Parallel Evolution of BiFC for Probing Protein-Protein Interactions in Living Cells
Yunlong Jia

To cite this version:

HAL Id: tel-03335719
https://tel.archives-ouvertes.fr/tel-03335719
Submitted on 6 Sep 2021

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.
Parallel Evolution of BiFC for Probing Protein–Protein Interactions in Living Cells

Évolution parallèle de la BiFC pour l’étude des interactions protéine-protéine dans les cellules vivantes

Devant le jury composé de :

<table>
<thead>
<tr>
<th>Nom</th>
<th>Titre</th>
<th>Institution</th>
<th>Rôle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mme Christine BRUN</td>
<td>DR</td>
<td>AMU</td>
<td>Rapporteure</td>
</tr>
<tr>
<td>M. René REZSOHAZY</td>
<td>Professeur</td>
<td>UCLouvain</td>
<td>Rapporteur</td>
</tr>
<tr>
<td>M. Didier AUBOEUF</td>
<td>DR</td>
<td>LBMC</td>
<td>Examinateur</td>
</tr>
<tr>
<td>Mme Corinne MAUREL-ZAFFRAN</td>
<td>MCF</td>
<td>AMU</td>
<td>Examinatrice</td>
</tr>
<tr>
<td>M. Samir MERABET</td>
<td>DR</td>
<td>IGFL</td>
<td>Examinateur</td>
</tr>
<tr>
<td>Mme Françoise BLEICHER-BARDELETTI</td>
<td>Professeure</td>
<td>UCBL</td>
<td>Directrice de thèse</td>
</tr>
</tbody>
</table>
Dedicated to

My mother, Yanli CHENG
My father, Xiaohua JIA
My wife, Yanxin ZHANG
My son, Xiaofan JIA
Acknowledgements

Simply put, the PhD training is hard. I can’t imagine if not you who guided me, when looking back at the four years. No acknowledgment is sufficient to thank my supervisors, Samir and Françoise, for your constant and remarkable mentorship. As a never-ending source of ideas, Samir is the kind of multi-hyphenate creative, enlightening me on all possible aspects of such exciting and challenging projects. You are a full-category music lover, rock drummer, scientist, all-round sports enthusiast, etc., and in awe for guiding my deep project reflection, making an attractive presentation, and granting me freedom in the project. Thank you for teaching me how to disagree productively by finding common ground, which will be one of the most important skills in my future. Concomitantly, a great thanks to Françoise. Your overarching support, approachability and upbeat nature have been greatly. I’ve never seen such a knowledgeable person like you. As long as I have questions, you can quickly answer them in detail and provide answers. Through your combined knowledge and practical expertise within the field and constant availability for any and all assistance I required, have made the research possible. Thank you both for creating a warm and tremendously caring lab environment for a long time, and driving my scientific development from the most basic, tangible level of technical competence to a lot of intellectual growth.

I would also like to thank the members of my CST, Dr. Didier AUBOEUF and Dr. Xavier MORELLI, for their helpful comments over the years. In particular I'd like to thank Didier for his yearly encouragement and genuinely helpful critiques of my work.

I am deeply grateful to all past and present members of the Samir lab for their support and friendship. Firstly, I'd like to thank Jonathan, with whom I have worked very closely for the past years. You taught me skills hand by hand ranging from basic molecular biology to experimental design and all relevant analyses. Without you, I would not have mastered as much and engaged myself quickly in project. Marilyne, you are like a mother who takes care of everyone in this family, understanding our personal situation, and always solving problems and giving care. Thank you, Fred, I will never forget our sincere communication, including in-depth discussions and suggestions on projects, and some thoughts on life and the future. Thanks to Guillaume, a hard rock boxer, for giving everything a sense of humor. Rachel, your strong affinity makes everyone want to be your friend from my first day in the lab. With your integrity and intelligence, I believe one day you will be one of the most capable scientists. Solène, do
you still need someone to take care of your cats? Please contact me, I love them. Don’t forget that we agreed to go fishing together with Marius. Thank you, Sylvie and Agnès. It had been an honor and pleasure working with you. I was impressed by your rigorous and dedicated attitudes that had a great impact on me. Thanks to Thiên, Trang and Yongshan, for showing me what my internship was like back then. Nawal, to be honest, you are the savior of my English. You came at the right time when I needed to learn English the most. Thanks to you, let me have to face various scenarios and think in English. I never thought that we helped each other, finally, from different fields. Thank you, Cindy, for your kindly thinking of me and birthday surprise. Thanks also to Benjamin and Sandrine from PSI, for our intellectually stimulating discussions and your patient assistance with NGS data.

All of these experiments would not have been possible without an efficient, well organized administrative system. For that reason, I would also like to thank Fabienne, Martine, and Fatiha, for always being there helping the team and dealing with a large and intricate paper work.

A special thanks to Crous Resto’U, for their careful care throughout my busy work, feeding me every day at a near-costless price, especially during my thesis writing in the race against time.

I should hate you, COVID-19. It is you who took away countless lives, disrupted our lives, and made people relationship with each other become more and more socially, even mentally distant. However, I am thankful to this ongoing pandemic, for slowing us down and having more time to reflect on the past and the future, with a come-to-mind question, “what is the most important and meaningful in your life for each of survivors like us?”

In the end, my most heartfelt thanks to everyone already mentioned, as well as Chérif, Christian, Michel, Karine, Wen Yue, Tingting and our IGFL community, for such a great network of supportive, capable people, who can always be counted on to help, advise, share, and talk whenever needed. I would not be the person I am today without the support of any of you.

As an old Chinese proverb says, all things must come an end. Ultimately, I managed to get out of this struggling final stage from my long-term educational journey. In turn, the end means a new beginning of the next. I affectionately wish you all the best in your next after me. Last but
never ending, thank you all and these four years of experience in this endeavor, which will be my future assets for life!

I dedicate this thesis to my parents, I fully understand and will keep in mind your full commitment from my birth to now, and my beloveds, Yanxin and Tamaru, I miss you~
"Only through tranquility will one obtain a deep and wide perspective on all matters"

- Zhuge Liang (181-234)
# Table of Contents

Abstract ............................................................................................................................... I  
Résumé ............................................................................................................................. III  
List of abbreviations ........................................................................................................ V  

Chapter I. Introduction ........................................................................................................ - 1 -  
1. Protein-protein interaction plays key role in biological processes .................. - 3 -  
2. Protein-protein interaction detection methods .................................................. - 7 -  
   2.1 Affinity-based methods ................................................................. - 8 -  
      2.1.1 AP-MS ....................................................................................... - 8 -  
      2.1.2 Co-IP ............................................................................................ - 10 -  
   2.2 Yeast-Two-Hybrid (Y2H) assay ...................................................... - 11 -  
   2.3 Protein complementation assays (PCAs) ........................................ - 14 -  
   2.4 FRET and BRET ............................................................................. - 15 -  
   2.5 Proximity-based labeling methods .............................................. - 17 -  
   2.6 In vitro methods ........................................................................... - 21 -  
   2.7 In silico methods ............................................................................. - 23 -  
   2.8 High-throughput PPI screenings .................................................. - 26 -  
   2.9 Summary ........................................................................................... - 29 -  
3. Bimolecular fluorescence complementation (BiFC) ...................................... - 31 -  
   3.1 Fluorescent proteins for BiFC assay .................................................. - 33 -  
   3.2 Advantages of BiFC assay ................................................................. - 37 -  
   3.3 Critical considerations ....................................................................... - 39 -  
      3.3.1 Topology of BiFC-tag fusions ....................................................... - 39 -  
      3.3.2 Linker selection ........................................................................... - 40 -  
      3.3.3 Improvement of signal-to-noise ratio .......................................... - 41 -  
      3.3.4 Controls for BiFC assay ............................................................... - 43 -  
      3.3.5 Irreversibility of BiFC ................................................................ - 46 -  
   3.4 Implementation of BiFC ................................................................. - 48 -  
      3.4.1 Low-throughput BiFC-based applications .............................. - 48 -  
      3.4.2 Large-scale applications of BiFC ............................................... - 50 -
4. Biology of Hox proteins

Chapter II. Results

Part 1. Bimolecular Fluorescence Complementation (BiFC) and Multiplexed Imaging of Protein–Protein Interactions in Human Living Cells (book chapter)

Part 2. Protein-protein interaction between HOX and TALE proteins

2.1 Human HOX Proteins Use Diverse and Context-Dependent Motifs to Interact with TALE Class Cofactors (article 1)

2.2 The human HOXA9 protein uses paralog-specific residues of the homeodomain to interact with TALE-class cofactors (article 2)

Part 3. Development and Optimization of Large-scale Bimolecular Fluorescence Complementation (BiFC) Screening

3.1 A cell protein complementation assay for ORFeome-wide probing of human HOX interactomes (article 3, manuscript)

3.2 Optimization for high-throughput BiFC screening

3.2.1 Introduction

3.2.2 Materials and Methods

3.2.3 Results

3.2.4 Discussion

Part 4. Development of a GFP nanobody-directed Proximity Biotinylation Assay based on the Bimolecular Fluorescence Complementation (BiFC) in living cells

4.1 Introduction

4.2 Materials and Methods

4.3 Results

4.3.1 TurboID-mediated protein biotinylation in mammalian cells

4.3.2 Development and performance of BibID method

4.4 Discussion

Chapter III. General Discussion and Perspectives

3.1 BiFC approach: direct or indirect interaction interrogation?

3.2 Limitation and future directions of current large-scale BiFC screening

3.2.1 Lack of rigorous negative control

3.2.2 Multicolor BiFC screening

3.2.3 The tag effect
Abstract

Proteins are crucial for most cellular functions and typically participate in biological processes in concert with other proteins. Hence, identification of key protein players and characterization of protein-protein interactions (PPIs) are highly important. Owing to substantial advances in current biotechnology, a wide range of methods has been developed to dissect the PPI landscape. Given their popularity and power, the bimolecular fluorescence complementation (BiFC) assay, based on the reconstitution of a fluorescent protein in vivo, has emerged as the most popular protein-fragment complementation method in cellular biology over the past years.

My team has previously established BiFC for probing different binary protein interactions in live Drosophila. During my PhD work, I expanded the utility of the BiFC in mammalian live cells. In particular, I standardized the protocol of the BiFC analysis to investigate protein binding affinities, in an applicable and simple manner. This quantitative BiFC approach was used in a systematic analysis of HOX/PBX/MEIS interaction properties in live cells and revealed novel interaction interfaces in several human HOX proteins.

Furthermore, I applied the BiFC from low to high throughput PPI detection. Pairing sequence-verified human ORF collections with next generation sequencing, I participated in the conception of a powerful tool for performing a large-scale BiFC interaction screen in live cells. Benefited chiefly from this approach, a synoptic view of comprehensive HOX interactomes was substantially contributed to the current limited knowledge on human HOX protein partners and provided a novel tool in the cell biology arsenal.

Along with the contemporary development of proximity labeling methods, in my side project, I depicted and tested a new cell-based PPI detection approach, which combines BiFC and BioID (proximity-dependent biotinylation identification) technologies, and allows deciphering the endogenous interactome of a protein complex.
In summary, my PhD work demonstrates that the BiFC is a versatile and powerful approach to study PPIs in the live cellular context, on either small or large scale. In addition, my work further enlarged the potential of BiFC applications by combining it with other tools.
Résumé

Les protéines sont cruciales pour la plupart des fonctions cellulaires et participent généralement aux processus biologiques de concert avec d'autres protéines. L'identification des acteurs protéiques clés et la caractérisation des interactions protéine-protéine (IPP) sont donc très importantes. Grâce aux progrès substantiels de la biotechnologie actuelle, un large éventail de méthodes a été développé pour disséquer le paysage des IPP. Compte tenu de leur popularité et de leur puissance, l'essai par complémentation de fluorescence biomoléculaire (BiFC), basé sur la reconstitution d'une protéine fluorescente, est apparu comme la méthode de complémentation la plus populaire pour analyser les IPP in vivo.

Mon équipe a établi la BiFC pour sonder différentes interactions binaires entre protéines dans l'embryon vivant de drosophile. Au cours de mon travail de Thèse, j'ai étendu l'utilité de la BiFC aux cellules vivantes de mammifères. En particulier, j'ai standardisé le protocole de l'analyse BiFC pour étudier les affinités de liaison des protéines, d'une manière simple et applicable. Cette approche quantitative de la BiFC a été utilisée dans une analyse systématique des propriétés d'interaction HOX/PBX/MEIS dans des cellules vivantes et a révélé de nouvelles interfaces d'interaction dans plusieurs protéines HOX humaines.

En outre, j'ai appliqué la BiFC de la détection des IPP à haut débit. En jumelant des collections d'ORF humains vérifiés par séquence avec le séquençage de nouvelle génération, j'ai participé à la conception d'un outil puissant permettant de réaliser un écran d'interaction BiFC à grande échelle dans des cellules vivantes. Grâce à cette approche, une vue synoptique des interactomes HOX complets a été réalisée. Ce travail a apporté des perspectives nouvelles sur les propriétés générales d’interaction des protéines HOX tout en fournissant de nouveaux outils permettant désormais des analyses interactomiques comparatives à large échelle.

Parallèlement au développement contemporain des méthodes de marquage de proximité, j'ai décrit et testé, dans le cadre de mes projets parallèles, une nouvelle approche pour la détection des IPP au niveau cellulaire, qui combine les technologies BiFC et BioID.
(identification par biotinylation dépendante de la proximité) et permet de déchiffrer l'interactome endogène d'un complexe protéique.

En résumé, mon travail de thèse démontre que la BiFC est un outil polyvalent et puissant pour étudier les IPP dans le contexte cellulaire vivant, à petite ou grande échelle. En combinant la BiFC à d'autres outils, mon travail ouvre également de nouveaux champs d'applications pour le futur. Le test BiFC a montré un potentiel de combinaison avec d'autres méthodes et a largement enrichi l'ensemble des outils en conférant des fonctionnalités améliorées ou nouvelles.
## List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>APEX-seq</td>
<td>RNA sequencing based on direct proximity labeling of RNA using the peroxidase enzyme APEX2</td>
</tr>
<tr>
<td>AP-MS</td>
<td>Affinity purification followed by mass spectrometry</td>
</tr>
<tr>
<td>BibID</td>
<td>BiFC-based TurboID</td>
</tr>
<tr>
<td>BiFC</td>
<td>Bimolecular fluorescence complementation</td>
</tr>
<tr>
<td>BiLC</td>
<td>Bioluminescence complementation</td>
</tr>
<tr>
<td>BIND</td>
<td>Biomolecular Interaction Network Database</td>
</tr>
<tr>
<td>BioGRID</td>
<td>Biological General Repository for Interaction Datasets</td>
</tr>
<tr>
<td>BioID</td>
<td>Proximity-dependent biotinylation identification</td>
</tr>
<tr>
<td>BioPlex</td>
<td>Biophysical interactions of ORFeome-based complexes</td>
</tr>
<tr>
<td>BRET</td>
<td>Bioluminescent resonance-energy transfer</td>
</tr>
<tr>
<td>CasID</td>
<td>dCas9-BirA* fusion-based proximity-labeling approach</td>
</tr>
<tr>
<td>C-BERST</td>
<td>dCas9-APEX2 biotinylation at genomic elements by restricted spatial tagging</td>
</tr>
<tr>
<td>CCSB</td>
<td>Center for Cancer Systems Biology</td>
</tr>
<tr>
<td>Cell-PCA</td>
<td>Cell Protein Complementation Assay</td>
</tr>
<tr>
<td>coBiFC</td>
<td>double BiFC combined method</td>
</tr>
<tr>
<td>CoFrac-MS</td>
<td>co-fractionation followed by MS</td>
</tr>
<tr>
<td>Co-IP</td>
<td>Co-immunoprecipitation</td>
</tr>
<tr>
<td>CORUM</td>
<td>Comprehensive resource of mammalian protein complexes</td>
</tr>
<tr>
<td>CrY2H-seq</td>
<td>Cre reporter-mediated Y2H coupled with next-generation sequencing</td>
</tr>
<tr>
<td>DAPA</td>
<td>DNA array to protein array</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>DBD</td>
<td>DNA-binding domain</td>
</tr>
<tr>
<td>DIP</td>
<td>Database of Interacting Proteins</td>
</tr>
<tr>
<td>e.g.</td>
<td>For example</td>
</tr>
<tr>
<td>EPR</td>
<td>Expression Profile Reliability</td>
</tr>
<tr>
<td>FRAP</td>
<td>Fluorescence-recovery after photobleaching</td>
</tr>
<tr>
<td>FRET</td>
<td>Fluorescence resonance energy transfer</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>HPRD</td>
<td>Human Protein Reference Database</td>
</tr>
<tr>
<td>HRP</td>
<td>Horseradish peroxidase</td>
</tr>
<tr>
<td>HuRI</td>
<td>Human binary protein interactions</td>
</tr>
<tr>
<td>IG1/IG2</td>
<td>Interaction Generalities Measure 1/2</td>
</tr>
<tr>
<td>IntAct</td>
<td>IntAct molecular interaction database</td>
</tr>
<tr>
<td>LC-MS/MS</td>
<td>Liquid chromatography tandem-mass spectrometry</td>
</tr>
<tr>
<td>mbSUS</td>
<td>Mating-based split-ubiquitin system</td>
</tr>
<tr>
<td>mcBiFC</td>
<td>Multicolor BiFC</td>
</tr>
<tr>
<td>M-NAPPA</td>
<td>Multiplexed nucleic acid programmable protein array</td>
</tr>
<tr>
<td>NAPPA</td>
<td>Nucleic acid programmable protein array</td>
</tr>
<tr>
<td>NGS</td>
<td>Next generation sequencing</td>
</tr>
<tr>
<td>ORFs</td>
<td>Open reading frames</td>
</tr>
<tr>
<td>PCA</td>
<td>Protein-fragment complementation assay</td>
</tr>
<tr>
<td>PDIs</td>
<td>Protein-DNA interactions</td>
</tr>
<tr>
<td>PISA</td>
<td>Protein in situ array</td>
</tr>
<tr>
<td>PL</td>
<td>Proximity-based labeling</td>
</tr>
<tr>
<td>PLM</td>
<td>Protein Localization Method</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>POIs</td>
<td>Proteins of interest</td>
</tr>
<tr>
<td>PPAs</td>
<td>Protein–protein associations</td>
</tr>
<tr>
<td>PPIs</td>
<td>Protein-protein interactions</td>
</tr>
<tr>
<td>PRIIs</td>
<td>Protein-RNA interactions</td>
</tr>
<tr>
<td>PVM</td>
<td>Paralogous Verification Method</td>
</tr>
<tr>
<td>q-AP-MS</td>
<td>quantitative dimension to AP-MS experiment</td>
</tr>
<tr>
<td>QUBIC</td>
<td>Quantitative BAC-GFP interactomics</td>
</tr>
<tr>
<td>RT-smCo-IP</td>
<td>real-time single-molecule Co-IP</td>
</tr>
<tr>
<td>splitFAST</td>
<td>split Fluorescence-activating and Absorption Shifting Tag</td>
</tr>
<tr>
<td>SRET</td>
<td>sequential BRET-FRET</td>
</tr>
<tr>
<td>S/N ratio</td>
<td>Signal-to-noise ratio</td>
</tr>
<tr>
<td>TAD</td>
<td>Trans-activation domain</td>
</tr>
<tr>
<td>TAP-MS</td>
<td>tandem affinity purification tag purification followed by mass spectrometry</td>
</tr>
<tr>
<td>TR-FRET</td>
<td>Time-resolved FRET</td>
</tr>
<tr>
<td>TriFC</td>
<td>Tripartite Fluorescence Complementation</td>
</tr>
<tr>
<td>WPR</td>
<td>Weighted positive ratio</td>
</tr>
<tr>
<td>Y2H</td>
<td>Yeast two-hybrid system</td>
</tr>
<tr>
<td>Y3H</td>
<td>Yeast three-hybrid system</td>
</tr>
</tbody>
</table>
Chapter I. Introduction
1. Protein-protein interaction plays key role in biological processes

Biological processes are modulated by complex networks of macromolecular interactions, including proteins, nucleic acids, lipids, chromatin, and low-molecular weight compounds. Characterizing the properties of these networks can provide important insights into cellular organization and underlying mechanism-level traits of life (Vidal et al., 2011). Sequencing the whole genome of many organisms has been a huge accomplishment with enormous scientific impact. However, genomic information alone is hard to explain cellular functions. Hence, in the post-genomic era, the research focus is moving towards the function analysis of encoded proteins and the decryption of their interaction networks.

Indeed, proteins control and execute most cellular processes. They rarely act on their own in vivo, instead fulfilling most of their biological tasks in complexes through interaction with other proteins (Alberts, 1998). Proteins that are composed of more than one subunit are found in many different classes of proteins. It has been revealed that over 80% of proteins do not operate alone but in complexes (Berggård et al., 2007). Protein complexes are groups of proteins that interact with each other (Figure 1), by direct or indirect associations. Given their functions, several significant properties of PPIs are listed, which had been enumerated previously (Phizicky and Fields, 1995). They can (1) alter the kinetic properties of enzymes; (2) perform as one common mechanism to confer substrate channeling; (3) form a new binding site for small effector molecules; (4) inactivate a protein; (5) change the specificity of a protein for its substrate via interaction with different binding partners; (6) play a regulatory role in either upstream or downstream level.

Therefore, identification of the interactors of a given protein, belonging to a multimeric protein complex, can help to predict its function for providing a deeper understanding of cell functions (Lalonde et al., 2008). To date, protein-protein interaction (PPI) is becoming one of the major objectives of system biology.
Figure 1. Types of protein complexes. (A) A variety of proteins occur as monomers, homodimers, homotrimers, homotetramers, or even larger homomeric complexes. (B) Proteins may assemble as hetero-oligomers consisting of homo- or hetero-oligomers. (C) Indirect interactions: scaffolding proteins can bind various monomeric or oligomeric proteins. Direct interactions can occur between scaffolding protein and each of other proteins. Adapted from figure 1 in (Lalonde et al., 2008).

The relationships between proteins can be modeled as graphical networks in which proteins are represented by ‘nodes’ and the interactions between them by ‘links’ (Vidal et al., 2011), as shown in Figure 2. Here the links between nodes represent permanent or transient bonds. In principle, these PPIs can be assigned to two different types, including constitutive and regulative interactions (Fujikawa et al., 2014). Concerning constitutive PPIs, they are relatively stable and mainly constitute subunits of protein complexes that perform structural functions in cells. For example, these proteins interact with high affinity, forming macromolecules, such as ribosomes, proteasomes, replisomes, and signalosomes (Morsy et al., 2008). Constitutive interactions are prone to stability and generally detected easily either by in vivo or in vitro methods.
**Figure 2. Two types of PPIs in cells.** Constitutive PPIs are stable and mostly form subunits of permanent complexes as part of cell structure. The cytoskeletal network and the membranous, organellar networks provide a general scaffold of the cell containing the physical interactions between cellular proteins. On the contrary, regulative PPIs are part of biochemical cascades and are often highly sensitive to regulatory stimuli and signalling events, as transient interactions in functional processes. In the signalling network elements of various signalling pathways are linked by the interactions between them. In the transcriptional regulatory network the elements are the transcription factors, genes and the connecting links are functional interactions between them. In the metabolic network we have the various metabolites as elements and the enzyme reactions as links. All these networks highly overlap with each other, and some of them contain modules of other networks. Adapted from figure 1 in [Korcsmáros et al., 2007].

In contrast, regulative PPIs exist widely in certain cellular or developmental contexts or in response to specific environmental stimuli. These PPIs participate in signaling cascades and are often highly sensitive to regulatory events. Regarding their interaction strength, these dynamic protein interactions are ultimately divided into weak and strong groups [Syafrizayanti et al., 2014]. In most cases, regulative PPIs are transient, showing a lower affinity binding, because proteins continuously associate with and dissociate from each other [Nooren and Thornton, 2003]. For example, protein-modifying enzymes, such as phosphatases, protein kinases, and cell surface receptors that are activated by dimerization [Pellicena and Kuriyan, 2006]. Lying in transient dynamic PPI changes, cells can rapidly respond to intra- and extracellular environmental stimuli. Specially, the detection of these interactions is more challenging, as their quick occurrence and easy destabilization.
Therefore, regulative interactions are highly recommended to be detected by *in vivo* methods, like protein-fragment complementation assay (PCA).

As mentioned above, the study of PPIs is important to infer the protein function within the cell. The function of unidentified proteins can be predicted based on the knowledge of their interacting protein partners, whose function is already revealed. The detailed study of PPIs has accelerated the modeling of functional pathways to illustrate the molecular mechanisms of cellular processes. Characterizing the PPIs in a given proteome will be phenomenal to outline the biophysical panorama of the cell (*Zhang, 2009*).
2. Protein-protein interaction detection methods

PPI networks are integral to nearly all aspects of cellular activity. In humans, the entire protein interactome is estimated to consist on average of 650,000 PPIs (Stumpf et al., 2008). Therefore, the study of PPIs is vital for understanding the mechanics of a specific biological process in an organism. Recent years have witnessed substantial advances in the development of PPI technology. These efforts, coupled with the availability of nearly complete human ORFeome collections of open reading frames (ORFs) (ORFeome Collaboration, 2016), laid the necessary foundation to enable the methods of PPI detection to expand from single experiments to high-throughput screens. Accordingly, direct PPIs as well as indirect protein–protein associations (PPAs) between the different proteins of the complex, can be identified. Two orthogonal approaches exist for experimentally identifying biophysical relationships between pairs of proteins: binary PPI approach, such as Y2H (yeast two-hybrid system), and protein complex approach, e.g. AP-MS (affinity purification mass spectrometry) and BioID (proximity-dependent biotin identification). Thus, the biological features of the interactions should be considered when a detection technique is selected for validation of the identified PPIs.

In this part, I review the commonly used techniques for detecting and characterizing PPIs, in line with their corresponding properties. More particularly, to raise awareness about the advent of new powerful approaches compared with conventional methods for probing PPIs, several featured studies will also be presented, which have undertaken different strategies to high-throughput PPI screenings. The BiFC method, as a special issue, will be mentioned in this part, but detailed further in Section 3, Chapter 1.
2.1 Affinity-based methods

2.1.1 AP-MS

Affinity-based purification (AP) methods allow the preparation of proteins of suitable purity and reduced complexity so that they can be interrogated efficiently by mass spectrometry (MS)-based protein identification approaches (Köcher and Superti-Furga, 2007). Nowadays, affinity purification followed by mass spectrometry (AP-MS) has become a fast, selective, and sensitive tool for the identification of PPIs under nearly native conditions (Van Leene et al., 2015). In general workflow, a protein of interest (POI) is fused to an affinity tag, allowing its isolation from the cell lysate by using a specific antibody (Figure 3). After several washing steps under mild, non-denaturing conditions, the tagged POI (bait) together with interacting prey proteins is eluted from the solid phase support and analyzed by liquid chromatography tandem-mass spectrometry (LC-MS/MS) (Fukao, 2012; Lee et al., 2017). Whereas, the affinity tag might influence the physiological properties of bait protein as well as the complex stoichiometry, resulting in steric hindrance of protein interactions. Reciprocally, it is applicable to use protein-specific antibodies against directly endogenous non-tagged bait proteins (Figure 4A), but the availability of specific antibodies and the one mentionable caveat that topology of bait protein in complexes may mask the antibody-targeting epitopes, leading to low AP efficiency or loss of bait protein binding, still require attention. As a consequence, the tagged bait is much more used in current research.
Different affinity tags correspond to different purification methods. For example, in single-step purifications (Figure 4B), the fluorescent proteins were commonly used as single tags, such as GFP, YFP and CFP. In the same way, double affinity tags were designed in a tandem affinity purification-based MS (TAP-MS) (Figure 4C). One of the most frequently applied TAP tag is the GS tag and its derivatives (Bürckstümmer et al., 2006; Van Leene et al., 2008). Original GS tag encompasses protein G part and a streptavidin-binding peptide (SBP), which is bridged by the TEV protease cleavage site. In the first purification, the TAP-bait is bound to IgG agarose beads, followed by TEV protease treatment. Subsequent 2nd purification will trap the bait complex by streptavidin-conjugated beads. Therefore, in comparison with single tagged AP-MS, TAP obtains cleaner samples with an increased signal-to-noise ratio. Notably, no matter which AP method is used, it is crucial to set up the suitable controls to remove non-specific background proteins (Figure 4D).
Figure 4. Overview of affinity purification strategies and the possible controls. (A) Immunoprecipitation of protein complexes with immobilized antibodies against the bait protein. (B) Single-step tag-based affinity purification with bead-immobilized antibody directed against a protein tag. (C) Tandem affinity purification with two consecutive purification steps by means of immobilized antibodies against two different tags. A protease cleavage is used to separate the tags. (D) Possible controls for the detection of background contaminants include the wild-type (WT) extract, purification from cells expressing the tag only, or unrelated proteins fused with a tag. FP, false positive; UP, unrelated protein. Adapted from figure 4 in (Struk et al., 2019).

2.1.2 Co-IP

Coimmunoprecipitation (Co-IP) is considered as the golden standard assay for the PPIs validation (Figure 5). Despite that it might only be able to capture stable or strong interactions and could not provide spatial and temporal information, it is still a popular and frequently used method to identify physiologically relevant PPIs. The key advantage of Co-IP is that proteins are present in their native context together with other bound proteins. In Co-IP, the target protein is precipitated from a whole cell lysis with an immobilized antibody and the putative interacting proteins can be detected by Western blotting or MS
analysis (Masters, 2004; Ransone, 1995). Besides the conventional engineered co-expression of POIs, the Co-IP can be also used to investigate directly the endogenous protein’s interactome with corresponding specific antibodies (Chinchilla et al., 2007; Xie et al., 2012). However, if so, this strategy finally will not provide information about the direct interaction between proteins and demonstrate only their coexistence in a near-native protein complexes.

Figure 5. Schematic representation of Co-IP. (A) In Co-IP, the cells are collected and lysed under non-denaturing conditions that preserve protein-protein interactions. The target protein is specifically immunoprecipitated from the cell extracts using target-specific antibodies, and the immunoprecipitates are fractionated by SDS-PAGE. Co-immunoprecipitated proteins are detected by western blotting, ELISA with an antibody directed against that protein, or mass spectrometry with high accuracy. Adapted from (Creative Biomart, 2021).

2.2 Yeast-Two-Hybrid (Y2H) assay

Since first reported in 1989, Y2H method, as one of the most prominent PPI assays, was rapidly adopted by the scientific community and used in various species and research fields (Fields, 2009; Fields and Song, 1989; Xing et al., 2016). In this ingenious binary system, target protein interactions are verified by the transcriptional activation of reporter gene, which allows the positive colonies’ growth on selective media or visibility through an enzymatic coloured reaction. The principle of classical Y2H is forthright and illustrated in Figure 6A. Two protein domains are required in the classical Y2H assay, namely the DNA-
binding domain (BD) and the trans-activation domain (AD) of the GAL4 transcription. When these domains are split, the factor is functionally disabled. If each domain is fused to two interacting proteins as hybrid proteins, the function of the factor can be rescued to transcribe reporter genes that then select for PPI events. During the last decade, the Y2H system has been well suited for high-throughput approaches and cDNA library screening (Figure 6B) (see also Section 2.7). It is noteworthy that, there are two of the biggest limitations for this system: the one is the obligatory nuclear localization of the interaction partners, which may prevent the large cytoplasmic proteins from entering the nuclear pores (Kriechbaumer et al., 2015); the other drawback is heterologous system, resulting in the inappropriate protein modification in a native-differed cellular environment. Due to these vices, the false positives that usually arise in a cDNA library screen are drawn from three different types of proteins and their interactions (Figure 6C). Despite a relatively high false-positive rate, Y2H can still be successfully used for discovering the new PPIs, such as the prestigious HuRI project (Luck et al., 2020). Furthermore, the Y2H system has been reported to detect ternary interaction, called yeast three-hybrid (Y3H) method (Maruta et al., 2016), in which the third protein can serve as a “bridge” to connect two not directly interacting proteins or to stabilize a weak interaction (Figure 6D). Given a little more versatility, inhibitor screening of dedicated known PPI was extended application of Y2H system, as reverse Y2H analysis (Figure 6E). Beyond this system per se, there are a range of further modifications and NGS-combined Y2H systems, such as Cre reporter-mediated Y2H coupled with next-generation sequencing, named CrY2H-seq (Trigg et al., 2017).
Figure 6. The yeast two-hybrid system and its modifications. (A) The original yeast two-hybrid analysis detects an interaction between two proteins fused to either the DNA-binding domain (BD; bait) or the DNA transactivation domain (AD; prey) of GAL4, respectively. An interaction reconstitutes the DNA-binding and transactivation domains to activate reporter genes such as ADE2, HIS3, and lacZ. (B) A cDNA library screen allows the identification of novel binding partners (green crayon-like shape), whereas non-interacting proteins (opaque square and sphere) fail to activate the reporter genes. (C) Examples of false-positive results that might be detected in yeast two-hybrid cDNA library screens that should be excluded by appropriate testing: (1) unrelated prey fusions that bind to the bait; (2) transcriptional activators that are sufficient to trigger reporter gene activity; and (3) enzymes that overcome the selection pressure on depleted medium by restoring prototrophy. (D) An extension of the classic yeast two-hybrid system is yeast three-hybrid analysis, in which two noninteracting or weakly interacting proteins (bait I and bait II) are bridged or stabilized by a third protein (cyan; prey/bridge). (E) Reverse yeast two-hybrid detects the interference of a known interaction couple and can be used to screen for inhibiting proteins or molecules (red cones). Transcript activation of a selection marker such as URA3 renders yeast sensitive to 5-fluoroorotic acid. If the interaction is prevented,
the reporter genes are not activated and the yeast can survive on 5-fluoroorotic acid. UAS, Upstream activating sequence. Reprinted from figure 3 in (Xing et al., 2016).

2.3 Protein complementation assays (PCAs)

Protein fragment complementation assays (PCAs) are based on halves of enzyme or FP reporters that are fused to two proteins of interest (proteins bait and prey). Upon target protein interaction, the reporter fragments are reconstituted, generating detectable signals for subsequent PPI analysis.

In split-enzyme assays, it is prevailing to use fragments of β-galactosidase (Rossi et al., 1997), dihydrofolate reductase (Pelletier et al., 1998; Remy and Michnick, 1999) or β-lactamase (Galarneau et al., 2002; Wehrman et al., 2002). The PPIs induce the catalysis of a specific substrate that can be visualized in a colorimetric assay or assessed by antibiotics resistance (Figure 7A). However, the split-enzyme PCAs permit PPI detection to be performed in various subcellular compartments, whereas high background signals are generated concomitantly. Alternatively, bimolecular fluorescence complementation (BiFC), as one of the most used PCAs with low-background, was developed (Ghosh et al., 2000). Through the formation of a fluorescent complex from auto-fluorescent proteins (AFPs) instead of substrate-dependent enzymes, BiFC assay is simple and rapid, providing the direct visualization of PPIs in living cells (Figure 7B). Moreover, PCAs can recruit split-luciferase as well to monitor PPIs, termed bioluminescence complementation (BiLC) assay. Similar to BiFC, the PPIs induce the reassembly and ligation of luciferase fragments to recover and irreversible luciferase activity (Figure 7C) (Ozawa et al., 2001). As complementation of BiFC, the BiLC overcomes the problem of cellular auto-fluorescence for some special biology materials, such as plant tissues. The bioluminescence reporter endows the ability to monitor PPI in living cells with extremely low background fluorescence (Paulmurugan and Gambhir, 2005). Collectively, in contrast to classical Y2H strategies, the main advantage of PCAs is that proteins of interest (POIs) are brought into a rather native context, enabling PPI detections within diverse subcellular compartments.
Figure 7. Protein complementation assays (PCAs). (A) Split-enzyme results in the generation of color from a chromogenic substrate (Blue circles) or fluorescence from a fluorescence substrate (yellow diamonds) upon reconstitution of the enzyme facilitated by bait-prey interaction; N-Enz, N-terminal half of enzyme. C-Enz, C-terminal half of enzyme. (B) Bimolecular fluorescent complementation (BiFC) results in the generation of fluorescence following light excitation upon reconstitution of the AFP facilitated by bait-prey interaction; AFP, auto-fluorescent protein. N-AFP, N-terminal half of AFP. C-AFP, C-terminal half of AFP. (C) Bioluminescence complementation (BiLC) results in the generation of fluorescence following excitation by a bioluminescence substrate (magenta parallelograms) upon reconstitution of the luciferase enzyme facilitated by bait-prey interaction; N-Luc, N-terminal half of Luciferase. C-Luc, C-terminal half of Luciferase. Adapted from figure 2 in (Morsy et al., 2008).

2.4 FRET and BRET

Like BiFC, Förster resonance energy transfer (FRET) is one of the most commonly employed PPI detection methods to visualize PPIs in a variety of model organisms (Hu and Kerppola, 2003; Kerppola, 2006a). It allows real-time quantitative analysis in living cells with the best spatial and temporal resolutions to study PPIs (Pollok and Heim, 1999). FRET is based on the principle that the non-radiative energy between two molecules can be transferred when a donor fluorophore has an emission spectrum that overlaps with the absorption spectrum of an acceptor (Figure 8A). Therefore, FRET can only occur when both donor and acceptor fluorophores are extremely close (<10nm), serving the detection of direct PPIs. As the signal detection is the most crucial part of FRET assay, to date, numerous methods are combined with FRET detection, including acceptor-sensitized
emission, acceptor photobleaching, and fluorescence lifetime imaging (FLIM). Among them, FRET-FLIM is one of the most powerful and popular approaches for determining time-resolved, quantitative, subcellular localization of specific PPIs within single living cells (Long et al., 2017; Somssich et al., 2015; Stahl et al., 2013). Thus far, FRET and its variations have not been used as widely as PCAs, due to that FRET signals usually need careful interpretation and multiple control experiments with advanced equipment (Bhat et al., 2006). Along with its low sensitivity and high labor-intensity, it further restricts the penetration of FRET in the general research field.

