chain locus. Eur. J. Immunol. *24*, 817–821.

Krokan, H.E., and Bjørås, M. (2013). Base excision repair. Cold Spring Harb. Perspect. Biol. *5*, a012583.

Laffleur, B., Bardet, S.M., Garot, A., Brousse, M., Baylet, A., and Cogné, M. (2014). Immunoglobulin genes undergo legitimate repair in human B cells not only after *cis*- but also frequent *trans*-class switch recombination. Genes Immun. *15*, 341–346.

Langerak, P., and Russell, P. (2011). Regulatory networks integrating cell cycle control with DNA damage checkpoints and double-strand break repair. Philos. Trans. R. Soc. B Biol. Sci. *366*, 3562–3571.

Latif, C., den Elzen, N.R., and O'Connell, M.J. (2004). DNA damage checkpoint maintenance through sustained Chk1 activity. J. Cell Sci. *117*, 3489–3498.

Lavin, M.F. (2004). The Mre11 complex and ATM: a two-way functional interaction in recognising and signaling DNA double strand breaks. DNA Repair *3*, 1515–1520.

Lee, J.-H., and Paull, T.T. (2005). ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. Science *308*, 551–554.

Lemaître, C., Grabarz, A., Tsouroula, K., Andronov, L., Furst, A., Pankotai, T., Heyer, V., Rogier, M., Attwood, K.M., Kessler, P., et al. (2014). Nuclear position dictates DNA repair pathway choice. Genes Dev. *28*, 2450–2463.

Lennon, G.G., and Perry, R.P. (1985). C mu-containing transcripts initiate heterogeneously within the IgH enhancer region and contain a novel 5'-nontranslatable exon. Nature *318*, 475–478.

Levin-Klein, R., Kirillov, A., Rosenbluh, C., Cedar, H., and Bergman, Y. (2014). A Novel Pax5-Binding Regulatory Element in the Igκ Locus. Front. Immunol. *5*.

Li, L., and Wang, Y. (2017). Cross-talk between the H3K36me3 and H4K16ac histone epigenetic marks in DNA double-strand break repair. J. Biol. Chem. *292*, 11951–11959.

Li, S.C., Rothman, P.B., Zhang, J., Chan, C., Hirsh, D., and Alt, F.W. (1994). Expression of I mu-C gamma hybrid germline transcripts subsequent to immunoglobulin heavy chain class switching. Int. Immunol. *6*, 491–497.

Li, Y., Gao, X., and Wang, J.-Y. (2011). Comparison of two POLQ mutants reveals that a polymerase-inactive POLQ retains significant function in tolerance to etoposide and γ-irradiation in mouse B cells. Genes Cells Devoted Mol. Cell. Mech. *16*, 973–983.

Li, Z., Otevrel, T., Gao, Y., Cheng, H.L., Seed, B., Stamato, T.D., Taccioli, G.E., and Alt, F.W. (1995). The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. Cell *83*, 1079–1089.

Libra, M., De Re, V., Gasparotto, D., Gloghini, A., Marzotto, A., Milan, I., Tirelli, U., Stivala, F., Carbone, A., and Boiocchi, M. (2002). Differentiation between non-Hodgkin's lymphoma recurrence and second primary lymphoma by VDJ rearrangement analysis. Br. J. Haematol. *118*, 809–812.

Lieber, M.R. (2008). The mechanism of human nonhomologous DNA end joining. J. Biol. Chem. *283*, 1–5.

Lieberson, R., Giannini, S.L., Birshtein, B.K., and Eckhardt, L.A. (1991). An enhancer at the 3' end of the mouse immunoglobulin heavy chain locus. Nucleic Acids Res. *19*, 933–937.

Lim, D.S., and Hasty, P. (1996). A mutation in mouse rad51 results in an early embryonic lethal that is suppressed by a mutation in p53. Mol. Cell. Biol. *16*, 7133–7143.

Liu, Z.-M., George-Raizen, J.B., Li, S., Meyers, K.C., Chang, M.Y., and Garrard, W.T. (2002). Chromatin structural analyses of the mouse Igkappa gene locus reveal new hypersensitive sites specifying a transcriptional silencer and enhancer. J. Biol. Chem. *277*, 32640–32649.

Lorat, Y., Schanz, S., Schuler, N., Wennemuth, G., Rübe, C., and Rübe, C.E. (2012). Beyond repair foci: DNA double-strand break repair in euchromatic and heterochromatic compartments analyzed by transmission electron microscopy. PloS One *7*, e38165.

Lorenz, M., Jung, S., and Radbruch, A. (1995). How cytokines control immunoglobulin class switching. Behring Inst. Mitt. 97–102.

Ma, Y., Pannicke, U., Schwarz, K., and Lieber, M.R. (2002). Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. Cell *108*, 781–794.

Marafioti, T., Hummel, M., Anagnostopoulos, I., Foss, H.D., Falini, B., Delsol, G., Isaacson, P.G., Pileri, S., and Stein, H. (1997). Origin of nodular lymphocyte-predominant Hodgkin's disease from a clonal expansion of highly mutated germinal-center B cells. N. Engl. J. Med. *337*, 453–458.

Martomo, S.A., Saribasak, H., Yokoi, M., Hanaoka, F., and Gearhart, P.J. (2008). Reevaluation of the role of DNA polymerase theta in somatic hypermutation of immunoglobulin genes. DNA Repair *7*, 1603–1608.

Masamichi Muramatsu, Kazuo Kinoshita, and Sidonia Fagarasan (2000). Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme. 553–563.

Masani, S., Han, L., Meek, K., and Yu, K. (2016). Redundant function of DNA ligase 1 and 3 in alternative end-joining during immunoglobulin class switch recombination. Proc. Natl. Acad. Sci. U. S. A. *113*, 1261–1266.

Mateos-Gomez, P.A., Gong, F., Nair, N., Miller, K.M., Lazzerini-Denchi, E., and Sfeir, A. (2015). Mammalian Polymerase Theta Promotes Alternative-NHEJ and Suppresses Recombination. Nature *518*, 254–257.

Matthias, P. (1998). Lymphoid-specific transcription mediated by the conserved octamer site: Who is doing what? Semin. Immunol. *10*, 155–163.

Mattiroli, F., Vissers, J.H.A., van Dijk, W.J., Ikpa, P., Citterio, E., Vermeulen, W., Marteijn, J.A., and Sixma, T.K. (2012). RNF168 Ubiquitinates K13-15 on H2A/H2AX to Drive DNA Damage Signaling. Cell *150*, 1182–1195.

McVey, M., and Lee, S.E. (2008). MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet. TIG *24*, 529–538.

Meyer, K.B., and Neuberger, M.S. (1989). The immunoglobulin kappa locus contains a second, stronger B-cell-specific enhancer which is located downstream of the constant region. EMBO J. *8*, 1959–1964.

Moshous, D., Callebaut, I., de Chasseval, R., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., Tezcan, I., Sanal, O., Bertrand, Y., Philippe, N., et al. (2001). Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. Cell *105*, 177–186.

Mowatt, M.R., and Dunnick, W.A. (1986). DNA sequence of the murine gamma 1 switch segment reveals novel structural elements. J. Immunol. *136*, 2674–2683.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell *102*, 553–563.

Nakamura, M., Kondo, S., Sugai, M., Nazarea, M., Imamura, S., and Honjo, T. (1996). High frequency class switching of an IgM+ B lymphoma clone CH12F3 to IgA+ cells. Int. Immunol. *8*, 193–201.