Benefiting from low-background bioluminescence, bioluminescence resonance energy transfer (BRET) was developed (Xu et al., 1999). The donor protein is replaced by a bioluminescent molecule, such as blue-light emitting Renilla luciferase, instead of using light emitted from an excited AFP, whereas the acceptor remains a fluorophore, either GFP or YFP (Figure 8A). Given that BRET and FRET are two quite similar methods, BRET is more amenable to high-throughput screening than FRET because no internal excitation light source is needed and the energy donor does not become photo-bleached (Morsy et al., 2008).

Interestingly, BRET and FRET can be integrated becoming a sequential BRET-FRET (SRET), which allows PPI analysis of a ternary protein complex (Figure 8B) (Carriba et al., 2008). Although SRET needs complicated optimization for both BRET and FRET, it could be an alternative yet attractive approach to examine protein hetero-dimerization under physiological conditions.
Figure 8. Illustration of FRET/BRET and SRET. (A) Illustration of FRET/BRET. Fluorescence resonance energy transfer (FRET) or bioluminescence resonance energy transfer (BRET) results in the generation of fluorescence following resonance energy transfer from emitted light from AFP or Luc from a bioluminescence substrate (magenta parallelograms) upon close physical proximity of the AFP or Luc facilitated by bait-prey interaction. Luc, Luciferase; AFP, auto-fluorescent protein. Adapted from figure 2b in (Morsy et al., 2008). (B) Scheme of SRET technique. Sequential BRET-FRET (SRET) for the chemokine receptors CKR1, CKR2, and CKR3 fused to Rluc, CFP, and YFP, respectively. Due to activation by its substrate DeepBlueC (blue dots), Rluc excites CFP (BRET), which in turn excites YFP (FRET) that is detected at 530 nm. Adapted from figure 3b in (Martínez-Muñoz et al., 2016).

2.5 Proximity-based labeling methods

In the recent past, proximity-based labeling (PL) methods were developed as alternative approaches for studying PPIs in living cells. These methods take advantage of promiscuous enzymes that are fused to the target protein and can directly label all proximal endogenous proteins with a covalent biotin tag. After the labeling reaction, cells are lysed and the biotinylated proteins are subsequently pulled down with streptavidin beads followed by mass spectrometry (Figure 9).
Figure 9. General workflow of proximity labeling followed by mass spectrometry with biotin ligase (A) or peroxidase (B). The protein of interest (bait) is fused to the reporter enzyme and expressed in cells. Supplying the enzymes with their substrates creates reactive intermediates that target amino acid side chains of proteins in proximity (prey). The covalently biotinylated proteins can be enriched by streptavidin beads. Subsequent on-bead digestion and identification of resulting peptides with mass spectrometry provides a candidate list of proteins in the vicinity of the bait. Reprinted from figure 1 in (Ummethum and Hamperl, 2020).

To date, along with unremitting efforts, there are three major enzymes that are used for proximity labeling: biotin ligase (BioID (Roux et al., 2012), BioID2 (Kim et al., 2016), BASU (Ramanathan et al., 2018), miniTurbo (Branon et al., 2018a), TurboID (Branon et al., 2018a)), horseradish peroxidase (HRP) (Kotani et al., 2008), and engineered ascorbate peroxidase (APEX (Martell et al., 2012), APEX2 (Lam et al., 2015)). These evolved PL enzymes are characterized and summarized in Table 1.
Table 1. Overview of available proximity labeling enzymes and their characteristics. Adapted from table 1 in (Ummethum and Hamperl, 2020).

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Type</th>
<th>Source organism</th>
<th>Amino acid mutations</th>
<th>Size in kDa</th>
<th>Labeling time</th>
<th>Substrate incubation time</th>
<th>Substrates</th>
<th>Labeling targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>BioID</td>
<td>Biotin ligase</td>
<td>E. coli</td>
<td>R118G</td>
<td>38</td>
<td>6-24 h</td>
<td>6-24 h</td>
<td>Biotin</td>
<td>Lys</td>
</tr>
<tr>
<td>BioID2</td>
<td>Biotin ligase</td>
<td>A. aeolicus</td>
<td>R40G</td>
<td>27</td>
<td>6-24 h</td>
<td>6-24 h</td>
<td>Biotin</td>
<td>Lys</td>
</tr>
<tr>
<td>BASLI</td>
<td>Biotin ligase</td>
<td>B. subtilis</td>
<td>13 mut, ΔN-tarn</td>
<td>29</td>
<td>30 min-12 h</td>
<td>30 min-12 h</td>
<td>Biotin</td>
<td>Lys</td>
</tr>
<tr>
<td>minTurbo</td>
<td>Biotin ligase</td>
<td>E. coli</td>
<td>12 mut, ΔN-tarn</td>
<td>28</td>
<td>10-60 min</td>
<td>10-60 min</td>
<td>Biotin</td>
<td>Lys</td>
</tr>
<tr>
<td>TurboID</td>
<td>Biotin ligase</td>
<td>E. coli</td>
<td>14 mut, ΔN-tarn</td>
<td>35</td>
<td>10-60 min</td>
<td>10-60 min</td>
<td>Biotin</td>
<td>Lys</td>
</tr>
<tr>
<td>HRP</td>
<td>Peroxidase</td>
<td>Horseradish</td>
<td>–</td>
<td>44</td>
<td>5-10 min</td>
<td>5-10 min</td>
<td>Biotin-phenol, fluorescein-arylazide</td>
<td>Tyr, Trp, Cys, His</td>
</tr>
<tr>
<td>APEX</td>
<td>Peroxidase</td>
<td>Pea</td>
<td>K14D, E112K, W41F</td>
<td>28</td>
<td>1 min</td>
<td>30-60 min</td>
<td>Biotin-phenol, -aniline, -naphtylamino</td>
<td>Tyr, Trp, Cys, His</td>
</tr>
<tr>
<td>APD2</td>
<td>Peroxidase</td>
<td>Soybean</td>
<td>K14D, E112K, W41F</td>
<td>28</td>
<td>1 min</td>
<td>30-60 min</td>
<td>Biotin-phenol, -aniline, -naphtylamino</td>
<td>Tyr, Trp, Cys, His</td>
</tr>
</tbody>
</table>

PL methods have become a valuable complement to classical PPI studies like AP-MS and ChIP. Nevertheless, some general considerations still need to be taken during their use. For example, the enzyme-fused bait protein may confer influence of its function or localization, compared to its wild type (Kim and Roux, 2016). Besides, the experimentally determined labelling radii are varied from 1-20nm for different enzymes (Kim et al., 2014; Martell et al., 2012; Mayer and Bendayan, 1997). Many relevant factors may address this issue, such as half-life of reactive enzymes (Rhee et al., 2013), properties of target proteins (Kim et al., 2014), diversity of flexible linkers (Kim et al., 2016), or different subcellular locations during labelling reaction (Hung et al., 2016). Owing to biotin-streptavidin bond-based pull-down process, another concern is the efficiency of biotinylated protein elution from beads, on which a harsh enough condition is needed to collect the majority of labelled proteins for subsequent analysis. In the course of designing a proximity labeling experiment, negative control is also an important point that must be considered (Lobingier et al., 2017). Because abundant false positives can be generated due to promiscuous association with the target protein, two types of control are necessary for reliable results with statistical significance of analysis: a technical control without PL reaction, which used to remove the technical
background, and a spatial control for patterning the PL reaction in specific cellular compartments. In classical BioID method, for instance, cells with no BirA*, BirA* alone, or BirA* fused to a target protein-mimicked location-specific tag are three most common controls. Furthermore, unlike the PCA methods, the PL methods investigate the co-complex of target protein, thus the quantification of biotin-labelled proteins does not necessarily reflect the strength of association (Minde et al., 2020).

However, PL methods have been conducted in various cell types and organisms for settling PPI mapping problems (reviewed in (Qin et al., 2021)). Recently, novel approaches combining PL and PCA were developed. Two POIs are fused to either half of a split PL enzyme. The PL enzyme is reconstituted only upon interaction of the POIs. Lying in PCA-like complementation of two target proteins, the PCA-PL method expands the applications of PL, which enables the co-complex profiling from single POI to two interacted POIs. As examples, split-BioID, split-APEX2 and split-TurboID have now been reported (Cho et al., 2020; Han et al., 2019; Munter et al., 2017; Schopp et al., 2017; Xue et al., 2017). By combining PL with dCas9, such as dCas9-BirA*, PL methods can also be extended to study PPIs at specific genome regions, which was originally termed CasID (Schmidtmann et al., 2016). Additionally, for protein-nucleic acid mapping, the PL-based methods have been reported in both studies of protein-DNA interactions (PDIs) and protein-RNA interactions (PRIs), like dCas9-APEX2 biotinylation at genomic elements by restricted spatial tagging (C-BERST) (Gao et al., 2018) and RNA sequencing based on direct proximity labeling of RNA using the peroxidase enzyme APEX2 (APEX-seq) (Fazal et al., 2019).

Accordingly, PL methods provide a powerful tool to investigate the proximity of a protein of interest, giving insight into potential interaction partners. Beyond the traditional PPI detection methods, advanced multiplexing PL enzymes and enrichment strategies could allow simultaneous molecular interactome mapping for multiple complexes at a time. Prospectively, continuing development of versatile PL methods may largely expand the scope of PL-based discoveries and open more intriguing biological conundrums.
2.6 In vitro methods

Several in vitro methods were developed and dedicated to investigate the binding affinity between two molecules, such as surface plasmon resonance (SPR) or microscale thermophoresis (MST). SPR is probably the simplest method to analyze thermodynamic and kinetic parameters of PPIs. SPR is based on an optical sensing technology that measures changes in the refractive index of the medium near the sensor surface, which can be influenced by a complex formation or its dissociation (Schuck, 1997). The interaction between the proteins will accumulate molecules on the surface that consequently changes the reflection angle of the polarized light (Figure 10A). Meanwhile, SPR is faithful to detect weak interactions, allows real-time kinetic studies, and generates valuable information about the binding affinities of protein complexes free of labelling. Moreover, commercial advanced instruments (e.g. BIAcore™) provide the analysis of interactions between multiple baits and prey at medium-scale level. Apropos MST, similar to SPR, is based on the physical principle of thermophoresis (Figure 10B). During the MST experiment, a temperature gradient in the sample is regulated by thermal elements that contact the glass capillaries, using a focused infrared (IR) laser as heat source. Therefore, relative modifications in the movement along the temperature gradient are detected and quantified by means of either covalently attached or intrinsic fluorophores. The data are then used to measure different parameters of an interaction, such as the dissociation constant, the stoichiometry, or the thermodynamics. The dominance of MST compared to SPR is that the experiment is executed in a solution, thereby obviating the affixation procedures and surface artifacts (Jerabek-Willemsen et al., 2011; Seidel et al., 2013). This method, ultimately, can be used in both artificial buffers and near-native conditions.
Figure 10. Principles of SPR and MST for PPI detection. (A) In SPR method, one of the tested proteins, designated the ligand, is immobilized on a dedicated sensor surface with a gold film. The binding partner, designated the analyte, is injected over the ligand-containing surface. The interaction is detected by changes in refraction of a polarized light of the medium close to the sensor surface upon addition of the analyte. Adapted from figure 5 in (Struk et al., 2019) (B) Layout of the MST instruments and measurement principles. (left) MST is measured in disposable capillaries that hold sample volumes of ~4μl. The sample temperature is regulated by thermal elements which directly contact these capillaries. A focused IR laser induces a local temperature gradient in the sample (typically in the order of 2–6K), triggering thermophoretic movement of molecules. Fluorescent molecules in the capillary are excited and detected through the same objective lens. (right) During an MST experiment, the fluorescence of molecules in solution (yellow dots) is detected over time. For simple FES experiments, detection of the initial fluorescence for 1–5s is sufficient. During a typical MST experiment, the infrared laser is activated after 5s, resulting in thermophoresis towards lower temperatures which can be quantified by measuring the fluorescence decay (in case of positive thermophoresis, as shown here), or fluorescence increases (negative thermophoresis). After a defined time, the infrared laser is switched off, resulting in re-equilibration of the solution by diffusion. Reprinted from figure 2 in (Alexander et al., 2014).

To pave the way for high-throughput PPI detection and quantification, protein microarrays provide an efficient and sensitive multiplex protein analysis, becoming a powerful tool to
probe PPIs. However, the additional optimizations are still needed to attain the accuracy levels of DNA microarrays (discussed further in Section 2.8). Alternatively, a real-time single-molecule Co-IP (RT-smCo-IP) has been developed (Figure 11), which also enables the quantification of the interaction kinetics, similar to the above, based on a cell-free system (Lee et al., 2013).

![Figure 11. Schematic representation of RT-smCo-IP.](image)

In RT-sm-Co-IP, antibodies directed against the bait are immobilized on the glass coverslip via a biotin–neutravidin interaction. After bait proteins are immobilized, cell extracts containing GFP-tagged prey proteins are added and the interaction between bait and prey is visualized with single-molecule fluorescence microscopy. Adapted from figure 6 in (Struk et al., 2019).

### 2.7 In silico methods

Bioinformatics techniques of PPI prediction strengthen and replenish the study of protein interactions, regarding the different facets, such as evolution, function or structure (Figure 12). While experimental in vitro and in vivo methods are widely accepted as the standard analysis for PPIs, in silico methods have emerged as complementary methods to overcome
the limitations of experimental techniques, through completing the missing pieces of experimental PPI data information and providing the clues of PPI mechanisms.

**Figure 12. In silico strategies for PPIs.** Diverse computational tools encompass various stages of PPI study, including the interpretation of protein network topology, the characterization of interface and hot spots, the exploration of PPI chemical spaces for lead discovery and optimization, and the elucidation of complex interactions and dynamics. Reprinted from figure 1 in *(Macalino et al., 2018)*.

In order to understand the total context of potential interactions, it is necessary to develop approaches that predict the full range of possible interactions between proteins *(Zhang, 2009)*. Thereby, a variety of *in silico* methods have been developed to support the known interactions, as well as the new PPI discovery. These include sequence-based approaches *(Hosur et al., 2011)*, structure-based approaches *(Berman et al., 2007)*, chromosome proximity *(Yamada et al., 2003)*, gene fusion *(Enright et al., 1999; Marcotte et al., 1999)*, *in silico* two-hybrid *(Pazos and Valencia, 2002)*, phylogenetic tree *(Sato et al., 2005)*, phylogenetic profile *(Srinivas, 2008)*, and gene expression-based approaches *(Grigoriev, 2001)*.
Asides from PPI predictions, in silico methods are very important to discriminate between true interactions and false positive results from high-throughput biological experiments. There are a number of verification methods that specially address this issue, such as Expression Profile Reliability (EPR) (Mora and Donaldson, 2012), Paralogous Verification Method (PVM) (Deane et al., 2002), Protein Localization Method (PLM) (Sprinzak et al., 2003), and Interaction Generalities Measures IG1 (Saito et al., 2003) and IG2 (Saito et al., 2002).

Last but not least, the computational methods, over decades, have been dedicated to compile information about experimentally-determined PPIs into databases or platforms, which can provide invaluable information with metadata, including different methods, interaction type, subcellular location, and other physiological aspects (Rivas and Fontanillo, 2010; Schaefer et al., 2013). For example, the primary databases are created by only the verified PPIs from individually published studies, such as the Biomolecular Interaction Network Database (BIND) (Bader et al., 2001), the Biological General Repository for Interaction Datasets (BioGRID) (Chatr-aryamontri et al., 2017), the Database of Interacting Proteins (DIP) (Salwinski et al., 2004), Human Protein Reference Database (HPRD) (Keshava Prasad et al., 2009), the IntAct molecular interaction database (IntAct) (Kerrien et al., 2012) or Comprehensive resource of mammalian protein complexes (CORUM) (Ruepp et al., 2010). The meta-databases contain experimentally validated PPIs collected from multiple primary databases and integrate them into one large data model. Furthermore, some of these meta-databases additionally integrated the PPI prediction results and gene expression profiles, which enable deep and multidimensional analysis for a given query, like GeneMANIA (Warde-Farley et al., 2010) and STRING (Szklarczyk et al., 2015). Nowadays, a one-stop comprehensive platform, on which people can interactively explore complex omics datasets via analysis and visualization functions, is urgently needed. The advent of Cytoscape (Shannon et al., 2003) has alleviated this problem to some extent. It is one of the most successful network biology analysis and visualization tools, supported by a large and vibrant community of app contributors.
2.8 High-throughput PPI screenings

Over the past decades, many of the above approaches have been improved and utilized in high-throughput screens, uncovering novel PPIs in diverse organisms. Among them, Y2H, MS-based methods, protein microarray, and PCAs, to date, are the most frequently applied.

Since the seminal study of Fields and Song (Fields and Song, 1989), the Y2H system has been utilized to detect PPIs in diverse cell types (Vidal and Fields, 2014). The first *S. cerevisiae* two-hybrid screening on a large scale was performed by making about 5,300 ORF bait strains and a pooled prey library, leading to the identification of 691 interactions (Uetz et al., 2000). Thereafter, Y2H and the related methods have been a popular tool for high-throughput studies of PPIs (Drees et al., 2001; Snider et al., 2013; Yu et al., 2008). Currently, a human ‘all-by-all’ reference interactome map of human binary protein interactions (HuRI) has been released, which was conducted through Y2H screens, with approximately 53,000 PPIs involving 8,275 proteins (Luck et al., 2020). In this landmark study, the newly established human ORFeome v9.1 upon integration with 17,408 protein-coding genes, encompassing over 150 million pairwise combinations, was commandeered, followed by Y2H screens, resulting in dataset HI-III-20 (Human Interactome obtained from screening Space III, published in 2020), referred as a reference map of HuRI. Together, combining HuRI with all previously published systematic Y2H screening efforts at the Center for Cancer Systems Biology (CCSB) yields 64,006 binary PPIs involving 9,094 proteins, including HI-III-20, HI-II-14 (Human Interactome obtained from screening Space II, published in 2014) (Rolland et al., 2014), and HI-I-05 (Human Interactome obtained from screening Space I, published in 2005) (Stelzl et al., 2005). The resulting HI-union dataset might be the most complete collection of binary PPI data available in human to data. Therefore, it demonstrates convincingly that Y2H can be operated at sufficient throughput for the compilation of proteome-wide interactome maps.

In recent years, the MS-based methods have achieved great improvement in the sensitivity of MS and bioinformatics approaches for accurate data analysis (Armean et al., 2013; Qu et al., 2017; Walton et al., 2015). Two approaches are widely used in large-scale studies at proteome scale: AP-MS and co-fractionation followed by MS (CoFrac-MS) (Havugimana
et al., 2012). In AP-MS, protein baits are purified from a cell lysate and copurified protein preys are detected by MS. Hein and colleagues used a label-free AP-MS strategy, termed quantitative BAC-GFP interactomics (QUBIC), which enables an unparalleled prospection of protein association strength at interactome scale (Hein et al., 2015). In this study, 1,125 GFP-tagged protein baits were affinity purified from HeLa cells. In a proteome-wide manner, 5,400 proteins with 28,500 interactions were resulted to assemble a large-scale map of the human interactome. Another study, for example, using high-throughput AP-MS, is denominated as Biophysical interactions of ORFeome-based complexes (BioPlex) (Huttlin et al., 2015, 2017). Initial experiments provided 23,744 interactions involving 7,668 proteins in HEK293T cells (Huttlin et al., 2015). Referring to CoFrac-MS, it is the only method, heretofore, that does not rely on the genetic manipulation of cells or organisms. CoFrac-MS has thus been able to predict endogenous and unmanipulated protein complexes on a considerably large scale (Havugimana et al., 2012; Wan et al., 2015) and to infer their PPIs (Drew et al., 2017). In CoFrac-MS, protein extracts are fractionated to separate protein complexes whose components are then detected by MS. In a cross-species study, Wan et al. identified and quantified 13,386 protein orthologues across 6,387 fractions obtained from 69 different experiments, generating a draft conservation map consisting of more than one million putative high-confidence co-complex interactions for 9 different species (Wan et al., 2015).

Protein microarray is another potential high-throughput method, allowing the study of thousands of proteins at a time in a single experiment (Zhu et al., 2001). Despite that this recombinant protein-based microarray permits the sensitive and immediate detection between two proteins, even weak and transient PPIs, it still suffers the laborious and costly protein production and purification (Struk et al., 2019). Several improvements have been addressed on this issue, hybridizing the protein in situ synthesis with sophisticated DNA microarray chips. In retrospect, array tools were developed with the implementation of DNA microarray technology as an accurate platform to quantify mRNA expression for thousands of genes on a chip scale (Taub et al., 1983). In the recent past, DNA microarrays have evolved towards protein microarrays, which comprise more than 1,000 elements per array in a high-density format (Angenendt et al., 2006). Thanks to in situ synthesis protein
microarray technologies, a large number of target genes can be tested, allowing a thousand-scale PPI detection at a lower cost and in less time (Jackson et al., 2004), such as nucleic acid programmable protein array (NAPPA) (Figure 13A) (Ramachandran et al., 2004) and multiplexed nucleic acid programmable protein array (M-NAPPA) (Figure 13B) (Yu et al., 2017). These improved technologies largely increased the array throughput, like M-NAPPA, as an ultra-high density proteome microarray, which could be performed on a scale of >10,000 proteins per slide.

Figure 13. Schematic illustration of NAPPA and M-NAPPA methods. (A) Diagram of NAPPA. Cell-free expression systems. RNA or DNA is deposited on the slide surface and rapidly expressing them just before an experiment (~2 h) through the use of various cell-free expression systems (e.g., lysate from wheat germ, insect cells, rabbit reticulocyte and human cells) (Jackson et al., 2004). Figure was adapted from (Wikipedia, 2017). (B) The schematic illustration of how M-NAPPA arrays are processed. Using a standard pin-based arrayer, each spot on M-NAPPA contains plasmids encoding for different proteins-of-interest with the same fusion tag. The genes are then transcribed and translated into recombinant proteins using a cell-free expression system, and captured to the slide surface in situ via a fusion tag antibody. The hits identified from
Because performing PCA assay is simple and inexpensive, it is suitable for large-scale screens for PPIs. Various vector systems for tagging fluorescent protein fragments to proteins of interest have been developed to date, and large-scale screens using PCAs, such as BiFC assays, have been performed in diverse species from yeast to mammalian cells, as detailed in Section 3.4.2.

In addition to the methods mentioned above, certain conventional PPI techniques have also been attempted to perform in a high-throughput fashion, such as surface plasmon resonance (SPR) (Faye et al., 2009), and FRET (You et al., 2006).

2.9 Summary

Over decades, a wide range of PPI detection approaches has been developed at frantic speed. As they are different from their specificity and sensitivity to a given PPI, each method has its own advantages and limitations in various scenarios (as summarized in Table 2). Therefore, the method selection is crucial during the experimental design. To date, no one single method assesses all the specific aspects of PPIs, thus application of multiple combined approaches is highly important to chart a more accurate and complete PPI network in living cells. Moving forward, continuous efforts on large-scale PPI study will further facilitate the protein interaction data generating and deciphering, heralding a new era of proteome-scale interactomics.
Table 2. Comparison of PPI techniques mentioned in this section.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Method</th>
<th>Interaction mode</th>
<th>PPI nature</th>
<th>High-throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Localization</td>
<td>Low-affinity</td>
</tr>
<tr>
<td>Affinity-based</td>
<td>AP-MS and its variants</td>
<td>Co-complex</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Co-IP</td>
<td>Co-complex/Binary</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Y2H and its variants</td>
<td>Binary/Tertiary PPI</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>PCAs</td>
<td>Split-enzyme</td>
<td>Binary PPI</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Binucleolar fluorescent complementation (BIFC)</td>
<td>Binary PPI</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Bioluminescence complementation (BILC)</td>
<td>Binary PPI</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>FRET or BRET</td>
<td></td>
<td>Binary PPI</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Proximity-based labeling (PL) methods</td>
<td></td>
<td>Co-complex</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>In vitro</td>
<td>surface plasmon resonance (SPR)</td>
<td>Binary PPI</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Microscale thermophoresis (MST)</td>
<td>Binary PPI</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Protein microarrays</td>
<td>Binary PPI</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>In silico</td>
<td>PPI prediction methods</td>
<td>Binary PPI</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Verification methods</td>
<td>Binary PPI</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>PPI databases</td>
<td>Co-complex/Binary</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
3. Bimolecular fluorescence complementation (BiFC)

Bimolecular fluorescence complementation (BiFC) is one of extensively used PCAs over the last decade, based on the reconstitution of a fluorescent protein (FP) to directly visualize PPIs *in vivo*. In practice, two proteins of interest (POIs) are fused to split non-fluorescent fragments derived from a FP and then expressed in living cells. If these two proteins interact with each other, the split fragments are assembled and reconstituted to active FP (Figure 14) (Kerppola, 2008). The BiFC signal can then be visualized by fluorescent microscopy or analyzed by flow cytometry. Additionally, the quantification of the BiFC signals can be easily performed in an appropriate manner (Part 1, Chapter II).

**Figure 14. Principle of BiFC.** (A) Three-dimensional structure of a fluorescent protein. The structure of the Venus yellow fluorescent protein (PDB ID: 1MYW from RCSB Protein Data Bank), with the β-strands and α-helices visualized in rainbow colors using PyMOL software (http:// pymol.sourceforge.net/). Two canonical split sites (between 7th and 8th β-strands, and between 8th and 9th β-strands) used for the BiFC assay are shown in white (and also indicated by scissors). (B) Folding topology of a fluorescent protein. The numbered green arrows and the orange boxes indicate β-strands and α-helices, respectively. The two closed circles indicate the positions of the canonical split sites, and the arrowhead indicates the split site between the 10th
and 11th β-strands. The star symbol indicates the fluorophore. (C) Principle of BiFC. The left structure represents the N-terminal BiFC fragment (VN155) fused to interacting protein X, with yellow indicating the fluorophore. The center structure represents the C-terminal BiFC fragment (VC155) fused to interacting protein Y. The α-helices of the bZIP domains of Jun (bJun) and Fos (bFos) are used as examples for the X and Y proteins, respectively. The right structure shows the reconstituted fluorescent protein (FP) and the X/Y protein complex (as a bJun/bFos dimer). Reprinted from figure 1 in *Kodama and Hu, 2012*.

Next, in this section, I describe first the development of FPs for BiFC, and then highlight BiFC advantages over other PPI detection methods. Meanwhile, I also discuss the critical considerations in the use of BiFC. To well illustrate the versatility of BiFC, the BiFC implementations at different throughputs are also presented, including the BiFC and its variants in single PPI experiments or large-scale studies.
3.1 Fluorescent proteins for BiFC assay

Since 1994, different fluorescent proteins and their variants have been discovered or developed (Chudakov et al., 2010; Heim et al., 1994; Shaner et al., 2004; Stepanenko et al., 2011). In the witness of the recent technical progress, many of them have been adopted for BiFC assays (Figure 15).

Figure 15. A timeline of major achievements in BiFC development.

The first application of BiFC can be dated back to the study of Regan and colleagues. Two fragments of GFP (NGFP and CGFP), split in a loop between residues 157 and 158, were fused to antiparallel leucine zippers in Escherichia coli (Ghosh et al., 2000). Subsequently, YFP was successfully reconstituted in mammalian cells using fragmented YFP fused to interacting transcription factors (Hu et al., 2002). Since then, the BiFC has been more widely used to visualize the PPIs in biological research, taking advantage of the different characteristics of fluorescent reporter proteins (Table 3).
Table 3. Main fluorescent proteins used in BiFC assay. nd = not determined.

<table>
<thead>
<tr>
<th>Fluorescent Protein</th>
<th>Excitation/Emission peak (nm)</th>
<th>Split site</th>
<th>Maturation time (min)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGFP</td>
<td>488/510</td>
<td>157/158</td>
<td>53 (in vitro)</td>
<td>(Ghosh et al., 2000)</td>
</tr>
<tr>
<td>EYFP</td>
<td>514/527</td>
<td>154/155</td>
<td>23 (in vitro)</td>
<td>(Hu et al., 2002)</td>
</tr>
<tr>
<td>ECFP</td>
<td>405/485</td>
<td>154/155 or 172/173</td>
<td>50 (S. cerevisiae)</td>
<td>(Hu and Korppola, 2003)</td>
</tr>
<tr>
<td>Cerulean</td>
<td>439/479</td>
<td>172/173</td>
<td>nd</td>
<td>(Shyu et al., 2006)</td>
</tr>
<tr>
<td>Venus</td>
<td>514/529</td>
<td>154/155 or 172/173</td>
<td>40 (in vitro)</td>
<td>(Shyu et al., 2006)</td>
</tr>
<tr>
<td>Citrine</td>
<td>516/528</td>
<td>154/155 or 172/173</td>
<td>nd</td>
<td>(Shyu et al., 2006)</td>
</tr>
<tr>
<td>mRFP1-Q66T</td>
<td>549/570</td>
<td>168/169</td>
<td>36 (in vitro)</td>
<td>(Jach et al., 2006)</td>
</tr>
<tr>
<td>mKG</td>
<td>494/507</td>
<td>168/169</td>
<td>nd</td>
<td>(Ueyama et al., 2008)</td>
</tr>
<tr>
<td>mCherry</td>
<td>587/610</td>
<td>159/160</td>
<td>15 (in vitro)</td>
<td>(Fan et al., 2008)</td>
</tr>
<tr>
<td>mNeptune</td>
<td>600/650</td>
<td>155/156</td>
<td>35 (in vitro)</td>
<td>(Lin et al., 2009)</td>
</tr>
<tr>
<td>mLumin</td>
<td>587/621</td>
<td>151/152</td>
<td>76 (in vitro)</td>
<td>(Chu et al., 2009)</td>
</tr>
<tr>
<td>Dronpa</td>
<td>503/518</td>
<td>164/165</td>
<td>nd</td>
<td>(Lee et al., 2010)</td>
</tr>
<tr>
<td>dRFP</td>
<td>690/713</td>
<td>120/123</td>
<td>nd</td>
<td>(Filonov and Vorkunsha, 2013)</td>
</tr>
<tr>
<td>Tripartite-Split GFP</td>
<td>488/530</td>
<td>193/194 and 212/213</td>
<td>nd</td>
<td>(Cabantous et al., 2013)</td>
</tr>
</tbody>
</table>

Following the initial development with GFP, a monomeric Kusabira-Green (mKG) with spectral characteristics similar to GFP has been developed for a BiFC assay (Ueyama et al., 2008). Otherwise, a notable advance in BiFC based on GFP is Tripartite split-GFP (Cabantous et al., 2013), in which superfolder GFP was used and will be further described in Section 3.4.1, Chapter I.

In addition to GFP-based BiFC, the YFP fragments for a BiFC assay were first demonstrated to visualize calcium-dependent PPIs in living cells (Nagai et al., 2001). Over
the past decade, different variants of YFP appeared in succession, such as EYFP (S65G, S72A, T203Y) \cite{Hu2002}, Citrine (a pH-resistant YFP variant) \cite{Shyu2006} and Venus (a rapidly-maturing YFP variant) \cite{Shyu2006}. A series of efforts have been witnessed on the development of CFP variants in BiFC assays. For example, ECFP split between amino acids 154 and 155 (CN155 and CC155) show fluorescence complementation when fused to bJun and bFos \cite{Hu2003}. An improved CFP variant, termed Cerulean (S72A, Y145A and H148D), has also been successfully applied for BiFC assay \cite{Shyu2006}.

In line with sequential application of GFP, YFP and CFP-based BiFC, the red fluorescent protein DsRed \cite{Matz1999} would further extend the detection range of BiFC. Continuously, arising from its strong tendency to oligomerize \cite{Baird2000}, a monomeric RFP, mRFP1, was thus generated \cite{Campbell2002}. As further better performance in BiFC, mRFP1-Q66T was developed with improved fluorescence intensity \cite{Jach2006}. The mRFP1-Q66T-based BiFC assay was sensitive enough to catch weak and transient PPIs. Another RFP-based BiFC system is the split mutant monomeric RFP, mCherry, with excitation and emission wavelengths at 587/610 nm \cite{Shaner2004}.

Moreover, the far-red fluorescent proteins are important for imaging deep tissue in animals. To achieve this purpose, a monomeric form of Katushka far-red protein, named mKate \cite{Shcherbo2007}, was chosen to develop a far-red-based BiFC. This system eventually exhibited high BiFC efficiency in COS-7 cells \cite{Chu2009}. Interestingly, the site-mutated mKate-S158A, mLumin, could increase 2-fold the mKate brightness in the same study \cite{Chu2009}. Further mutation of mKate is called Neptune, which is the first bright fluorescent protein with an excitation peak reaching 600 nm. The monomeric variant of Neptune, mNeptune \cite{Lin2009}, has been successfully used for BiFC assays in animal tissues. For whole animal imaging, the near-infrared-based BiFC has been developed in the near past \cite{Filonov2013}. The near-infrared protein iRFP-based BiFC exhibits high fluorescence intensity and low cytotoxicity and
utilizes endogenous concentrations of biliverdin chromophore to acquire fluorescence (Filonov et al., 2011).

Generally, tracking BiFC signals is hindered by repeated capturing. To settle this matter, Dronpa, an artificial GFP-like fluorescent protein cloned from Pectiniidae (Ando et al., 2004), was used as a BiFC reporter. Dronpa has a reversible photo-switching activity between the fluorescent and non-fluorescent states. The Dronpa-based BiFC was successfully performed in HEK293 cells (Lee et al., 2010). It will therefore enable the study of protein complex translocation between various cellular compartments.

Nowadays, the single-molecular PPIs in living cells is coming into focus. Nonetheless, current BiFC is limited by the brightness and photo-stability of fluorescent proteins, resulting in insufficient resolution for single molecule tracking. BiFC coupled with photo-activated localization microscopy (BiFC-PALM) (Betzig et al., 2006; Hess et al., 2006), largely alleviates this problem, and allows the imaging and tracking of single-molecule PPIs at sub-diffraction resolution in crowded PPI background of living cells, by using split of photo-switchable (mEos3.2) (Liu et al., 2014) and photo-convertible (PAmCherry1) (Nickerson et al., 2014) fluorescent proteins. Likewise, the newly reported TagBiFC (Shao et al., 2021), which leveraged the split HaloTag system for single-molecule PPI in living cells via super-resolution imaging, provides an alternative approach to tackle this important issue.
3.2 Advantages of BiFC assay

BiFC is a very sensitive method with minimum background (Kerppola, 2008), thus it enables direct visualization of PPIs and has been successfully applied in a wide variety of cell types and organisms(Kerppola, 2006b). Since the inception of PPI study, diverse techniques have been developed, such as FRET, Y2H, and AP-MS. By comparison, BiFC presents several advantages per se over other PPI detection methods.

FRET and BiFC are the two most commonly employed PPI detection methods in cells. Like BiFC, FRET also employs the FPs as reporter signals. It needs energy transfer from an excited donor fluorophore to an acceptor at angstrom distances (10–100Å) (Sekar and Periasamy, 2003) and in a permissive orientation. Therefore it is applicable only to analyze bimolecular, direct PPIs (Sun et al., 2013). In contrast, BiFC assays may generate positive signals, as long as two tag-fused proteins are present within the same protein complex, including direct and indirect interactions between the two proteins. FRET, moreover, requires confocal image capture of two different FPs, as well as accurate and elaborate computation via time-correlated single-photon counting to predict protein interactions (Sekar and Periasamy, 2003).

In an Y2H assay, the system addresses only the qualitative question, yes or no on protein associations. Its experiments drive expression of the target proteins into a rather unnatural context, especially for non-yeast proteins, differing from their native situation. The genetic-manipulated proteins, accordingly, may be misfolded due to the absence of mediating factors. Reciprocally, BiFC conducts context-dependent interaction studies with the native context from which the target proteins derive.

AP-MS method and its variants capture the bait protein complex in the native cellular context, but the weak PPIs will be disrupted by the cell lysis and purification steps (Miteva et al., 2013; Roux et al., 2012). Consequently, AP-MS precludes the weak or transient interaction partners from its final candidate list. The proximity-labelling methods, like BioID, have revitalized the detection of transient and low-affinity interactions for the AP-based methods (Roux et al., 2012). Nevertheless, paired controls are still necessary to
reduce the final false positives (Gingras et al., 2019; Varnaitė and MacNeill, 2016), which would be labor- and time-consuming, as compared to BiFC.

Collectively, these merits make BiFC as one of the most popular methods for the study of PPIs. A whole spectrum of FPs can be used for BiFC analyses, which underpins the multicolor visualization of different protein binding partners at the same time and in the same cell (Kerppola, 2013). Throughout the years, BiFC assays have been utilized in high-throughput screens, uncovering novel PPIs in yeast, plant and mammalian cells (Section 3.4.2, Chapter 1). However, each of these methods has its advantages and limits that make them best suited methods in certain fields. The continuing efforts on BiFC improvement will further extend its application in more extensive circumstances.
3.3 Critical considerations

Despite these advantages mentioned above, there are some potential problems and critical factors that one should consider when performing BiFC.

3.3.1 Topology of BiFC-tag fusions

A wrong BiFC-tag orientation could mask the true PPI-based BiFC signals and change the protein localization, as well as inducing abnormal protein expression. Therefore, it is advised to test, firstly, the full-length FP-fused protein in a same vector backbone to the BiFC-vectors (Waadt et al., 2014), for confirming the target protein localization and expression level. If there is no available localization information, both N- and C-terminal fusion should be made for a subsequent co-localization test. Then choose the one same to the locus of its known interacting protein. When both tag orientations are available, the C-terminal tag fusion is preferable, as it has been reported that N-terminal tag fusion may generate higher background (Waadt et al., 2014). Once the tag orientation is determined, a corresponding split BiFC-tag will be used to generate final BiFC constructs and test the fluorescence intensity of reconstructed FP.