Nikaido, T., Nakai, S., and Honjo, T. (1981). Switch region of immunoglobulin Cmu gene is composed of simple tandem repetitive sequences. Nature *292*, 845–848.

Nikaido, T., Yamawaki-Kataoka, Y., and Honjo, T. (1982). Nucleotide sequences of switch regions of immunoglobulin C epsilon and C gamma genes and their comparison. J. Biol. Chem. *257*, 7322–7329.

Nitschke, L., Kestler, J., Tallone, T., Pelkonen, S., and Pelkonen, J. (2001). Deletion of the DQ52 element within the Ig heavy chain locus leads to a selective reduction in VDJ recombination and altered D gene usage. J. Immunol. Baltim. Md 1950 *166*, 2540–2552.

Nussenzweig, A., Chen, C., da Costa Soares, V., Sanchez, M., Sokol, K., Nussenzweig, M.C., and Li, G.C. (1996). Requirement for Ku80 in growth and immunoglobulin V(D)J recombination. Nature *382*, 551–555.

Pan, Q., Petit-Frére, C., Dai, S., Huang, P., Morton, H.C., Brandtzaeg, P., and Hammarström, L. (2001). Regulation of switching and production of IgA in human B cells in donors with duplicated α1 genes. Eur. J. Immunol. *31*, 3622–3630.

Pan-Hammarström, Q., Lähdesmäki, A., Zhao, Y., Du, L., Zhao, Z., Wen, S., Ruiz-Perez, V.L., Dunn-Walters, D.K., Goodship, J.A., and Hammarström, L. (2006). Disparate roles of ATR and ATM in immunoglobulin class switch recombination and somatic hypermutation. J. Exp. Med. *203*, 99–110.

Pan-Hammarström, Q., Zhao, Y., and Hammarström, L. (2007). Class Switch Recombination: A Comparison Between Mouse and Human. In Advances in Immunology, (Academic Press), pp. 1–61.

Patenaude, A.-M., Orthwein, A., Hu, Y., Campo, V.A., Kavli, B., Buschiazzo, A., and Di Noia, J.M. (2009). Active nuclear import and cytoplasmic retention of activation-induced deaminase. Nat. Struct. Mol. Biol. *16*, 517–527.

Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina San-Martin, B., Barreto, V., et al. (2010). Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. Cell *143*, 122–133.

Pawlitzky, I., Angeles, C.V., Siegel, A.M., Stanton, M.L., Riblet, R., and Brodeur, P.H. (2006). Identification of a Candidate Regulatory Element within the 5' Flanking Region of the Mouse Igh Locus Defined by Pro-B Cell-Specific Hypersensitivity Associated with Binding of PU.1, Pax5, and E2A. J. Immunol. *176*, 6839–6851.

Perlot, T., Alt, F.W., Bassing, C.H., Suh, H., and Pinaud, E. (2005). Elucidation of IgH intronic enhancer functions via germ-line deletion. Proc. Natl. Acad. Sci. U. S. A. *102*, 14362–14367.

Péron, S., Metin, A., Gardès, P., Alyanakian, M.-A., Sheridan, E., Kratz, C.P., Fischer, A., and Durandy, A. (2008). Human PMS2 deficiency is associated with impaired immunoglobulin class switch recombination. J. Exp. Med. *205*, 2465–2472.

Péron, S., Laffleur, B., Denis-Lagache, N., Cook-Moreau, J., Tinguely, A., Delpy, L., Denizot, Y., Pinaud, E., and Cogné, M. (2012). AID-Driven Deletion Causes Immunoglobulin Heavy Chain Locus Suicide Recombination in B Cells. Science *336*, 931–934.

Pettigrew, C.A., and Cotter, T.G. (2009). Deregulation of Cell Death (Apoptosis): Implications for Tumor Development. Discov. Med. *8*, 61–63.

Pfister, S.X., Ahrabi, S., Zalmas, L.-P., Sarkar, S., Aymard, F., Bachrati, C.Z., Helleday, T., Legube, G., La Thangue, N.B., Porter, A.C.G., et al. (2014). SETD2-dependent histone H3K36 trimethylation is required for homologous recombination repair and genome stability. Cell Rep. *7*, 2006–2018.

Pierce, A.J., Hu, P., Han, M., Ellis, N., and Jasin, M. (2001). Ku DNA end-binding protein modulates homologous repair of double-strand breaks in mammalian cells. Genes Dev. *15*, 3237–3242.

Pinaud, E., Aupetit, C., Chauveau, C., and Cogné, M. (1997). Identification of a homolog of the C alpha 3'/hs3 enhancer and of an allelic variant of the 3'IgH/hs1,2 enhancer downstream of the human immunoglobulin alpha 1 gene. Eur. J. Immunol. *27*, 2981–2985.

Pinaud, E., Khamlichi, A.A., Le Morvan, C., Drouet, M., Nalesso, V., Le Bert, M., and Cogné, M. (2001). Localization of the 3' IgH locus elements that effect long-distance regulation of class switch recombination. Immunity *15*, 187–199.

Pinaud, E., Marquet, M., Fiancette, R., Péron, S., Vincent-Fabert, C., Denizot, Y., and Cogné, M. (2011). Chapter 2 - The IgH Locus 3' Regulatory Region: Pulling the Strings from Behind. In Advances in Immunology, F.W. Alt, K.F. Austen, T. Honj, F. Melchers, J.W. Uhr, and E.R. Unanue, eds. (Academic Press), pp. 27–70.

Porter (1973). STRUCTURAL STUDIES OF IMMUNOGLOBULINS. Science 713–716.

Prakash, R., Zhang, Y., Feng, W., and Jasin, M. (2015). Homologous Recombination and Human Health: The Roles of BRCA1, BRCA2, and Associated Proteins. Cold Spring Harb. Perspect. Biol. *7*, a016600.

Prasad, R., Lavrik, O.I., Kim, S.J., Kedar, P., Yang, X.P., Vande Berg, B.J., and Wilson, S.H. (2001). DNA polymerase beta -mediated long patch base excision repair. Poly(ADP-ribose)polymerase-1 stimulates strand displacement DNA synthesis. J. Biol. Chem. *276*, 32411–32414.

Preud'homme, J.L., Petit, I., Barra, A., Morel, F., Lecron, J.C., and Lelièvre, E. (2000). Structural and functional properties of membrane and secreted IgD. Mol. Immunol. *37*, 871–887.

R L Coffman, J Ohara, M W Bond, J Carty, A Zlotnik, and W E Paul (1986). B cell stimulatory factor-1 enhances the IgE response of lipopolysaccharide-activated B cells. J. Immunol.

Rada, C., Williams, G.T., Nilsen, H., Barnes, D.E., Lindahl, T., and Neuberger, M.S. (2002). Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. Curr. Biol. CB *12*, 1748–1755.

Rada, C., Di Noia, J.M., and Neuberger, M.S. (2004). Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. Mol. Cell *16*, 163–171.

Rajewsky, K. (1996). Clonal selection and learning in the antibody system. Nature *381*, 751–758.

Reina-San-Martin, B., Chen, J., Nussenzweig, A., and Nussenzweig, M.C. (2007). Enhanced intra-switch region recombination during immunoglobulin class switch recombination in 53BP1-/- B cells. Eur. J. Immunol. *37*, 235–239.

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A., et al. (2000). Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). Cell *102*, 565–575.

Rijkers, T., Van Den Ouweland, J., Morolli, B., Rolink, A.G., Baarends, W.M., Van Sloun, P.P., Lohman, P.H., and Pastink, A. (1998). Targeted inactivation of mouse RAD52 reduces homologous recombination but not resistance to ionizing radiation. Mol. Cell. Biol. *18*, 6423–6429.

Robert, I., Dantzer, F., and Reina-San-Martin, B. (2009). Parp1 facilitates alternative NHEJ, whereas Parp2 suppresses IgH/c-myc translocations during immunoglobulin class switch recombination. J. Exp. Med. *206*, 1047–1056.

Robertson, A.B., Klungland, A., Rognes, T., and Leiros, I. (2009). DNA repair in mammalian cells: Base excision repair: the long and short of it. Cell. Mol. Life Sci. CMLS *66*, 981–993.

Rodriguez-Abreu, D., Bordoni, A., and Zucca, E. (2007). Epidemiology of hematological malignancies. Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. *18 Suppl 1*, i3–i8.

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.

Rooney, S., Sekiguchi, J., Zhu, C., Cheng, H.L., Manis, J., Whitlow, S., DeVido, J., Foy, D., Chaudhuri, J., Lombard, D., et al. (2002). Leaky Scid phenotype associated with defective V(D)J coding end processing in Artemis-deficient mice. Mol. Cell *10*, 1379–1390.

Rooney, S., Chaudhuri, J., and Alt, F.W. (2004). The role of the non-homologous end-joining pathway in lymphocyte development. Immunol. Rev. *200*, 115–131.

Roos, W.P., and Kaina, B. (2006). DNA damage-induced cell death by apoptosis. Trends Mol. Med. *12*, 440–450.

Rouaud, P., Saintamand, A., Saad, F., Carrion, C., Lecardeur, S., Cogné, M., and Denizot, Y. (2014). Elucidation of the enigmatic IgD class-switch recombination via germline deletion of the IgH 3' regulatory region. J. Exp. Med. *211*, 975–985.

Saintamand, A., Rouaud, P., Garot, A., Saad, F., Carrion, C., Oblet, C., Cogné, M., Pinaud, E., and Denizot, Y. (2015). The IgH 3' regulatory region governs μ chain transcription in mature B lymphocytes and the B cell fate. Oncotarget *6*, 4845–4852.

Saribasak, H., Maul, R.W., Cao, Z., McClure, R.L., Yang, W., McNeill, D.R., Wilson, D.M., and Gearhart, P.J. (2011). XRCC1 suppresses somatic hypermutation and promotes alternative nonhomologous end joining in Igh genes. J. Exp. Med. *208*, 2209–2216.

Sartori, A.A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., and Jackson, S.P. (2007). Human CtIP promotes DNA end resection. Nature *450*, 509–514.

Schrader, C.E., Vardo, J., and Stavnezer, J. (2003). Mlh1 Can Function in Antibody Class Switch Recombination Independently of Msh2. J. Exp. Med. *197*, 1377–1383.

Schrader, C.E., Guikema, J.E.J., Linehan, E.K., Selsing, E., and Stavnezer, J. (2007). Activation-induced cytidine deaminase-dependent DNA breaks in class switch recombination occur during G1 phase of the cell cycle and depend upon mismatch repair. J. Immunol. Baltim. Md 1950 *179*, 6064–6071.

Schroeder, H.W., Zemlin, M., Khass, M., Nguyen, H.H., and Schelonka, R.L. (2010). Genetic control of DH reading frame and its effect on B-cell development and antigen-specifc antibody production. Crit. Rev. Immunol. *30*, 327–344.

Schubart, D.B., Rolink, A., Kosco-Vilbois, M.H., Botteri, F., and Matthias, P. (1996). B-cell-specif ic coactivator OBF-1/OCA-B/Bob1 required for immune response and germinal centre formation. Nature *383*, 538–542.

Schwering, I., Bräuninger, A., Klein, U., Jungnickel, B., Tinguely, M., Diehl, V., Hansmann, M.-L., Dalla-Favera, R., Rajewsky, K., and Küppers, R. (2003). Loss of the B-lineage-specific gene expression program in Hodgkin and Reed-Sternberg cells of Hodgkin lymphoma. Blood *101*, 1505–1512.

Singleton, B.K., Torres-Arzayus, M.I., Rottinghaus, S.T., Taccioli, G.E., and Jeggo, P.A. (1999). The C Terminus of Ku80 Activates the DNA-Dependent Protein Kinase Catalytic Subunit. Mol. Cell. Biol. *19*, 3267–3277.

Smith, L.B. (2010). Nodular lymphocyte predominant Hodgkin lymphoma: diagnostic pearls and pitfalls. Arch. Pathol. Lab. Med. *134*, 1434–1439.

Snapper CM, Peschel C, and Paul WE (1988). IFN-gamma stimulates IgG2a secretion by murine B cells stimulated with bacterial lipopolysaccharide. J. Immunol.

Solinger, J.A., and Heyer, W.-D. (2001). Rad54 protein stimulates the postsynaptic phase of Rad51 protein-mediated DNA strand exchange. Proc. Natl. Acad. Sci. *98*, 8447–8453.

Soni, A., Siemann, M., Grabos, M., Murmann, T., Pantelias, G.E., and Iliakis, G. (2014). Requirement for Parp-1 and DNA ligases 1 or 3 but not of Xrcc1 in chromosomal translocation formation by backup end joining. Nucleic Acids Res. *42*, 6380–6392.

Soulas-Sprauel, P., Le Guyader, G., Rivera-Munoz, P., Abramowski, V., Olivier-Martin, C., Goujet-Zalc, C., Charneau, P., and de Villartay, J.-P. (2007). Role for DNA repair factor XRCC4 in immunoglobulin class switch recombination. J. Exp. Med. *204*, 1717–1727.

Stavnezer, J., Guikema, J.E.J., and Schrader, C.E. (2008). Mechanism and regulation of class switch recombination. Annu. Rev. Immunol. *26*, 261–292.

Stavnezer, J., Björkman, A., Du, L., Cagigi, A., and Pan-Hammarström, Q. (2010). Chapter 3 - Mapping of Switch Recombination Junctions, a Tool for Studying DNA Repair Pathways during Immunoglobulin Class Switching. In Advances in Immunology, F.W. Alt, K.F. Austen, T. Honjo, F. Melchers, J.W. Uhr, and E.R. Unanue, eds. (Academic Press), pp. 45–109.

Stewart, G.S., Wang, B., Bignell, C.R., Taylor, A.M.R., and Elledge, S.J. (2003). MDC1 is a mediator of the mammalian DNA damage checkpoint. Nature *421*, 961–966.

Stucki, M., and Jackson, S.P. (2006). gammaH2AX and MDC1: anchoring the DNA-damage-response machinery to broken chromosomes. DNA Repair *5*, 534–543.

Tang, H., and Sharp, P.A. (1999). Transcriptional regulation of the murine 3' IgH enhancer by OCT-2. Immunity *11*, 517–526.

Tang, J., Cho, N.W., Cui, G., Manion, E.M., Shanbhag, N.M., Botuyan, M.V., Mer, G., and Greenberg, R.A. (2013). Acetylation limits 53BP1 association with damaged chromatin to promote homologous recombination. Nat. Struct. Mol. Biol. *20*, 317–325.