Another strategy is using the split BiFC-tag to determine optimal positions for the fusions (Kerppola, 2006b). FP fragments and both terminal ends of each interaction partner will be fused to test for all eight possible combinations (Figure 16). Moreover, the available combinations should go through the protein localization test and expression level comparison, by immunofluorescence or Western blot analysis.
Figure 16. Combinations of fusion proteins to be tested for BiFC. Fusion proteins that produce an optimal signal must generally be empirically determined. Multiple combinations of fusion proteins should be tested for BiFC. N- and C-terminal fusions can be used to test eight distinct combinations (A through H). Although it may appear that combinations E through H might not be favorable for bimolecular complex formation, this will depend on the precise structures and flexibilities of the fusion proteins, which are difficult to predict. For true interaction partners, it is virtually always possible to find fusion proteins that produce a detectable signal. Adapted from figure 2 in (Kerppola, 2006b).

3.3.2 Linker selection

In general, split BiFC-tags and target proteins are separated by a linker sequence to provide the flexibility of flanking proteins. The linker is crucial for independent motion of the FP fragments and the interaction partners, facilitating FP fragment reconstitution and proper folding of fusion proteins (Wouters et al., 2019). As a result, different linkers with various lengths have been designed, based on the sequences from natural multi-domain proteins (Gokhale and Khosla, 2000). The most frequently applied linkers in BiFC are composed of non-polar glycine and polar serine amino acids (GS) series, developed by Argos (Argos, 1990). These flexible GS-based linkers have been widely applied by employing the (GGGGS)n or (GGSGG)n template, where n indicates the GS copy number. For BiFC
linkers, empirically, GS-based linkers include about 8-15 amino acids, like DGGSGGS (Carayon et al., 2014), GGGSGGGS (Armando et al., 2014), or GSSGGGSGGGSSG (Cannaert et al., 2016). Sometimes, rigid linkers are also implemented to keep a fixed distance between the domains and to maintain their independent functions, for instance ATGLDLELKASNSAVDGTAGPVAT (Kerppola, 2006b), as well as RSIAT and RPACKIPNDLKQKVMNH linker sequences, which have been used in many fusion constructs for BiFC analysis (Hu and Kerppola, 2003; Hu et al., 2002). In addition, researchers have designed many other empirical linkers for some special purposes. Therefore, the linker selection may largely vary on a case-by-case basis.

3.3.3 Improvement of signal-to-noise ratio

The fluorescence intensity of reconstituted fragments should be comparable to that of independent non-fluorescent fragment self-assembly, which makes the true BiFC bright enough to be distinguished from background signals. This signal ratio between true BIFC and background is the so-called signal-to-noise (S/N) ratio, influenced by three main aspects: property of FPs, split-site, and hydrophobic interactions between split non-fluorescent fragments.

All proteins can be split into non-fluorescent fragments, but that does not mean they can all be used in BiFC; those recommended for BiFC analysis are summarized in Figure 15 and Table 3. To reduce the S/N ratio, the spontaneous self-assembly of split fragments, which is endowed by the inherent property of FPs to a large extent, should be avoided. Against this backdrop, two FPs, EYFP and Venus, have been most extensively used in BiFC (Kerppola, 2008; Waadt et al., 2014).

Notably, applying different split-sites within same or different FPs shows concomitant variances of S/N ratio. Several studies reported that GFP could be split at different positions, which could lie in a loop between the 6th and 7th, the 7th and 8th, or the 8th and 9th β-strands (Baird et al., 1999). Given the structural similarity between GFP and EYFP,
fragments of EYFP truncated at residue 155 (designated YN155 and YC155), which are split at a position between the 8th and 9th β-strands, exhibit a relatively high S/N ratio, when used in BiFC (Hu et al., 2002). Another split site at residue 173, generating YN173 and YC173, cut in the loop between 7th and 8th β-strands (Hu and Kerppola, 2003). Similarly, Venus, as a mutated GFP with high fluorescence intensity, can be also truncated at either residue 155 or 173 (Shyu et al., 2006). A great advantage of Venus-based BiFC is that it avoids a short incubation at 30°C to facilitate fluorophore maturation, which is generally required for split YFP fragments (Kerppola, 2006b). Moreover, from a split-site screening study, Venus can be split at the loop between the 10th and 11th β-strands, with lower background fluorescence in BiFC application (Ohashi et al., 2012). Compared to previously published sfGFP-based BiFC (Zhou et al., 2011), the Venus-based BiFC assay had the lowest background fluorescence, albeit the same split-site imposed. Prospectively, this differential effect of the same split site might be also linked to the target proteins used in the test.

Decreasing the hydrophobic interactions between split non-fluorescent fragments has been reported to weaken the BiFC background fluorescence, further improving the S/N ratio (Nakagawa et al., 2011). In an informative study, Venus was split in two fragments, VN155 and VC155, which were used to test bFos-bJun interaction in the BiFC assay. To test the impact of hydrophobicity in BiFC, four residues L201, L207, V150, and I152 were replaced with other hydrophobic amino acids (L, I, V, and A respectively) by site-directed mutagenesis (Figure 17A). As a result, the V150A mutation increased by 8.6-fold the S/N ratio in the BiFC assay of bFos-bJun interaction, significantly improving the S/N ratio (Figure 17B).
**Figure 17. Improvement test of S/N ratio in Venus-based BiFC.** (A) Mutation sites that increase the S/N ratio in Venus-based BiFC. The three-dimensional structure of Venus (PDB ID: 1MYW from RCSB Protein Data Bank) is shown as generated by the PyMOL software (http://pymol.sourceforge.net/). The 7th and 10th β-strands are shown in yellow and green, respectively. The V150, I152, L201, and L207 residues are indicated in magenta. Reprinted from figure 3 in (Kodama and Hu, 2012). (B) Relative BiFC Efficiencies of the Venus Variants. L201 and L207 of bFos-VC155 and V150 and I152 of bJun-VN155 were replaced with amino acids L, I, V, and A by PCR-based site-directed mutagenesis to obtain four substitutions at each position. The replaced bJun-VN155/bFos-VC155 plasmids were transfected into mouse C3H10T1/2 cells. The bFosΔZIP-VC155 plasmid, lacking the carboxy-terminal half of the bFos leucine zipper domain, was used as negative control (background). The mCherry plasmid was also included in the transfection cocktail at a molar ratio of 1:3 to the BiFC plasmids, and was used as internal control for protein expression. Relative BiFC efficiency was calculated from the number of co-fluorescent BiFC cells among mCherry fluorescent cells. The positive BiFC fluorescence signal and the negative backgrounds of the various Venus variants were obtained with bJun-VN155/bFos-VC155 and bJun-VN155/ bFosΔZIP-VC155 respectively. The BiFC efficiency of wild-type bJun-VN155/bFos-VC155 was taken to be 100. Standard deviations were calculated from three independent experiments. Reprinted from figure 3 in (Nakagawa et al., 2011).

### 3.3.4 Controls for BiFC assay

As mentioned previously, the true BiFC fluorescence signals are brought when the two tagged-proteins are interacting with each other and generating a functional FP. Equally, the random collisions of two FP fragments can also occur when they are co-expressed in a limited space of the subcellular compartment. These random encounters can subsequently generate PPI-independent and non-specific signals. Moreover, many BiFC assays, which were routinely implemented by transfection, involve transient overexpression of the candidate proteins. Coupling with the irreversibility of the BiFC, the artificial non-specific signals will be fixed and then measured together with the true PPI-driven fluorescence, contributing to the overall fluorescence signal in the BiFC assay. Accordingly, it is important to design and choose rigorous controls (summarized in **Figure 18**) for determining this non-specific fluorescence or technical background signal, and faithfully detecting interactions that are truly biologically relevant.
Figure 18. Summary of Possible Negative Controls in BiFC Experiments. (A) Interaction of proteins A and B mediates efficient fluorescence complementation and reconstitution of the FP. (B) to (F) Appropriate negative controls (green background). (B) Interaction domain mutated in A, thereby abolishing interaction with B. (C) Interaction domain mutated in B, thereby abolishing interaction with A. Verification that the mutated protein is similarly stable as the wild-type form is additionally required. (D) Ax is closely related to A (e.g., a member of the same protein family) but does no interact with B. (E) Bx is closely related to B but does not interact with A. (F) If none of the controls in (B) to (E) is possible, an unrelated protein, localized in the same subcellular compartment as the proteins of interest, can be used as the last resort. In this case, it is necessary also to provide evidence for this unrelated protein being part of an established interaction (X and Y) that can be reproduced by BiFC. (G) to (N) Inappropriate negative controls (red background). (G) Expression of either the N- or the C-terminal FP fragment alone. (H) N- and C-terminal FP fragments are coexpressed, but without fusion to the proteins of interest. (I) N-terminal FP fragment fused to protein A is expressed alone. (J) C-terminal fragment fused to protein B is expressed alone. (K) N-terminal FP fragment fused to protein A is coexpressed with the unfused C-terminal FP fragment. (L) C-terminal FP fragment fused to protein B is coexpressed with the unfused N-terminal FP fragment. (M) and (N) Unrelated protein Z with different subcellular localization and no positive interaction control for Z and a partner protein (see [F]) is coexpressed with A or B. (O) and (P) Possible orientations of the protein fusions in BiFC assays. It is
important to note that the orientation can have a strong impact on the propensity of spontaneous FP reconstitution (Bracha-Drori et al., 2004; Horstman et al., 2014). Hence, it is essential that, for the negative controls, exactly the same orientations are used as for the positive interaction. N, N-terminal fragment of split FP; C, C-terminal fragment of split FP; *, mutation in the interaction site; Ø, no partner protein present; red outline, no interaction with expressed partner protein possible. Reprinted from figure 1 in (Kudla and Bock, 2016).

One of the most stringent controls in BiFC assay is the use of a mutant negative control where a single mutation or small deletion is introduced into one of the two interacting proteins (Figures 18B and 18C). In most cases, nevertheless, it will be challenging to find an appropriate mutated control, due to lack of protein biochemical or structural information, when examining an interaction between two new novel proteins. If so, a very similar protein from the same protein family can be a surrogate negative control (Figures 18D and 18E). In case this control protein is not available, an unrelated protein can be used as a negative control, when meeting certain conditions. It is desirable that this unrelated control protein co-localizes with the protein of interest in the same subcellular compartment. Moreover, the proof that this unrelated protein can interact with its known interactors using BiFC method, should be reported (Figure 18F). The last negative control assay is, BiFC competition analysis (Hu et al., 2002). One of the binding partners is expressed as an untagged protein, which serves as a competitor when coexpressed with the two fusion proteins, to measure the signal difference with original BiFC, validating the true interactions. Retrospectively, a large number of BiFC-based studies have been published with inappropriate controls (Figures 18G to 18N). For example, the most frequently used controls are combinations of the individual split FP fragments with the complementary split-fragment tagged proteins (Figures 18K and 18L). In addition, the orientation of protein fusions (Figures 18O and 18P), should be taken into consideration, when applying BiFC control, to ideally keep the same topology to the tested interaction. Overall, appropriate controls are crucial to validate BiFC data and establish specificity of the observed PPIs.
3.3.5 Irreversibility of BiFC

The irreversibility of BiFC has been frequently reported (Shyu and Hu, 2008), as a long-standing limitation. Related to this issue, most fluorescent protein-based BiFC complex formation is irreversible, which largely impedes the analysis of dynamic interactions. Instead, this seeming blemish facilitates the visualization of transient or weak protein-protein interactions (Morell et al., 2007) and offers a significant advantage for BiFC-based PPI screening (Ding et al., 2006; Remy and Michnick, 2004a). Indeed, the reversible BiFC development would expand the BiFC application in monitoring the dynamic PPI, providing one alternative method in this domain. As a response to this need, a reversible BiFC system (Lönn and Landegren, 2017), based on a reconstituted infrared fluorescent protein IFP1.4, has been used to study the spatiotemporal dynamics of protein complexes in yeast and mammalian cells (Figure 19A). Another system named splitFAST (Tebo and Gautier, 2019), which specifically and reversibly binds fluorogenic hydroxybenzylidene rhodanine (HBR) analogs, demonstrated rapid and reversible complementation, allowing the real-time PPI visualization (Figure 19B).

Figure 19. Reversible BiFC systems based on split FPs that use exogenous chromophores. (A) IFP-based system, is a split version of IFP1.4 that is reported to be reversible. Reconstitution of IFP1.4 due to the interaction of X and Y activates the fluorescence of the bound biliverdin molecule. (B) splitFAST, a reversible split fluorescent reporter that allows real-time monitoring of both formation and dissociation of a protein assembly. This split system was engineered from the fluorescence-activating and absorption shifting tag (FAST), a small protein of 14 kDa that specifically and reversibly binds HBR analogs, like HMBR. HBR, hydroxybenzylidene rhodanine. HMBR, 4-hydroxy-3-methylbenzylidene rhodanine, which provides green-yellow fluorescence. Modified from figure 5 in (Wiens and Campbell, 2018).
In summary, a good experimental design is a prerequisite for a successful assay. BiFC is a technically straightforward and time-saving PPI detection method. Nonetheless, there are still step-to-step or case-to-case differences in practice. To set up a general standard guide as reference, one protocol was given in Part 1, Chapter II, which focuses on the visualization of protein interactions in cultured mammalian cells via the BiFC assay, using a true example coupling with published data.
3.4 Implementation of BiFC

In the last few years, although several other genuine methods have been developed to analyze PPIs, classical methods and their variants are still widely used by scientists, such as the BiFC-based methods.

3.4.1 Low-throughput BiFC-based applications

BiFC method has been widely applied to detect binary PPIs in many living systems, as discussed previously, especially EYFP- and Venus-based BiFC assays (Figure 20A). In addition to its generic usage, the use of two distinct FPs with different spectra, such as YFP-based BiFC combined with mCherry-based BiFC, enables the visualization of quaternary protein complex in living cells, or, at least simultaneous visualization of two independent binary PPIs (Kodama and Wada, 2009). This double BiFC combined method is designated as coBiFC (Figure 20B).

Besides coBiFC, a lot of attention has been paid on multicolor BiFC, named mcBiFC (Figure 20C), which provides an effective assay to compare the subcellular distributions of protein complexes formed with different binding partners (Kerppola, 2008). In several studies, the split-Venus and Cerulean were used to construct the N-terminal part of Cerulean (1-172aa, Cerulean\textsuperscript{N173}) fused Protein A, the C-terminal part of Cerulean (155-238aa, Cerulean\textsuperscript{C155}) fused Protein B, and the N-terminal part of Venus (1-172aa, Venus\textsuperscript{N173}) fused Protein C (Dard et al., 2018; Vidi et al., 2008, 2010). Interaction of proteins A/B produces a Cerulean signal, whereas proteins B/C interaction generates a Venus signal. With imaging via different excitation and emission wavelengths, the mcBiFC allows studying ternary complexes, and investigates the interactions between three different proteins within the same cells. Alternatively, BiFC-based BRET or FRET (Figure 20D), which involves co-expression of two interacting proteins tagged to YFP- or Venus fragments with one interacting protein tagged to Renilla reniformis luciferase (RLuc) or Cerulean, can be also used for the analyses of ternary protein complexes (Kwaaitaal et al., 2010; Rebois et al., 2006). To take a step further, theoretically, the luciferase can still be
split for BiLC assay, enabling a secondary binary PPI detection. By combination of BiLC and BiFC-based BRET (Figure 20E), reconstituted *Gaussia princeps* luciferase (GLuc) was used for BRET on reconstituted Venus and enabled analyses of quaternary protein complex (Rebois et al., 2008).

Moreover, the applications of sfGFP have to be mentioned here. Since the discovery of its third split site (Cabantous et al., 2005), tripartite split-sfGFP have been reported in studies of both binary and ternary PPI analyses. sfGFP was split into three parts: sfGFP1-9, sfGFP10, and sfGFP11. Each part can be fused to one of the target proteins. The reconstitution of the FP requires all three parts to be brought into proximity, and then demonstrating a ternary PPIs (Figure 20F) (Kellermann et al., 2013; Waldo and Cabantous, 2010). When binary PPIs is needed, only two twenty amino-acids long GFP tags, GFP10 and GFP11, are fused to interacting protein partners, and coexpressed the GFP1-9 fragment as complementary BiFC tag to finally fulfill the GFP-based BiFC investigation (Cabantous et al., 2013). In addition to detecting PPIs, split-sfGFPs (sfGFP1–10 and sfGFP11) were also used for self-assembly, allowing visualization of single protein localization and imaging (Avilov and Aleksandrova, 2018).
Figure 20. Different examples for BiFC applications. BiFC can be applied to investigate (A) interactions of single PPI pairs (BiFC), (B) co-localization of different PPI pairs (coBiFC), (C) competition of two protein complex formations enabling visualization of ternary complex (mcBiFC), (D) analyses of ternary protein complex formations (BiFC-based FRET or BRET) and (E) the analysis of quaternary protein complexes (by combining BiLC and BiFC-BRET), as well as (F) tripartite split-GFP, in which sfGFP is split into three parts that are only capable of forming their chromophore when all three are brought together. The protein fragment and/or donor/acceptor combinations recommended for each application and the resulting excitation and emission maxima or substrate catalysis of the respective (complemented) proteins are indicated. The proteins to be investigated for PPI are labeled from A – D. Adapted from figure 1 in (Waadt et al., 2014).

3.4.2 Large-scale applications of BiFC

A large number of high-throughout studies have been performed for scrutinizing complex protein interactomes in diverse organisms, thanks to current advances in various technologies, including that of BiFC-based screening method.

Since its invention in 2000, upon advantages of simplicity and low-cost, BiFC has become a widely used approach for PPI detections, and thereby suitable for large-scale screens. Indeed, several efforts have been witnessed over the past decade, coupling with the availability of nearly complete hORFeome collections (Lamesch et al., 2007; Luck et al., 2020; Rual et al., 2004; Yang et al., 2011), which enables prospecting PPIs at unparalleled scale in two different formats, arrayed or pooled BiFC screens. To date, large-scale screens using BiFC assays have been reported in yeast, plant and mammalian cells (Table 4).

The facile genetic manipulation on yeast is one of main advantages for BiFC assay. As the first effort on genome-wide screen, Sung et al. initially developed BiFC-tagged fusion plasmids, which allow expression of tagged target proteins in S. cerevisiae (Sung and Huh, 2007). Subsequently, the construction of a S. cerevisiae fusion library expressing each ORF fused with the N-terminal fragment of Venus (VN) was achieved by Huh lab (Sung et al., 2013). To perform a genome-wide BiFC screen for the SUMO interactome, 5,911 VN-tagged fusion (≈95% known yeast proteins) strains were mated with the strain expressing VC-tagged Smt3. Finally, 367 out of 5911 ORFs were identified as Smt3-interacting candidates, by fluorescence microscopy in arrayed format. Similarly, these BiFC-based screenings
were also applied to interrogate ABC (ATP-binding cassette) transporter and TORC1 interactomes in yeast (Chang et al., 2021; Snider et al., 2013).

In the plant, a different BiFC-tagged ORF delivery system was used in a study of core cell cycle protein interactions (Boruc et al., 2010). The GFP-fragment fused constructs were transiently co-expressed in leaf epidermal cells of tobacco by A. tumefaciens-mediated leaf infiltration. Then the high-throughput BiFC assays were performed to test a total of 917 PPIs for 58 cell cycle-related proteins. As result, 341 PPIs were identified as BiFC-positive. In another BiFC-based screen, an Arabidopsis cDNA library comprising ~2×10^5 cDNAs was fused to C-terminal fragment of YFP (CYFP) for PPIs screening of subsets of NYFP-fused baits in Arabidopsis leaf protoplasts (Lee et al., 2012). This screen identified single cDNA clones encoding proteins that interact with bait proteins, VirE2 and VirD2, by co-transfection manner in an arrayed format.

Different from the large-scale screen by BiFC with fluorescence microscopy in yeast and plant, many BiFC screens in the mammalian cells were coupled with fluorescence-activated cell sorting (FACS). For instance, one split GFP-based BiFC system was designed to identify proteins interacting with protein kinase B (PKB), in which the BiFC-positive cells were collected by FACS (Remy and Michnick, 2004b). In this study, C-terminal fragment of GFP (CGFP) was fused to bait PKB, proceeding a large-scale screen with a human cDNA library that expressed NGFP fusions in COS-1 cells, by pooled cotransfection. The resulting DNA, including genome DNA and ORF-containing plasmids, was extracted from BiFC-positive cells, followed by bacterial transformation, and then the single colony derived plasmids were retransfected in COS-1 cells, to perform the second BiFC for removal of false positive candidates. However, usage of the pooled DNA transfection greatly simplified the screening process compared to the one-by-one arrayed format, albeit further improvement is needed. Moreover, application of viral BiFC vectors allows efficient BiFC-based analysis in mammalian cells. Adenoviral BiFC vectors have been generated based on adenovirus high-throughput system and have been used to monitor G protein-coupled receptor (GPCR) activation in human cells by an adenovirus-based β-arrestin BiFC assay (Song et al., 2014). In parallel, a retrovirus-based protein-fragment
complementation assay, termed RePCA, was developed to identify protein–protein interactions in mammalian cells (Ding et al., 2006). In this study, a host cell line was made for stably expressing the N-terminal fragment of Venus (VN) fused protein AKT1. The screen was executed by infecting this stable bait cell line with prey VN-fusion lentivirus, following single fluorescent cell sorting by FACS. Large-scale screens by BiFC may also facilitate drug discovery. As published, mKG-based BiFC was used to screen for PPI inhibitors in a natural product library based on a cell-free system (Hashimoto et al., 2009), which provided a deeper understanding of potential drug actions, therefore demonstrating great potential for high-throughput BiFC screening in drug discovery.

Taken together, recent application of BiFC in large-scale studies has demonstrated its potential for uncovering protein interactomes in live cells. In particular, BiFC-tagged cDNA or ORFeome-based cell libraries have allowed BiFC assays to be more widely applied in high-throughput studies. Upon the well-developed genome-wide screens, such as pooled overexpression screen, and drop-out screens (e.g. RNAi/shRNA screen or CRISPR-Cas9 screen), the arrayed large-scale BiFC screens are more popularly used, relying on combining multiple individual BiFCs in a microplate rather than a pooled format. Though, in few studies, pooled transfection or infection was used in BiFC screen, which is far away from one-gene-one-cell output, the current next-generation sequencing (NGS) is hardly implemented during the deconvolution step. This to-be-improved pooled screening to some degree impaired its original intention. Whereas continuous efforts have paved the future for a versatile and easy-taking BiFC-based screening, there is a long way to realize a rigorous quantification-based pooled method.
Table 4. Summary of BiFC-based large-scale PPI studies in various contexts. Different colors indicate various organisms, including yeast, plant, mammalian cells or cell-free system.

<table>
<thead>
<tr>
<th>Fluorescent Protein</th>
<th>Bait</th>
<th>Prey</th>
<th>High-throughput Strategy</th>
<th>Pooled screening</th>
<th>Experimental cells or Organisms</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Venus</td>
<td>VC-SUMO</td>
<td>VN-tagged strains (5911 ORFs)</td>
<td>Individual mating (array screening)</td>
<td>No</td>
<td>Yeast (Sung et al., 2013)</td>
<td></td>
</tr>
<tr>
<td>Venus</td>
<td>17 VC-ABC transporter genes</td>
<td>VN-tagged strains (209 ORFs)</td>
<td>Individual mating (array screening)</td>
<td>No</td>
<td>Yeast (Snider et al., 2013)</td>
<td></td>
</tr>
<tr>
<td>Venus</td>
<td>3 VC-Proteins from TORC1</td>
<td>VN-tagged strains (5911 ORFs)</td>
<td>Individual mating (array screening)</td>
<td>No</td>
<td>Yeast (Chang et al., 2021)</td>
<td></td>
</tr>
<tr>
<td>mKG</td>
<td>3 pairs of mKG_N- or mKG_C-ProteinA/mKG_C-ProteinB</td>
<td>123599 potential PPI inhibitors</td>
<td>Protein solution mix (array screening)</td>
<td>No</td>
<td>in vitro (Cell-free) (Hashimoto et al., 2009)</td>
<td></td>
</tr>
<tr>
<td>GFP</td>
<td>58 eGFP-core_cell_cycle proteins</td>
<td>58 nGFP-core_cell_cycle proteins</td>
<td>Transiently coexpressed in leaf by infiltration</td>
<td>No</td>
<td>Plant (tobacco epidermal cells) (Boruc et al., 2010)</td>
<td></td>
</tr>
<tr>
<td>mYFP</td>
<td>YN-CPK3</td>
<td>YC-cDNA plasmid library</td>
<td>Pooled co-transfection of YN-bait and YC-library</td>
<td>Yes</td>
<td>Plant (Arabidopsis) (Berendzen et al., 2012)</td>
<td></td>
</tr>
<tr>
<td>YFP</td>
<td>19 YN- or YC-proteins (G-proteins)</td>
<td>33 YN- or YC-proteins (G-proteins)</td>
<td>Individual BiFC (cotransfection)</td>
<td>No</td>
<td>Plant (Arabidopsis) (Klopffleisch et al., 2011)</td>
<td></td>
</tr>
<tr>
<td>YFP</td>
<td>VirE2/VirD2/CTE-nYFP</td>
<td>eYFP-tagged cDNA array library (~100,000 clones)</td>
<td>Cotransfection in microplate (array screening)</td>
<td>No</td>
<td>Arabidopsis leaf protoplasts (Lee et al., 2012)</td>
<td></td>
</tr>
<tr>
<td>YFP</td>
<td>6 nYFP-tagged core telomere proteins</td>
<td>eYFP-tagged retroviral array library (hORFeome v3.1)</td>
<td>Coinfection in microplate (array screening)</td>
<td>No</td>
<td>HTC75 cells (Lee et al., 2011)</td>
<td></td>
</tr>
<tr>
<td>Venus</td>
<td>VC-GPCRs</td>
<td>VN-β-arrestin 2</td>
<td>adenovirus-based cotransduction (array screening)</td>
<td>No</td>
<td>U-2 osteosarcoma (OS) cells (Song et al., 2014)</td>
<td></td>
</tr>
<tr>
<td>GFP</td>
<td>GFP[2]-PKB</td>
<td>GFP[1]-tagged human brain cDNA library</td>
<td>Pooled co-transfection of bait and prey</td>
<td>Yes</td>
<td>COS-1 cells (Remy and Michnick, 2004b)</td>
<td></td>
</tr>
<tr>
<td>Venus</td>
<td>Stable bait cell line expressing VN-AKT1</td>
<td>VC-tagged prey lentivirus</td>
<td>Pooled infection using VN-AKT1-expressing stable cell line</td>
<td>Yes</td>
<td>Hela cells (Ding et al., 2006)</td>
<td></td>
</tr>
<tr>
<td>YFP</td>
<td>Stable bait cell line expressing ROP18/ ROP18II-nYFP</td>
<td>eYFP-tagged prey retrovirus (hORFeome v7.1)</td>
<td>Individual infection (array screening)</td>
<td>No</td>
<td>HTC75 cells (Xia et al., 2018)</td>
<td></td>
</tr>
<tr>
<td>Venus</td>
<td>VN/VCAHCY</td>
<td>VN/VCAHCY hORFs</td>
<td>Cotransfection (array screening)</td>
<td>No</td>
<td>HEK293T cells (Lepur et al., 2016)</td>
<td></td>
</tr>
</tbody>
</table>
4. Biology of Hox proteins

In 1978, Homeobox genes were first identified as key regulators of embryogenesis in the fruit fly *Drosophila melanogaster* (*Lewis, 1978*). It was not until 1990 that “HOX” genes and their protein structures were confirmed in humans (*Kamps et al., 1990; Nourse et al., 1990*).

The HOX genes, encoding transcription factors, share a common 120bp DNA sequence, named homeobox, which codes for a highly conserved homeodomain (HD) that mediates DNA binding (*Gehring et al., 1994*). Given the illustration of the Hox-PBC-DNA complex (*Figure 21A*), the HD of HOX has a three-helix structure, with a N-term arm that binds into the minor groove of DNA. The helix 3 in HD, contacts the major groove of DNA, also termed the recognition helix, which has a more conserved sequence signature than other regions in HOX HD (*Figure 21B*). Additionally, in most of the Hox proteins, there exist a short conserved motif, called hexapeptide (HX), mediating contact with members of the PBC class of proteins, such as Extradenticle (Exd) in *Drosophila* (*Mann and Chan, 1996*).

![Figure 21. Structure and sequence signature of Hox proteins.](image)

(A) Structure of Hox-PBC-DNA complex, here illustrated by the Ubx-Exd-DNA crystal structure (*Passner et al., 1999*). (B) The hexapeptide (HX) motif and the homeodomain (HD) are the most prominent sequence signatures of Hox proteins. The sequence of HD is highly conserved, as illustrated here by aWebLogo (established usingweblogo.berkeley.edu software) derived from a compilation of human/mouse and *D. melanogaster* sequences. The HD contains three alpha helices and contacts DNA in the major (via helix 3) and minor (via the N-terminal arm) grooves. The HX, a short motif located upstream of the HD, contacts protein partners of the PBC class, which are also HD-containing proteins, allowing the cooperative assembly of a Hox-PBC-DNA complex. Adapted from figure 1A in (*Rezsohazy et al., 2015*).
In vertebrates, genetic and molecular analyses revealed that HOX genes are arranged in genomic clusters that have undergone duplications (Duboule, 2007), such that most vertebrates have four clusters. For example, humans have 39 HOX genes arranged in four different clusters (Figure 22). The linear organization of HOX genes within each cluster correlates with their spatial and temporal expression profile along the anterior–posterior axis of the body, displaying thirteen paralogue groups (PGs). Paralogous Hox proteins also show high levels of sequence conservation. According to their between-similarity of sequence, PGs are further grouped in anterior, central and posterior classes, which reflects their expression profile in patterning the body plan of bilaterian embryos (Wellik, 2007).

![Genomic organization of the HOX genes of the Drosophila and vertebrate.](image)

**Figure 22.** Genomic organization of the HOX genes of the *Drosophila* and vertebrate. The arrangement 5′ to 3′ is conserved across these species. Members of the same paralogue group are colored identically. Modified from figure 1a in (Jia, 2017).

In contemporary studies, functions of HOX proteins have been dissected extensively. As transcription factors, HOX is critically involved in morphogenetic and
organogenesis processes during animal development (Hombria and Lovegrove, 2003), through the control of various cellular functions including differentiation, proliferation, migration or death (Sánchez-Herrero, 2013). Given that Hox functions truly require a high degree of specificity, the highly conserved Hox HDs alone are not sufficient for dictating the stringent genetic specificity from the distinct phenotypes. One well-established solution to this paradigm is that Hox proteins do not act alone, but cooperate with different factors to fulfill their functions (Mann et al., 2009). The transcription factors, interacting with HOX proteins in the transcriptional process can be divided into three groups (Figure 23A), including Hox protein per se, cofactors that have DNA binding domains, which can form Hox/cofactor complex and bind DNA, and general factors that do not bind DNA, but are part of multiprotein-DNA complex and contact with Hox or cofactors. The two well documented Hox interaction proteins are PBC (Extradenticle/Pbx) and HMP (Homothorax/Meis/Prep) (Figure 23B). These proteins contain a different HD with the presence of three extra residues in between helix 1 and helix 2 (Penkov et al., 2000), as known as Three Amino-acid Loop Extension (TALE) subclass of homeodomains (Bertolino et al., 1995). Consequently, PBC and HMP cofactors have the capacity to form dimeric or trimeric complexes on DNA with Hox proteins, improving the DNA-binding specificity of Hox proteins. Besides these two classes of cofactors, only few additional transcriptional cofactors of a specific HOX have been identified. A Y2H screen with the drosophila Ubx protein led to the characterization of less than 15 TFs as interacting partners (Bondos et al., 2006). A similar approach with the mouse Hoxa1 protein also led to the identification of about 30 TFs (Lambert et al., 2012). Taking advantage of the BiFC technique applied to Drosophila embryo, Baëza et al. identified 35 new cofactors for 5 drosophila Hox paralogs (antp, abdA, abdB, ubx and scr) (Baëza et al., 2015). More recently, the interactome of Hox transcription factor Ubx was scrutinized using BioID, in a multi-level and lineage-specific manner. This study demonstrated that Ubx can interact with only a few lineage-restricted factors, in spite of the fact that hundreds of Ubx interaction partners were identified in different biological contexts (Carnesecchi et al., 2020).

Moreover, the Y2H assay on Hoxa1 identified about 20 non TFs interactors opening the field of HOX functions. Then, Hox proteins are also linked to mRNA translation, DNA repair, initiation of DNA replication, and possibly modulation of signal transduction (Rezsohazy, 2014). For comprehensive discussion of Hox interacting
proteins, it is recommended to refer to excellent reviews by Rezsohazy and his colleagues (Rezsohazy, 2014; Rezsohazy et al., 2015). Concerning current known Hox-interaction partners in human, the initial short list has grown considerably over the last decade, so far, an updated exhaustive Hox-interacting protein list is accessible in BioGRID (https://thebiogrid.org/) (Oughtred et al., 2021) with a study-to-study annotation for each known candidate, including both low- and high-throughput approaches. From this source, 428 interactors for the HOXA, 289 interactors for the HOXB, 254 interactors for the HOXC and 143 interactors for the HOXD paralogs are identified to date with some discrepancies regarding Hox proteins. For example, only 1 partner is reported for HOXA4 or HOXA6 whereas 300 interactors are listed for HOXA1.

Figure 23. (A) Diagram of a generalized Hox transcription complex. The Hox protein (brown) is shown binding to DNA, as are cofactors (blue) that interact both with the Hox protein and adjacent DNA elements, while general factors (grey) are recruited solely by means of protein–protein interactions. (B) Diagram of Hox, PBC, and HMP family proteins. Hox, PBC, and HMP proteins are shown schematically with N-termini to the left and C-termini to the right. The diagram is intended to represent one generic member of each family and is not drawn to scale. Interacting factors are listed above each protein family at the approximate location reported to contain the binding site for that factor. The dashed
line in PBC indicates the longer C-terminus present in PbxA. Note that all members of a family may not engage in all interactions shown. For instance, Prep proteins may not interact with CBP or TORC. Also, interacting proteins for which binding sites have not been mapped are not shown, for instance, PBC proteins are reported to bind CBP, but the exact binding site in PBC has not been delineated. Adapted from figure 1 in *Ladam and Sagerström, 2014*.

Reasonably, to give a glance at the application of high-throughput BiFC (HT-BiFC) screen in Hox-related study, following an investigation of the PubMed database using keywords, such as (HOX) AND ((BiFC) OR (Bimolecular fluorescence complementation) OR (PCA) OR (Protein-Fragment Complementation Assay)) AND ((high-throughput) OR (large-scale) OR (pooled screen) OR (arrayed screen) OR (ORFeome)), the HT-BiFC screen has not been applied on this issue.

Therefore, during my PhD, I first applied the BiFC technique to decipher the mode of interaction of Hox proteins with their canonical TALE partners in living cells. We identified new paralog-specific TALE-binding sites that are used in a highly context-dependent manner. In addition, I developed a high-throughput BiFC approach that enabled the investigation of systematic HOX interactomes in living cells, which further represented a paradigm for understanding common and specific interactions between different Hox proteins.
Chapter II. Results
Part 1. Bimolecular Fluorescence Complementation (BiFC) and Multiplexed Imaging of Protein–Protein Interactions in Human Living Cells *(book chapter)*

**Yunlong Jia**¹, Françoise Bleicher¹, Jonathan Reboulet¹ and Samir Merabet¹,#.

¹ Institut de Génomique Fonctionnelle de Lyon, UMR5242, Université de Lyon, Université Lyon 1, CNRS, Ecole Normale Supérieure de Lyon, 46, allée d’Italie 69364 Lyon Cedex 07 France

# To whom correspondence should be addressed.

Tel: +33(0)6.22.82.11.88; E-mail: samir.merabet@ens-lyon.fr

*(In press) Multiplexed Imaging: Methods and Protocols, Methods in Molecular Biology, 2021,*

Bimolecular Fluorescence Complementation (BiFC) and Multiplexed Imaging of Protein–Protein Interactions in Human Living Cells

Yunlong Jia, Françoise Bleicher, Jonathan Reboulet, and Samir Merabet

Abstract

Deciphering protein–protein interactions (PPIs) in vivo is crucial to understand protein function. Bimolecular fluorescence complementation (BiFC) makes applicable the analysis of PPIs in many different native contexts, including human live cells. It relies on the property of monomeric fluorescent proteins to be reconstituted from two separate subfragments upon spatial proximity. Candidate partners fused to such complementary subfragments can form a fluorescent protein complex upon interaction, allowing visualization of weak and transient PPIs. It can also be applied for investigation of distinct PPIs at the same time using a multicolor setup. In this chapter, we provide a detailed protocol for analyzing PPIs by doing BiFC in cultured cells. Proof-of-principle experiments rely on the complementation property between the N-terminal fragment of mVenus (designated VN173) and the C-terminal fragment of mCerulean (designated CC155) and the partnership between HOXA7 and PBX1 proteins. This protocol is compatible with any other fluorescent complementation pair fragments and any type of candidate interacting proteins.

Keywords BiFC, Multicolor, Living cells, Protein–protein interaction, mVenus, mCerulean

1 Introduction

Studying protein–protein interactions (PPIs) is central for the understanding of the molecular mechanisms underlying protein function. These molecular contacts will change from cell to cell, occurring in different places and with various affinities within the cell. Understanding protein function therefore requires capturing underlying PPIs in the correct cell context and in native conditions [1].