Taylor, G.C.A., Eskeland, R., Hekimoglu-Balkan, B., Pradeepa, M.M., and Bickmore, W.A. (2013). H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction. Genome Res. *23*, 2053–2065.

Truong, L.N., Li, Y., Shi, L.Z., Hwang, P.Y.-H., He, J., Wang, H., Razavian, N., Berns, M.W., and Wu, X. (2013). Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. Proc. Natl. Acad. Sci. U. S. A. *110*, 7720–7725.

Tsuzuki, T., Fujii, Y., Sakumi, K., Tominaga, Y., Nakao, K., Sekiguchi, M., Matsushiro, A., Yoshimura, Y., and MoritaT, null (1996). Targeted disruption of the Rad51 gene leads to lethality in embryonic mice. Proc. Natl. Acad. Sci. U. S. A. *93*, 6236–6240.

Vincent-Fabert, C., Fiancette, R., Pinaud, E., Truffinet, V., Cogné, N., Cogné, M., and Denizot, Y. (2010). Genomic deletion of the whole IgH 3' regulatory region (hs3a, hs1,2, hs3b, and hs4) dramatically affects class switch recombination and Ig secretion to all isotypes. Blood *116*, 1895–1898.

Wang, M., Wu, W., Wu, W., Rosidi, B., Zhang, L., Wang, H., and Iliakis, G. (2006). PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. Nucleic Acids Res. *34*, 6170–6182.

Wuerffel, R., Wang, L., Grigera, F., Manis, J., Selsing, E., Perlot, T., Alt, F.W., Cogne, M., Pinaud, E., and Kenter, A.L. (2007). S-S Synapsis during Class Switch Recombination Is Promoted by Distantly Located Transcriptional Elements and Activation-Induced Deaminase. Immunity *27*, 711–722.

Xu, Y., Davidson, L., Alt, F.W., and Baltimore, D. (1996). Deletion of the Ig kappa light chain intronic enhancer/matrix attachment region impairs but does not abolish V kappa J kappa rearrangement. Immunity *4*, 377–385.

Xue, K., Rada, C., and Neuberger, M.S. (2006). The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in msh2−/− ung−/− mice. J. Exp. Med. *203*, 2085–2094.

Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H., Robbiani, D.F., McBride, K., Nussenzweig, M.C., and Casellas, R. (2011). Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. Nat. Immunol. *12*, 62–69.

Yan, C.T., Boboila, C., Souza, E.K., Franco, S., Hickernell, T.R., Murphy, M., Gumaste, S., Geyer, M., Zarrin, A.A., Manis, J.P., et al. (2007). IgH class switching and translocations use a robust non-classical end-joining pathway. Nature *449*, 478–482.

Yousefzadeh, M.J., Wyatt, D.W., Takata, K., Mu, Y., Hensley, S.C., Tomida, J., Bylund, G.O., Doublié, S., Johansson, E., Ramsden, D.A., et al. (2014). Mechanism of Suppression of Chromosomal Instability by DNA Polymerase POLQ. PLoS Genet. *10*.

Yu, X., and Baer, R. (2000). Nuclear localization and cell cycle-specific expression of CtIP, a protein that associates with the BRCA1 tumor suppressor. J. Biol. Chem. *275*, 18541–18549.

Zachau, H.G. (2000). The immunoglobulin kappa gene families of human and mouse: a cottage industry approach. Biol. Chem. *381*, 951–954.

Zan, H., Tat, C., Qiu, Z., Taylor, J.R., Guerrero, J.A., Shen, T., and Casali, P. (2017). Rad52 competes with Ku70/Ku86 for binding to S-region DSB ends to modulate antibody class-switch DNA recombination. Nat. Commun. *8*.

Zannini, L., Delia, D., and Buscemi, G. (2014). CHK2 kinase in the DNA damage response and beyond. J. Mol. Cell Biol. *6*, 442–457.

Zarrin, A.A., Tian, M., Wang, J., Borjeson, T., and Alt, F.W. (2005). Influence of switch region length on immunoglobulin class switch recombination. Proc. Natl. Acad. Sci. U. S. A. *102*, 2466–2470.

Zhang, Z.Z., Pannunzio, N.R., Han, L., Hsieh, C.-L., Yu, K., and Lieber, M.R. (2014). The strength of an Ig switch region is determined by its ability to drive R loop formation and its number of WGCW sites. Cell Rep. *8*, 557–569.

Zheng, K., Wu, R., He, Y., Xiao, S., Zhang, J., Liu, J., Hao, Y., and Tan, Z. (2014). A competitive formation of DNA:RNA hybrid G-quadruplex is responsible to the mitochondrial transcription termination at the DNA replication priming site. Nucleic Acids Res. *42*, 10832–10844.

Zhou, X., Xiang, Y., and Garrard, W.T. (2010). The Igκ gene enhancers, E3' and Ed, are essential for triggering transcription. J. Immunol. Baltim. Md 1950 *185*, 7544–7552.

Zhou, X., Xiang, Y., Ding, X., and Garrard, W.T. (2012). A new hypersensitive site, HS10, and the enhancers E3' and Ed, differentially regulate Igκ gene expression. J. Immunol. Baltim. Md 1950 *188*, 2722–2732.

Zhuang, Y., Soriano, P., and Weintraub, H. (1994). The helix-loop-helix gene E2A is required for B cell formation. Cell *79*, 875–884.

ANNEXES

Mes résultats de thèse ont été présentés lors de congrès internationaux.

Anomalie de la Recombinaison suicide du locus IgH dans la maladie de Hodgkin

H. Boutouil (1) ; J. Moreau (1) ; J. Feuillard (2) ; M. Cogné (1) ; S. Peron (1)

(1) CRIBL, UMR CNRS 7276 INSERM 1262, Limoges; (2) Hématologie Biologique, CHU Limoges, Limoges.

Le **BCR** (pour récepteur des lymphocytes B) et son précurseur, le pré-BCR, sont cruciaux pour le développement des lymphocytes B (LB) des premiers stades jusqu'à leur maturation terminale et pour la fonction efficace des LB dans les réponses immunitaires. La **recombinaison suicide du locus IgH** (**LSR**) est une recombinaison entre la région de commutation (S) et la région répétitive like-switch (LS) dans la région régulatrice en 3') du locus IgH (Peron et al., Science 2012) qui induit la perte d'expression du BCR par le LB qui la subit. La LSR apparait comme la contrepartie nuisible de la commutation isotypique des Ig (CSR) dans les LB activés. La LSR semble être un élément de la régulation de l'homéostasie des LB.

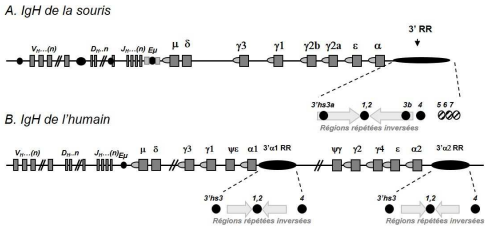


Figure 1. Locus IgH de la souris (A) et de l'humain (B). Représentation schématique des régions variable (segments V,D et J) et constante (gènes m, d, g, e et a). Les éléments régulateurs sont représentés par les disques noirs et rayés. (Figure adaptée à partir de Pinaud et al., Adv. In Immunol. 2011).

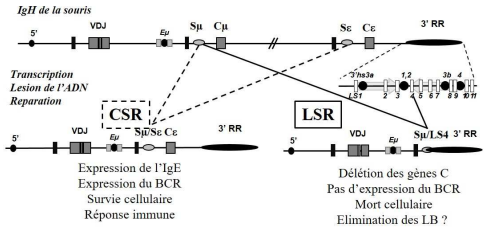
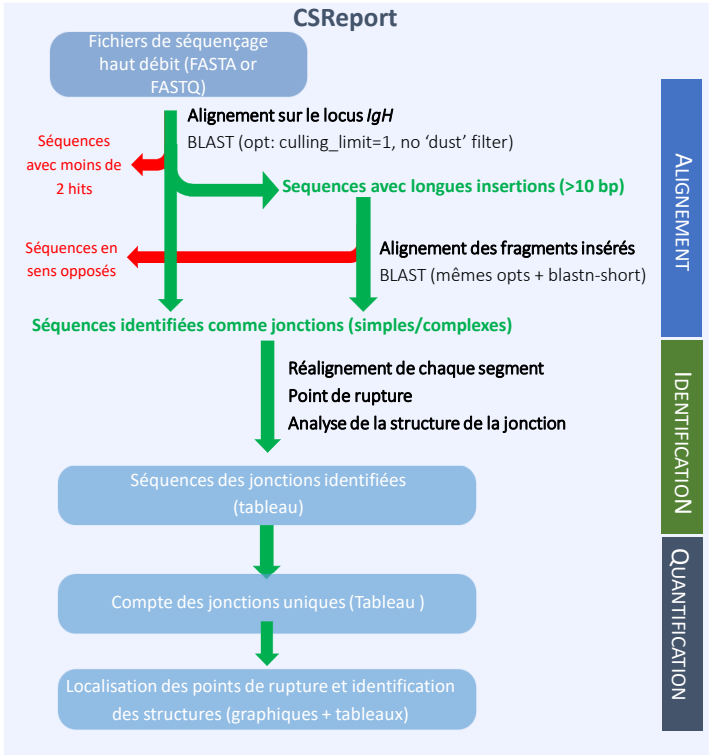


Figure 2. LSR et CSR. La CSR est une recombinaison génétique entre deux regions S et permet l'expression d'un isotype d'Ig autre que l'IgM. La LSR est un rearrangement de l'ADN entre les régions S et LS, localisée au sein de la 3'RR, qui résulte en l'excision de tous les gènes qui codent pour la partie constante de l'Ig.

Le lymphome de Hodgkin est d'origine cellulaire B monoclonale et se caractérise par l'absence d'expression de BCR sur les cellules tumorales (Marafioti et al., Blood, 2000). La perte d'expression des Ig dans les cellules de Reed-Sternberg est le résultat de mutations non sens dans les gènes codant pour la partie variable de l'IgH. Des anomalies régulatrices de la transcription dans le gène IgH ont également été incriminées. Nous étudions la possibilité que le LSR entraîne une perte d'expression de l'IgH et l'absence d'expression de la BCR par les cellules de lymphome de Hodgkin.

Matériel et méthodes

Des échantillons de ganglions lymphatiques congelés de la maladie de Hodgkin, avec une grande infiltration de cellules, ont été utilisés pour préparer l'ADN à partir duquel nous amplifions et quantifions les jonctions LSR en utilisant notre méthode de séquençage à haut débit et l'outil CSReport (Boyer et al., Journal of Immunology 2017).



Résultats

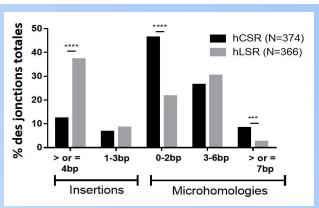


Figure 3. Structure de réparation des jonctions de CSR et de LSR chez l'humain. Pour la CSR (en noir) et la LSR (en gris) sont représentés les % de jonctions de réparation caractérisées par la présence d'insertions ou de microhomologies. Ces résultats ont été obtenus par amplification des jonctions de LSR et CSR à partir de PBMC stimulés *in vitro*, leur séquençage sur Ion Proton (Life) et l'analyse des données par le logiciel CSReport. Test statistique c2, *** : P≤0.005, **** : P≤0.001.

Echantillons sains	Nbre jonctions LSR	Moyenne Jonctions LSR	Echantillons de maladie de Hodgkin	Nbre jonctions LSR	Moyenne Jonctions LSR
Tonsil 89	14	52,9	HL 17	0	11,4
			HL 18	0	
Tonsil 90	61		HL 19	1	
			HL 20	1	
Tonsil 91	104		HL 21	2	
Tonsil 92	23		HL 55	11	
			HL 56	0	
Tonsil 93	143		HL 57	7	
Tonsil 94	21		HL 58	10	
Tonsil 55	11		HL 59	0	
Tonsil 56	56		HL 60	2	
Tonsil 57	26		HL 61	0	
Tonsil 58	30		HL 62	2	
Tonsil 59	17		HL 63	1	
Tonsil 60	79		HL 64	1	
Tonsil 61	95		HL 65	47	
Tonsil 62	12		HL 66	6	
Tonsil 63	25		HL 67	8	
Tonsil 64	129		HL 68	118	

Figure 4. Analyse du nombre de jonctions de LSR. Le nombre de jonctions de LSR a été déterminé à partir d'amygdales de témoins sains ou de patients atteints de maladie de Hodgkin. L'analyse par CSReport permet d'énumérer le nombre de jonctions uniques de LSR après séquencage des produits de PCR amplifiés à partir de l'ADN de chaque prélèvement.

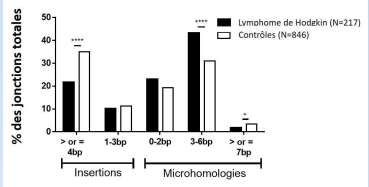


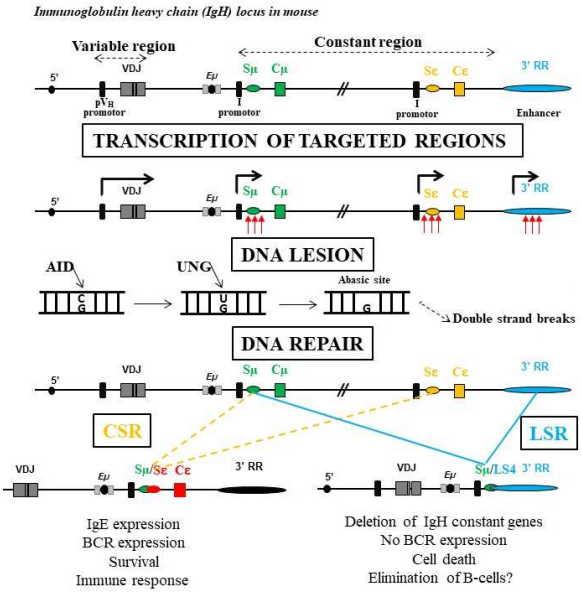
Figure 5. Structure de réparation des jonctions de LSR dans des ganglions sains et de patients atteints de Lymphome de Hodgkin. Les pourcentages de jonctions de réparation LSR sont représentées en fonction de leur structure avec insertions ou microhomologies. Ces résultats ont été obtenus par amplifications des jonctions de LSR et CSR à partir d'ADN de coupes de ganglions de témoins sains (blanc) ou de patients malades de lymphome de Hodgkin. Les jonctions ont été séquencées sur Ion Proton (Life) et les données ont été analysées avec le logiciel CSReport. Test statistique c2, *: P≤0.05, **** : P≤0.001.