Over the past decade, different methods have been developed for PPI visualization in living cells, but the most commonly used hitherto are bimolecular fluorescence complementation (BiFC) and fluorescence resonance energy transfer (FRET), which can be applied in a variety of model organisms [2, 3].
BiFC is based on the reconstitution of a monomeric fluorescent protein (FP) such as cyan fluorescent protein (CFP) [4] and yellow fluorescent protein (YFP) [5] from two complementary nonfluorescent subfragments (N- and C-terminal subfragments) upon spatial proximity. Interaction between a bait protein and a prey protein fused to such complementary subfragments is sufficient to lead to the reconstitution of the FP, resulting in an emission signal upon excitation (Fig. 1) [6]. BiFC allows not only the detection of PPIs but also has the ability to determine the PPI subcellular location and PPI affinity in the live cell. BiFC is however of irreversible nature, which, unlike FRET, forbids analyzing dynamic complex formation and dissociation [7, 8].

FRET and BiFC have distinct advantages and limitations, which make the two methods complementary. FRET usually requires high expression levels of fusion proteins and depends on the very close proximity (<30 nm) and the orientation of the two chromophores. These criteria are not always easily met, especially with large proteins or when prior structural knowledge is lacking. In contrast, BiFC stabilizes weak PPIs, allowing their detection upon normal expression levels. Moreover, FRET requires fluorescence measurement parameters that are not simple (fluorescence lifetime or precise intensity measurements), whereas BiFC signals can directly be detected under a fluorescence microscope or fluorescence-activated cell sorting (FACS).

As a simple and cost-efficient strategy, BiFC is a powerful and sensitive tool for protein interaction detection in living cells [2, 7, 9]. In this protocol, we focus on the visualization of HOXA7-PBX1 interactions in HEK-293 cells. HOX proteins are homeodomain (HD) transcription factors that play critical roles during embryonic development and in adults, specifying cell fates and organs in all tissues [10]. PBX proteins are able to interact with the large majority of HOX proteins and as such are considered as generic HOX cofactors for gene regulation in development and disease [11]. Structural studies revealed that HOX-PBX interactions are mediated by a short conserved HOX protein motif called hexapeptide (HX) [12]. In addition, recent work showed that the HOXA7-PBX1 interaction is also dependent on conserved residues of the HOXA7 HD [13]. This molecular knowledge on HOXA7-PBX1 interaction properties makes the HOXA7/PBX1 partnership an ideal model system for providing a detailed BiFC protocol in human live cells (Fig. 2).
2.1 BiFC Vector Designing and Cloning Constructs

The vector design and FP fragment choice are a crucial step in the BiFC assay. We take advantage of the properties of the split fluorescent protein mVenus [5] and mCerulean [4]. mVenus, as a brighter and more photostable version of yellow fluorescent protein (YFP), is reported to be a rapidly maturing monomer (see Note 1) with moderate acid sensitivity. It can be separated into two nonfluorescent and slightly overlapping fragments: the N- and C-terminal fragment of mVenus (VN1-172 and VC155-238). Similar to mVenus, mCerulean also derives from the original green fluorescent protein (GFP) and is a rapidly maturing monomer. It can be divided as N-terminus (CN1-172) and C-terminus (CC155-238). All related FP fragments are summarized in Table 1.

In our test, fusions with the N-terminal fragment of mVenus (VN173) and C-terminal fragment of mCerulean (CC155) were...
Fig. 2 Investigation of interaction properties between HOXA7 and PBX1 proteins. (a) Structure representation of motifs and domains involved in the HOXA7/PBX1 partnership. HOX proteins use hexapeptide (HX, W-containing) motif to interact with PBX homeodomain (HD, PYP-containing). PBX domains, PBCA and PBCB, are also indicated. In HOX/PBX dimerization, PYP residues of the PBX HD participate in the formation of a hydrophobic pocket that interacts with the W residue of HOX HX motif, as determined by the HOX/PBX crystal structure [14]. (b) Representation of the HOXA7 proteins used in this protocol. Shaded regions represent the domains with mutation.

Table 1
Fluorescent protein (fragments) used in BiFC assay

<table>
<thead>
<tr>
<th>t.1</th>
<th>FP fragments</th>
<th>Dissection point</th>
<th>Ex/Em (filter)</th>
</tr>
</thead>
<tbody>
<tr>
<td>t.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t.3</td>
<td>VN173/VC155 (mVenus)</td>
<td>1–172aa; 155–238aa</td>
<td>515/527 (mVenus)</td>
</tr>
<tr>
<td>t.4</td>
<td>CN173/CC155 (mCerulean)</td>
<td>1–172aa; 155–238aa</td>
<td>433/475 (mCerulean)</td>
</tr>
<tr>
<td>t.5</td>
<td>VN173/CC155 (mVenus/mCerulean)</td>
<td>1–172aa; 155–238aa</td>
<td>504/513 (mVenus)</td>
</tr>
<tr>
<td>t.6</td>
<td>mCherry (as BiFC signal normalizer)</td>
<td>–</td>
<td>587/610 (mCherry)</td>
</tr>
</tbody>
</table>

used for BiFC (see Note 2). The excitation and emission spectra of VN173/CC155 complementation are green-shifted in comparison to mVenus (Fig. 3). This slight shift is caused by the absence of the
Fig. 3 Excitation and emission spectra of the different FP fragment combinations. (a) The excitation spectra for each pair of FP fragment, shown as solid lines; (b) the emission spectra for each pair of FP fragment, shown as dashed lines. (Adapted from reference (2))

T203Y mutation in mCerulean [15]. However, signals resulting from the VN/CC complementation are efficiently detected, using mVenus setting. Moreover, this combination makes available the multicolor BiFC assay to investigate two different PPIs simultaneously (see Note 3).
Table 2
References of sequences used in BiFC assay

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pcDNA3 plasmid</td>
<td><a href="https://www.addgene.org/vector-database/2092/">https://www.addgene.org/vector-database/2092/</a></td>
</tr>
<tr>
<td>mCherry</td>
<td><a href="https://www.addgene.org/vector-database/6496/">https://www.addgene.org/vector-database/6496/</a></td>
</tr>
<tr>
<td>mVenus</td>
<td><a href="https://www.addgene.org/browse/sequence_vdb/6524/">https://www.addgene.org/browse/sequence_vdb/6524/</a></td>
</tr>
<tr>
<td>mCerulean</td>
<td><a href="https://www.snapgene.com/resources/plasmid-files/?set=fluorescent_protein_genes_and_plasmids&amp;plasmid=mCerulean">https://www.snapgene.com/resources/plasmid-files/?set=fluorescent_protein_genes_and_plasmids&amp;plasmid=mCerulean</a></td>
</tr>
<tr>
<td>HOXA7 (coding region)</td>
<td>CCDS5408.1</td>
</tr>
<tr>
<td>PBXI (coding region)</td>
<td>CCDS1246.1</td>
</tr>
</tbody>
</table>

*The Consensus CDS (CCDS, https://www.ncbi.nlm.nih.gov/CCDS) project is a collaborative effort to identify a core set of human and mouse protein coding regions that are consistently annotated and of high quality.

1. All constructs are cloned into pcDNA3 plasmid. The final plasmids used in BiFC are as follows, with reference sequences listed in Table 2.

   pcDNA3-VN173-HOXA7s, refer to Fig. 2b, including:
   (a) pcDNA3-VN173-HOXA7 as positive control.
   (b) pcDNA3-VN173-HOXA7[dGA] as negative control (see Note 4).
   (c) pcDNA3-CC155-PBX1.
   (d) pcDNA3-mCherry, as internal control and signal normalizer (see Note 5).

2. The pcDNA3-VN173-HOXA7s vectors harbor a 2X FLAG epitope tag between VN fragment and HOX protein sequence (see Note 6). For pcDNA3-CC155-PBX1, there are two amino acids separating the CC and PBX1, which are generated from a cloning site, as a pseudo-linker (see Note 7). The details about protein fusion vectors are shown in Fig. 4.

2.2 Cell Culture and Plasmid Transfection

1. HEK-293 cells were obtained from the American Type Culture Collection (ATCC) through LGC Standards Sarl (FR). Cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM-GlutaMAX-I, Gibco by Life Technologies) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS) and 1% (v/v) penicillin-streptomycin (5000 U penicillin and 5 mg streptomycin/mL), incubating at 37 °C, in an atmosphere of 5% CO₂.

   One or 2 weeks before transfection, a stock of HEK-293 cells should be maintained. Cells must remain below 80% confluent (see Note 8).
Fig. 4 Plasmids designed for BiFC. (a) pcDNA3-VN173-HOXA7s plasmid. Different HOXA7 constructs and VN173 fragment are cloned into EcoRI and Xhol sites, the expression being driven by CMV promoter. 2X FLAG epitope tag is used for linking VN and HOX proteins; (b) pcDNA3-CC155-PBX1 plasmid. CC155 fragment is inserted between EcoRI and Xhol sites, followed by PBX1 sequence cloned at Xhol and Xbal sites; (c) pcDNA3-mCherry plasmid. mCherry sequence is located between Xhol and Xbal sites, as further internal control.

2. Flat bottom cell culture 6-well plates for cell seeding.
3. Malassez counting chamber or equivalent.
4. Microscope coverslip 22 × 22 mm for cells growing on, stored in 100% ethanol (see Note 9).
5. JetPRIME (Polyplus transfection, France) was used according to the manufacturer’s instructions for transfecting HEK-293 cells.

2.3 Confocal Imaging
1. A confocal microscope system (see Note 10). In our case, we used Zeiss LSM780 confocal system.
2. Microscope settings for image collection (see Note 11):
Argon as laser source (MBS 488/561/633 nm), using the 561 nm and 488 nm laser lines for the excitation of mCherry and VN/CC BiFC, respectively.

Emission filters BP590–630 nm and RP499–543 nm were used to detect mCherry and reconstructed mVenus-like (VN/CC) signals, respectively.

Select an image size of 1024 × 1024 pixels and pinhole of 1.0 AU (see Note 12).

Set detector gain as 750 and laser power as 1.0% (mCherry detection) or 1.5% (VN/CC BiFC detection).

3. Microscope slides 76 × 26 mm for mounting cell-coated coverslip under microscope.

2.4 Data Analysis

1. Mainstream desktop computer.

2. Freely available Fiji imaging software (a distribution of ImageJ, https://imagej.net/Fiji) [17].

3 Methods

3.1 Transfection of BiFC Plasmids

1. In a cell culture hood, prepare the sterile coverslips, and place them on blotting paper for air-drying 15 min.

2. Meanwhile, prepare two 6-well plates, and distribute 2 mL of culture medium for each well.

3. Once coverslips are fully dried, carefully place coverslip on the media surface to each well. Floating coverslip can be pressed down using one 200 µL sterile pipet tip. Ensure that there are no air bubbles between coverslip and well bottom, and then gently shake the plate, and circle around the media to cover the coverslip.

4. Prior to seeding cells, verify stock cell status in the incubator. The cells should have enough confluence (60–80%), and viability of cells must be over 90% using the trypan blue exclusion method.

5. Subsequently, seed 500,000 HEK-293 cells (see Note 13) in each coverslip-containing well.

6. Incubate cells in 6-well plates at 37 °C and 5% CO₂ for 16–24 h or until cells reach a confluence of 60–80%, which is optimal for transfection.

7. Before transfection, prepare plasmid DNA diluted solution (see Note 14), at the following concentrations: 0.5 μg/µL, pcDNA3-VN173-HOXA7s. 0.5 μg/µL, pcDNA3-CC155-PBX1. 1 μg/µL, pcDNA3-mCherry.
8. A three-plasmid transfection system is used in BiFC assay for each HOXA7 construct condition. Prepare a 1.5 mL Eppendorf with 200 µL of jetPRIME buffer for each transfection condition. Add 1 µL of each VN/CC/mCherry-containing plasmid solution for a total 2 µg DNA/reaction. Mix by vortexing for 10 s and spin down.

9. Add 4 µL jetPRIME reagent, vortex for 5 s, and spin down briefly.

10. Incubate for 15 min at room temperature.

11. Gently add 200 µL of transfection mix to HEK-293-containing well, and distribute evenly.

12. Incubate the plate for 20 h (see Note 15) at 37 °C and 5% CO₂, and then image the transfected cells.

3.2 Detect Fluorescent Signals

1. Mount carefully the cell-coated coverslip on a glass slide for image capture under microscope.

2. Image all samples using identical settings (refer to “2.3 Confocal Imaging” in “2. Materials” part) for each FP channel, using a 20× objective.

3. Begin focus on the z-stack that produces the greatest BiFC signal using the fine adjustment knob on the microscope, and then move the focus along the Z-axis to define the first and last sections.

4. Take at least 3 images from different areas of each slide, ensuring that each image includes at least 100 cells (see Note 16).

3.3 Semi-quantitative BiFC Analyses

1. Open original images in a software that allows “maximum intensity projection.” In this protocol, we use the ZEN 2.3 SPI FP1 black edition software as example.

2. In the Processing tab, create a maximum projection (see Note 17) for each image (Method → Maximum intensity projection); choose Z in Method Parameters.

3. Save all newly created maximum projection images, and then open them in Fiji, choosing Hyperstack, Grayscale, and Auto-scale (see Note 18) in Bio-Formats Import Options (Fig. 5a).

4. In the menu bar, go to Analyze and click Set Measurements. Select on Area, Integrated density, Mean gray value, Stack position, and Display label, and keep redirect to “None” in the dropdown list (Fig. 5b, c). This setting will be applied to all image measures.

5. Measure the mean gray value for mVenus (BiFC) and mCherry channels, by pressing “M” on keyboard. The mean values will appear in a new floating window, “Results.”
Fig. 5 Screenshots showing how to semiquantitate BifC fluorescence based on acquired images. (a) Import the original images to Fiji. (b, c) Set appropriate measure parameters. (d) Enable brush selection for background mean gray value measures. (e) Background mean gray value measure on mCherry channel. (f) Background mean gray value measure on BiFC channel.

6. Then the background fluorescence for each channel needs to be measured, respectively.

7. In the main interface of Fiji, double-click the second graph icon. Tick on *Enable selection brush*, and set proper pixel size (Fig. 5d).

8. Reselect the image, and on mCherry channel, place the cursor in regions devoid of cells. Holding the “Shift” key on the keyboard, single-click to draw hollow dots as much as possible (≥20) (Fig. 5e).

9. Measure mean value of selected regions by pressing “M,” as background value of mCherry channel.

10. Roll the middle mouse button to change the channel to BiFC channel; the selected region is fixed as previous location; measure the background value of BiFC channel by pressing “M” (Fig. 5f).

11. Repeat steps 1–10 to measure all images. For each image, four different values are measured, global BiFC mean value, global mCherry mean value, BiFC mean background (BG) value, and mCherry mean BG value.
**HOXA7a/PBX1 interaction in HEK-293**

![Graph showing relative BIFC intensities](image)

**Fig. 6** Quantification of BIFC with mutated HOXA7 proteins in HEK-293. Normalized BIFC values were standardized by positive control VN173-HOXA7/CC155-PBX1 from images of HEK-293 cells with the indicated HOXA7 construct combinations using a 20X objective. Compared to positive control, BIFC signal of negative control HOXA7(dGA)/PBX1 exhibited only about 22% BIFC intensities. Significance is shown relative to BIFC with wild-type HOXA7 and was evaluated by t test (**p < 0.001; ns, nonsignificant). Bars represent mean ± SD (n = 3).

12. All values could be gathered in the same “Results” window.

13. Export all data to Microsoft Excel for further analysis.

14. The measured fluorescence should be corrected by subtracting the respective background values from both the mCherry normalizer and BIFC signals. The final normalized BIFC value is calculated by the BIFC/mCherry normalizer fluorescence ratio, according to the following formula:

\[ NF_B = \frac{FM_B - BG_M_B}{FM_R - BG_M_R} \] (see Note 19).

where \( NF_B \) = normalized BIFC fluorescence intensities, \( FM_B \) = measured BIFC global mean gray value, \( BG_M_B \) = measured BIFC background mean gray value, \( FM_R \) = measured mCherry global mean gray value, and \( BG_M_R \) = measured mCherry background mean gray value.

15. The final normalized value should be averaged from at least three independent images for each condition.

16. To facilitate the comparison of all conditions, the final BIFC value will be scaled using positive control VN173-HOXA7/CC155-PBX1 interaction as the reference (Fig. 6).
3.4 Data Interpretation

In this study, interactions between different HOXA7 constructs and PBX1 were investigated in living HEK-293 cells. Despite the limitation of the transfection system (see Note 20), BiFC could be used to measure different interaction affinities, allowing identification of key HOXA7 residues involved in the interaction with PBX1. In our example, the BiFC signal was reduced by 80% with HOXA7 [dGA] (negative control) compared to the wild type HOXA7. We also demonstrated that HOXA7/PBX1/MEIS1 forms a trimeric complex using bicolor BiFC (see Note 3 and Fig. 7). For statistical significance, results should be obtained from at least three independent experiments, and the different interactions should be measured from the same pool of transfected cells. This kind of consideration allows limiting the fluctuating effect from the transfection efficiency. In addition, using vectors with inducible promoters could allow controlling more tightly the expression level of each construct, which was not the case here with the pcDNA3 vector (with the constitutive CMV promoter).

4 Notes

1. The most currently used FPs for BiFC are derived from the GFP, including YFP, Venus, CFP, and Cerulean. The GFP itself is prone to form noncovalent dimers [18], which could lead to incorrect localization and function of the BiFC complex in case of increased expression levels in the cell. FPs used in this protocol inherit an A206K mutation [19], preventing the dimerization, therefore these monomeric FPs are suitable for BiFC.

2. The choice of the fusion topology (fusion with the FP fragment at the N- or C-terminus of the targeted protein) requires prior knowledge of protein characteristics. In many cases, these characteristics are not known. Therefore, both N- and C-terminally tagged proteins need to be tested independently.

3. The C-terminal fragments of Venus (VC155) and Cerulean (CC155) display 96% of identity. This allows using one or the other for doing mVenus-like BiFC with the complementary VN173 fragment. In addition, the CC155 fragment can complement with the N-terminal fragment of Cerulean (CN173) for making mCerulean-like BiFC. This property allows visualizing two different PPIs simultaneously by doing mVenus- and mCerulean-like BiFC with three fusion proteins. The CN173/CC155/VN173 has proven to be the best combination for multicolor BiFC [15]. The multicolor BiFC enables visualization of distinct interactions in the same protein complex [2]. For example, the bicolor BiFC for HOXA7/PBX1/MEIS1 can be done by
Fig. 7 Bicolor BiFC in HEK-293 cells. (a) Schematic representation of VN-HOXA7/CC-PBX1/CN-MEIS1; (b) illustrative confocal pictures of BiFC between HOXA7/PBX1/MEIS1. Scale bar, 10 mm

co-expressing CC-PBX1, CN-MEIS1, and VN-HOXA7 (Fig. 7). The protocol is very similar to classical BiFC, which is detailed in the main text. In the three-plasmid transfection system (see step 8 in “3.1 Transfection of BiFC Plasmids”), replace pcDNA3-mCherry by adding 0.5μg pcDNA3-CN-MEIS1 for a total 1.5 μg DNA/reaction. Add 3 μL jetPRIME reagent for final transfection mix.

4. It is important to have a negative control of the BiFC, even when the expression of fusion proteins is tightly controlled in the cell (e.g., by using an inducible promoter). Several negative controls can be used. When possible, proteins with point mutations predicted to disrupt the interaction should be tested: these mutations should affect the BiFC signal when compared to BiFC with the corresponding wild-type proteins under the same parameters of expression. When such knowledge is not available, another good negative control will consist in doing competition against the BiFC complex. In this context, one of the two protein partners will be co-expressed with the two fusion
proteins as a cold competitor: this cold protein partner A will compete with the same fusion protein partner A for interacting with the fusion protein partner B, leading thus to a diminution of the BiFC signal.

Unfused FP fragments like CC or VN can be regarded as inappropriate controls for BiFC experiments [20], as the ability of the FP fragments to trigger artificial interaction should be considered in the context of their fusion with the candidate proteins. Despite this fact, they are still frequently seen in the literature [21], especially in BiFC-based screening studies [22–24].

5. To compensate for differences in transfection efficiencies across conditions and cell-to-cell variations, an internal control FP with distinct spectral properties is co-expressed with the BiFC fusions and used to normalize BiFC signals [25], thus considered as a signal normalizer. However, for a credible BiFC test, the internal FP should be expressed at least in 50–70% of the cells.

6. For further orthogonal methods to verify the PPI, like Co-IP, an epitope tag is necessary. This implies that one of the protein partners is only fused to a classical epitope tag for co-IP (like HA or FLAG tag), while the other partner remains unchanged (fused to the FP fragment). Revelation of the interactor upon Co-IP with the anti-HA or anti-FLAG antibody can be done with polyclonal anti-GFP antibodies that recognize both N- or C-terminal fragments of Venus and Cerulean.

In our lab, rabbit anti-GFP polyclonal antibody (Invitrogen, # A-11122) is frequently used for both CC155 and VN173 immunoassay.

7. To attenuate the effect of the fusion, a short flexible linker is preconized for the construct, which may provide sufficient freedom to the N- and C-terminal FP fragments for complementation. [GGSGG]_n linker is one of the most commonly flexible linker that is widely used in fusion protein construct. This linker should not be too long to avoid revealing indirect PPIs.

8. The cell population should be less than 80% confluent and under exponential growth phase. The exponential phase is optimal for efficiency of transfection experiments. Experiments should also be performed with low-passage cell culture (no more than 15 passages) to avoid any possible genetic drift changes in genotype that may affect the final result.

9. When using confocal microscopy, cells must be plated on glass coverslips because high numerical aperture objectives of confocal microscopy will not go through the plastic dishes or glass slides. It is recommended to use dishes with embedded
coverslips with an optical thickness of 0.17 mm. Coating coverslip could be better for cell attachment. This is however not necessary for HEK-293 cells, which are highly adherent cells. Coating can be done by treating the coverslip with polylysine or collagen prior to adding cells.

Classical protocols for sterilizing coverslips are based on ethanol in concentrations ranging from 70% to 100%, either with or without further sterilization steps [26]. Using ethanol alone is convenient for manipulation. In case of potential undesired side effects for the cell culture, other thorough methods like UV or autoclaving can be envisaged.

10. Images can be taken by confocal microscopy or by standard fluorescence microscopy equipped with appropriate excitation and emission filters or laser units to image CFP, YFP, and RFP, with 20× objectives and a CCD camera.

11. The imaging settings for mCherry and BiFC can be different; however, the settings used in each different condition must remain the same for consistent comparison among all the captured images.

12. The size of the pinhole can affect the resolution and light throughput. For a pinhole size of ~1 AU, 86% of the light collected from a point-like source passes through the pinhole. For most applications in confocal microscopy, a pinhole size between 0.8 and 1.0 AU is optimal. Alternatively, using a pinhole greater than 1.0 AU can increase image brightness but reduce image resolution.

13. The seeding cell number may differ from cell types. For example, HEK and MDA-MB-231 could be seeded with 300,000 cells/well, for appropriate confluence after 24 h incubation.

14. Use ultrapure PCR-grade water or jetPRIME buffer to dilute your plasmid solution. The concentration of the plasmid solution should not be too diluted, as it will notably change the transfection mix volume, leading to low transfection efficiency.

15. After transfection, a long time culture will lead to fusion protein overaccumulation. To minimize this effect, a short culture time post-transfection is recommended, usually around 20/24 h for having enough maturation time for BiFC signals.

16. Individual cells express different levels of fusion proteins upon transient transfections. It is therefore essential to measure BiFC signals in large cell populations to dilute cell-to-cell variation. Even if the internal transfection-control plasmid is used to reduce this bias, it is highly recommended to measure BiFC in a high number of fluorescent cells (minimum 100 cells). In addition, images should be captured from at least three different fields in each condition.
17. Maximum projection means that the algorithm chooses, for each pixel, the highest value found in any of the z sections. ZEN allows “stacking” all z-stacks into a single image. This process will much facilitate the subsequent quantitative analysis.

18. “Autoscale” is an optional choice for image import. It will make the fluorescent signal more clear and visible than original image but without altering underlying values in the image. This option is very useful for weak fluorescence cellular location display.

19. The total measured fluorescence for each channel is calculated as follows:

\[ F_B = \text{FM}_B \times \text{FS}_B \]
\[ F_R = \text{FM}_R \times \text{FS}_R \]

where \( F_B \) = total measured BiFC fluorescence, \( \text{FM}_B \) = measured BiFC global mean gray value, \( \text{FS}_B \) = total measured BiFC area surface, \( F_R \) = total measured mCherry fluorescence, \( \text{FM}_R \) = measured mCherry global mean gray value, and \( \text{FS}_R \) = total measured mCherry area surface.

The total measured fluorescence for background of each channel is calculated as follows:

\[ \text{BGF}_B = \text{BGM}_B \times \text{BGS}_B \]
\[ \text{BGF}_R = \text{BGM}_R \times \text{BGS}_R \]

where \( \text{BGF}_B \) = total measured BiFC background fluorescence, \( \text{BGM}_B \) = measured BiFC background mean gray value, \( \text{BGS}_B \) = total BiFC background area surface, \( \text{BGF}_R \) = total measured mCherry background fluorescence, \( \text{BGM}_R \) = measured mCherry background mean gray value, and \( \text{BGS}_R \) = total mCherry background area surface.

The measured fluorescence should be corrected by subtracting the respective background values from both the mCherry normalizer and BiFC signals, according to the following formula:

\[ \text{CF}_B = F_B - \text{BGF}_B \]
\[ \text{CF}_R = F_R - \text{BGF}_R \]

where \( \text{CF}_B \) = corrected total measured BiFC fluorescence and \( \text{CF}_R \) = corrected total measured mCherry fluorescence.

The final formula is generated as follows:

\[ \text{NF}_B = \text{CF}_B / \text{CF}_R \]
\[ = (F_B - \text{BGF}_B) / (F_R - \text{BGF}_R) \]
\[ = (\text{FM}_B \times \text{FS}_B - \text{BGM}_B \times \text{BGS}_B) / (\text{FM}_R \times \text{FS}_R - \text{BGM}_R \times \text{BGS}_R) \]
As the total measured fluorescence area is equal to the area surface of background:

\[ F_{SB} = BGS_B = F_{SR} = BGS_R = S \]

thus:

\[ NF_B = (F_{M_B} - BGM_B)/(F_{M_R} - BGM_R) \]

20. It is difficult to ensure that each fusion protein will be equally expressed in all cells when using three plasmids for transient transfection. More highly expressed proteins will interact more frequently and/or at higher levels, which can induce false-positive results. Statistical reproduction of the experiment can eliminate these effects that are by definition nonreproducible. Alternatively, stable cell line could be generated for a more constant protein expression [27]. The viral 2A self-cleaving peptide was also applied for combined protein translation in one mRNA [28]. However, the 2A self-cleavage is not always 100% efficient [29].

Acknowledgments

Research in the laboratory of S. Merabet is supported by Association pour la Recherche sur le Cancer (ARC, PIA20141202007); Fondation pour la Recherche Médicale (FRM, DEQ 20170336732); Ligue Nationale Contre le Cancer, Centre National de Recherche Scientifique (CNRS); CEFIPRA (5503-2); CNRS; and Ecole Normale Supérieure (ENS) de Lyon. We further thank the China Scholarship Council (CSC, File No. 201708070003) for the doctoral grant to Y. Jia.

References

cells using bimolecular fluorescence complementation (BiFC) analysis. Curr Protoc Cell Biol Chapter 21:Unit 21.3


30. 21
Part 2. Protein-protein interaction between HOX and TALE proteins

2.1 Human HOX Proteins Use Diverse and Context-Dependent Motifs to Interact with TALE Class Cofactors (article 1)

Amélie Dard,1,2 Jonathan Reboulet,1,2 Yunlong Jia,1,2 Françoise Bleicher,1,2 Marilyne Duffraisse,1 Jean-Marc Vanaker,1 Christelle Forcet,1 and Samir Merabet1,3,*

1 Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, Université Lyon 1, CNRS, Ecole Normale Supérieure de Lyon, 46, allée d’Italie 69364 Lyon cedex 07 France

2 These authors contributed equally

3 Lead Contact

* Correspondence: samir.merabet@ens-lyon.fr


Declaration of Contribution: I conducted BiFC experiments and concomitant analysis. I designed and engineered cDNA plasmids used in the study.
Human HOX Proteins Use Diverse and Context-Dependent Motifs to Interact with TALE Class Cofactors

Amelie Dard,1,2 Jonathan Reboulet,1,2 Yunlong Jia,1,2 Françoise Bleicher,1,2 Marilyne Duffraisse,1 Jean-Marc Vanaker,1 Christelle Forcat,1 and Samir Merabet1,3,*
1Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, Université Lyon 1, CNRS, Ecole Normale Supérieure de Lyon, 46, allée d’Italie 69364 Lyon Cedex 07, France
2These authors contributed equally
3Lead Contact
*Correspondence: samir.merabet@ens-lyon.fr
https://doi.org/10.1016/j.cellrep.2018.02.070

SUMMARY

HOX proteins achieve numerous functions by interacting with the TALE class PBX and MEIS cofactors. In contrast to this established partnership in development and disease, how HOX proteins could interact with PBX and MEIS remains unclear. Here, we present a systematic analysis of HOX/PBX/MEIS interaction properties, scanning all paralog groups with human and mouse HOX proteins in vitro and in live cells. We demonstrate that a previously characterized HOX protein motif known to be critical for HOX-PBX interactions becomes dispensable in the presence of MEIS in all except the two most anterior paralog groups. We further identify paralog-specific TALE-binding sites that are used in a highly context-dependent manner. One of these binding sites is involved in the proliferative activity of HOXA7 in breast cancer cells. Together these findings reveal an extraordinary level of interaction flexibility between HOX proteins and their major class of developmental cofactors.

INTRODUCTION

HOX genes code for homeodomain (HD)-containing transcription factors (TFs) that regulate numerous developmental processes during embryogenesis (Hueber and Lohmann, 2008). The diverse and specific transcriptional activities of HOX proteins often depend on the presence of two families of key cofactors that belong to the TALE (three amino acids loop extension) class of HD-containing TFs: PBC and MEINOX (Mukherjee and Bürglin, 2007). In vertebrates, the PBC family comprises Pbx1–4 proteins, while the MEINOX family includes Meis1–3 and Prep1 and 2 proteins (Longobardi et al., 2014). TALE cofactors are known to be important for several HOX developmental functions in vertebrates and invertebrates, regulating cell processes as diverse as apoptosis (Domsch et al., 2015), differentiation (Cerdà-Esteban and Spagnoli, 2014), or proliferation (Chen et al., 2016). Not surprisingly, these cofactors are also involved in the oncogenic potential of HOX proteins in various solid cancers (Bhatkar et al., 2014) and leukemia (Shinawa et al., 2008). Given its major impact on development and disease, the partnership among HOX, PBX, and MEIS proteins has been the subject of numerous studies (Mann et al., 2009). Interaction between PBX and MEIS depends on evolutionarily conserved domains located in the N-terminal region of both proteins (Figure 1A). This interaction uncovers a nuclear localization signal in PBX, allowing its translocation into the nucleus in vivo (Berthelsen et al., 1999; Rieckhof et al., 1997). Original studies showed that mouse HOX proteins, with the exception of posterior paralog groups 11–13, can associate with PBX on DNA (Chang et al., 1995; Knöpfel and Kamps, 1995). HOX-PBX interactions involve a short HOX protein motif located upstream of the HD and called hexapeptide (HX) or W-containing motif (Figure 1A). This motif is present in all HOX members from paralog groups 1–10, and it is characterized by an invariant Trp residue located within a hydrophobic environment and followed by basic residues at +2 to +5 positions (in de Ried et al., 2004). Crystal structures from Drosophila (Joshi et al., 2007; Pasnér et al., 1999) or mouse (LaRonde-LeBlanc and Wolberger, 2003; Piper et al., 1989) HOX/PBX complexes indicate a preponderant role of the Trp residue in establishing strong contacts within the hydrophobic pocket formed in part by the three extra residues of the HD of PBX. In contrast to these solid molecular analyses, the role of the HX as a key TALE interaction motif has only been demonstrated at the functional level in the case of the mouse HOX-A1 protein (Renacl et al., 2004).

HOX proteins also form trimeric complexes with both PBX and MEIS cofactors (Shanmugam et al., 1999). These trimeric complexes are surprisingly much less studied although they represent an important fraction of HOX regulatory functions in vivo (Amin et al., 2015; Penkov et al., 2013). Recent work showed that five of the 39 murine HOX proteins (HOXB6, HOXB7, HOXB8, HOXA9, and HOX10) could form trimeric complexes with PBX1 and MEIS1 in the absence of the HX motif (Hudry et al., 2012). The dispensability of the HX motif for complex formation with TALE cofactors was also observed in several Drosophila HOX proteins (Hudry et al., 2012). Moreover, the Drosophila Ultrabithorax (Ubx) and Abdominal-A (AbdA) proteins were shown to accomplish Pbx-dependent functions without the HX motif and by using another short motif in vivo (Foos et al., 2015; Merabet et al., 2007, 2011).
Altogether, these observations led to the proposition that molecular diversification in Hox-Pbx-Mei interaction properties could underlie paralog-specific functions of Hox proteins (Merabet and Mann, 2016). Although very appealing, this model awaits further experimental validation. In particular, the capacity to interact with TALE-class cofactors in the absence of the HX motif has not been demonstrated for Hox proteins of all paralog groups, and alternative TALE interaction motifs have not been described outside the Drosophila Ubx and AbdA proteins. Therefore, it remains to be established whether interactions between mammalian Hox proteins and the Pbx and Mei cofactors could rely on diverse and specific Hox protein signatures and whether this molecular complexity could constitute a general rule among all Hox paralog groups.

Here, we describe a systematic analysis of the interaction properties of 16 human and 3 mouse Hox proteins with the Pbx1 and Mei1 cofactors, taking into account all mammalian Hox paralog groups (Figures 1A and 1B). Our work relies on the molecular dissection of Hox-Pbx1-Mei1 interactions in different cell lines and on various types of DNA-binding sites in vitro (Figures 1C and 1D). Results show that Hox proteins from all except anterior paralog groups 1 and 2 are able to interact with Pbx1 and Mei1 in the absence of the HX motif. Analysis of complex formation with human Hox proteins from three different paralog groups further reveals the existence of specific TALE-binding sites that are evolutionarily conserved to different degrees. These TALE-binding sites are used with the HX motif in a DNA-binding site and cell-specific manner. One of these molecular signatures uncovers a motif that was functionally validated in the context of the proliferative activity of HOXA7 in breast cancer cells. This work definitively establishes the role of various and specific TALE-binding sites in human Hox proteins that helps us to understand Hox/TALE function during normal development and oncogenesis.

RESULTS

HOX Proteins from All but Paralog Groups 1 and 2 Interact with TALE Cofactors in the Absence of the HX Motif

The role of the HX motif was systematically analyzed by doing two complementary experimental approaches: biomolecular fluorescence complementation (BIFC) in live HEK293T cells, which
derive from human embryonic kidney, and electrophoretic

Figure 2. Interaction Properties among HOX, PBX1, and MEIS1 Proteins
(A) BIFC between wild-type or HX-mutated HOX proteins and PBX1 in live HEK293T cells, as indicated. An illustrative confocal picture is given for anterior, central, and posterior HOX representatives. Quantification of BIFC signals (green, upper pictures) with HX-mutated HOX proteins is provided as a percentage of the signal measured with the corresponding wild-type HOX protein. Quantification takes into account the efficiency of transfection, which is assessed by systematically co-transfecting a red fluorescent reporter. Scale bar, 10 μm.

(A') Quantification of BIFC with HX-mutated HOX proteins. BIFC is strongly affected upon the HX mutation in the case of HOXA1, HOXB1, and HOXB2. The mean and SD are shown from three independent experiments. Significance is shown relative to BIFC with the corresponding wild-type HOX protein and was evaluated by t test (**p < 0.001; ns, nonsignificant).

(B) Illustrative band shift experiments with wild-type or HX-mutated HOX proteins and PBX1 (P) and MEIS1 (M), as indicated. An illustrative gel is given for anterior, central, and posterior HOX representatives. Colored arrows indicate the monomer binding when present. Gray and black arrows depict dimeric and trimeric complexes with PBX1 or PBX1 and MEIS1, respectively. Black arrowhead indicates the PBX1/MEIS1 dimer. Gels have voluntarily been separated for each HOX protein to better illustrate the different complexes of different sizes. *F* indicates anti-FLAG antibody to supershift the FLAG-tagged HOX protein from the trimeric complex (star). The mean value of dimeric (black dotted boxes) or trimeric (yellow dotted boxes) complex formation with HX-mutated HOX proteins is given as a percentage of the complex obtained with the corresponding wild-type HOX protein.

(B') Quantification of trimeric complex formation with HX-mutated HOX proteins. The HX mutation does not affect complex formation in the presence of MEIS1 for all tested HOX proteins, except for HOXA1, HOXB1, and HOXB2. Bars represent mean ± SD of three independent experiments.

HOXB3 to assess for a potential redundant function between different TALE interaction interface(s).

In vitro analysis on the consensus ant/cent DNA-binding site showed that the dC deletion did not change interaction properties of wild-type or HX-mutated HOXB3 with PBX1 and MEIS1 (Figure S2A). By contrast, the d100 deletion significantly affected trimeric complex formation but only when the HX was also mutated (Figure S2A). To identify alternative TALE-binding sites in the first 100 residues of HOXB3, we generated additional deletions that progressively removed the different predicted SLIMs (HOXB3d61, HOXB3d91, and HOXB3d21; Figure S3A). Results showed that removing the first 61 or 31 residues of HOXB3 was sufficient to affect trimeric complex formation with PBX1 and MEIS1 in the context of the HX mutation (Figure S2A). Interestingly, the short d21 deletion did not affect complex formation of HX-mutated HOXB3 (Figure S2A). The region in between d21 and d31 deletions contains a predicted conserved SLIM with the core NGF Gould sequence (Figure S3A). We found that this motif was important for trimeric complex formation, but only when combined with the HX mutation and the d21 deletion (construct B3d21NG, Figure S2A), suggesting that the first 21 residues of HOXB3 have also a redundant TALE interaction activity with the NGF Gould and HX motifs. This region contains two predicted SLIMs: one at the extreme N
Figure 3. Deciphering Alternative TALE Interaction Properties in the Anterior Human HOXB3 Protein

(A) Global organization of HOXB3 with conserved predicted short linear motifs (SLMs, green peaks and green arrow), ordered domains (brown blocks), and long disorganized regions (blue waves). Red peaks indicate the level of conservation of each residue among vertebrate species. A schematic representation of HOXB3 is shown above the picture, with deletions generated to identify TALE interaction regions. SLMs are schematized by small boxes. Boxes surrounded by a black line depict SLMs that are involved in the HOXB3-TALE interaction. The sequence of the N-terminal part of HOXB3 is indicated with the two alternative TALE interaction SLMs (KATYYD and NGFGFD motifs, surrounded). Prediction was obtained by using SIMPred0.9 (http://bioware.ucd.ie/~ompasa/biowareweb/Server_pages/simpred_legacy.php).

(B–D) Quantifications of trimeric complex formation with mutated HOXB3 proteins on the ant/(cent), ant/(cent-MESinr) (C), or cent/post (D) nucleotide probe, as indicated. Bars represent mean ± SD of three independent experiments. Mutations affecting the HOXB3-TALE interaction are highlighted in red. E and F Quantification of BFC with mutated HOXB3 proteins in HEK (E) or HeLa (F) cells, as indicated. Mutations affecting the interaction are highlighted in red. Significance is shown relative to BFC with wild-type HOXB3 and was evaluated by t test (*** p < 0.001; ns, nonsignificant). Bars represent mean of three independent experiments.

(G) Models recapitulating the context-specific use of the HX motif and alternative TALE interaction motifs in HOXB3. In context 1, the KATYYD, NGFGFD, and HX motifs have a redundant activity for HOXB3-TALE interaction, as observed on the ant/(cent) nucleotide probe in vitro and in live HEK cells. In context 2, the HX motif has a preponderant and independent role while the KATYYD and NGFGFD motifs act redundantly and have a minor contribution for HOXB3-TALE interaction, as observed on the ant/(cent-MESinr) nucleotide probe in vitro. In context 3, the HX motif has a preponderant activity that is in part redundant with the KATYYD and NGFGFD motifs for HOXB3-TALE interaction, as observed on the cent/post nucleotide probe in vitro. In context 4, the KATYYD and NGFGFD motifs act redundantly and in addition to an unknown motif (indicated by the question mark) that is inhibited by the HX motif for HOXB3-TALE interaction, as observed in HeLa cells.
Figure 4. Deciphering Alternative TALE Interaction Properties in the Central Human HOXA7 Protein

(a) Global organization of HOXA7. Color code and symbols as in Figure 3. The core sequence of the HX and GAGA motifs of HOXA7 is indicated.
(b–d) Quantifications of trimeric complex formation with mutated HOXA7 proteins on the cent/post (b), cent/post-MEIS1 (c), or ant/post (d) nucleotide probe, as indicated. Bars represent mean ± SD of three independent experiments. Mutations affecting the HOXA7-TEL interaction are highlighted in red. Significance is shown relative to BIFC with wild-type HOXA7 and was estimated by t test (**p < 0.01, ns, nonsignificant). Bars represent mean of three independent experiments.
(e) Models recapitulating the context-specific use of the HX and GAGA motifs in HOXA7. In context 1, the GAGA motif has a preponderant role for HOX7-TEL interaction, while the HX motif has an inhibitory role, potentially against another TALE-binding site that remains to be characterized (indicated by the question mark). This context was observed on the cent/post nucleotide probe in vitro and in HEK cells. In context 2, the HOXA7-TEL interaction relies on a TALE-binding site(s) other than the HX or GAGA motif (indicated by the question mark). This context was observed on the cent/post-MEIS1 nucleotide probe in vitro. In context 3, the GAGA and HX motifs have a minor and redundant contribution for the HOXA7-TEL interaction, which suggests the role of other TALE-binding site(s) in HOXA7 (indicated by the question mark). This context was observed on the ant/post nucleotide probe in vitro. In context 4, the GAGA and HX motifs have an independent and important contribution for the HOXA7-TEL interaction. This context was observed in HeLa cells.

The combined loss of the HX and/or GAGA motif had no significant effect on the cent/post-MEIS1 probe (Figures 4C and S4B), while the two motifs have a redundant activity on the ant/post probe (Figures 4D and S4C). Together, these observations highlight that the HX and GAGA motifs are used in a DNA-binding site-specific manner for complex assembly with PBX1 and MEIS1 (summarized in contexts 2 and 3 of Figure 4G, respectively).
Interaction properties were then analyzed at a large scale level by doing BIFC in live HEK cells. These cells expressed very weak levels of endogenous HOX7, and representative HOX7 constructs were expressed at a similar range in the nucleus (Figures S4D–S4E). We observed that the d47 deletion or the micro-deletion of the GAGA motif had a strong effect on BIFC, with a global loss of 80% (Figures 4E, S4F, and S4F). The simultaneous mutation of the HX motif also led to a significant rescue of complex formation, with BIFC levels comparable to levels obtained upon the single HX mutation (with 69% to 53% of remaining signals; Figures 4E, S4F, and S4F). Thus, effects observed in HEK cells recapitulate effects observed in vitro on the consensus cent/post probe (context 1 in Figure 4G).

Finally, BIFC in HeLa cells showed that the HX and GAGA motifs have an independent contribution for HOX7-TALE interaction (Figures 4F and S4G, summarized in context 4 of Figure 4G) and therefore distinct molecular properties when compared to HEK cells.

Altogether, in vitro and BIFC analyses showed the existence of a paralog-specific TALE-binding site in HOX7. As previously noticed for HOX8, the usage mode of TALE-binding sites in HOX7 is highly sensitive to the context, with redundant, opposite, or independent contributions, depending on the type of the DNA-binding site and cell environment.

The Central HOXC8 Protein Contains Alternative TALE-Binding Sites That Are Used in a Context-Dependent Manner

The last HOX protein that we dissected was HOXC8, which has an intermediary intra-molecular organization when compared to HOXB3 and HOX7, with three conserved predicted SLIMs in addition to the HX and two small organized domains in addition to the HD (Figure 5A). Two predicted SLIMs are located in the N-terminal part of HOXC8. One of these motifs is conserved in other paralog groups and known to be important for the transcriptional activation potential of Hox proteins (the MSSYF motif; Tour et al., 2005). The third predicted SLIM is located immediately downstream of the HD and is followed by a repetition of Gin residues (Figure 5A). According to these predictions, we generated two different deletions removing either the N-terminal (deletion d106) or the C-terminal (deletion dG34) part of HOXC8 (Figure 5A). Deletions were also analyzed in the context of wild-type and HX-mutated HOXC8.

In vitro analyses on the consensus cent/post nucleotide probe showed that the N-terminal deletion had minor effects on trimeric complex assembly, even in the absence of the HX motif (Figure S5A). By comparison, the C-terminal deletion led on its own to a global loss of 30% of trimeric complex formation, and the effect was dramatically enhanced in the context of the HX mutation, with only 23% of remaining trimeric complexes on average (construct HOXC8d106dC34; Figure S5A). Thus, the HX motif and C-terminal region have a redundant role for complex formation with TALE cofactors on the cent/post probe.

The C-terminal region of HOXC8 contains a predicted conserved SLIM with a core KLPG sequence. To assess whether this motif could be involved in HOXC8-TALE interaction, we generated two additional constructs affecting either the KLPG motif (mutation into Aaa, construct C6A) or the Gin stretch (deleted construct C6E). Analysis on the cent/post probe showed that the loss of either the KLPG motif or Gin stretch led to a loss of around 60% of trimeric complex when combined with the HX mutation (constructs C83ddX and C83ddE; Figures S5B, S5A, and S5A'). This result shows that the KLPG motif and Gin stretch act redundantly with the HX motif and participate in two independent TALE interaction platforms for HOXC8-TALE complex formation on the consensus cent/post probe (corresponding to context 1 in Figure 5G).

HOXC8-TALE interaction properties were then dissected by doing BIFC in HEK cells. These cells expressed very weak levels of endogenous HOXC8, and representative HOXC8 constructs were expressed at a similar range in the nucleus (Figures S5D–S5E). N- and C-terminal deletions confirmed the preponderant role of the C-terminal part of HOXC8 for the interaction with TALE cofactors, with around 50% loss of BIFC in the context of the HX mutation (Figures 5E and S5G). Surprisingly, removing the Gin stretch had no effect with or without the HX mutation, while the simultaneous mutation of the HX and KLPG motifs led to a global loss of 44% of BIFC (Figures 5E, S5F, and S5G). This result indicates that the HD and KLPG motifs act redundantly for complex formation with TALE cofactors in HEK cells (context 2 in Figure 5G). The fact that BIFC was not completely lost upon the HD and KLPG mutations also suggests that additional TALE-binding site(s) could exist in HOXC8.

Given the context-specific mode of activity for TALE-binding sites in HOXB3 and HOX7, we also analyzed TALE interaction properties of HOXC8 on two additional nucleotide probes in vitro and in HeLa cells. We observed that the KLPG and HX motifs had a redundant activity, while the Gin-rich stretch was also independently required for trimeric complex formation on the cent/post-MET/Siv (Figures 5C and S5B, summarized in context 3 of Figure 5G). The relationships between TALE-binding sites of HOXC8 was different on the ant/cen probe, with a preponderant and unique role of the Gin-rich stretch (Figures 5D and S5C, summarized in context 4 of Figure 5G), and in HeLa cells, with a redundant activity of the HX motif and Gin repetition (Figures 5F and S5H, summarized in context 5 of Figure 5G). Together these results confirm that HOXC8 is able to use various combinations of different TALE-binding sites depending on the DNA-binding site and cell context.

Finally, given that the KLPG motif could be redundant with the HX motif of HOXC8 in several instances, we asked whether this alternative TALE interaction motif could also act in trans and rescue TALE interaction properties of HX-mutated HOX1A. To this end, we generated HOX1A-HOXC8 chimeric proteins, with the same rationale as previously described for HOXB3 and HOX7. Here, the N-terminal fragment of HOX1A, mutated in the HX motif and containing the HD, was fused to the wild-type or KLPG-mutated C-terminal part of HOXC8 (Figure S5E). Results showed that the C-terminal part of HOXC8 could rescue the HX mutation of HOX1A in vitro and in HEK cells (Figures S5F–S5H). Importantly, the KLPG mutation significantly affected the interaction (around 50% or 40% loss in vitro or in HEK cells, respectively; Figures S5G–S5H), highlighting that
A

the KLPG motif is important for the rescued activity of HX-mutated HOXA1.

The Paralog-Specific GAGA Motif of HOX7 Is Important to Promote PBX-Dependent Proliferative Activities in Human Breast Cancer Cells

Since HOX members of the paralog group 7 have been shown to be involved in the progression and resistance to treatment of breast cancer cells (Jin et al., 2012), we tested whether the growth-promoting activity of HOX7 on breast cancer-derived MCF7 cells could depend on the GAGA motif. MCF7 cells expressed endogenous PBX1 but a very low level of MEIS1 (Figure S1).

The role of the GAGA motif was first analyzed by doing BIFC in live MCF7 cells. Given the very low nuclear expression level of endogenous MEIS1, we performed bi-color BIFC by co-expressing CC-PBX1, CN-MEIS1, and VN-HOX7 (Figure 6A) or Venus-based BIFC by expressing CC-PBX1 and VN-HOX7 (Figure 6B). The same type of results was obtained in both cases, with no effect of the HX mutation while the GAGA mutation led to a strong decrease of fluorescence. Interestingly, the combined mutation of the HX and GAGA motifs led to a rescue of the HOX7 interaction potential (Figures 6A‘, 6A”, 6B‘, and 6B”). Of note these different effects were not due to variations in the expression level of HOX7 constructs (Figure 6C). Altogether these observations confirmed that the GAGA motif is important for HOX7-PBX1 interaction in MCF7 cells.

We next investigated the role of the GAGA motif on HOX7 proliferative activities. The experiment was realized without co-transfected PBX1 or MEIS1 to measure HOX7 proliferative activities in the presence of endogenous TALE cofactors. Transfection of wild-type HOX7 led to a significant increase of cell proliferation, as previously described (Figure 6D; Zhang et al., 2013). The HX-mutated form of HOX7 was even a more potent inducer of cell proliferation, suggesting that the HX motif could inhibit HOX7 proliferative activities in this particular context (Figure 6D‘). Importantly, transfection of the GAGA mutated HOX7 protein did not increase the number of MCF7 cells when compared to the empty vector, highlighting that this motif is important for HOX7 proliferative activities (Figure 6D). Interestingly, the combined HX and GAGA mutation led to a rescue of HOX7 proliferative activities, thus recapitulating previous BIFC observations with PBX1 (Figure 6D). Thus, the proliferative activity of HOX7 in MCF7 cells seems to be controlled by a subtle balance between different interfaces that are able to promote or inhibit the interaction with PBX1.

DISCUSSION

Our work revealed the presence of highly diverse alternative TALE-binding sites that occupy various positions within the HOX protein. The usage mode of the different TALE-binding sites is also involving intra-molecular inhibitory loops, highlighting that the activity of the different TALE-binding sites is tightly controlled. This regulation in cis will differ depending on the cell environment and topology of DNA-binding sites. For example, the inhibitory role of the HX motif in HOX7 was observed in HEK and MCF7 cells, but not in HeLa cells.

We also noticed that alternative TALE-binding sites have a different degree of evolutionary conservation within the paralog group and among vertebrate lineages. This evolutionary plasticity does not only apply between different HOX proteins but also within the same HOX protein (Figure 7). For example, the KATYYD motif is more ancient and more widely conserved among vertebrate lineages than the NGTGYF/ 

Short protein motifs, or SLMs, are by definition highly dynamic during evolution and are described to establish context-specific and low-affinity interactions (Davey et al., 2012). It is interesting to note that Drosophila Hox/TALE complexes were described to recognize low-affinity and divergent DNA-binding sites in vivo (Crockier et al., 2015). Alternative TALE interaction SLMs could, therefore, be of great interest in this context, allowing HOX/TALE complexes to adapt their conformation on different types of DNA-binding sites. Such a molecular mode of action was noticed for the three dissected HOX proteins, which displayed a distinct requirement of their alternative TALE interaction motifs depending on the DNA-binding sites and cell context. Altogether, these observations underline that the choice of the TALE-binding site in the HOX protein is highly sensitive to the DNA and protein content environment.

In all tested HOX proteins, interaction properties with PBX1 are strongly remodelled in the presence of MEIS. Future exciting challenges will be to decipher how the MEIS partner could induce such a strong molecular remodelling and how it could eventually impact on the HOX/TALE function in vivo. Interestingly, we observed that PREP1 could not induce HX-independent interaction modes between PBX1 and HOXB3, HOX7, or HOX8 in vitro (Figure S9). Thus, the role of MEIS for diversifying HOX-PBX interaction modes appears very specific among the TALE family.

Cumulative data have established that HOX and TALE proteins could cooperate in several solid cancers (Bhatkar et al., 2014) and leukemia (Sitwala et al., 2008). Disrupting this partnership could, therefore, constitute a promising strategy for therapeutic approaches. Accordingly, an HX-mimicking peptide has been described to induce apoptosis in several cancer cell lines, including breast cancer cells (Morgan et al., 2012). Interestingly, it was also noticed that MCF7 cells were relatively insensitive to the HX-mimicking peptide when compared to other breast cancer-derived cell lines (Morgan et al., 2012). Our work establishes that the majority of HOX proteins use additional TALE interaction interfaces that could act in a redundant, independent, or antagonistic way with the HX motif. In addition, the HX motif was recently shown to be important for promoting or inhibiting interactions with other types of cofactors (Baizza et al., 2015), suggesting that the effect of the HX-mimicking peptide could not be restricted to HOX-PBX interactions.

The identification of specific TALE interaction motifs could open new promising avenues for future therapeutic approaches against HOX/TALE-induced cancers. First, targeting these specific motifs is a way to inhibit the collaborative oncogetic activity of a specific HOX/TALE complex in cancer. We have identified
Figure 7. Sequence Conservation of TALE-Binding Sites in Vertebrate HOX Proteins

(A) Alignment of the N-terminal sequence of HOXB5 from different vertebrate species. The NGFDF(V)/D/E motif is specifically conserved in HOXB3 paralog members among vertebrate lineages, while the KATYDY motif is more widely conserved in all vertebrate HOX3 members.

(B) Schematic representation of the NGFDF(E)/D and KATYDY motifs in the mammalian HOX3 paralog group.

(C) Alignment of the GAGA-containing region of HOX3 sequences from different vertebrate species. The GAGA motif is not conserved outside mammalian lineages.

(D) Sequence conservation suggests that the GAGA motif was acquired among the paralog group 7 in mammalian lineages during vertebrate evolution.

(E) Schematic representation of the GAGA motif in the mammalian HOX7 paralog group.

(F) Alignment of the C-terminal sequence of HOX8 from different vertebrate species.

(G) Sequence conservation suggests that the HOX8-specific KLPG motif and Gin-rich region were acquired in the tetrapoda ancestor during vertebrate evolution.

(H) Schematic representation of the KLPG motif and Gin stretch peptide in the mammalian HOX8 paralog group.

Alignments were performed with the CLUSTALW program (http://www.genome.jp/tools/clustalw/). Sequences are from Homo sapiens (Hs), Canis lupus (C), Mus musculus (Mm), Galus (Gg), Xenopus laevis (Xl), and Takifugu rubripes (Tr).

total amount of 2 μg DNA: 500 ng VM-HOX fusion vector, 600 ng VC-PBX1 or CC-PBX1 fusion vector, and 1 μg PCMV-N-Myc or CN-ME1 construct. Coverslips were taken 20 hr after transfection, which allowed having the fluorescence level below saturation with each tested wild-type HOX protein. Analysis was performed with a Zeiss LSM780 confocal microscope. Pictures correspond to the z projection of stacks, using the Zen software.

Statistical Methods

Quantification of fluorescence was performed from four to six different fields containing 150 cells on average, and it was repeated from three independent experiments in each condition. Data were analyzed using Microsoft Excel and are represented as mean ± SD. Statistical significance was determined by Student’s t test. Values of p < 0.05 were considered significant.
SUPPLEMENTAL INFORMATION

Supplemental information includes Supplemental Experimental Procedures, six figures, and two tables and can be found with this article online at https://doi.org/10.1016/j.icep.2018.02.070.

ACKNOWLEDGMENTS

We thank Robert K. Slany, Jacqueline Deschamps, and René Rezzahzay for reagents and Sylvie Di Ruscio for technical assistance. We thank René Rezzahzay and Yosine Grahe for helpful comments on the manuscript. This work was supported by the Association pour la Recherche sur le Cancer (ARC, PJ2014-41020097). Fondation pour la Recherche Médicale (FRM, DREG20170326732), Ligue Nationale contre la Cancer, Centre National de Recherche Scientifique (CNRS), CEPFRA (55532-2), and Ecole Normale Superieure (ENS) de Lyon.

AUTHORS CONTRIBUTIONS


DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 27, 2017
Revised: January 12, 2018
Accepted: February 19, 2018
Published: March 13, 2018

REFERENCES


Figure S1, related to Figure 2: (A) RT-qPCR of endogenous PBX1 and MEIS1 in HEK-293T, HeLa and MCF7 cells. Values are given in Ct (cycles threshold) and normalized to endogenous RPLPO (36B4) mRNA. A reference values for a transcription factor encoding gene known to be expressed and active in HEK cells, LSD1, is given (Carnescechi et al., 2017). (B) Fluorescent immunostaining against PBX1 (in red) in HEK293T, HeLa and MCF7 cells, as indicated. Dapi (blue) stains the nucleus. Note that PBX1 is weakly expressed in the nucleus and present in the cytoplasm of MCF7 cells. (C) Fluorescent immunostaining against MEIS1 (in red) in HEK293T, HeLa and MCF7 cells, as indicated. Dapi (blue) stains the nucleus. Note the absence of nuclear staining in MCF7 cells. (D-F) Venus-based BiFC between VN-HOXA1 and the VC fragment. (D), VC-PBX1 and the VN fragment (E), or VN-HOXA1 and VC-PBX1 mutated in the residue 54 of the HD (F), a mutation known to affect the interaction with HOX proteins (Hudry et al., 2011). Bars represent mean of three independent experiments. Significance is shown relative to BiFC obtained between VN-HOXA1 and VC-PBX1 and was evaluated by t-test (**P<0.001). (G-G’) Example of BiFC normalisation with HOX1 and PBX1. Quantifications are shown on the enlargement of a confocal capture for better clarity. (G) The mean value of the mean value of transfection (here 28) is obtained from quantifications in all green fluorescent cells (white-dotted cells). The mean value of transfection (here 1025: right panel) serves as a reference (100%) to calculate the effect of protein mutations on BiFC.
(G') An example is provided with the HX mutation in HOXA1, which leads to a mean BiFC/Transfection efficiency value of 0.58. This value corresponds to 56.6% of the fluorescence normally obtained with wild type HOXA1 (graph). Such analysis was systematically performed for the entire field of three to six different acquisitions (each field contains around 150 cells on average) and was repeated for three independent experiments in each case. (H) Expression of endogenous MEIS1 analysed by RT-qPCR after transfection with two different siRNAs (siM1 and siM2) targeted against MEIS1-3 isoforms relative to a control siRNA (sic). Values are mean ± SD of two independent experiments performed in triplicate. Note that double transfection with the two different siRNAs leads to a better silencing. (I) siM1 and siM2 trigger cytoplasmic localisation of co-transfected VC-PBX1 (in red) construct in HEK cells. Co-transfection with control siRNA does not affect the nuclear localisation of VC-PBX1. (J-J’) Control siRNA neither affects the nuclear localisation of VC-PBX1, nor changes BiFC properties between VC-PBX1 and wild type or HX-mutated VN-HOXB3 in HEK cells. The same result was obtained with VN-HOXC8 (not shown). (K-K’) Effects of siRNA against MEIS on BiFC between VC-PBX1 and wild type or HX-mutated VN-HOXB3 in HEK cells. (L-L’) Effects of siRNA against MEIS on BiFC between VC-PBX1 and wild type or HX-mutated VN-HOXC8 in HEK cells. Graphs illustrate quantification of BiFC (first column) and nuclear PBX1 immunostaining (second column) with HX-mutated HOX proteins. Significance is shown relative to the corresponding wild type HOX protein and was evaluated by t-test (**P<0.001; ns, nonsignificant). Mean values of BiFC or PBX1 immunostaining are indicated on illustrative confocal captures. Note that co-transfection with the HOX protein attenuates the effect of siMEIS on the nuclear localisation of VC-PBX1 (with more VC-PBX1 in the nucleus; compare with Figure S4B). Scale bar is 10 μm.
Figure S2, related to Figure 3: (A) Illustrative band shift experiments between HOXB3 constructs and PBX1 and MEIS1 on the anticent nucleotide probe. (A') Quantification of trimeric complexes with the different HOXB3 constructs on the anticent nucleotide probe. Bars represent mean +/- SD of three independent experiments. (B-C) Illustrative band shift experiments between wild type or mutated HOXB3 constructs and PBX1 and MEIS1 on the anticent-MEISinv (B) or cent/post (C) nucleotide probe, as indicated. Red-dotted boxes highlight significantly affected trimeric complexes. (D) RT-qPCR for endogenous HOXB3 in HEK-293T and HeLa cells. Values are given in Ct as in Figure S1. (E) Immunostaining against the wild type or mutated form of HOXB3 in HEK cells, as indicated. (E') Quantification of the VNHOXB3d13^GFP^ construct in the nucleus of HEK cells. (F) Illustrative confocal captures of BiFC between the different HOXB3 constructs and PBX1 in HEK cells, as indicated. (F') Statistical quantification of BiFC. Bars represent mean of three independent experiments. Significance is shown relative to the corresponding wild type HOX protein and was evaluated by t-test (**p<0.001; ns, nonsignificant). Constructs with affected TALE interaction properties are highlighted in red. (G) Illustrative confocal captures of BiFC between the different HOXB3 constructs and PBX1 in HeLa cells, as indicated. The mean percentage of BiFC is indicated in the merge picture for each construct. Scale bar is 10 μm.
Figure S3, related to Figures 3 and 5: (A) Scheme of the two chimeric HOXB3-HOXA1 proteins. Mutations are indicated in red and HOXB3 or HOXA1 part is illustrated by their respective colour code. (B) Illustrative band shift experiment between chimeric HOXB3-HOXA1 proteins and PBX1 and MEIS1 on the antisense nucleotide probe, as indicated. Colour code is as in Figure 2. (C) Illustrative confocal captures of BiFC between chimeric HOXB3-HOXA1 proteins and PBX1 in HEK cells, as indicated. The mean percentage of BiFC is indicated in the merge picture. (D) Statistical quantification of BiFC. Bars represent mean of two independent experiments. Significance was evaluated by t-test (***P<0.001). (D’) Immunostaining of chimeric HOXB3-HOXA1 proteins in HEK cells, as indicated. (D’) Quantification of the nuclear immunostaining of the inactive chimeric protein, as indicated. (E) Scheme of the two chimeric HOXA1-HOXC8 proteins. Mutations are indicated in red and HOXA1 or HOXC8 part is illustrated by a different colour. (F) Illustrative band shift experiment between chimeric HOXA1-HOXC8 proteins and PBX1 and MEIS1 on the antisense nucleotide probe, as indicated. Colour code is as in Figure 2. (G) Illustrative confocal captures of BiFC between chimeric HOXA1-HOXC8 proteins and PBX1 in HEK cells, as indicated. The mean percentage of BiFC is indicated in the merge picture. (G’) Statistical quantification of BiFC. Bars represent mean of three independent experiments. Significance was evaluated by t-test (***P<0.001). (H) Immunostaining of chimeric HOXA1-HOXC8 proteins in HEK cells, as indicated. (H’) Quantification of the nuclear immunostaining of the KLPG-mutated chimeric protein, as indicated. Scale bar is 10 mm.
Figure S5, related to Figure 5: (A) Illustrative band shifts experiments with various HOXC8 constructs on the cent/post probe, as indicated. Star indicates super-shifted trimeric complexes with an anti-GFP (box “γ”) recognising VN-HOXC8 constructs. (A') Quantification of super-shifted trimeric complexes (yellow- and red-dotted boxes) on the cent/post nucleotide probe. (B-C) Illustrative band shifts experiments with wild type and mutated HOXC8 constructs on the cent/post-MEISnov (B) or ant/cen (C) probe, as indicated. All EMSAs were voluntary exposed for a longer time to better detect and quantify the band corresponding to the super-shift. Colour code and mean values are as in Figure S2. (D) RT-qPCR for endogenous HOXC8 in HEK and Hela cells. Values are given in Ct as in Figure S1. (E) Immunostaining against wild type or mutated HOXC8 in HEK cells, as indicated. (E') Quantification of mutated HOXC8 immunostaining. (F) Illustrative confocal pictures of BiFC between relevant HOXC8 constructs and PBX1 in HEK cells, as indicated. Values correspond to the mean percentage of fluorescent signals when compared to wild type HOXC8. (F') Statistical quantification of BiFC. Significance is shown as in Figure S2. Mutations affecting BiFC are highlighted in red. (G) Illustrative confocal pictures of BiFC between HOXC8 constructs and PBX1 in HeLa cells, as indicated. Scale bar is 10 μm.
**Figure S6. Related to Figures 3-5.**
PREP1 does not induce HX-independent interaction modes between HOX and PBX1 proteins.

(A) Illustrative band shift experiment between wild or HX-mutated HOXB3, HOXA7 or HOXC8 proteins, as indicated. Colour code is as in Figure 2, except for the black box that indicates the presence of PREP1 in the binding reaction (Pr). Trimeric complexes are highlighted by yellow-dotted boxes on the gel. (B) Immunostaining against PREP1 (red) in HEK, HeLa and MCF7 cells, as indicated. No significant staining is found in the nucleus of the three cell lines.
Supplemental Experimental Procedures

BiFC analysis in live cells

Human HEK293T cells were maintained in Dulbecco’s modified Eagle’s medium (1 g/L glucose, Invitrogen) supplemented with 10% foetal bovine serum (Invitrogen), 20 IU/ml penicillin, and 20 mg/ml streptomycin (Invitrogen) at 37°C in a humidified 5% CO2 atmosphere. 24 h before transfection, 5.10^6 cells were plated on glass coverslips. Transfections were carried out using the JetPRIME reagent (Polyplus), with a total amount of 2 µg of DNA: 500 ng of the VN-HOX fusion vector, 500 ng of the VC-PBX1 or CC-PBX1 fusion vector, and 1 µg of the pCMV-mCherry or CN-MEIS1 construct. Coverslips were taken 20h after transfection, which allows having fluorescence level below saturation with each tested wild type HOX protein. Analysis was performed with a Zeiss LSM780 confocal microscope. Pictures correspond to the Z projection of stacks, using the Zen software. Four to six different fields of cells were acquired under the same confocal parameters at the 20x objective from three independent experiments in each condition. Each field represents around 150 cells on average. Quantification of green (BiFC), red (transfection efficiency) or blue (CC-PBX1/CN-MEIS1) fluorescence in nuclei of each acquired field was realized by using the histogram function of ImageJ software. All green or red fluorescent cells were selected for quantification and a mean ratio of green/red or green/blue signals was established for each acquisition. Values obtained with wild type proteins were used as the reference for each transfection experiment. See also Figure S2.

MCF7 proliferation assays

MCF7 (human mammary carcinoma) cells were cultured in DMEM supplemented with 10% FCS, 10U/ml penicillin and 10µg/ml streptomycin. For proliferation analysis, 10^4 cells were seeded in 96-well plates. After 12hrs, cells were transfected with 100ng DNA using the JetPrime transfection reagent (Polyplus) under the conditions recommended by the manufacturer. Cell index was analysed 48 or 72hrs after transfection using the CellTiterGlo kit (Promega) in conditions recommended by the manufacturer. Results are expressed relative to control-transfected cells (i.e. with pSG5-GFP) and represent the average of three independent experiments performed in triplicate. Statistical analysis was performed using Student t-test.

RNA Extraction and Real-Time PCR.
For expression and siRNA analyses, $3 \times 10^5$ cells were seeded in 6 well plates. After 12hrs, cells were transfected with 2µg DNA using JetPrime reagent. 48hrs after transfection, RNAs were extracted using the TriPure solution (Roche). 1µg of total RNA was converted to first strand cDNA using iScript cDNA synthesis kit (Biorad). Real-time PCR were performed in a 96-well plate using the IQ SYBR Green Supermix (Biorad). Data were quantified by the ΔΔ-Ct method and normalized to RPLP0 (36b4) mRNA expression for plasmid expression (Carnesecchi et al., 2017) or to endogenous MEIS1 for siRNA (Table S2). Two different siRNAs against MEIS1 and a control siRNA (medium GC Stealth RNA interference negative control duplexes, Invitrogen) were used.

Data for each HOX and PBX1 construct are given relative to the endogenous expression level of the corresponding HOX or PBX1 gene observed after transfection with empty pcDNA3 vector. Transfection efficiency between the different conditions was normalized by co-transfecting a luciferase-expressing vector in each case. Sequences of the primers are shown in Table S2. Bars represent the average relative expression of the indicated gene with error bars showing S.D. (n=3). Statistical analysis was performed using Student t-test.

**Western Blot Analysis.**

Cells were lysed in Nonidet P-40 or RIPA buffer supplemented with Protease Inhibitor Mixture (Sigma-Aldrich). Proteins (25–50 µg) were resolved on 8–15% SDS/PAGE, blotted onto PVDF membrane (GE-Healthcare), and probed with specific antibodies after saturation. The antibodies used were: rabbit anti-GFP (Invitrogen A11122, 1/2000), recognizing VN to reveal VN-HOXA7 constructs and mouse anti-Hsp90 (Enzo, ADI-SPA-830, 1/1000) for the loading control.

**Immunostaining**

Rabbit anti-GFP (Invitrogen A11122, 1/200), mouse MEIS1/2/3 (Santa Cruz sc-101850, 1/200) and mouse PREP1 (Abcam ab55603, 1/200) antibodies were used to quantify nuclear protein expression in cell culture. Fluorescent revelation was realised by using anti-rabbit (ThermoFisher A32732, 1/500) or anti-mouse (ThermoFisher A32727, 1/500) secondary antibodies coupled to Alexa fluor 555.
2.2 The human HOXA9 protein uses paralog-specific residues of the homeodomain to interact with TALE-class cofactors (article 2)

Amélie Dard¹⁺, Yunlong Jia¹⁺, Jonathan Reboulet¹⁺, Françoise Bleicher¹⁺, Catherine Lavau² and Samir Merabet¹⁺

¹ Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, Université Lyon 1, CNRS, Ecole Normale Supérieure de Lyon, 46, allée d’Italie 69364 Lyon cedex 07 France

² Pediatric Hematology Oncology, Duke University Medical Center, Research Drive, LSRC building, Room B-365, Box 91001, Durham, NC 27710, USA.

* co-first authors

 Correspondence and requests for materials should be addressed to S.M. (email: samir.merabet@ens-lyon.fr)

Published in Scientific Reports, 05 April 2019, doi: 10.1038/s41598-019-42096-y

Declaration of Contribution: I performed and analyzed BiFC measurements in human HEK, HeLa and MCF7 cells. I designed and engineered cDNA plasmids used in the study.
The human HOXA9 protein uses paralog-specific residues of the homeodomain to interact with TALE-class cofactors

Amélie Dard1, Yunlong Jia2, Jonathan Reboulet3, Françoise Bleicher3, Catherine Lavau4 & Samir Merabet1

HOX proteins interact with PBX and MEIS cofactors, which belong to the TALE-class of homeodomain (HD)-containing transcription factors. Although the formation of HOX-PBX complexes depends on a unique conserved HOX motif called hexapeptide (HX), the additional presence of MEIS induces a remodeling of the interaction, leading to a global dispensability of the HX motif for trimeric complex formation in the large majority of HOX proteins. In addition, it was shown that the anterior HOXB3 and central HOXA7 and HOXB8 proteins could use different alternative TALE interaction motifs, with or without the HX motif, depending on the DNA-binding site and cell context. Here we dissected the molecular interaction properties of the human posterior HOXA9 protein with its TALE cofactors, PBX1 and MEIS1. Analysis was performed on different DNA-binding sites in vitro and by using Bimolecular Fluorescence Complementation (BiFC) in different cell lines. Notably, we observed that the HOXA9-TALE interaction relies consistently on the redundant activity of the HX motif and two paralog-specific residues of the HOXA9 HD. Together with previous work, our results show that HOX proteins interact with their generic TALE cofactors through various modalities, ranging from unique and context-independent to versatile and context-dependent TALE binding interfaces.

HOX genes encode for homeodomain (HD)-containing transcription factors (TFs) that are involved in the control of numerous processes during embryonic development. This evolutionary conserved family of developmental regulators is classified into anterior, central and posterior paralog groups (PGs), reflecting their organization and function along the anterior-posterior (AP) axis of the embryo. Mutations affecting this organization and resulting in the inappropriate expression of a HOX gene product along the AP axis lead to the famous homeotic transformations. These phenotypes emphasize that each HOX protein has specific functions during embryogenesis.

HOX proteins are known to regulate distinct sets of target genes in vivo, which is at odds with their highly conserved HD and similar DNA-binding properties in vitro. This paradox has in part been solved by the identification of the PBC-class cofactors, which belong to the TALE family of HD-containing TFs. PBC representatives include Extradenticle (Exd) in Drosophila or PBX1-4 in vertebrates. All PBC proteins form dimeric complexes with HOX proteins from PGs 1-10. Importantly, dimeric HOX/PBC complexes display distinct DNA-binding properties, with increased specificity and affinity when compared to the HOX monomer binding. Crystal structures of several vertebrate and invertebrate HOX/PBC complexes have also shown the preponderant role of a canonical HOX peptide motif called Hexapeptide (HX) or W-containing motif (because of the presence of an invariant conserved W residue in all HX motifs) in establishing strong contacts with particular residues of the PBC HD. More recently, the interaction between HOX and PBC was described as revealing a “latent specificity”, allowing paralog-specific residues of the HOX protein to recognize a typical shape of the DNA minor groove.
Crystal structures have so far been obtained with incomplete HOX and PBC proteins, and important information could therefore be lacking. In addition, no structure has been solved in the presence of the third partner MEIS, which forms trimeric complexes with HOX and PBC proteins\textsuperscript{13}. MEIS proteins also belong to the TALE family and are required for the nuclear translocation of PBC, a role that is evolutionary conserved in the animal kingdom\textsuperscript{14,15}. MEIS has long been considered as a “simple” nuclear carrier of PBC but several characterized HOX target enhancers contain MEIS binding sites, highlighting that MEIS could also directly collaborate with HOX and PBC proteins in vivo\textsuperscript{16}. Accordingly, MEIS forms cooperative dimeric DNA-binding complexes with PBC and with posterior mammalian HOX proteins\textsuperscript{20,21}. MEIS was also shown to interact more generally with several HOX proteins in the absence of DNA, but the functional relevance of these interactions remains to be determined\textsuperscript{21}.

What about HOX/PBC/MEIS complexes? Several HX-mutated HOX proteins have been described to perform PBC-dependent functions in vivo, which led to reconsider the HOX/PBC interaction model\textsuperscript{22-25}. In particular, it was found that MEIS is important for revealing HX-independent interaction between PBX1 and mouse or human HOX proteins\textsuperscript{15,26}. These results suggested that HOX proteins could contain alternative interfaces to form trimeric complexes with PBC and MEIS cofactors. It was proposed that the use of different TALE interaction motifs could be at the heart of the functional diversity and specificity of different HOX/TALE complexes in vivo\textsuperscript{17,27,28}. This hypothesis was first supported by the identification of an alternative and specific TALE interaction motif in the Drosophila Hox protein Ultrabithorax (Ubx)\textsuperscript{24,25}.