L'analyse quantitative des résultats déterminera si la perte d'expression de BCR dans le lymphome de Hodgkin peut être attribuée à la recombinaison de LSR. Par ailleurs, l'anomalie de la réparation de la LSR observée dans les échantillons de ganglions de la maladie de Hodgkin sera mieux documentée afin de déterminer si elle est spécifique des cellules de Reed-Sternberg et si elle reflète des dérégulations de la réparation de l'ADN associées à la transformation tumorale.

Péron S, Laffleur B, Denis-Lagache N, Cook-Moreau J, Tinguely A, Delpy L, Denizot Y, Pinaud E, Cogné M. AID-driven deletion causes immunoglobulin heavy chain locus suicide recombination in B-cells. Science. 2012; Pinaud E, Marquet M, Fiancette R, Péron S, Vincent-Fabert C, Denizot Y, Cogné M. The IgH locus 3' regulatory region: pulling the strings from behind. Adv Immunol. 2011; Marafioti T, Hummel M, Foss HD, Laumen H, Korbjuhn P, Anagnostopoulos I, Lammert H, Demel G, Theil J, Wirth T, Stein H. Hodgkin and Reed-Sternberg cells represent an expansion of a single clone originating from a germinal center B-cell with functional immunoglobulin gene rearrangements but defective immunoglobulin transcription. Blood. 2000 ; Boyer F, Boutouil H, Dalloul I, Dalloul Z, Cook-Moreau J, Aldigier JC, Carrion C, Herve B, Scaon E, Cogné M, Péron S. CSReport: A New Computational Tool Designed for Automatic Analysis of Class Switch Recombination Junctions Sequenced by High-Throughput Sequencing. J Immunol. 2017.

Class Switch Recombination and Locus Suicide Recombination both target the IgH locus but employ different DNA DSB repair pathways

H. Boutouil (1), F. Boyer, J. Moreau (1), M. Cogné (1), S. Peron (1).
(1) CRIBL, UMR CNRS 7276 INSERM 1262, Limoges.

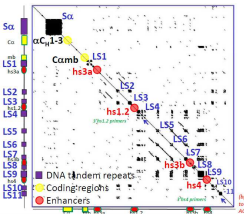


In mature B-cells, *immunoglobulin heavy chain* (*IgH*) locus DNA remodeling can occur during several **genetic rearrangements**:

-**Class switch recombination** (CSR) intends for B-cells to ensure their functions in immune responses

-**Locus suicide recombination** (LSR) eliminates B-cells due to death induced by loss of IgH expression (Péron et al., Science 2012).

CSR and LSR operate by the **same molecular mechanisms**: 1) non-coding transcription of target regions, 2) DNA lesions by Activation-Induced Cytidine Deaminase (AID) leading to DNA double-strand breaks (DSB), and 3) DNA repair.



"Like Switch" repeats within the mouse 3'RR.

The mouse IgH locus 3' regulatory region (3'RR) contains four enhancers (hs3a, 1-2, 3b and 4) with strong B-lineage specificity and contains Switch like sequences targeted by AID.

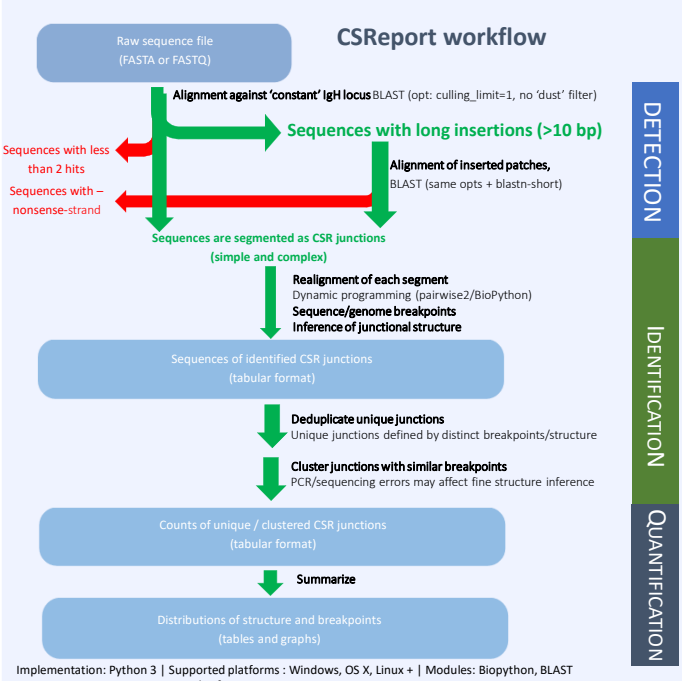
Like Sγ, where 5bp repeats are arranged within higher order 49bp repeats, dot plot analysis of the mouse 3'RR revealed eleven 0.5-1kb long "like Sγ" stretches (LS regions) where 49bp repeats appeared with a threshold set above 60% identity.

We performed similar analysis on the 3'RR from human, rabbit and dog : in all four species, LS regions flank both sides of the 3'RR and are interspersed with enhancers.

Analysis was done with the program Nucleic Acid Dot plot (http://www.vivo.colostate.edu/molkit/dnadot/index.html).
(Péron et al., Science 2012)

Material and methods

Using high throughput sequencing of LSR and CSR junctions and the computational tool CSReport we developed to analyze data (Boyer et al., Journal of Immunol. 2017), we analysed structural features of LSR junctions.



Implementation: Python 3 | Supported platforms : Windows, OS X, Linux + | Modules: Biopython, BLAST
Running time (laptop): few hours / 10^6 sequences (both strands)

Results

We analysed LSR and CSR junctions using high throughput sequencing method and the tool CSReport. DNA from tonsils from healthy controls (N=6 and N=16), from lymph node samples from Hodgkin disease patients(N=19) or from peripheral blood mononuclear cells from chronic lymphoïd leukemia (CLL) samples (N=8) was used to amplify LSR and CSR junctions. In vivo activated B-cells from wild type murine spenocytes (N=4) were used to amplify LSR and CSR junctions. LSR and CSR amplifications were used to prepare sequencing libraries (Ion Torrent technology, Ion proton sequencer, Life Technology).

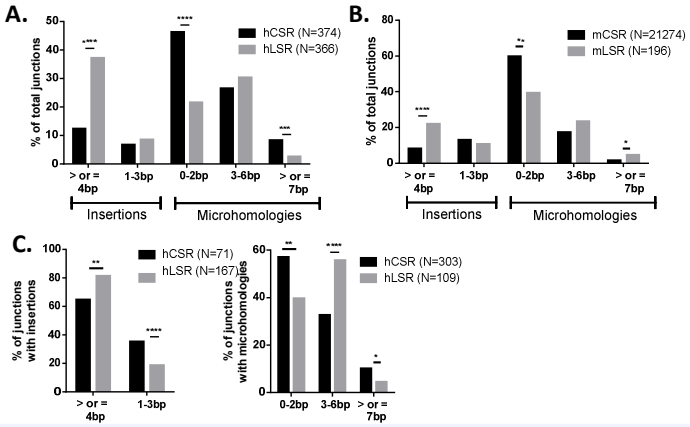


Figure 2. Structural features of LSR and CSR junctions.
We examined the structural features of LSR and CSR junctions in human and murine samples. Surprisingly, we observed similar results in human (**A.**) and murine (**B.**) activated B-cells of different junction structure features between LSR and CSR. **C.** In LSR blunt junctions and junctions with microhomologies less than 3bp frequencies were decreased and junctions with insertions longer than 4bp were increased compared to CSR.