Recent work revealed alternative and specific TALE interaction motifs in the human anterior HOX3B, and central HOX7 and HOX8 proteins\textsuperscript{26}. Interestingly, these motifs are used in different combinations, depending on the DNA-binding site topology and cell context, demonstrating that HOX-TALE interactions are not rigid.

Here we dissected the molecular interaction properties of the human posterior HOXA9 protein with the PBX1 and MEIS1 cofactors (Fig. 1A). Compared to anterior and central HOX proteins, HOXA9 has a more divergent HD which binds DNA with higher affinity and recognizes a distinct preferential consensus nucleotide sequence\textsuperscript{41}. HOXA9 also has a divergent HX motif, with a single conserved W residue, while the core Y/FPWM sequence is found in anterior and central HOX members\textsuperscript{30,31}. Finally, HX-mutated HOXA9 was shown to interact with PBX1 in the presence of MEIS1\textsuperscript{18}, suggesting that HOXA9 could also contain alternative TALE interaction interfaces.

We found that the formation of HOXA9-PBX1-MEIS1 trimeric complexes relied on the redundant activity of the divergent HX motif and two paralog-specific residues of the HOXA9 HD, which was obtained in the context of different DNA-binding sites and cell types. Thus, HOXA9-TALE interactions are relatively insensitive to the DNA-binding site topology and protein environment. Together with previous work, these results show that HOX proteins interact with their generic TALE cofactors through various modalities, ranging from unique and context-independent to versatile and context-dependent TALE binding interfaces.

**Results**

The homeodomain (HD) of HOXA9 is necessary and sufficient for complex formation with PBX1 and MEIS1. Previous work showed that HX-mutated HOXA9 can form a complex with PBX1 and MEIS1 on a consensus DNA-binding site called CENT/POST (Fig. S1 and Supplementary Fig. S1). HX-independent interaction with PBX1 was also observed by Bimolecular Fluorescence complementation (BiFC) with PBX1 in HEK cells, which express endogenous MEIS1\textsuperscript{15}. Given the global structure of HOXA9 (Fig. 1B), we generated a long N-terminal deletion to remove most of the predicted short peptide motifs (also called SLIMs for Short linear interaction motifs\textsuperscript{22}), organized domains and disorders regions (Fig. 1C). The resulting 86 residues long fragment was tested either intact (construct A9IN187) or with the HX mutation (construct A9IN187\textsuperscript{H}).

Electromobility shift assays (EMSA) on the consensus CENT/POST nucleotide probe confirmed that the formation of dimeric HOXA9/PBX1 complexes but not that of trimeric HOXA9/PBX1/MEIS1 complexes was dependent on the HX motif (Fig. 2A-A). Thus, HOXA9 can use alternative TALE interaction interface(s) in the presence of MEIS1. The A9IN187 protein fragment could also form a dimeric or trimeric complex with PBX1 or PBX1 and MEIS1, respectively (Fig. 2B-B). Importantly, the dimeric but not the trimeric complex was strongly affected by the HX mutation, highlighting that other residues lying within the A9IN187 fragment could replace the HX motif in the presence of MEIS1 (Fig. 2B-B).

Given that the A9IN187 fragment did not contain any obvious molecular signature with the exception of the HX, we generated another construct only containing the HD (construct A9HD in Fig. 1C). Results showed that the HOXA9 HD could strongly interact with PBX1 or PBX1 and MEIS1 (Supplementary Fig. S2).

The critical role of the HOXA9 HD for complex formation with TALE cofactors was further confirmed by generating a chimeric protein consisting of full length HOXA9 containing the HOXA1 HD, with or without the HX mutation (constructs A9HDA1 and A9HDA1\textsuperscript{H}, in Fig. 1C). In this context, monomer binding, dimeric and trimeric complexes could be observed on the CENT/POST nucleotide probe, but the HX mutation led to a loss of both dimeric and trimeric complexes (Fig. 2C-C). This result shows that the HOXA1 HD is not able to rescue the loss of the HX motif, as observed in HOXA9\textsuperscript{18}.

We also tested the inverse chimeric protein consisting of full-length wild type or HX-mutated HOXA1, with its HD swapped with that of HOXA9 (constructs A1HDA9 and A1\textsuperscript{H}HDA9 in Fig. 1C). Results showed that the HX-mutated chimeric protein was still able to form a trimeric complex on the CENT/POST probe (Fig. 2D-D’), demonstrating that the HOXA9 HD was sufficient to rescue the effect of the HX mutation in HOXA1\textsuperscript{22}.

Interaction properties of the aforementioned constructs with PBX1 were also analyzed in live HEK cells by conducting BiFC, as previously described (Fig. 2 and Methods). We observed that the HX mutation had no effect in the context of full length or truncated HOXA9 (Fig. 2E-F’). In addition, BiFC with wild type and HX-mutated chimeric HOXA1-HOXA9 proteins confirmed that the HOXA9 HD was necessary and sufficient (in the context of HOXA1) for the interaction with PBX1 in HEK cells (Figs 2G-H and S1). BiFC also revealed considerable
Figure 1. Identification of alternative TALE interaction interface(s) in the human HOXA9 protein. (A) Schematic representation of HOXA9, PBX1 and MEIS1, with protein domains and motifs involved in the formation of dimeric or trimeric complexes. The hexapeptide (HX) is indicated. The three PYP residues of the PBX1 homeodomain (HD) participate in the formation of an hydrophobic pocket that interacts with the conserved Trp (W) residue of the HX motif, as determined by the HOXA9/PBX1 crystal structure[13]. Domains of interaction between PBX1 (PBxA and PBxB) and MEIS1 (MEISA and MEISB) proteins are also indicated. The right panel illustrates the formation of dimeric (upper) and trimeric (lower) complexes between HOXA9 and PBX1 or HOXA9, PBX1 and MEIS1, respectively. The characterized interaction between the conserved Trp residue of the HOXA9 HX and the hydrophobic pocket formed in part by the PYP residues of PBX1 is indicated. Question mark highlights the role of additional uncharacterized binding interface(s) in HOXA9 that could be involved in the interaction with PBX1 and MEIS1 in the trimeric complex. (B) Schematic diagram of HOXA9 that represents predicted short linear interaction motifs (SLIMs, green bars), organized domains (brown blocks) and disordered regions (blue waves). The level of conservation of each residue among vertebrate HOXA9 sequences is also indicated (red bars). A diagram of HOXA9 summarizes the prediction of SLIMs (bars) and organized domains (white boxes). The deletion of the first 187 residues is indicated (dn187). This structure prediction was obtained with SlIMPred[12]. (C) Schematic representation of the HOXA9 constructs analyzed with the TALE cofactors in this study. Fusions with Venus fragments are voluntarily not indicated (see also Table 1). Mutations are indicated and highlighted with a black bar. HOXA9 and HOXA1 protein fragments are in red or blue, respectively.
<table>
<thead>
<tr>
<th>Constructs</th>
<th>Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>VNHOXA9</td>
<td></td>
</tr>
<tr>
<td>VNHOXA9.HX</td>
<td>W→A</td>
</tr>
<tr>
<td>VNHOXA9.DN187</td>
<td></td>
</tr>
<tr>
<td>VNHOXA8.DN187.HX</td>
<td>W→A</td>
</tr>
<tr>
<td>VNHOXA9.HD.A1</td>
<td></td>
</tr>
<tr>
<td>VNHOXA9.HD.A1.HX</td>
<td></td>
</tr>
<tr>
<td>VNHOXA9.HD.A1.HX</td>
<td></td>
</tr>
<tr>
<td>VNHOXA9.HD.A1.HX</td>
<td></td>
</tr>
<tr>
<td>VNHOXA1.HD.A9</td>
<td></td>
</tr>
<tr>
<td>VNHOXA1.HD.A9.HX</td>
<td></td>
</tr>
<tr>
<td>VNHOXA1.HD.A9.D29M56</td>
<td>W→A HD: LTDRRR→LTRAARR RMKMKKIN→RMWKWKKIN</td>
</tr>
<tr>
<td>VNHOXA9.K4A</td>
<td></td>
</tr>
<tr>
<td>VNHOXA9.HX.K4A</td>
<td>W→A HD: TRKRCPTT→TRKARCPYT</td>
</tr>
<tr>
<td>VNHOXA9.G6Q</td>
<td></td>
</tr>
<tr>
<td>VNHOXA9.HX.G6Q</td>
<td>W→A HD: TRKRCPTT→TRKARCPYT</td>
</tr>
<tr>
<td>VNHOXA9.P7G</td>
<td></td>
</tr>
<tr>
<td>VNHOXA9.HX.P7G</td>
<td>W→A HD: TRKRCPTT→TRKARCPYT</td>
</tr>
<tr>
<td>VNHOXA9.M24R</td>
<td></td>
</tr>
<tr>
<td>VNHOXA9.D29A</td>
<td></td>
</tr>
<tr>
<td>VNHOXA9.HX.D29A</td>
<td>W→A HD: FNMYLDRDRR→FNRYLDRRR</td>
</tr>
<tr>
<td>VNHOXA9.M56W</td>
<td></td>
</tr>
<tr>
<td>VNHOXA9.HX.M56W</td>
<td>W→A HD: RMKMKKIN→RMWKWKKIN</td>
</tr>
<tr>
<td>VNHOXA9.D29AM56W</td>
<td></td>
</tr>
<tr>
<td>VNHOXA9.HX.D29AM56W</td>
<td>W→A HD: LTRDRR→LTRAARR RMKMKKIN→RMWKWKKIN</td>
</tr>
</tbody>
</table>

**Table 1.** List of constructs used in this study. Residues resulting from the mutation are highlighted in italic. HD denotes residues of the homeodomain.

Reduction in interaction between HOXA9 and a mutated form of PBX1 that cannot bind DNA (Supplementary Fig. S2), showing that the formation of HOXA9/TALE complexes is DNA-binding dependent. In contrast, BiFC was not affected when HOXA9 was tested with a PBX1 mutant in which the amino acids known to interact with the HX W residue, based on previous crystal structures, were altered (P1P residues of the HD3, Supplementary Fig. S2). This suggests that other residues of PBX1 could be involved in the interaction with HOXA9. Finally, we also performed BiFC between wild type or HX-mutated HOXA9 and PBX1 in the presence of a siRNA targeting endogenous MEIS1 in HEK cells (as described in 93). In this context, the HX mutation led to a significant reduction (50%) of the fluorescent signals, confirming the important role of MEIS1 in promoting alternative molecular interaction properties between HOXA9 and PBX1 (Supplementary Fig. S2).

Altogether, EMSAs and BiFC demonstrate that the HOXA9 HD constitutes an alternative TALE interaction interface for trimeric complex formation with PBX1 and MEIS1 cofactors.

**Paralog-specific residues of the HOXA9 HD act redundantly with the HX motif for complex formation with PBX1 and MEIS1.** In order to identify residues of the HOXA9 HD that could be involved in the interaction with the TALE cofactors, we performed a sequence alignment of the mouse HOXA9 HD and the HD of additional human HOX proteins from PG9 and other PGs (Fig. 3A). This analysis revealed two residues that are specifically conserved in PG9, and four residues conserved in HOX proteins of PG9 and PG10. These residues are all in positions compatible with protein-protein interactions (blue boxes above the sequences in the Fig. 3A). The two residues that are specifically conserved in PG9 proteins are located in the helix 2 and in the recognition helix (residues D29 and M56, respectively: Fig. 3A). Three out of the four residues conserved between HOXA9 and HOXAI0 are located in the N-terminal arm (residues K4, C6 and P7; Fig. 3A), which is a region of the HD that is known to be important for HOX functional specificity94. Based on these attributes, we decided to mutate each of these conserved HOXA9 HD residues into a residue normally found in an anterior or central HOX protein (see Methods and Table 1). Mutants were analyzed in the context of wild type or HX-mutated HOXA9 (Fig. IC) and the resulting effect on the TALE interaction potential of HOXA9 was assessed by EMSAs and BiFC, as described above.
Figure 2. The HOXA9 homeodomain (HD) is necessary and sufficient for trimeric complex formation with PBX1 and MEIS1. (A–D) Band shift experiments of wild type or HX-mutated forms of HOXA9 with PBX1 or PBX1 and MEIS1 on the CENT/POST nucleotide probe, as indicated. Dimeric and trimeric complexes are highlighted (black- and white-dotted boxes, respectively). Values at the bottom indicate the quantification of dimeric and trimeric complexes as a percentage of complexes formed with the corresponding wild type HOX construct on the illustrative gel. Red/blue, black and grey arrows indicate the binding of the HOX monomer, HOX/PBX1 and HOX/PBX1/MEIS1 complexes, respectively. Black arrowhead indicates binding of PBX1 (P)/MEIS1 (M) complexes. (A’–D’) Quantification of trimeric complexes with the HX mutated forms of HOXA9 construct from three independent experiments (see Materials and Methods). Numbers above each bar correspond to the average value. (E–H) Illustrative confocal captures of BiFC (green) of wild type or HX-mutated forms of HOXA9 with PBX1 in HEK cells, as indicated. The quantification of HX-mutated forms is shown as a percentage of BiFC obtained with the corresponding wild type forms. The red fluorescent reporter
is used to normalize results to transfection efficiency (see Materials and Methods). Note that BiFC occurs in the presence of endogenous MEIS1\(^{19}\). (E-H) Quantification of BiFC between the different HX mutated forms of HOXA9 and PBX1 from three independent experiments. Significance is shown relative to BiFC with the corresponding wild type form and was evaluated using t test (\(* * * p < 0.001; \text{ns}, \text{nonsignificant}\)). See also Supplementary Figs S1 and S2.

EMSAs on the CENT/POST consensus probe showed that none of the mutations in the N-terminal arm, alone or combined with the HX mutation, affected trimeric complex formation (Fig. 3B-F). Along the same line, the M24 mutation, with or without the HX mutation, had negligible effects on complex formation with TALE cofactors, with or without the HX mutation (Fig. 3B-B). By contrast, the D29 and M56 mutations led to a significant loss of trimeric complex formation (respectively 60% and 70% loss: Fig. 3B-B), but only when combined with the HX mutation. This observation shows that the D29 and M56 residue act redundantly with the HX motif. Combining the HX, D29 and M56 mutations simultaneously did not further decrease trimeric complex formation (Fig. 3B-B), suggesting that the D29 and M56 residues form two independent TALE-binding interfaces with the HX motif. The role of the D29 and M56 residues was also confirmed in the context of the HX-mutated HOXA1 chimeric protein, since their mutation strongly affected the rescue activity of the HOXA9 HD (Supplementary Fig. S3).

BiFC in live HEK cells confirmed that the three residues of the N-terminal arm and the M24 residue are not required for HOXA9-TEL interaction (Fig. 3C-C). Also, surprisingly, the M56 mutation did not affect BiFC either (with or without the HX mutation: Fig. 3C-C), while the D29 mutation was not neutral when combined with the HX mutation (with a global loss of 50%: Fig. 3C-C). Combining the HX, D29 and M56 mutations led to an additive effect with a global loss of 80% of BiFC when compared to wild type HOXA9 (Fig. 3C-C). Of note, these mutated constructs were all expressed at similar levels in nuclei of HEK cells (Supplementary Fig. S1).

Together EMSAs and BiFC show that the D29 and M56 residues of the HOXA9 HD constitute two important alternative TALE-binding interfaces that act redundantly with the HX motif.

**Role of alternative TALE-binding residues of the HOXA9 HD in different contexts.** Alternative TALE interaction interfaces identified in anterior and central HOX proteins were described as being used in a highly context-specific manner with the HX motif\(^{26}\). We thus asked whether this could also apply to HOXA9. To this end, we analyzed two additional nucleotide probes for EMSAs and the HeLa and MCF7 cells for BiFC assays. The two nucleotide probes, called CENT/POST-MEISinv and ANT/CENT, diverge from the CENT/POST probe by containing an inversed MEIS binding site or a consensus Hox/Pbx binding site for anterior and central Hox proteins, respectively (Supplementary Fig. S1 and\(^{26}\)). HeLa and MCF7 cells are derived from cervical and breast cancers, respectively, as opposed to kidney derived HEK cells used above for the BiFC assays.

EMSAs showed that the D29 and M56 mutations only affected trimeric complex formation on the two probes when they were coupled with the HX mutation (Figs 4A,B and S3). These results demonstrate that the formation of HOXA9/TALE complexes on the CENT/POST-MEISinv and ANT/CENT binding sites relies on the redundant activity between the HX motif and the D29 and M56 residues.

BiFC in HeLa and MCF7 cells also revealed that the TALE interaction potential of HOXA9 was only significantly affected upon the simultaneous mutation of the HX motif and the D29 and M56 residues (Fig. 4C-CD-D). Thus, the three TALE binding sites also behave redundantly in HeLa and MCF7 cells. Together with the previous observations, these results demonstrate that the HX motif and D29 and M56 residues are required for HOXA9-TEL interaction across different DNA binding sequences and cell contexts.

**Discussion**

The identification of PBC and MEIS members as generic Hox cofactors in several developmental and oncogenic contexts has in part solved the Hox paradox to explain how a family of TFs displaying poor DNA-binding specificity *in vitro* could regulate distinct sets of target genes *in vivo*. PBC and MEIS members belong to the TALE-class of HD-containing TFs and interact with the large majority of Hox proteins. It was long considered that this interaction relied on the unique and canonical HX motif, which was somewhat at odds with the selective activity of each different HOX/TALE complex, raising the question of how Hox proteins could have different functions by interacting with the same set of TALE cofactors.

Several studies have shown that Hox proteins interact with TALE cofactors without the HX motif\(^{27,28}\), a property that is dependent on the presence of MEIS in most of the cases studied (some *Drosophila* Hox proteins being the exception\(^{28}\)). This observation indicates that Hox-PBC interactions are likely to involve considerable remodeling in the presence of MEIS. In addition, recent work identified alternative TALE interaction motifs in three different human HOX proteins, HOXB3, HOXA7 and HOX8\(^{29}\). Interestingly, these motifs are evolutionarily conserved to different extents and used independently, redundantly or even in opposition to the HX motif, depending on the DNA-binding site topology and the cell context\(^{26}\). Thus, Hox proteins can use versatile combinations of different TALE-binding motifs, showing a remarkable level of interaction flexibility with the same set of cofactors.

Here we dissected TALE interaction properties of the human posterior HOXA9 protein, which is characterized by the presence of a divergent HX motif when compared to Hox proteins of anterior and central PBCs. Our results show that removal of the first 187 HOXA9 residues, which contain most of the predicted SLIMs, did not affect the interaction between HOXA9 and TALE cofactors. The HOXA9 HD was found to be sufficient to interact with Pbx1 and MEIS1, not only in the context of HOXA9, but also in the context of HX-mutated HOXA1. Conversely, HX-mutated HOXA9 could not interact with the TALE cofactors when its HD was swapped with
Figure 3. Paralog-specific residues of the HOXA9 HD are important for the interaction with TALE cofactors. (A) Sequence alignment of the HD of HOXA9 and other human (*Homo sapiens*, *Hs*) or mouse (*Mus musculus*, *Mm*) HOX proteins. The global structure and orientation of the alphabetic chain of each residue is indicated above the sequences. Blue and white boxes symbolize residues that are accessible or not for protein-protein interactions, respectively (based on 26). Residues that are conserved in paralog groups 9 and 10 or only in paralog group 9 are highlighted in light orange or red, respectively. (B) Band shift experiments between HOXA9 constructs and PBX1, or PBX1 and MEIS1 on the CENT/POST nucleotide probe, as indicated. Color code and quantifications of HOXA9/TALE protein complexes are as in Fig. 2. (B') Quantification of trimeric complexes with the different mutated forms of HOXA9 from three independent experiments. (C) Illustrative confocal pictures of BiFC between different HOXA9 constructs and PBX1 in HEK cells, as indicated. Color code is as in Fig. 2. (C') Quantification of BiFC between the different mutated forms of HOXA9 and PBX1 from three independent experiments. Significance is shown relative to BiFC with wild type HOXA9 and was evaluated using t test (**p < 0.001; *p < 0.01; ns, nonsignificant).
the HX-binding hydrophobic pocket of PBX1 was not necessary in the presence of MEIS1. Although we could not exclude the existence of indirect protein-protein interactions, this result again highlights the important molecular remodeling occurring between dimeric and trimeric complexes. In addition, HOXA9 can interact with MEIS1 on DNA in vitro. The both HOXA9/MEIS1 and HOXA9/PBX1/MEIS1 interactions are restricted in the context of HOXA9/MEIS1 trimers. Given the spatial proximity between the three protagonists, the only way to get insights into the nature of HOXA9-TALE molecular contacts is by obtaining a crystal structure of the trimeric complex. This structure should be solved with at least full-length PBX1 and MEIS1 proteins to reveal all potential alternative HOXA9-TALE contacts, which remains a technically challenging issue.

The D29 and M56 residues were not identified as being important for interaction in previous crystal structures of HOXA9/PBX1. Of note, the K88 residue of the HOX9-HD was described as establishing a hydrogen bond with the S34 residue of PBX1 in the HOX9/PBX1 crystal structure. Interestingly, this contact is also observed in the AbdB/Exd complex, but only when the structure is solved on a particular (the highest affinity) DNA-binding site. In addition, the absence of effect observed with the K4, C6 or P7 mutations is in agreement with the less important role of the N-terminal arm of the HOXA9 or AbdB HD to recognize specific minor groove width minima when compared to other Hox proteins such as Scm or Dfd.

The D29 and M56 residues display different levels of evolutionary conservation among PG9 members: D29 is specifically conserved in vertebrates and tunicates, while M56 is conserved in all animal lineages that have been looked at (Figs 5A-C and S4). This observation suggests that the redundant role of the D29 and M56 residues as alternative TALE-binding interfaces could constitute a recent acquisition in PG9 Hox proteins during animal evolution. As a corollary, one cannot exclude the possibility that orthologous HOX3 proteins could use other alternative TALE binding interfaces in other animal lineages. Of note, whether the D29 and M56 residues are also important in human HOX9 and HOX9 remains to be demonstrated. Along the same line, it will be interesting to know whether HOX PG10 members could also use the same strategy to interact with the TALE cofactors. HOX10 members contain conserved residues (E29 and L56) that have similar chemical properties to D29 and M56 residues and that could therefore potentially play a similar role as the one observed in HOX9.

The 3D modeling indicates that the D29 and M56 residues are part of two separate interaction interfaces, suggesting that their respective alpha helix chains could contact different portions in PBX and/or MEIS1 (Supplementary Fig. S5). Moreover, EMSA on different DNA-binding sites and BifC in different cell contexts showed that these two residues were consistently used in redundancy with the HX motif. Thus, HOXA9-TALE interactions appear relatively insensitive to the DNA- and protein environment, which contrasts with the interaction properties of the anterior HOXB3 and central HOX7 and HOXC8 proteins. This observation highlights that different human HOX proteins use distinct molecular strategies to interact with TALE cofactors, from a unique HX-dependent interaction mechanism (PBX1-2), to consistently redundant interactions (HOX3-8) or context-specific (HOXB3, HOX7 and HOXC8) TALE interaction modes (Fig. S5D). This range of molecular strategies emphasizes that HOX proteins have developed distinct levels of interaction plasticity for complex formation with their generic TALE cofactors. Understanding how different levels of molecular plasticity could potentially be linked to functional diversity and specificity requirements of each HOX/TALE complex in vivo is the next challenging issue to tackle in order to crack the HOX paradoxy in the future.

Materials and Methods
Protein constructs. HOXA9 and PBX variants were generated by PCR from full-length complementary DNAs and restriction enzyme-cloned alone or in fusion with the N-terminal (VN) of Venus (for HOXA9 constructs), or the C-terminal (CC) fragment of Cerulean (for PBX1 constructs) in the pcDNA3 vector, respectively. See Table 1 for a complete list of all constructs. Complementation between VN and CC produces a Venus-like fluorescent signal, as previously described. Primers used for constructs are available upon request. A short linker of two amino acids separates the Venus or Cerulean fragment from HOXA9 or PBX1 in all fusion constructs. The linker region corresponds to the Flag-encoding sequence in the case of the HOXA9 HD construct. The sequence of all constructs was verified before use.

Protein Expression and Electrophoretic Mobility Shift Assays. Constructs cloned in the pcDNA3 vector were produced with the TNT T7-coupled in vitro transcription/translation system (Promega). Production yields of wild type and mutant counterpart proteins were estimated by 35S-methionine labelling. EMSAs were performed as described previously. Shortly, between 3 and 6 microliters of programmed lysate was used for each protein (100 ng) of proteins were produced on average). PBX1 and MEIS1 were co-produced together (0.5ng of each plasmid was used for the in vitro transcription/translation reaction). Supershift against the flag-tagged HD was performed by adding the anti-Flag antibody after 15 minutes in the binding reaction. Each band shift experiment was repeated at least three times and the quantification of protein complexes was done by using the histogram function of ImageJ software, using the complex with wild type proteins on the same gel as the reference value. Significance for each average quantification value could not be calculated because of the too small number (three to four) of samples that are considered.

BifC analysis in live cells. BifC in human HEK, HeLa and MCF7 cells was performed as previously described. Briefly, transfections were carried out using the JetPRIME reagent (Polyplus), with a total amount of 2 µg of DNA: 500 ng of the VN-HOX fusion vector, 500 ng of the VN-PBX1, and 1 µg of the pCMV-mCherry. Coverslips were taken 20 hours after transfection, which allows having fluorescence level below saturation with each tested HOX construct. Analysis was performed with a Zeiss LSM780 confocal microscope. Pictures correspond to the z projection of stacks, using the Zen software. For 14 different fields of cells were acquired under the same confocal parameters at the 20x objective from three independent experiments in each condition. Quantification of green (BifC) and red (for transfection efficiency) fluorescence in all nuclei of each acquired field was realized
Figure 5. Conservation of TALE-binding residues in HOXA9 and different modes of TALE interaction among different HOX paralog groups. (A) The D29 and M56 residues are conserved in all human HOX members of the paralog group (PG) 9. (B, C) Conservation of the D29 and M56 residues in HOXA9 proteins from different species. The D29 residue is conserved in vertebrates and tunicates, while the M56 residue is conserved in all Bilaterian (B) lineages. Protein sequences were obtained from UniProt (http://www.uniprot.org).

Representative Deuterostome (D) species are: Homo sapiens (Hs), Heterodius francisci (Hf), Okapia johnstoni (Oj), Branchiostoma lanceolatum (Bl), Strongylocentrotus purpuratus (Sp) and Saccoglossus kowalevskii (Sk).

Representative Protostome (P) species are: Strigamia maritima (Sm), Drosophila melanogaster (Dm) and Lineus stagnatilis (Ls). See also Supplementary Figs S4 and S5. (D) HOX proteins from different PCGs use distinct molecular strategies to interact with TALE cofactors. A representative example is given in each case (HOXA1, HOXA9 and HOXA7). Models result from this study and from previous work. The light-blue gradient illustrates the number of possible interaction modalities with TALE cofactors. HOXA9 uses a unique HX-dependent interaction mode. HOXA9 uses the HX motif and two paralog-specific residues of the HD. HOXA7 uses different alternative TALE-binding motifs in various combinations with the HX motif, depending on the DNA and/or protein environment. The importance of each TALE-binding interface within the HOX protein is symbolized by the number of signs “+” and the width of each corresponding arrow.

by using the histogram function of ImageJ software. A mean ratio of green/red signals was established for each acquisition. Values obtained with wild type proteins were used as the reference for quantification.

Immunostaining. Rabbit anti-GFP (Invitrogen A11122, 1/500) antibody was used to quantify nuclear expression of the different VN-HOXA9 fusion constructs in cell culture. Fluorescent labelling was done with an anti-rabbit (ThermoFisher A32732, 1/500) secondary antibody coupled to Alexa Fluor 555.

References

Acknowledgements

Research in the laboratory of S. Merabet is supported by Association pour la Recherche sur le Cancer (ARC, PIA20141202007), Fondation pour la Recherche Médicale (FRM, DEQ, 20170336732), Ligue Nationale Contre le Cancer, Centre National de Recherche Scientifique (CNRS), CEFIPRA (5503-2), CNRS et Ecole Normale Supérieure (ENS) de Lyon. We thank Laurent Terradot for his help with the PyMol Viewer software and Brian Ford for proof-Reading of the last version of the manuscript.

Author Contributions

A.D., J.J. and J.R. performed the cloning and BiFC analyses in cell lines. F.B. performed EMSAs. C.L. performed cloning. C.L. and S.M. wrote the paper. S.M. designed and supervised the project.
Additional Information
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-42096-y.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019
SUPPLEMENTARY INFORMATION

Supplementary Figure S1. A. Sequence of the nucleotide probes used for the band shift experiments. The orientation of HOXA9, PBX1 and MEIS1 binding sites is indicated. The two central nucleotides of the HOX/PBX binding sites involved in DNA-binding preferences are in bold. B-C. Immuno-staining of the HOXA9-HOXA1 chimera with an anti-GFP (green) recognizing the VN fragment at the N-terminus. D. Immunostaining of mutated HOXA9 constructs, as indicated. Plasmids were co-transfected with a mCherry encoding vector to assess for transfection efficiency in each condition (red). Graphs on the right show quantification of the immunostaining with the different mutated forms. Quantification is relative to the immunostaining with the corresponding wild type protein and takes into account transfection efficiency (ns: nonsignificant).

Supplementary Figure 2. A. Band shift experiment with the homeodomain (HD) of HOXA9 and the PBX1 and MEIS1 cofactors. Symbols are as in Figure 2. Asterisks denote supershifts with an antibody (γ) recognizing the Flag epitope fused to the HD of HOXA9. The two bands correspond to shifted HD/PBX1 or HD/PBX1/MEIS1 complexes. B. Illustrative confocal captures of BiFC between the HD of HOXA9 and PBX1 in HEK cells. The quantification is normalized to BiFC obtained with full length HOXA9 (ns:nonsignificant). C. Illustrative confocal captures of BiFC between HOXA9 and PBX1 mutated in the residue 54 of the HD in HEK cells. The quantification is normalized to BiFC obtained with wild type PBX1 (**p<0,01). D. Illustrative confocal captures of BiFC between HOXA9 and PBX1 mutated in the residue PYP of the HD in HEK cells. The quantification is normalized to BiFC obtained with wild type PBX1. E-F. Illustrative confocal captures of BiFC of wild type or HX-mutated HOXA9 with PBX1 in condition of controlled RNAi (E) or RNAi against endogenous MEIS in HEK cells (see Materials and Methods and 26). The quantification is normalized to BiFC obtained with HOXA9 (**p<0,001, ns: nonsignificant).
Supplementary Figure 3. A. Band shifts experiments of chimerix HX-mutated HOXA1 proteins with PBX1 and MEIS1 on the CENT/POST nucleotide probe, as indicated. Note that the D29 and M56 mutations strongly affect the rescue efficiency of the HOXA9 HD in the HX-mutated HOXA1 chimeric protein. A’. Quantification of the trimeric complex with the HD-mutated form on the CENT/POST nucleotide probe. B. Band shifts experiments of wild type and mutated HOXA9 constructs with PBX1 or PBX1 and MEIS1 on the CENT/POST-MEISinv nucleotide probe, as indicated. C. Band shifts of wild type and mutated HOXA9 constructs with PBX1 or PBX1 and MEIS1 on the ANT/CENT nucleotide probe, as indicated. Color code and symbols are as in Figure 2.

Supplementary Figure 4. Alignment of HOXA9 HD sequences from different vertebrate and invertebrate species. Protein sequences were obtained from UniProt (http://www.uniprot.org). Representative Deuterostome (D) species are: Homo sapiens (Hs), Heterodontus francisci (Hf), Oikopleura dioica (Od), Branchiostoma lanceolatum (Bl), Strongylocentrotus purpuratus (Sp) and Saccoglossus kowalevskii (Sk). Representative Protostome (P) species are: Strigamia maritima (Sm), Drosophila melanogaster (Dm) and Lineus sanguineus (Ls). Sequence alignment was obtained with CLUSTALW (https://www.genome.jp/tools-bin/clustalw).

Supplementary Figure 5. Position and orientation of paralog-specific residues of the HOXA9 HD involved in the interaction with TALE cofactors. A-A’. Side chain positioning of the D29 and M56 (PG9-specific), and K4, C6 and P7 (PG9/10-specific) residues (all highlighted in orange) in two differently rotated HOXA9 HDs. The Trp (W) residue of the HX motif lying upstream of the HD is indicated in one orientation (right panel). B-C. Side chain positioning and chemical properties of wild type (B) or mutated (C) D29 and M56 residues (surrounded by a dotted circle). Non-charged/hydrophobic residues are in white, positively and negatively charged residues are in red or blue, respectively. The position, orientation and chemical properties of the D29 and M56 residues of the HD suggest that they could be involved in different interaction interfaces with TALE cofactors.
Supplementary Figure S1
Supplementary Figure S2
Part 3. Development and Optimization of Large-scale Bimolecular Fluorescence Complementation (BiFC) Screening

3.1 A cell protein complementation assay for ORFeome-wide probing of human HOX interactomes (article 3, manuscript)

Yunlong Jia¹, Jonathan Reboulet¹, Benjamin Gillet¹, Sandrine Hughes¹, Françoise Bleicher¹ and Samir Merabet¹,*

¹ Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, Université Lyon 1, CNRS, Ecole Normale Supérieure de Lyon, 46, allée d’Italie 69364 Lyon cedex 07 France

* Correspondence: samir.merabet@ens-lyon.fr
Yunlong Jia, Jonathan Reboulet, Benjamin Gillet, Sandrine Hughes, Françoise Bleicher and Samir Merabet.

A live cell protein complementation assay for ORFeome-wide probing of human HOX interactomes

Abstract

Biological pathways rely on the formation of intricate protein interaction networks called interactomes. Getting a comprehensive map of these interactomes implies developing tools that allow capturing transient and low affinity protein-protein interactions (PPIs) in live cells. Here we present an experimental strategy called Cell-PCA (Cell Protein Complementation Assay), which is based on BiFC (Bimolecular Fluorescence Complementation) and high throughput sequencing for ORFeome-wide analyses of different interactomes in the same live cell context. Cell-PCA was established by using HOX proteins as a proof-of-concept, revealing interactomes with unexpected molecular properties. Taken together, our work demonstrates that Cell-PCA is applicable for capturing and comparing interactomes between different bait proteins.
Introduction
Organismal development and fitness depend for a large part on their cell protein content. Proteins are versatile molecules that work in a crowded environment, establishing a number of interactions with other surrounding proteins. These protein-protein interactions (PPIs) form intricate networks, called interactomes, which are changing from cell-to-cell and stage-to-stage.

Characterizing these dynamic molecular networks is a key issue for understanding protein function and implies developing highly sensitive tools. One major experimental approach to capture PPIs is the yeast two-hybrid system (Y2H), which relies on the indirect readout of a reporter gene to reveal any interaction between a bait protein and candidate partners (Paiano et al., 2019). Although very popular, Y2H presents the major inconvenient of being performed in a heterogeneous context (i.e., in the yeast for mammalian proteins and in the context of proteins fused to heterologous DNA-binding and activation domains).

Recent approaches based on biotin-ligase enzymes and Liquid Chromatography-Mass Spectrometry (LC-MS) identification of biotinylated partners constitute promising alternatives for the characterization of interactomes in specific cell- and tissue types (Varnaite & MacNeill, 2016; Carnecchi et al., 2020). These approaches are appropriate for the identification of endogenous interactomes but are not compatible for more systematic and high throughput interrogations of binary PPIs with dedicated candidate libraries.

In addition, the use of fluorescent (Miller et al., 2015) and non-fluorescent (Remy et al., 2007) reporters for protein-fragment complementation assays (PCAs) represent the simplest and most sensitive tools for screening PPIs in live cell conditions. These approaches rely on the property of the reporters to be reconstituted from separate N- and C-terminal fragments upon spatial proximity (Fig. 1A). However, all fluorescent PCA-based screens described to date relied on Open Reading Frames’ libraries (ORFeomes) that were transiently used in the cell, either by co-transfection (Remy & Michnick, 2004; Berendzen et al., 2012) (Fig. 1B) or transduction of lentiviral particles (Ding et al., 2006; Lee et al., 2011; Simon E. Coopera, 2015) (Fig. 1C-C’). These strategies do not allow comparing interactomes between different bait proteins since the libraries were not stably conserved and screened in one shot with no replicates in the cell population. Here, we propose an alternative experimental strategy that we called cell-based protein complementation assay (Cell-PCA, in comparison to Retrovirus (Re)-based PCA used in previous screens (Ding et al., 2006; Lee et
al, 2011; Simon E. Coopera, 2015) to circumvent this issue (Fig. 1D). More specifically, the strategy consists in the establishment of cell lines expressing a PCA-compatible human ORFeome. These cell lines can be amplified and used several times to simultaneously screen PPIs for different bait proteins, eventually allowing comparing interactome properties against the same ORFeome in a uniform cell context (Fig. 1D).

As a proof-of-concept, we applied this novel experimental strategy to the HOX protein family, which is involved in the regulation of numerous processes during embryonic development (Pearson et al, 2005; Moisés Mallo, 2018) and adult life (Seifert et al, 2015). HOX proteins are transcription factors (TFs) and therefore act by regulating the expression of downstream target genes in vivo. Several cofactors have been identified for different individual HOX proteins in various cell and developmental contexts (Carnescchi et al, 2020; Lambert et al, 2012; Baeëza et al, 2015; Bischof et al, 2018), but no systematic large-scale interaction screening has been performed for several HOX members in the same biological system so far. As a consequence, very little is known about their general and specific interactome properties. For example, the question of HOX cofactor specificity remains poorly understood: is there a large proportion of specific versus common cofactors between different HOX proteins? Along the same line, work with mouse HOXA1 showed that a number of cofactors were not traditional TFs, suggesting that HOX proteins could act at different regulatory levels (Lambert et al, 2012). Whether this property could apply more largely to other HOX protein members remains to be investigated.