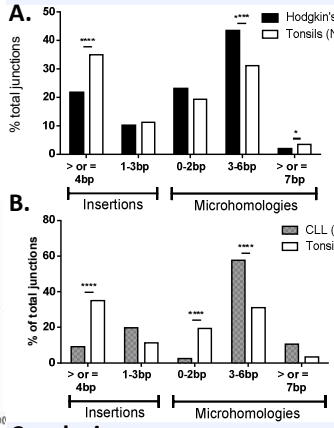
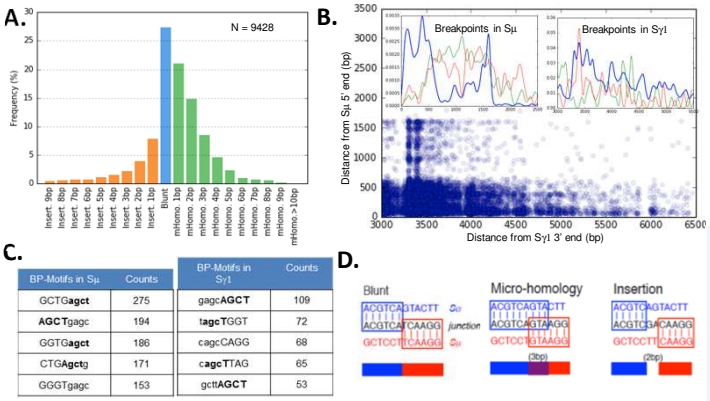


Figure 3. Abnormal LSR junction structure in B-cell lymphomas.
In samples from Hodgkin disease (A.) and CLL (B.), HTS and CSReport analysis revealed structural features of LSR junctions different from LSR junctions in healthy controls.

Frequency of junctions with small microhomologies (3bp-6bp) was increased and junctions with insertions longer than 4bp were decreased in lymphomas compared to normal samples.



BP-Motifs in Sμ	Counts	BP-Motifs in Sγ1	Counts
GCTG**agct**	275	gagc**AGCT**	109
AGCTgagc	194	**tagc**TGGT	72
GGTG**agct**	186	cagcCAGG	68
CTG**Agctg**	171	**cagc**TTAG	65
GGGTgagc	153	gctt**AGCT**	53

Figure 1. Statistics of Sμ-Sγ1 junctions using CSReport. Murine splenocytes were stimulated *in vitro* with LPS+ IL4. After 3 days, cells were recovered and lysed for DNA extraction and then CSR junctions were amplified and sequenced by HTS. Sequences were analyzed using CSReport. **A.** Structural profile of Sμ-Sγ1 junctions. **B.** Distribution of breakpoints in Sμ and Sγ1. **C.** Most frequent 8nt-motifs at breakpoints in Sμ and Sγ1. D. Identification of junction structure from two pairwise alignments. Break points are determined with 1-bp resolution for accurate structure assignment.

Conclusion

LSR appears to be the detrimental counterpart of CSR in activated B-cells, so an important question to address is the regulation of LSR. **Our preliminary data suggest that DNA DSB repair implicated during LSR is different from that for CSR**.

We will determine which DSB repair pathway is channeled during LSR by studying its occurrence in mouse models invalidated for DSB repair components (**XRCC4, 53BP1, Polymerase theta**) and analysing their recruitment to LSR target sequences by **ChIP experiments**. We will also document the **chromatin marks** known to orient the DSB DNA repair pathway to the 3'RR in activated B-cells.

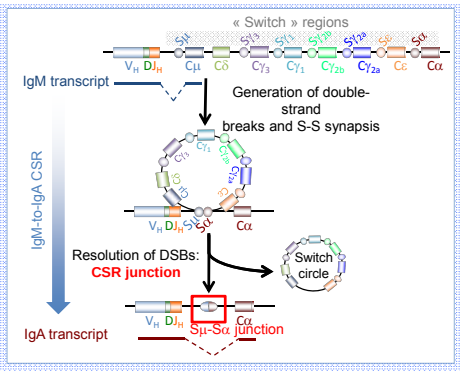
Péron S, Laffleur B, Denis-Lagache N, Cook-Moreau J, Tinguely A, Delpy L, Denizot Y, Pinaud E, Cogné M. AID-driven deletion causes immunoglobulin heavy chain locus suicide recombination in B-cells. Science. 2012;
Boyer F, Boutouil H, Dalloul I, Dalloul Z, Cook-Moreau J, Aldigier JC, Carrion C, Herve B, Scaon E, Cogné M, Péron S. CSReport: A New Computational Tool Designed for Automatic Analysis of Class Switch Recombination Junctions Sequenced by High-Throughput Sequencing. J Immunol. 2017.

CSReport : a new software tool designed for automatic analysis of Class Switch Recombination Junctions sequenced by High Throughput Sequencing

François Boyer†*, Hend Boutouil†*, Iman Dalloul†, Zeinab Dalloul†, Jeanne Cook-Moreau†, Jean-Claude Aldigier†, Claire Carrion†, Michel Cogné†‡ and Sophie Péron†.
† CRIBL, Université de Limoges - UMR CNRS 7276, Limoges, France.
‡ Institut Universitaire de France, F-75000 Paris, France.

Class switch recombination…

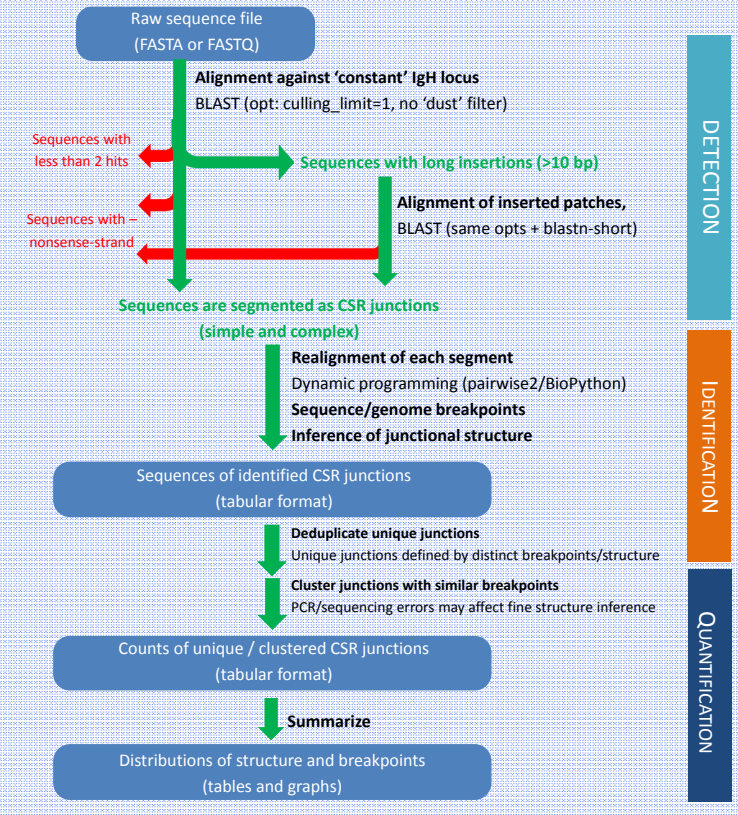
B-cells ensure humoral immune responses due to the production of antigen (Ag) -specific memory B-cells and antibody secreting plasma cells. In secondary lymphoid organs, Ag-driven B-cell activation induces terminal maturation and immunoglobulin (Ig) isotype class switch (class switch recombination, CSR). CSR creates a virtually unique *IgH* locus in every B cell clone by intra-chromosomal recombination between two switch (S) regions upstream of each constant (C) region genes. Amount and structural features of CSR junctions reveal valuable information about the CSR mechanism.