To tackle the issue of HOX interactome properties, we present the first large-scale screening of PPIs for eight different human HOX proteins in the same immortalized cell populations. Our results showed that TFs are generally not employed as HOX-specific cofactors, but instead used in different combinations in the different interactomes underlying common expected functions, such as RNA-Pol-II dependent transcriptional regulation. Surprisingly, our results revealed new functional interactomes that are also largely shared among HOX members, such as the regulation of the circadian rhythm or translation regulation. These interactomes contain more HOX-specific interactions with non-TFs, highlighting that HOX proteins could display various molecular strategies depending on the cellular or physiological process to regulate.

Taken together, this work establishes new tools for assessing the issue of interactome specificity in live cells.
Results

Cell-PCA screen design

The fluorescence-based complementation approach, also called BIFC (Bimolecular Fluorescence Complementation), relies on the property of hemi-fragments of monomeric fluorescent proteins such as the GFP (Green Fluorescent Protein) or Venus, to reconstitute a functional fluorescent protein upon spatial proximity (Miller et al, 2015). Here we used a library of 8200 ORFs and fused these ORFs to the C-terminal fragment of the blue fluorescent protein Cerulean at the 5’ end (fragment CC, Fig. 2A and Materials and Methods). This fragment can complement with the N-terminal fragment of Venus (VN, leading to a Venus-like fluorescent signal) or Cerulean (CN, leading to a Cerulean-like fluorescent signal), enabling to simultaneously visualize interactions of two different bait proteins with a common cofactor (Bischof et al, 2018; Hu & Kerppola, 2003).

The CC-ORF library was cloned in a lentivirus vector, downstream of regulatory sequences (Tet-Responsive Element (TRE), Fig. 2A and Fig. S1) that respond to the tTA (tetracycline-controlled transactivator) factor in the presence of Doxycycline. The CC-ORF pooled plasmid library was used for producing lentiviral particles and subsequent infection of HEK-293T cells (see Materials and Methods). Referring to the functional titer, the pooled lentiviral libraries were transduced at a low multiplicity of infection (MOI) to achieve only one stably-integrated CC-ORF for most cells (see Materials and Methods). The two resulting cell lines were named CC-HEK-1 and CC-HEK-2, which respectively encompasses 5799 and 5549 ORFs (Fig. 2A and Dataset EV1-2). The presence and basal expression of the CC-ORF constructs were further verified in the established CC-HEK cell lines by immunostaining against the CC fragment (Fig. S1).

The VN fusion bait protein used for the screen was under the control of the constitutive CMV promoter (see Materials and Methods). Transfecting the VN-fusion plasmid into the established CC-HEK cells will lead to BIFC-positive signals only when interaction occurs between the VN-fused bait and the CC-fused prey protein encoded by the integrated CC-ORF (Fig. 2B). The fluorescent cells are then specifically sorted using flow cytometry, from which the genomic DNA was extracted to prepare a sequencing library with specific oligonucleotides matching the CC-ORF construct (Fig. 2B, Fig. S2 and Materials and Methods). The presence and relative abundance of the integrated CC-ORFs from sorted cells
were assessed using a dedicated next-generation sequencing (NGS) approach (see Materials and Methods). This targeting approach allows to reduce the sequencing effort to only the beginning of the inserted CC-ORF, instead of the complete genome or insert fragments (that are highly variable in size). This strategy improves the sequencing coverage together with a reduced cost.

**Proof of concept Cell-PCA screen for HOXA9 interactomes**

As a proof-of-concept, we performed a high throughput interaction screen with the human HOXA9 protein, whose interaction with two known cofactors, PBX1 and MEIS1, has been extensively described by using BiFC in HEK and other cell lines (Dard et al, 2019a). This previous work established BiFC as a specific and sensitive method for deciphering HOXA9 interaction properties, therefore establishing the appropriateness of our tools in the context of a large-scale BiFC interaction screen. We also considered a mutant form of HOXA9 as a supplementary control of the screen (VN-HOXA9\(^{W}\)). This form is mutated in a conserved Trp residue that mediates the interaction with the PBX cofactor in the context of HOX/PBX dimeric complexes (Dard et al, 2019a; LaRonde-LeBlanc & Wolberger, 2003). This conserved Trp residue was also shown to have additional and versatile activities that varied depending on the cell context, suggesting that it could interact with other cofactors (Dard et al, 2019a). Thus, the Trp-mutant form constitutes a good specificity control for assessing the sensitivity of our experimental tools in the context of two highly similar bait proteins.

The pilot screens with VN-HOXA9 and VN-HOXA9\(^{W}\) were sequentially performed in the two different CC-HEK-1 and CC-HEK-2 cell lines for two main reasons. First, transduction was performed with a low MOI to get only one CC-ORF construct in the majority of cells (see Materials and Methods). This transduction condition resulted in only partial (70%) integration of the CC-ORFeome library. Second, the two different CC-HEK cell lines were used as biological replicates for assessing the reproducibility of interactions among their common pool of integrated CC-ORFs (72% of the integrated CC-ORFs were common between the two CC-HEK cell lines, Fig. 3A).

The basal expression of the TRE promoter was used for the BiFC screen to not only simplify the protocol but also have a minimum expression level of the CC-ORF library, allowing screening BiFC signals in more stringent conditions with the transfected VN-fusion bait protein. Under these conditions, fluorescent signals were only observed upon co-
transfection of VN-HOX9 (Fig. S3), and this pattern was systematically obtained in the subsequent different screens. Finally, two types of selection criteria were applied for selecting most relevant candidate interactions, depending on scored enrichment of deconvolved CC-ORFs (whether it was specific of a CC-HEK cell line or present in the common 4751 CC-ORFs: see Materials and Methods).

According to our experimental protocol and selection criteria, we found that HOXA9 had a total number of 525 (8%) positive interactions, among which 121 were common to the two CC-HEK cell lines (Fig. 3B and Dataset EV3). To verify that our post-NGS selection criteria were correct, we randomly selected 20 CC-ORFs with a board range of enrichment scores (log2FC) and that were positive either in one (9 CC-ORFs) or the two CC-HEK cell lines (11 CC-ORFs) for doing individual BiFC tests with VN-HOXA9 (Fig. 3C). Three negative TFs from the screen were also picked-up (C6orf201, FHL5 and UBE2T). The known VN-HOXA9/CC-PBX1 interaction was used as a positive PPI control and BiFC system reference, as well as calibrator of microscope parameters, allowing repeating and comparing individual assays between the different biological replicates (see Materials and Methods). These individual assays confirmed the three negative interaction status and the positive interaction status for all but the interaction with CC-CRY1 (Cryptochrome Circadian Regulator 1; Fig. 3D-E).

Interestingly, these analyses revealed different interaction profiles in live cells, highlighting both the sensitivity and specificity of the BiFC signals (Fig. 3D). Collectively, these observations confirmed that the applied filtering criteria were appropriate for selecting positive interactions from the large-scale cell-PCA screen.

We next performed the ORFeome-wide BiFC screen in the two CC-HEK cell lines with VN-HOXA9W. Surprisingly, this single mutation led to strong differences when compared to wild type HOXA9. In total, 461 (7%) interactions were captured with VN-HOXA9W (Fig. 4A and Dataset EV4), among which 152 were common with VN-HOXA9 (36% of the interactions, Fig. 4B). Accordingly, several overrepresented biological functions were present in both VN-HOXA9 and VN-HOXA9W interactomes, including functions linked to transcriptional regulation or morphogenesis (Fig. 4C). Other less characterized functions were also common to VN-HOXA9 and VN-HOXA9W, such as regulation of the circadian rhythm or dsRNA processing (Fig. 4C). Moreover, the Trp mutation also led to the loss of a number of important HOX-related functions, like pattern specification, skeletal muscle cell differentiation or chromatin remodeling (Fig. 4D). Interestingly, transcriptional misregulation
in cancer was also lost with HOXA9\textsuperscript{W} (Fig. 4D), an effect that has previously been reported in several studies for HOXA9 (Dickson \textit{et al}, 2013; Ando \textit{et al}, 2014). Finally, the Trp mutation led to ectopic interactions involved in novel functions (not find with HOXA9), such as positive regulation of neuron differentiation, rRNA modification in the nucleus and cytosol, or thyroid cancer and apoptosis (Fig. 4D).

Together, results obtained with VN-HOXA9 and VN-HOXA9\textsuperscript{W} validated the proof-of-concept ORFeome-wide interaction screen for capturing and distinguishing interactomes between two highly related HOX proteins, which encouraged us to test this strategy in a systematic HOX interactome exploration. To this end, we next applied the same experimental strategy for capturing the interactome of seven additional human HOX proteins, tackling the general issue of human HOX interactome specificity in the same biological system.

\textbf{Using Cell-PCA for a global comparison of HOX interactomes}

HOX members belonging to anterior (HOXA1 and HOXA2), central (HOXC6, HOXA7 and HOXC8) and posterior (HOXA9, HOXD10 and HOXB13) paralog groups were chosen for the ORFeome-wide comparison between different HOX interactomes (Fig. 5A). HOX proteins were fused to the VN fragment, as previously described (Dard \textit{et al}, 2018a, 2019b), and each VN-HOX encoding plasmid was transfected in the two CC-HEK-1 and CC-HEK-2 cell lines for the ORFeome-wide BiFC screen. We applied the same selection criteria as previously described with HOXA9 for sorting fluorescent cells and selecting candidate interaction partners.

Results showed that each HOX member had a comparable number of positive interactions (between 9\% and 6.5\% of positive interactions, Fig. 5B and Datasets EV5). The majority of these interactions are not unique, being also found with one or more additional HOX proteins (Fig. 5C). Still, each HOX protein showed a specific cluster of interactions (Fig. 5C). Interestingly, we found that an important part of the interactions (between 35\% and 55\%) did not correspond to TFs (Fig. 5D). Moreover, non-TFs were particularly enriched among HOX-specific interactions. As a corollary, TFs were more enriched in non-HOX specific interactions (interactions with two or more HOX proteins, Fig.5D). The analysis of TF DNA-binding domains further showed that the same TF classes were globally found in non-HOX-specific TFs (including SMAD, Homeodomain, C2H2 Zinc Fingers and bZIP domains: Fig. S4).
In contrast, few individual classes were found in the HOX-specific TFs, as observed for HOXA1 (ARID/BRIGHT class), HOXA2 (E2F class), C6 (CxxC class), A7 (HSF class) and D10 (Rel class) HOX proteins (Fig. S4).

As expected, the heatmap of the top-20 enriched functions among all HOX interactors revealed functions linked to transcriptional regulation and embryonic development (Fig. 5E). We next analyzed more precisely the biological functional networks that were enriched in HOX interactors (see also Materials and Methods). This analysis confirmed that networks involved in transcriptional regulation, embryonic development, and other functions described in the literature (for example DNA repair or protein modification pathways such as SUMOylation and ubiquitination) were largely distributed throughout HOX protein members (Fig. S5). Intriguingly, additional functions were largely distributed yet poorly or never described in the literature, such as fat cell differentiation, chemokine/cytokine signaling pathway, lipid metabolic process, aging/senescence/longevity, circadian rhythm and pathways involved in mRNA processing and translation regulation (Fig. S5). Other biological functions were found to be more specific, like cilium organization, TOR signaling, kidney or liver development (Fig. S5). Altogether these observations highlight that HOX proteins can establish several major interaction networks that are not obligatory involved in transcriptional regulation.

Finally, we analyzed representative networks involved in common or specific HOX functions in terms of cofactor type composition (TF versus non-TF and HOX-specific versus non-HOX-specific). The first two networks analyzed are involved in transcriptional regulation by RNA-Pol-II or epigenetics, and were found with all HOX proteins. As expected, the RNA-Pol-II interactome contains a large majority of TFs, therefore a majority of non-specific interactions (Fig. 6A). The epigenetic interactome contains more non-TFs that were in addition often specific (Fig. 6B). The interactome underlying circadian rhythm, which a novel common Hox function, also contains a majority of non-specific TFs and few specific non-TFs (Fig. 6C). Finally, the interactome underlying more specific functions, like DNA-damage response (found for HOXA2, HOXC6 and HOXA9) was not particularly more enriched in HOX-specific interactions when compared to the previous largely distributed functions (Fig. 6D). Altogether, these analyses reveal that HOX interactome specificity mostly occurs at the level of specific combinations with non-specific cofactors, rather than at the level of interactions with several HOX-specific cofactors.
Discussion

Deciphering and comparing interactomes in the same live cell context

Protein interactomes are versatile networks involving hundreds of transient and low affinity interactions. Over the last years, several experimental strategies based on PCA systems have been developed to capture these molecular interactions, leading to promising alternatives in addition to LC-MS or yeast-two-hybrid-based approaches. In this context, BiFC-based PCA is particularly well-adapted for revealing pair-wise interactions in live cells and has been applied in several screen strategies to study interactomes of different bait proteins (Remy & Michnick, 2004; Berendzen et al., 2012; Ding et al, 2006; Miller et al, 2015). Although these screens were based on an off/on readout, with no enrichment scores neither replicates, they have dissected sets of specific interactions that were further confirmed by alternative molecular and functional assays. Altogether, this previous work established BiFC as a powerful approach for performing specific and sensitive large-scale protein interaction screens. Our work further enriches the repertoire of applicability of BiFC for large-scale protein interaction screens, in particular by proposing an experimental setup that allows using the same cell line for different screens and getting an additional level of information for comparing different interactomes.

Our novel strategy relies on the establishment of a cell line that has integrated a BiFC-compatible ORFeome. This cell line can be amplified and used several times for BiFC interaction screens with different bait proteins. As a proof of concept, we used the HOX protein family and two different CC-HEK cell lines, and proposed stringent filtering parameters for selecting best candidate interaction partners. The screen was voluntary performed in conditions of low expression level for each CC-ORF prey construct, allowing getting specific BiFC signals with the transfected (therefore under multiple copies) VN-bait protein. The specificity of the screen was further confirmed by individual BiFC assays, which confirmed the negative or positive interaction status of 23 out of 24 candidate cofactors randomly picked from the screen with HOXA9, and additionally revealed specific interaction profiles within the cell.

In conclusion, our approach is advantageous in performing ORFeome-wide interaction screens upon simple transfection of a bait protein, which not only simplifies the protocol (the screen can be performed in a classical A2 laboratory environment since it does
not rely on systematic transduction, as previously described (Ding et al., 2006)), but also enables testing different bait proteins in the same batch of cells, therefore providing a unique level of information for comparative interactome analyses.

**Role of the HOXA9 Hexapeptide (HX) motif: a global interaction motif?**

As a proof-of-concept experiment for validating the specificity of our tools, we compared wild type and Trp-mutated HOXA9. This residue is part of the so-called Hexapeptide (HX) motif that is also found, together with additional conserved residues, in other vertebrate and non-vertebrate HOX proteins. Originally described to ensure the interaction with the generic PBC-class of HOX cofactors, the HX motif was later found to be involved in a number of interactions with various TFs in Drosophila Hox proteins (Baeëza et al., 2015). The HX motif was also recently described to promote the interaction with the exportin CRM1/Emb protein (Duffraisse et al., 2020), highlighting that it constitutes a unique platform for establishing highly diverse types of PPIs.

We found that the single Trp mutation led to drastic changes in the HOXA9 interactome, most of them being related to transcriptional regulatory processes. Surprisingly, other functional networks were not affected (like endocrine system development), enriched (like dsRNA processing) or even novel (like positive regulation of neuron differentiation, megakaryocyte differentiation or apoptosis). These results show that the HX motif is not only important for promoting but also for inhibiting PPIs, recalling previous observations with Drosophila Hox proteins (Baeëza et al., 2015). They also established that our experimental tools were sensitive and specific enough for comparing interactomes between two highly related bait proteins.

**HOX interactomes: importance of non-transcriptional networks and non-TFs for HOX specific functions**

The analysis of eight different HOX interactomes revealed several unexpected and interesting molecular features. For example, there was a high proportion of non-TFs in the overall pool of HOX interacting proteins. This result confirms previous observations that were so far limited to HOXA1 (Lambert et al., 2012; Taminiau et al., 2016). Interestingly, TFs were mostly engaged in non-HOX specific interactions, whereas the proportion of non-TFs was increased in HOX-specific interactions. Accordingly, HOX interactomes related to
transcriptional regulatory processes contained distinct combinations of non-specific interacting TFs. This observation underlines that HOX transcriptional specificity mostly relies on the establishment of specific combinations of interactions with TFs that have the potential to interact with several HOX proteins. This molecular mode of action has already been proposed for *Drosophila* Hox interactomes, suggesting that it could be a general and conserved feature underlying Hox transcriptional specificity (Baeëza et al., 2015).

In contrast, we found that non-TFs were more frequently engaged in HOX-specific interactions. This novel type of interactions illustrates the ability of HOX proteins to be engaged in the regulation of several post-transcriptional regulatory process, a level that has poorly been investigated so far. A similar observation has been noticed for the *Drosophila* Ubx protein, which was shown to engage tissue-specific interactions with cofactors involved in translational regulation (Carnesecchi et al., 2020).

In conclusion, our work confirmed that HOX proteins are probably involved in so far poorly investigated post-transcriptional regulatory processes and that our understanding of their specific molecular mode of action requires considering more the potential role of non-TF partners.

**Figure legends**

**Figure 1. Principle of the Protein Complementation Assay (PCA) and its applications in large-scale interaction screens.**

A. Application of fluorescent-based PCA for revealing interaction between two candidate partners. The N (FPN)- or C (FPC)-terminal fragment of the fluorescent protein (FP) is fused to one of the two putative interaction partners (bait and prey proteins). The interaction between the bait and prey proteins allows the reconstitution of the fluorescent protein and the emission of fluorescent signals upon excitation. This principle of complementation has also been developed with enzymes for large-scale interaction screens (see for example (Remy et al., 2007)).

B-D. Application of PCA-based strategies for large-scale interaction screens in living cells. The Cotransfection (Co)-PCA screening strategy relies on the transitory expression of the FPN-fusion bait protein and the FPC-fusion human ORFeome library upon co-transfection in the cell line (Remy & Michnick, 2004) (Berendzen, K.W., Böhmer, M., Wallmeroth, N., Peter, S., Vesić, M., Zhou, Y., Tiesler, F.K.E., Schleifenbaum, F., and Harter, 2012) (B). The Retrovirus (Re)-PCA screening strategy relies on the use of a cell line stably expressing the FPN-fusion bait protein and infected by
retroviruses containing either the FPC-encoding fragment in the three possible open reading frames (Ding et al, 2006) (C), or a human ORFeome library fused to the FPC fragment (Lee et al, 2011)(Simon E. Coopera, 2015) (C'). In the first generation, the FPC fragment will randomly insert in the genome and produce endogenous FPC-prey proteins that could interact with the FPN-bait protein. In the second generation, the FPC-fusion library is artificially expressed with a constitutive promoter. The FPN-bait protein is constitutively expressed with an artificial promoter in both screens. The Cell-PCA screening strategy relies on the use of cell lines stably expressing the FPC-fusion library (D). These cell lines can be used multiple times for screening for interacting partners of different FPN-fusion bait proteins upon transfection.

**Figure 2. Experimental procedure for the Cell-PCA-based screen.** A. A pool of ~8200 hORFs derived from the hORFeome v3.1 was cloned en masse by Gateway® LR reaction into the lentiviral vector pLV-CC (Fig. S1), subsequently generating the CC-ORFeome plasmid library (pLV-CC-hORFs). The final expression constructs (~8000 ORFs) were used to produce lentivirus and infect HEK293T cells to generate two cell lines (CC-HEK-1 and CC-HEK-2). B. Each CC-HEK cell line can be transfected with the VN-HOX-encoding plasmid. Any interaction with a CC-ORF leads to fluorescent cells that are collected using flow cytometry. Genomic DNA (gDNA) is extracted from the fluorescent sorted cells and interacting ORFs are identified through a next generation sequencing (NGS) dedicated approach. CC: C-terminal fragment of mCerulean (residues 155-238). VN: N-terminal fragment of mVenus (residues 1-172). MOI: multiplicity of infection.

**Figure 3. Establishing the Cell-PCA screening strategy with HOXA9.** A. Venn diagram depicting the number of integrated ORFs in the CC-HEK-1 (blue) and CC-HEK-2 (red) cell lines. B. Venn diagram showing the number of HOXA9-positive ORFs in the two CC-HEK cell lines. VN-HOXA9 is schematized above the Venn diagram (with the Trp-containing motif -W- and the homeodomain -HD-). C. Plot of the 553 selected HOXA9-interacting candidates, ranked from the most to the lowest enriched in the Cell-PCA assay. Among them, 20 CC-ORFs were randomly picked for individual validation by BiFC, using two criteria: the 9 red dots were unique to one CC-HEK cell line with a log2 fold change (FC) superior to 9; the 11 blue dots were present in the two CC-HEK cell lines with a log2FC superior to the
background. See also Materials and Methods. **D-E.** Illustrative confocal pictures of BiFC between HOXA9 and the positive (D) or negative (E) candidates selected from the screen in live HEK293T cells. Pictures are illustrative of two independent biological replicates. BiFC between HOXA9 and PBX1 and the mCherry reporter (merge panels) were systematically used as a positive control for assessing transfection efficiency. BiFC was confirmed between HOXA9 and all candidates but CRY-1. Note the various intra-cellular BiFC profiles with different candidates. Interactions were defined as negative when the normalized (with mCherry) fluorescence intensity of the BiFC was below 15% of the fluorescence intensity resulting from HOXA9/PBX1 BiFC on average. Scale bar, 10µm.

**Figure 4. Cell-PCA reveals distinct interactomes for HOXA9 and HOXA9W.** A. Venn diagram of HOXA9W-interacting ORFs in the two CC-HEK cell lines. The Trp (W) mutation into an Ala (A) is shown in the schematized VN-HOXA9W protein above the Venn diagram. B. Venn diagram showing the comparison between HOXA9 and HOXA9W interactomes. C. Heatmap for the top 20 enriched biological functions in both HOXA9 and HOXA9W interactomes. One row per function, using a discrete color scale to represent statistical significance (from high (dark red) to no (gray) significance). Red asterisk and circles highlight functions involved in transcriptional regulation or morphogenesis, respectively. D. Heatmap showing the specific biological functions underlying HOXA9 and HOXA9W interactomes (considering 373 HOXA9-specific and 272 HOXA9W-specific interactions).

**Figure 5. Application of Cell-PCA for global HOX interactome screening and comparison.** A. Schematic arrangement of the 39 human HOX genes. HOX genes belong to anterior (green), central (purple) and posterior (blue) paralog groups. The 8 HOX genes used in the screen are highlighted (framed in black). B. Pie chart illustrating the number of interacting ORFs identified for each HOX protein in the Cell-PCA screens. C. Heatmap of interacting ORFs identified for each HOX protein in the Cell-PCA screens. Hierarchical clustering was performed on both column and row, according to the Pearson distance based on log2 fold change (FC) values, using average method. Scale bar indicates enrichment score (log2FC) for each HOX-interacting ORFs. D. Distribution of transcription factors (TFs) among three different categories of HOX interactors (full set of interactors: light gray; non-HOX specific interactors: gray; HOX-specific interactors: dark gray). Note that TFs constitute around 50%
of the total interactions on average and are enriched in the non-HOX-specific category. E. Heatmap of the top-20 enriched functional profiles of the different HOX-interacting proteins. Hypergeometric p-values and enrichment factors were calculated and used for filtering. A hierarchical clustering was performed on both column and row based on Kappa-statistical similarities among their gene memberships. A discrete color scale is used to represent statistical significance. Gray color indicates a lack of significance. Functions involved in transcriptional regulation or embryonic development are denoted with a red asterisk or circle, respectively.

Figure 6. Representative HOX functional interactomes based on MCODE clusters. Interaction networks underlying representative common (A-C) and specific (D) HOX functions. Transcription factors are boxed in a rectangle whereas non transcription factor partners are boxed in a rhomb. Yellow-colored boxes represent specific interactors of one HOX in the described function. The same color code as in Fig. 5 applies for anterior, central and posterior HOX proteins.

Supplementary Figures

Supplementary Figure 1. Map of the lentiviral expression vector and basal expression level of the CC-ORF constructs in the established CC-HEK cell line. A. A lentiviral destination vector (pLV-CC-Gateway) was generated to clone the human ORFeome library in frame with the C-terminal fragment of Cerulean (CC), leading to the pLV-CC-ORF vector. The expression of the CC-ORF is under the control of a first-generation TRE promoter. This system (Tet-Off) requires the co-expression of a Tet transactivator (tTA) for full activation of the promoter. B. The basal activity of the TRE promoter (without adding Dox and the tTA-coding vector) is sufficient to detect the expression of the inserted CC-ORFs in the cell population. Immunostaining was performed with an anti-GFP recognizing the CC fragment (green). DAPI (blue) stains for nuclei. Scale bar = 50 μm.

Supplementary Figure 2. Schematic of the ORF Capture-Sequencing method. Libraries were constructed using our own designed proprietary protocol in order to enrich in sequences covering the beginning of all the CC-ORFs inserted in the genome. Roughly, the DNA extracted from sorted cells (1) is fragmented using the Covaris S220 Ultrasonicator (2).
Then, the Ion adapter P1 is non-directionally ligated to both ends of the DNA fragments according to the standard blunt-end ligation protocol for Ion Torrent libraries (3). Because the P1 libraries molecules containing hORF sequences are underrepresented, an enrichment by a first PCR is performed by using a biotinylated forward primer located on the plasmid sequence upstream the hORFs and a reverse primer located at the end of the P1 adapter (4). Biotinylated PCR fragments, containing now at each end the beginning of the hORFs and the P1 adapter respectively, are then captured using streptavidin-coupled magnetic beads (5, 6). These molecules are used as a template for the final hemi-nested PCR amplification using a forward fusion primer containing a barcoded Ion A adapter and the same reverse primer as before (7). The hemi-nested PCR reinforces the specificity of the fragments to be sequenced since the primer fused to A and the barcode is located closer to the hORF beginning on the plasmid. After a size selection using SPRI beads to meet Ion Torrent requirements (8), the qualified and quantified barcoded libraries are multiplexed in an equimolar manner and sequenced on the Ion Proton sequencer using a P1 chip following the manufacturer’s recommendations (9).

**Supplementary Figure 3. Testing for specific BiFC signal and FACS gate for BiFC-positive cell population.** A. Schematic representation of the cold (non-transfected) CC-HEK cell line. A’. Illustrative confocal picture of non-transfected cells upon excitation wave length for BiFC. Transmission light capture shows the cells. A”. FACs gates for the GFP channel. No fluorescent signal can be observed in the cold/non-transfected CC-HEK cell population. B. Schematic representation of VN-HOX9 transfection in the CC-HEK cell line. B’. Illustrative confocal picture of transfected cells upon excitation wave length for BiFC. Fluorescent cells are indicated (white arrows). Transmission light capture shows the cells. B”. FACs gates for the GFP channel. Fluorescent signals can be observed in the CC-HEK cell lines upon transfection of VN-HOXA9. Compared to A”, the specific GFP-positive population can be isolated from the negative/non-fluorescent cells. This population represents 2.67% of the total cell population. A similar percentage of fluorescent cells (between 1,5 and 3%) was obtained with all HOX proteins.
Supplementary Figure 4. Analysis of the DNA binding domain (DBD) of HOX interacting TFs in the non-HOX specific (upper panel) or HOX-specific (lower panel) category. A different color illustrates the different classes of TFs, as indicated.

Supplementary Figure 5. Heatmap of enriched biological functions retrieved from HOX MCODE clusters. In HOX protein interaction network formed by their interacting-candidates, the MCODE algorithm (Bader and Hogue, 2003) is iteratively applied to each connected network components, generating different HOX MCODE clusters. GO enrichment analysis was applied to each MCODE network to assign “meanings” to the network component. The functional enrichment analysis for MCODE clusters is represented as a heatmap. The color code on the left represents diverse functional categories, which were manually curated. Gradient blue color indicates different significance level (“−log10” transformed p-values of enriched biological functions) and non-significant enrichment is shown in white. Asterisks highlight largely distributed yet unknown functions of HOX proteins.

Supporting Information

Dataset EV1. List of 5799 ORFs expressed in CC-ORFeome HEK cell line-1.
Dataset EV2. List of 5549 ORFs expressed in CC-ORFeome HEK cell line-2.
Dataset EV3. List of HOXA9-interacting proteins identified by Cell-PCA screening.
Dataset EV4. List of HOXA9W-interacting proteins identified by Cell-PCA screening.
Dataset EV5. Full list of HOX-interacting proteins identified by Cell-PCA screens.

Materials and Methods

Cell Lines

HEK-293T cells were purchased from European Collection of Authenticated Cell Culture (ECACC) through the biological resource center Anira-AGC platform of the SFR Biosciences UAR3444/US8 of Lyon. Cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM-
GlutaMAX-I, Gibco by Life Technologies) supplemented with 10% (v/v) heat inactivated fetal bovine serum (FBS) and 1% (v/v) Penicillin-Streptomycin (5,000U penicillin and 5mg streptomycin/mL), incubating at 37°C, in an atmosphere of 5% CO₂. HEK-293T-CC-ORFs were cultured as above with 0.3 μg/ml of puromycin (Gibco, Cat No. A1113803) in their culture medium.

**Plasmids**

The bait plasmids pcDNA3-VN-HOXs (expressing VN-HOXs) were made as described previously (Dard et al, 2019b, 2018b). The lentiviral plLV-CC-ORFs vector collection, representing about 8 200 ORFs from the V3.1 version of the hORFeome, was kindly provided by P. Mangeot (CIRI, ENSL, France). ORFs were fused at the 5' end to the C-terminal part of the mCerulean gene (encoding the last 155-238 aa) using the Gateway® technology. For individual BiFC tests, constructs were cloned into the pLIX _403 vector (a gift from David Root, Addgene plasmid # 41395; http://n2t.net/addgene:41395 ; RRID:Addgene_41395). DNA sequencing of all constructions were carried out at GENEWIZ Company (Germany). All vectors are available upon request.

**Lentivirus Preparation and Infection**

The pooled lentiviral constructs plLV-CC-ORFs were packaged into lentivirus particles at the AniRA-Vectorology core facility (SFR Biosciences UAR3444/US8, Lyon, France). HEK-293T cells were transduced in independent replicates with two batches of lentivirus (CC-ORF library 1 and CC-ORF library 2) at a low multiplicity of infection (0.3) to achieve approximately one-gene-one-cell condition(Yea, K., Zhang, H., Xie, J., Jones, T.M., Yang, G., Song, B.D., and Lerner, 2013), with ≥500X representation. Culture medium was supplemented with 8 μg/mL polybrene (Sigma) at the time of transduction, and was changed the next day. Two days after transduction, cells were selected with 0.5 μg/mL Puromycin (Gibco, Cat No. A1113803) for 4 days, until the uninfected control cells completely died and the selected cells reach near confluence. Final amplified transduced cells were split into aliquots of 4x10⁶ cells each and stored in liquid nitrogen for future screens.

**Immunostaining of CC-ORF transduced HEK cells**
1x10^5 cells were seeded on glass coverslips in 24-well plates. Twenty-four hours after plating, cells were fixed in 4% paraformaldehyde in PBS for 15 min at room temperature, permeabilized in 0.3% Triton X-100 for 10 min, and rinsed in PBS. The cells were preincubated in 3% Bovine serum albumin-PBS at room temperature for 1 h and incubated overnight at 4°C with primary antibodies against CC-ORF (rabbit anti-GFP, polyclonal, Invitrogen A1122, 1:1000). The sections were then incubated with the corresponding fluorescein-conjugated secondary antibodies (Alexa Fluor 488 anti-rabbit, Invitrogen A11008, 1:1000) for 2 h at room temperature. Coverslips were mounted in VECTASHIELD Antifade Mounting Medium with DAPI (VECTOR, Cat No. LS-J1033-10). They were analyzed via confocal microscopy (Zeiss LSM780).

**Cell-PCA Screen**

8.10^6 CC-HEK cells (~800X representation) were thawed and passaged for 2 population doublings to recover. For each screen, aliquots of 6 x10^6 recovered cells were seeded in two 6-well plate (500k cells/well) and grown for 24 hours for achieving a final confluence around 80%. The basal expression of the TRE promoter was used for the expression of the CC-ORFs in the cell population. The transfection of different bait plasmids pcDNA3-VN-HOXs was performed using the jetPRIME reagent (Polyplus, Ref 114-15) following manufacturer’s instruction. After 18 h of transfection, all transfected cells were pooled and BiFC-positive cells were subsequently sorted using a BD FACS Aria II Cell Sorter (AniRA-Cytometry core facility of the SFR Biosciences UAR3444/US8, Lyon, France). After each screen, sorted cells were harvested and genomic DNA was extracted using PureLink Genomic DNA mini kit (Invitrogen, Cat No. K182001), according to manufacturer’s instruction. The genomic DNA was used for library preparation and subjected to next-generation sequencing at in-house NGS sequencing platform (PSI, IGFL, Lyon, France).

**Next Generation Sequencing and identification of the positive hORFs**

Libraries were constructed using our own designed proprietary protocol in order to enrich in sequences covering the beginning of all the hORFs inserted in the genome (see details in Fig. S2). After a size selection using SPRI beads to meet Ion Torrent requirements, the qualified and quantified barcoded libraries were multiplexed in an equimolar manner and sequenced
on the Ion Proton sequencer using a P1 chip following the manufacturer’s recommendations.

NGS raw data were analyzed with the Galaxy instance (Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., 2018) of the ENS of Lyon and maintained by the Centre Blaise Pascal (CBP, ENS Lyon). A dedicated Galaxy pipeline was created to identify the hORFs detected by sequencing and their associated read counts for each barcoded sample. After demultiplexing, the raw reads were trimmed to remove very low-quality bases in the 3’ and 5’ ends using a sliding window process. By construction, the libraries are oriented and all the reads begin with the same short plasmid sequence that is present upstream of any inserted hORF. This sequence was removed by Cutadapt (Martin, 2011), by allowing a maximum error rate of 0.15. Only the trimmed reads beginning with ATG were retained for further analysis. This last step removes any reads that could result from a non-specific PCR amplification. Because read length is variable with the Ion Torrent technology, the reads were then trimmed to 50bp to have all the same length. These reads were next compared to the hORFeome.V3.1 database by BLAST tool (Cock, P.J.A., Chilton, J.M., Grüning, B., Johnson, J.E., and Soranzo, 2015) using strict conditions (only one hit retained with at least 98% of identity on 95% of the query coverage; matches starting to position 1 of the hORFs). For each hORF that obtains hits, the number of reads matching this hORF was counted and sum up in a table for further analyses. According to the criteria used, only hORFs that have more than 3 bases of difference in their first 50bp can be differentiated. Moreover, when the beginning of the hORFs is identical or nearly identical (mainly hORFs corresponding to different isoforms of the same gene), the read is assigned to only one of the possible alternatives, usually always the same.

**Identification of candidate HOX-interacting ORFs**

The resulting sequencing reads for each ORF in each library were normalized to 10M. The present ORFs were denoised using a chosen threshold of at least 479 reads. Then, the number of reads in sorted cells was divided by the number of reads in the control and the log2FC was calculated as an enrichment score (ES) for each ORF. To define which ORF is significantly enriched in each replicate, we calculated a theoretical threshold for enrichment, by dividing the total number of present ORFs in the control by that in sorted cells and transforming the result as a log2.
To generate our final list of HOX-interacting candidates, we combined the ORFs that were present in both replicates with an ES>= theoretical threshold, assigning the higher ES to each ORF. To get an extensive view of potential interactions, we also considered the top-enriched genes (with a stringent ES>=9) in each replicate.

**Individual BiFC validation in live cells**

For transfection, 3 \times 10^5 cells were seeded on glass coverslips in 6-well plates and incubated for 24h. Then, cells were transfected with jetPRIME (Polyplus, Ref 114-15) following manufacturer’s instruction. A total of 1.75ug of plasmid DNA were transfected per well: 750 ng of plix-VN-HOXA9, 750 ng of plix-CC-ORF and 250 ng plix-mCherry plasmids. After 18 hours of incubation in the presence of doxycycline (100 ng/ml final), the cell-coated coverslip was taken and mounted carefully on a glass slide for image capture under confocal microscopy (Zeiss LSM780). All samples were imaged using identical settings and quantified as previously described (Dard et al, 2018b). Two biological replicates were systematically done, using the interaction between HOXA9 and PBX1 as a positive BiFC control and the mCherry reporter for assessing transfection efficiency.

**Functional enrichment and Interactome analysis**

Both functional and interactome analyses were performed with Metascape (https://metascape.org/) (Zhou et al., 2019) using custom analysis settings. Subsequently, Cytoscape v3.8.2 (Shannon et al., 2003) was conducted to visualize representative HOX functional interactomes.

In functional enrichment analysis, the HOX-interacting protein candidates were searched against GeneOntology Biological Processes, KEGG pathways, CORUM and Reactome databases. A p-value cutoff ≤0.01 was used to determine significant functional terms. They were then hierarchically clustered into a tree based on Kappa-statistical similarities among their gene memberships. A kappa score ≥0.3 was applied as the threshold to cast the tree into term clusters. We selected the term with the best p-value within each cluster as its representative term and display them in the heatmaps.

Physical PPIs from multiple data sources were captured for construction of the interaction networks. Homo sapiens were selected as the organism for subsequent analysis. Min network size 3 was regarded as cut-off criterion for network visualization and disconnected
nodes was hidden. The complex identification algorithm, MCODE (Bader, G.D., and Hogue, 2003), was used to identify highly interconnected clusters in the network. The most important protein complex clusters in the PPI network were extracted, with default settings in MCODE, degree cutoff = 2, node score cutoff = 0.2, Max depth = 100, and k-score = 2. For each complex, it further applied function enrichment analysis and used significantly enriched terms for annotation of its biological roles. Following manual curation, the similar terms were combined and classified into non-redundant parent functions and categories, which was visualized by heatmap. Based on combined data set, all representative HOX functional interactomes were generated and visualized by Cytoscape.