…with a deeper insight

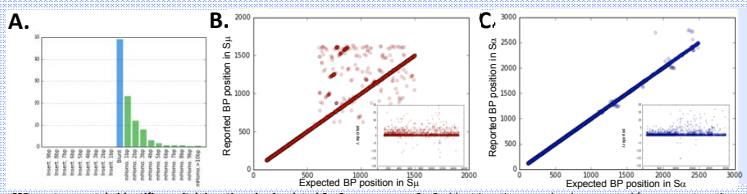
To provide an automated tool able to analyze large data sets of CSR junction sequences produced by high throughput sequencing (HTS) we designed **CSReport,**, a software progarm dedicated to support analysis of CSR recombination junctions sequenced with a HTS-based protocol (Ion Torrent technology).

CSReport workflow

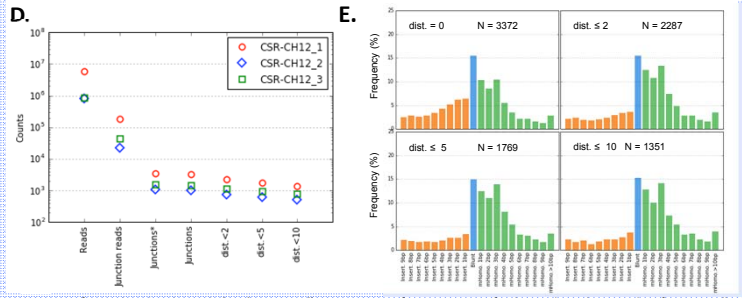


Implementation: Python 3 | Supported platforms : Windows, OS X, Linux + | Modules: Biopython, BLAST
Running time (laptop): few hours / 10^8 sequences (both strands)

Assessment of CSReport



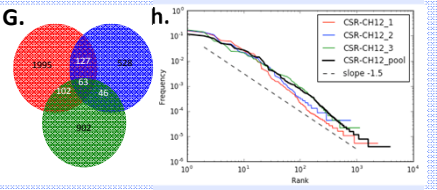
CSReport accurately identifies switch junctions in simulated Sμ-Sα sequences. Sμ-Sα blunt junctions, randomly generated from concatenation of germline segments, were used for assessment of CSReport. **A.** Structural profile of reported switch junctions. **B-C.** Reported breakpoint (BP) positions in Sμ (B) and Sα (C) are compared to expected positions (known from computer-generation). Insets: Differences between reported and expected breakpoint positions for more precise quality control.



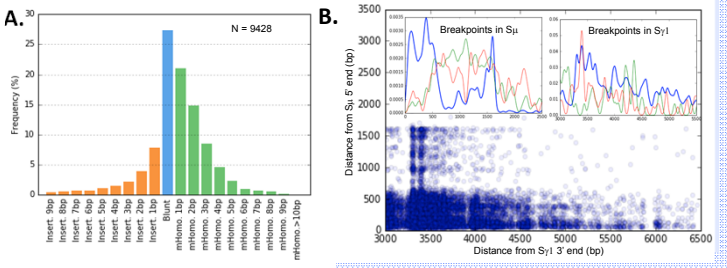
Assessment of Sμ-Sα junctions in CSR-induced CH12F3 cells. D. Overview of junction counts for three independent samples of CH12F3 cells. *Reads*: sequences in raw fastq file (Ion Proton output), *Junction reads*: sequences that can be segmented in a strand-aware consistent junction, *Junctions** : uniquely defined by two BP positions and one structure, *Junctions* : same as before and >10bp-insertions are excluded; *dist<x* : junctions with close BP positions are counted only once. **E.** Structural profile of sample CSR-CH12_1 for four different values of the BP-distance clusterization parameter (similar trends were observed for the other samples).

F.

BP-distance clustering ≤ 2			
Sample	Junction reads	Unique junctions	Inverse Simpson index
CSR-CH12_1	186158	2287	13.8
CSR-CH12_2	22928	764	14.3
CSR-CH12_3	45270	1113	16.1
CSR-CH12_pool	254356	3763	23.3



Statistics of junction diversity. F. Sμ-Sα junction counts and diversity index (Inverse Simpson index) obtained using CSReport on HTS sequences from three samples of stimulated CH12F3 cells with clusterization of BP-distance ≤ 2. **G.** Venn diagram comparing the three independent samples. **H.** Rank-frequency curves of unique junctions are compared to a generalized Zipf law $f = k^{-1.5}$.



C.

BP-Motifs in Sμ	Counts	BP-Motifs in Sγ1	Counts
GCTG**agct**	275	gagc**AGCT**	109
AGCTgagc	194	**tagc**TGGT	72
GGTG**agct**	186	cagcCAGG	68
CTG**Agctg**	171	**cagc**TTAG	65
GGGTgagc	153	gctt**AGCT**	53

Statistics of Sμ-Sγ1 junctions using CSReport. Murine splenocytes were stimulated *in vitro* with LPS+ IL4. After 3 days, cells were recovered and lysed for DNA extraction and then CSR junctions were amplified and sequenced by HTS. Sequences were analyzed using CSReport. **A.** Structural profile of Sμ-Sγ1 junctions. **B.** Distribution of breakpoints in Sμ and Sγ1. **C.** Most frequent 8nt-motifs at breakpoints in Sμ and Sγ1.

Conclusions and perspectives

We studied CSR junctions using a protocol based on HTS (Ion Torrent technology, Life). This technique challenges currently used methods that are implemented to analyse extremely large amount of collected data . To address the lack of automated tools for CSR analysis, we designed CSReport software. To evaluate errors in junction analysis induced by CSReport, we subjected it to analysis of artificially designed blunt-ended Sμ-Sα junctions with known breakpoints. Obtained results showed that CSReport interpretation induced effects which are likely to be present in previous methods of junction analysis as well. As HTS data are more prone to sequencing errors, it appeared that clustering of unique junctions can lower the abnormal rate of detection of junctions with insertions. Based on our results we recommend using CSReport with a BP-distance ≤2 bp.

We propose CSReport as an automated software tool able to support analysis of large data sets from CSR junction high throughput sequencing. Widespread use of CSReport will level differences among studies analyzing switch junctions as it provides a single pipeline providing breakpoint positions and junction structure (blunt-end junctions or junctions with insertions or microhomologies). Another interesting point is that CSReport is able to support Fastq and Fasta files regardless of the sequencing technology used as well as the species studied. Finally, in addition to the count and characterization of CSR junctions, CSReport presents results in the form of clear and easily manipulated tabular files of valuable information. CSReport facilitates and accelerates the study of CSR junctions and harmonizes analysis and interpretation.

Sophie Péron, CNRS UMR 7276, CBRS, Faculté de Médecine, 2 rue du Docteur Marcland, Limoges 87000, France. Phone: +33 519 564 2 10, fax: +33 555 435 847, e-mail: sophie.peron@inserm.fr.