Acknowledgements

We thank Dr. P. Mangeot (CIRI, Lyon, France) for providing the pLV-CC-ORF plasmid library, the AniRA-Cytometry, AniRA-Vectorology and AniRA- Genetic and Cellular Analysis platforms of the UAR3444/USB (Lyon, France), and the Blaise Pascal Center (ENSL) for computer resources. We thank Frédéric Marmigère for careful reading of the manuscript and helpful comments.

This work was supported by the CNRS, ENS-Lyon, Fondation pour la Recherche Médicale (FRM #160896), Association pour la Recherche contre le Cancer (ARC #PJA20191209567), Ligue Régionale contre le Cancer (#119030) and Centre Franco-Indian pour la Promotion de la recherche Avancée (Cefipra #5503-P). Y.J. received a PhD grant from the China Scholarship Council (grant #201708070003).

Competing of interests

Authors declare that no competing interests exist.

References


Ando H, Natsume A, Senga T, Watanabe R, Ito I, Ohno M, Iwami K, Ohka F, Motomura K,


simple survival protein-fragment complementation assay (PCA) based on the enzyme dihydrofolate reductase. *Nat Protoc* 2: 2120–5

Remy I & Michnick SW (2004) Regulation of Apoptosis by the Ft1 Protein, a New Modulator of Protein Kinase B/Akt. *Mol Cell Biol* 24: 1493–1504


Varnaité R & MacNeill SA (2016) Meet the neighbours: mapping local protein interactomes by proximity-dependent labelling with BioID. *Proteomics* 1–16


Figure 1
Figure 2
Figure 3
Figure 4
Figure 5
Figure 6
Supplementary Figure 1
Supplementary Figure 3
3.2 Optimization for high-throughput BiFC screening

Yunlong Jia¹, Jonathan Reboulet¹, Françoise Bleicher¹, Agnès Dumont¹, Sylvie Di Ruscio¹,
Benjamin Gillet¹, Sandrine Hughes¹, and Samir Merabet¹

¹ Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, Université Lyon 1, CNRS, Ecole
Normale Supérieure de Lyon, 46, allée d’Italie 69364 Lyon cedex 07 France
3.2.1 Introduction

Since the original Cell-PCA screen was developed and demonstrated a robust and economical high-throughput approach to systematic dissection of cellular interactomes, more recently it has been implemented in a series of projects engaging with many target proteins in various biological contexts, such as MYC, PER2, ERK, etc. Nowadays, biotechnological advances are leapfrogging ahead with new instrumentation and state-of-the-art molecular toolkits. For example, versatile gain-of-function (ORF-overexpression based) and loss-of-function (RNAi or CRISPR/CAS9) screenings have blown the way of functional protein investigation. Though our Cell-PCA screen performed an appropriate readout in different studies, even leaky ORF expression existing in the cold CC-ORF cell library, some optimizations are still needed, for not only a known-issue patch, but also aiming to keep Cell-PCA a comparable method alongside of ever-changing technical progress. To this end, stepwise efforts have been done during my PhD.

3.2.2 Materials and Methods

3.2.2.1 Cell line and Plasmids

HEK-293T cells were purchased from Europen Collection of Authenticated Cell Cutures (ECACC) through the biological resource center CelluloNet (AniRA platform of Lyon). Cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM-GlutaMAX-I, Gibco by Life Technologies) supplemented with 10% (v/v) heat inactivated fetal bovine serum (FBS) and 1% (v/v) Penicillin-Streptomycin (5,000U penicillin and 5mg streptomycin/mL), incubating at 37°C, in an atmosphere of 5 % CO2.

The plix-403 lentiviral vector was bought from addgene (plasmid # 41395). This plasmid was subsequently modified by digestion and ligation with MCS (multiple cloning site sequence) oligo. The resulting plix-MCS was further used as backbone of plix-CC-mCherry and plix-CC-PBX1. Plasmid plix-VN-HOXA9 as mentioned in previous part (Part 3.1, Chapter II) was used to generate the plix-VN-HOXA9-PGK-BleoR-mCherry, with PURO element replaced by BleoR sequence. All restriction enzymes were purchased
from New England Biolabs. The subsequent ligation reaction were performed by T4 DNA ligase (Promega, France) and transformed into one shot Top10 competent cells (Invitrogen, Cat No. C404010). DNA sequencing were carried out at GENEWIZ Company (Germany). All described vectors are freely available for academic use and can be requested from the authors.

3.2.2.2 Transfection

For transfection, $3 \times 10^5$ cells were seeded on glass coverslips in 6-well plates. Twenty-four hours after plating, cells were transfected with jetPRIME (Polyplus, Ref 114-15) following manufacturer’s instruction. 1ug of each plasmid DNA were transfected per well, according to different test conditions. After 18 hours of incubation in the presence of doxycycline (100 ng/ml final), the cell-coated coverslip were taken and mounted carefully on a glass slide for image capture under confocal microscopy (Zeiss LSM780). All samples were imaged using identical settings and quantified as previously described (Dard et al., 2018, 2019).

3.2.2.3 Generation of Stable cell line

For stable transfection, HEK-293T cell line was used. The cells were seeded in 6-well plate one day before transfection. 3μg of plix-VN-HOXA9-PGK-BleoR-mCherry construct were transfected, containing the BleoR gene (resistant to Zeocin), using with jetPRIME (Polyplus, Ref 114-15) following manufacturer’s instruction. Seven hours post-transfection, change the transfection medium by new fresh cell growth medium and turn the dish to the incubator. After 40h, check the cell confluence, if cell confluence $\geq 80\%$, the selection was initiated with 200μg/ml of Zeocin (Invitrogen, France). Bleo-resistant cells were dominated after 5 days of Zeocin selection and propagated using the same medium, transfer established cells into a big vessel, such as T75 flask. The Zeocin-selection
lasted for about 3 weeks. Final polyclonal stable cell line was cryopreserved in liquid nitrogen for further usage.

### 3.2.2.4 Functional titration by qPCR

A Biorad CFX96 (Biorad, France) was used for all qPCR measurements. To each reaction (10 µl) containing 5µl iTaq™ Universal SYBR® Green Supermix (2x) (BioRad) and 500 nM of each primer,

WPRE [reaction 1] :

*Forward:* ACAATTCCGTGGTGTTGTCG  
*Reverse:* AAGGGACGTAGCAGAAGGAC

RPPH1 [reaction 2]:

*Forward:* AATGGGCGGAGGAGTAGT  
*Reverse:* AGCTTGGAACAGACTCACGG

Plus, we added 4 µl diluted genomic or plasmid DNA (50ng). The temperature profile was 95°C for 3 min followed by 40 cycles of amplification (95°C for 5s, 60°C for 30 s). All samples were analyzed by melting curve analysis (65-95°C, 0.5°C increments at 2-5 s/step). Vector copy numbers in cells are normalized to human RPPH1 gene copies and presented as proviral copies per genome equivalent. Calculate titers (Transdution units per ml, TU ml⁻¹) according to the following formula:

\[ TU \text{ ml}^{-1} = \frac{C \times N \times D \times 1,000}{V} \]

where \( C = \) proviral copies per genome, \( N = \) number of cells at time of transduction, \( D = \) dilution of vector preparation, \( V = \) volume (ul) of diluted vector added in each well for transduction. The detailed protocol is available on request.
3.2.2.5 New protocol of pooled plasmid DNA preparation and transformation

The protocol of pooled plasmid DNA preparation was provided by the Genetic Perturbation Platform (GPP) (https://portals.broadinstitute.org/gpp/public/resources/protocols).

The collection of 1837 TF ORFs was divided in 7 minipools of average 262 hORFs (the true ORF number varies from 140 to 362). For each minipool, the hORFs were cloned en masse from the pDONR223 into the lentiviral expression vector plix-CC-Gateway by Gateway LR reaction (Invitrogen, CAT. 11791020) following manufacturer’s instruction.

To select for the successfully cloned CC-ORFs, each reaction mixtures were transformed into Endura™ electro-competent cells (Lucigen, CAT. 60242-1) and transformed bacteria were selected on LB + Ampicillin plates. Finally, collect all bacterial colonies by scraping plate and proceed to lentiviral plix-CC-TF plasmid extraction using one maxi-prep reaction for each minipool.

3.2.3 Results

3.2.3.1 Design and Performance of new lentiviral vector

Firstly, given the leaky TRE (Tet Response Element) promoter of lentiviral vector pLV-CC-ORF, the 2nd generation Tet-On (Tet-Advanced) system was used in the new lentiviral vector, plix-403. To assess the performance of this new vector, the reporter mCherry was used to constitute plix-CC-mCherry and pLV-CC-mCherry, respectively (Figure II-1A). The basal expression of mCherry was evaluated by transfection of HEK-293 cells in a Dox-free complete media. In Tet-On system, herein plix-CC-mCherry, new transactivator rtTA (reverse tetracycline-controlled transactivator) was created by fusing rTetR with VP16, which reversed the phenotype and created a reliance on the presence of tetracycline for induction, rather than repression. The rtTA will not bind TRE promoter to induce the gene expression, in absence of Dox. In contrast, pLV-CC-mCherry belongs to part of the original first generation of Tet-Off Systems, which is tTA (tetracycline-controlled transactivator)-
dependent to promote expression. The result demonstrated that pLV-CC-mCherry has approximately 7-fold higher basal expression than that of plix-CC-mCherry (Figure II-1B). New lentiviral plasmid plix-403 offered a significant improvement over the original first generation doxycycline-inducible system, such as previous pLV, with significantly reduced basal expression, mitigating the adverse effect from potential genetic perturbations.

Figure II-1. (A) Illustration of plix-CC-mCherry and pLV-CC-mCherry constructs. P_{Tight-TRE}, Tet-responsive tight promoter as 2nd generation doxycycline-inducible version, consisting of seven tet operator sequences followed by the minimal CMV promoter. P_{TRE}, original Tet-responsive promoter. CC, C-terminal fragment of mCerulean, 155-238aa. hPGK, human phosphoglycerate kinase 1 promoter. IRES, internal Ribosome Entry Site, allows for initiation of translation from an internal region of the mRNA. PURO, puromycin-resistant element. T2A, self-cleaving 2A peptides. (B) Comparison of basal fluorescent signals between plix-CC-mCherry and pLV-CC-mCherry in HEK cells. Images were taken by Zeiss confocal microscope at x20 objective. mCh, channel mCherry. TL, channel transmitted light. Scale bar = 50 μm. (Jonathan Reboulet, unpublished data)
Moreover, in new plix-403 plasmid, hPGK promotor confers an independent moderate expression of PURO, which is favorable to select and maintain the established stable cell line without Dox, eliminating the further mutual influence with puromycin during a long-term culture. Another advantage of plix-403 is using a 3rd-generation vectors for lentivirus packaging system, in which including a chimeric 5’LTR removes the requirement for the HIV Tat protein, thus decreasing the probability of creating replication-competent lentivirus in your target cells with a lower biosafety risk.

Accordingly, we decided to replace the pLV- with plix-403-based cell library preparation. Therefore, to know the plix-403 performance at different concentration of Dox is important when using this Tet-On system. Concentration gradient experiment was successful to prove the sensitivity and rigorousness of plix-based system (Figure II-2). Henceforward, the transfected gene expression can be easily adjusted by Dox concentration as different test conditions need.

Figure II-2. Testing Tet-On system using plix-CC-mCherry. HEK cells were transfected by plix-CC-mCherry and incubated with complete media containing 0 to 1000 ng/mL Dox for 18 h. Images were taken by Zeiss confocal microscope at x20 objective. Scale bar = 50 μm.
The plix-based plasmid, concerning BiFC, can be used for both VN-bait and CC-prey proteins. For optimizing the final BiFC-positive cell sorting and collection, reporter fluorescent protein mCherry was chosen as transfection marker. The proof-of-principle test was carried out on VN-fused HOXA9 bait protein (Figure II-3A). The plix-VN-HOXA9-mCherry plasmid was first tested by transfection with plix-CC-PBX1, which showed a good colocalization profile for BiFC fluorescence signals and mCherry, indicating all the VN-HOXA9/CC-PBX1 reassembly co-occurred with expression of mCherry marker (Figure II-3B). Furthermore, the attempting stable cell line harboring endogenous mCherry marker was established by plix-VN-HOXA9-mCherry. The single CC-PBX1 prey was tested by transfection and the BiFC assay displayed clear and pertinent protein interaction loci as with double transfection system (Figure II-3C), implying the feasibility of future transfection-free BiFC screening (BleoR is secondary antibiotic selection marker, working independently of PURO).
Figure II-3. Construction and performance of plix-VN-HOXA9-mCherry. (A) Illustration of plix-VN-HOXA9-hPGK-BleoR-mCherry. BleoR, confers resistance to bleomycin, phleomycin, and Zeocin, (B) BiFC performance test of plix-VN-HOXA9-mCherry cotransfected with plix-CC-PBX1. The test was executed in live HEK cells at Dox concentration of 100ng/mL. Images were taken by Zeiss confocal microscope at x10 objective. Scale bar = 100 μm. (C) BiFC test in plix-VN-HOXA9-mCherry stable cell line. One plasmid transfection was performed by plix-CC-PBX1 in live cells at Dox concentration of 100ng/ml. Images were taken by Zeiss confocal microscope at x20 objective. Scale bar = 50 μm.

3.2.3.2 Lentivirus functional titration

Functional titers measure how many viral particles can infect your target cells, which is critical for subsequent one-ORF-copy-to-one-cell infection. The popular FACS-based titration is only available when viral vectors carrying fluorescent markers. Consequently, our previous pLV-CC-ORF lentiviral library titer was measured by antibiotic-dependent colony counting method. However, it may underestimate viral titer and can not handle the multiple integration events. In addition, the experiments are time-consuming and prone to large inaccuracy. As such, the most accurate method of titration, qPCR, was applied to the plix-CC-ORF lentiviral library. For example of functional titration by qPCR, the result from one newly made plix-CC-ORF pooled lentivirus was shown in Figure II-4. The absolute copy numbers of WPRE and RPPH1 were calculated based on standard curve (Figure II-4A), in a range of virus dilutions (from 0ul to 80ul). The final functional titer can vary from different diluted conditions (TU, transduction units; from 5.48 X 10^6 to 1.23 X 10^7 TU/mL) (Figure II-4B, right). The acceptable titer should meet these criteria: the calculated titer should be similar between different DNA quantities in the same dilution condition; the ratio of maximal and minimal titer should not exceed five. Either max or min titer could be used for subsequent stable cell line establishing procedure, but the MOI (Multiplicity of Infection) and cell number may be adjusted according to your final titer choosing. Moreover, the virus copy number per cell is another important information, which could be obtained by this qPCR titration method (Figure II-4B, left).
**Figure II-4. Functional titration of plix-CC-ORF viral library by qPCR.** (A) Standard curve of WPRE and RPPH1. WPRE, Woodchuck Hepatitis Virus (WHV) Posttranscriptional Regulatory Element, as marker gene of lentivirus for copy number determination. RPPH1, the RNA component of the RNase P ribonucleoprotein, a gene that exists as a single copy per haploid genome (or 2 copies per human cell). (B) Quantification of lentiviral copy number per cell and functional titer. The values in grey cells are discarded, due to big discrepancies.

### 3.2.3.3 Custom sub-library: hORFeome v8.1-based human transcription factor library

Complete sets of cloned protein-encoding open reading frames (ORFs), as known as ORFeomes, are essential tools for large-scale proteomics and systems biology studies. Human ORFeome clone collection is currently the largest publicly available resource of full-length human ORFs, which is created by the OC (http://www.orfeomecollaboration.org/). To achieve a nearly entire set of ORFs, expansion of hORFeome library is still ongoing, for example the latest version v9.1 containing 17,408 protein-coding genes (Luck et al., 2020). Likewise, a recent update was made for our previous hORFeome stock, which was now replaced by hORFeome v8.1 purchased from BioCat GmbH. This collection represents almost 12,000 unique genes, and NGS-confirmed ORFs as a set of Gateway Entry clones ready for transfer to Gateway-compatible
expression vectors, assigning high accuracy and quality to this bacterial glycerol stock. Based on this new hORFeome, a sublibrary comprising all known human TFs was made by a multi-functional robotic platform. There are two reasons to undertake this sublibrary. First, TFs are of the most interest in biological community as potential research subjects of various gene function-related studies. The screening dedicated to known TFs is highly needed. Second, the new protocols of cloning and pooled plasmid DNA production were used, which will further improve the ORF representation in final pooled library and enable a less biased deep sequencing.

The known TFs were manually curated, referring to two previous studies (Chawla et al., 2013; Lambert et al., 2018). 1837 unique TF ORFs were cherry-picked from whole hORFeome v8.1 glycerol stock by high-throughput robotic liquid handling (Figure II-5A). For the sake of more efficient pooled transformation, new electro-competent cells were used, which fulfilled transformant colony number at least 1000x greater than the number of constructs in the pooled library (Figure II-5B).
Figure II-5. (A) Multi-functional robotic platform, FREEDOM EVO 200. The largest workstation in the Freedom EVO series, it provides an extensive work area and variable configurations, with a choice of liquid handling and robotic arms. (B) New transformation performance, illustrated by 268 TFs-minipool transformation. 1ul of deactivated Gateway LR reaction used in new competent cells according to the manufacturer’s instructions. The transformant was plated on a square Bioassay dish (left). The resulting colony number was calculated by petri dish colony counting, on which contains 1/1000 of original transformant. The representation was calculated attaining outperformed X3881.
3.2.4 Discussion

Apart from some of the representative results that were given above, there are still other optimizations addressing our BiFC screening strategy. For instance, the illumina NextSeq 500 system was used in our ORF deconvolution instead of ever Ion Torrent PGM sequencing, which enables a highly-multiplexing barcode sequencing with a much greater depth. Benefited from this new NGS platform, our new plix-CC-ORF pooled plasmid was sequenced in a full plasmid manner, as it was previously reported that this strategy yielded smaller coverage variability (Figure II-6) (Yang et al., 2011). Furthermore, to alleviate the side effect of cell death during FACS that largely affects the sorted population purity and limits the cell sorting speed, one reagent, named CellCover, is undergoing testing. It claimed that this reagent is compatible with all cell fixation-based downstream applications, without chemical crosslinking, including FACS and (single cell-based) NGS, protein sequencing, and more. However, it is worth noting that not one single method assesses all the specific and true positive PPIs, thus application of orthogonal methods is highly important to chart a more accurate PPI readout. For this purpose, co-IP, as the gold standard assay for PPIs, is considered to be performed as a complementary method for HT-BiFC putative candidates. Owing to recent technical advances, Jess automates traditional Western blotting while maximizing multiplexing with multiple detection channels, also compatible for a mid-throughput co-IP assay. Truly, automation of protein separation and immune-detection eliminates many of the tedious, error-prone steps of traditional blotting that limit data quality. This system is recently available in our region (Protein Science Facility, SFR Biosciences, Lyon), which is worth trying in the next step.
Figure II-6. Pilot experiments to optimize pooling strategy for next generation sequencing of ORF clones. (A) Schematic of pilot. Two conditions were tested. Reaction 1 consisted of 142 unmanipulated ORF plasmids. Reactions 2-3 were duplicate reactions in which ORF inserts from 376 ORF plasmids were enzymatically purified from plasmid backbones. (B) Coverage across the length of 2 representative ORFs. Adapted from Supplementary figure 1 in (Yang et al., 2011).

Reflecting on the optimized screening and to-do list, collectively, a highly-improved HT-BiFC screening method is expected in the near future, which will undoubtedly contribute to the endeavor of promising rigorous and robotic multi-omics study.
Part 4. Development of a GFP nanobody-directed Proximity Biotinylation Assay based on the Bimolecular Fluorescence Complementation (BiFC) in living cells

Yunlong Jia¹, Jonathan Reboulet¹, Françoise Bleicher¹, Nawal Hajj Sleiman¹, and Samir Merabet¹

¹ Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, Université Lyon 1, CNRS, Ecole Normale Supérieure de Lyon, 46, allée d’Italie 69364 Lyon cedex 07 France
4.1 Introduction

Protein-protein interactions (PPIs) are crucial for most cellular functions and typically participate in various biological processes. A large number of methods have been developed for PPI detection, enabling a deep and mechanistic understanding of cellular functions in different biological contexts. Currently, there are two types of approaches existing for experimentally identifying the target protein-associated interaction partners or interactome: binary interaction mapping, such as yeast-2-hybrid (Y2H) or Cell-PCA (described in Part 3.2, Chapter II), and protein-complex mapping, which aims to capture a snapshot of target protein-associated multimeric complex and identify the set of belonging proteins.

Indeed, along with the continuous efforts on nearly complete ‘ORFeome’ collections of ‘ready-to-be-expressed’ human open reading frames (ORFs) (Lamesch et al., 2007; Luck et al., 2020; Rual et al., 2004; Yang et al., 2011), the binary interaction mapping methods have achieved great progress in an efficient high-throughput fashion. Nevertheless, the binary PPIs are commonly assessed in a synthetic genetic condition that exogenous tag-fused proteins may perform differently to their wild-types considering physiochemical properties, abundance, and subcellular localizations. As a consequence, false positive candidates are inevitable. Moreover, the binary detection method loses associated binding partners that belong to the same protein complex, like molecular chaperones, which is crucial for protein network analysis and potential cellular function prediction (Roux et al., 2012). Similarly, the conventional protein-complex mapping method, such as AP-MS, suffers loss of candidates through protein insolubility and transient or weak interactions, resulting in an incomplete investigation of protein interactome.

Due to these limitations, the recent development of several proximity-dependent labeling (PL) methods in living cells partially addresses these deficiencies. These methods take advantage of promiscuous enzymes that are fused to the target protein and can directly label all proximal endogenous proteins, including transient and weak partners, with a covalent biotin tag. After the labeling reaction, cells are lysed and the biotinylated proteins are subsequently pulled down with streptavidin beads, allowing harsher wash conditions
to be employed, which reduces non-specific binding. The resulting purified biotinylated proteins are then identified by mass spectrometry (Ummethum and Hamperl, 2020). To date, there are three major enzymes that are used for proximity labeling: biotin ligase (BioID (Roux et al., 2012), BioID2 (Kim et al., 2016), BASU (Ramanathan et al., 2018), miniTurbo (Branon et al., 2018a), TurboID (Branon et al., 2018a), horseradish peroxidase (HRP) (Kotani et al., 2008), and engineered ascorbate peroxidase (APEX (Martell et al., 2012), APEX2 (Lam et al., 2015)). Amongst them, the BioID labelling is simple and non-toxic, as the most used PL enzyme, since its advent. More recently, TurboID, as a directed-evolution variant of BioID, was reported to be much greater efficient than BioID, and enabled 10-minute PL in cells instead of previous 18-24h labeling (Branon et al., 2018a). In addition, to increase the versatility of TurboID, split forms of TurboID (split-TurboID) have been developed and used for contact-dependent PL in cells (Cho et al., 2020). This conditional proteomics method could enable not only to probe the proteomic composition in a spatiality-specific locus, such as ER–mitochondria contact sites, but also the labelling of additional vicinal proteins belonging to the corresponding binary protein complex, like split-BioID (Schopp et al., 2017), as one of PCA-based PL methods. The powerful split-TurboID extends the PL method from single bait protein-based to a binary protein complex-based proteomic approach. However, the utilization of split-TurboID was impeded by two main limitations. The one is the activity loss of reconstituted split-TurboID, which is relatively time-consuming and less sensitive compared to full-length TurboID that needed only 1min for specific proximity labeling in control test (Cho et al., 2020). On the other hand, the split-TurboID (or split-PL enzymes) is not possible to monitor and visualize the putative binary PPI when the concomitant labelling occurs in a native cellular environment, which is crucial to deliver the specific and defined protein complex-dependent labeling for a conditional proteomics method.

In this study, we provide a novel BiFC assay integrated with the proximity-dependent biotinylation technique TurboID, called BibID (or BiFC-based TurboID). Here two non-fluorescent fragments are targeted to two known interacting proteins respectively, like classical BiFC assay. To achieve a binary PPI-specific PL, GFP-nanobody was recruited for recognizing and binding GFP-derived FPs and relevant reconstituted BiFC tags. When
engineered to fuse with TurboID enzyme, upon the interaction of two target proteins, the GFP nanobody will target the reassembled BiFC tags and drive TurboID to execute an efficient BiFC-specific PL for additional vicinal proteins surrounding the binary complex in live cells. We assessed this system using a well-studied PPI, HOXA9/PBX1 interaction, as proof-of-principle. As a multi-modular system, the TurboID-based PL was first performed in HEK cells, and then the TurboID was fused to GFP nanobody, finally enabling the GFP-like protein targeted PL in context of BiFC assay. Aiming to verify the applicability of BibID, the HOXA9/PBX1-specific PL and a series of control tests were set up to demonstrate the restricted biotinylation activity of BibID on only defined binary protein complexes. Collectively, our BibID system is a highly specific approach that allows in a single and simple assay to both experimentally visualize binary PPI and to unbiasedly study additional interacting factors, taking full advantage of the TurboID system.
4.2 Materials and Methods

4.2.1 Cell culture

HEK-293T cells were purchased from European Collection of Authenticated Cell Cultures (ECACC) through the biological resource center CelluloNet (AniRA platform of Lyon). Cells (passage number < 20) were cultured as a monolayer in Dulbecco’s modified Eagle’s medium (DMEM-GlutaMAX-I, Gibco by Life Technologies) supplemented with 10% (v/v) heat inactivated fetal bovine serum (FBS) and 1% (v/v) Penicillin-Streptomycin (5,000U penicillin and 5mg streptomycin/mL), incubating at 37°C, in an atmosphere of 5 % CO2.

4.2.2 Construction of plasmids

Genetic constructs used in this study, were listed in Table V1, Chapter V, with construct designs, linkers, epitope tags, and functional inserts. For cloning, PCR fragments were amplified using Q5 polymerase (New England BioLabs). All restriction enzymes were purchased from New England Biolabs. The vectors were double-digested using standard enzymatic restriction digestion and ligated to gel purified PCR products by T4 DNA ligation (Promega, France). Ligated plasmid products were introduced by heat shock transformation into one shot Top10 competent cells (Invitrogen, Cat No. C404010). DNA sequencing was carried out at GENEWIZ Company (Germany).

4.2.3 Immunofluorescence

For transfection, 50k cells were seeded on glass coverslips in 24-well plates. Twenty-four hours after plating, cells were transfected with jetPRIME (Polyplus, Ref 114-15) following manufacturer’s instructions. 0.5ug of each plasmid DNA were transfected per well, according to different test conditions. After 14 hours of incubation in the presence of doxycycline (50ng/ml final), replace old medium with fresh complete medium without Dox and return to incubator, then prepare the new medium containing 50uM biotin. 7hr after incubation, replace the medium for biotin-present condition wells and incubate cells with biotin-contained medium for 10min at 37°C.
For immunostaining, the biotin-treated cells were subsequently washed 2 times by 1X PBX and fixed with 4% (v/v) paraformaldehyde in PBS at 4°C for 15 min. Cells were then washed three times with PBS and permeabilized with cold methanol at −20°C for 5 min. Cells were then washed three times with PBS, and then incubated with primary antibody in PBS supplemented with 3% (w/v) BSA for 2 hr at RT. After washing three times with PBS, cells were then incubated with Anti-Mouse-Alexa Fluor 647 for HA-tag, and Alexa Fluor 555 Streptavidin for biotinylated proteins, in PBS supplemented with 3% (w/v) BSA for 1 hr at RT. Cells were then washed three times with PBS. Finally, cell-coated coverslips were mounted in VECTASHIELD Antifade Mounting Medium with DAPI (VECTOR, Cat No. LS-J1033-10), and imaged by confocal microscopy (Zeiss LSM780).
4.3 Results

4.3.1 TurboID-mediated protein biotinylation in mammalian cells

As the biotin ligase functions as a key component in our BibID system, we first tested the ability of TurboID to catalyze HOXA9-specific biotinylation in HEK cells. To enable an adjustable ligase expression level, the doxycycline-inducible Tet-On system was used in previously described plix-403 vector to drive the HOXA9-fused TurboID expression (HOXA9-TurboID). Meanwhile, two controls were designed for evaluating the specificity of HOXA9-TurboID-based biotinylation, including nuclear localization sequence fused TurboID (NLS-TurboID), as spacial control that imports TurboID into the cell nucleus by nuclear transport (Mahato et al., 1999), for patterning the PL reaction of non-specific proteins in nucleus, and TurboID only, as system control to remove the general background. Moreover, all TurboID-containing constructs were finally HA-tagged, which facilitates the visualization of TurboID expression in different conditions. Biotin concentration and incubation time are two crucial factors that affect biotin ligase efficiency in cultured cells (Branon et al., 2018b; Kim and Roux, 2016; Roux et al., 2012). As such, the corresponding TurboID-containing plasmids were transfected to HEK cells and then 10min of labelling was initiated with the addition of 50 μM exogenous biotin in complete culture medium. We kept this empirical setting, which was reported and well performed in HEK cells (Branon et al., 2018b), for our preliminary test, though case-by-case variance and further condition optimization should be concerned. The TurboID-derived biotinylation was eventually stained by streptavidin-AlexaFluor555 (Strep-A555) and observed by confocal microscopy. In comparison between biotin-absent (Biotin -) and biotin-present (Biotin +) conditions (Figure II-7A vs. 7B), the TurboID-based labeling exhibits a highly biotin-dependent profile. HOXA9-TurboID and TurboID test conditions show strong biotinylation activity in their corresponding cell compartments, colocalizing with TurboID ligases, as expected (Strep-A555 and α-HA channels, Figure II-7B). Interestingly, the NLS-TurboID expression is diffused to cytoplasm, irrespective of its nucleus-specific property, albeit a little more nucleus-concentrated (α-HA channel, Figure II-7A and 7B). In contrast, the NLS-TurboID-based PL still keeps the nucleus restriction, indicating a
suitable special control for nucleus-localized HOXA9-PL (Strep-A555 channel, Figure II-7B). We hypothesize that under 50uM biotin treatment, NLS-TurboID consumes the biotin to label the proteins in nucleus more effectively than those in cytoplasm, due to the higher intensity of NLS-TurboID and nucleus-localized proteins, as evidence that the cytoplasmic PL is still dimly visible. However, our result shows that the TurboID-based proximity system can be successfully used in specific PL of target protein in living cells.
Figure II-7. Immunofluorescence of TurboID-mediated proximal biotinylation in HEK cells. Cells were fixed and stained with streptavidin-AlexaFluor555 (Strep-A555) to detect biotinylated proteins, and anti-HA.
antibody (α-HA) to detect ligase expression. Nuclear DNA was labelled with DAPI stain. Non transfected HEK cells as technical control. Scale bar, 20 μm. (A) Biotin-absent PL test. (B) Biotin-present PL test. Exogenous 50uM biotin was prepared in complete culture medium and cells were treated for 10min before fixation.

4.3.2 Development and performance of BibID method

Given the target protein-specific PL of TurboID, we conceived a BiFC-based TurboID system, BibID, which enabled the binary protein complex-specific PL in living cells. This method takes advantage of both GFP-nanobody and TurboID, schematically represented in Figure II-8. The nanobody was first reported in 1993, as a specific class of light chain-deleted antibodies in camels (Hamers-Casterman et al., 1993). A nanobody is strictly monomeric, highly stable, and generally smaller, thus it is far more efficient than classical antibodies. Besides, as GFP is the most important genetic marker for biological research, several different GFP nanobodies (GBP) have been developed for targeting and binding GFP or its variants (Fridy et al., 2014; Kubala et al., 2010; Twair et al., 2014). More recently, conditionally stable GFP-binding nanobody (csGBP) was described, which enabled detection of less noisy GFP-tagged proteins, notably, along with their analysis showing that unbound csGBP was efficiently degraded by the proteasome (Ariotti et al., 2018). By combining the csGBP with TurboID, the BibID system facilitates an improved signal-to-noise (S/N) ratio for a specific PL detection. Furthermore, csGBP straddles the split site in commonly used BiFC pairs (Kubala et al., 2010), such as the one recruited in our system, CC155/VN173. The folding induced by resulting BiFC is absolutely required for csGBP binding, leading that recognition of the unfolded halves by csGBP will not occur theoretically.
Figure II-8. Schematic representation of BibID system. The binding between csGBP and folded CC/VN stabilises TurboID-csGBP, which leads to proximity biotinylation around the CC/VN-tagged POIs, here referred as A/B protein complex. The unbound TurboID-csGBP will be degraded by the ubiquitin proteasome system, which would largely alleviate non-specific labelling when csGBP-CC/VN binding saturates. The biotin labelled proteins can be subsequently isolated by biotin affinity purification and identified by LC-MS/MS analysis. HA, HA epitope tag. csGBP, conditionally stable GFP-binding nanobody. CC, C-terminal fragment of mCerulean (155-238aa). VN, N-terminal fragment of mVenus (1-172aa).

As proof-of-concept, N-terminal split mVenues (VN173) and C-terminal split mCerulean were used to construct BiFC system with one well-studied PPI, HOXA9/PBX1, generating VN-HOXA9 and CC-PBX1. Meanwhile, the modular TurboID-csGBP fusion was made and tagged by HA-tag. To validate the system efficiency, a preliminary immunofluorescence test was carried out in HEK cells by transfection comparing with several control settings (Figure II-9). The EGFP was coexpressed with TurboID-csGBP as system positive control. VN-HOXA9 and CC-PBX1 were exchanged in reciprocal experiment, cotransfected with TurboID-csGBP, as two negative controls of csGBP. One
system negative control was also designed to monitor the background of PL, using only TurboID-csGBP alone transfection. The complete BibID system utilized a 3-plasmid transfection, similar to previous single TurboID test, showing a highly biotin-dependent PL profile (Strep-A555 channel, Figure II-9B). The resulting PL was observed for tight association of VN-HOXA9/CC-PBX1 interaction or EGFP expression, as opposed to just one-half of the interaction pair, such as VN-HOXA9 or CC-PBX1 only (Strep-A555 vs. BiFC/EGFP, Figure II-9B). Interestingly, in the absence of folded CC/VN or EGFP, the TurboID-scGBP expression was still generally visible in the control conditions regardless of biotin addition. These background-like weak HA signals are likely caused by TurboID-csGBP en route to proteasomal degradation. Plus, it is noted that the biotin-present condition exhibited a stronger TurboID-csGBP signal than that in the biotin-absent condition. It seemed that the biotin favours to stabilize the free unbound TurboID-csGBP, especially for CC-PBX1 and VN-HOXA9 control conditions, which have a clear detectable HA signal, implying the potential non-specific binding between csGBP and one of the split halves (CC or VN) (α-HA channel, Figure II-9B). If so, surprisingly enough, only subtle PL can be detected in these negative controls, further justifying our former assumption, regarded as one of biotin side effects. proteasomes. To sum up, these results clearly demonstrate that GFP or GFP-derived BiFC can be effectively visualized and proximity-labeled using the BibID system at the presence of biotin in a native cell context.
**Figure II-9. Immunofluorescence of BibID test performed in HEK cells.** Cells were fixed and stained with streptavidin-AlexaFluor555 (Strep-A555) to detect biotinylated proteins, and anti-HA antibody (α-HA) to detect ligase expression. BiFC/EGFP belongs to endogenous fluorescence. Nuclear DNA was labelled with DAPI stain. Non transfected HEK cells as technical control. Scale bar, 20 μm. (A) Biotin-absent PL test. (B) Biotin-present PL test. Exogenous 50uM biotin was prepared in complete culture medium and cells were treated for 10min before fixation.

Concerning the relative low efficiency of 3-plasmid transfection, we next substituted the BibID system with 2-plasmid system, leveraging a powerful bidirectional vector with Tet-responsive bidirectional promoter (Bi-P_{TRE}) (**Figure II-10A**). In this new system, VN-HOXA9 and CC-PBX1 were assembled into the same plasmid, and cotransfected with a 2nd plasmid TurboID-csGBP in practice of BibID using. To assess its performance, the immunostaining test was similarly made in HEK cells, with or without biotin (**Figure II-10B, 10C**). As expected, in the presence of biotin, the BiFC-specific PL was obtained by 2-plasmid transfection, possessing the equivalent or better performance than previously. This improvement is not only facilitating the transfection, but also benefit the further stable cell line making for true proteomics analysis.
Figure II-10. Design of two-plasmid BibID system and performance test. (A) Illustration of 2-plasmid BibID system construction. All constructs were controlled by Tet-On system, driving the gene expression in presence of Dox. PTRE, Tet-responsive bidirectional promoter. Bi-PTRE, Tet-responsive bidirectional promoter. (B, C) In immunofluorescence test, after transfection, cells were fixed and stained with streptavidin-AlexaFluor555 (Strep-A555) to detect biotinylated proteins, and anti-HA antibody (α-HA) to detect ligase expression. BiFC/EGFP belongs to endogenous fluorescence. Nuclear DNA was labelled with
DAPI stain. Non transfected HEK cells as technical control. Scale bar, 20 μm. (B) Biotin-absent PL test. (C) Biotin-present PL test. Exogenous 50uM biotin was prepared in complete culture medium and cells were treated for 10min before fixation.

4.4 Discussion

Biotin-based PL is a unique method to screen for physiologically relevant protein interactions that occur in living cells. TurboID is the most active and outperformed other ligases, which makes possible the study of dynamic or transient processes that occur on the timescale of minutes or even a few hours. Overall, we have developed a BibID system that can be used for spatially binary protein complex-specific PL in cells, despite additional validation, such as western-blotting, being ongoing. For rapidity and conveniency, we used only empirical biotin concentration and labeling time in our preliminary test. The further improvement should be drawn in subsequent blotting and LC-MS/MS analysis, upon the abundance of each protein across samples. Our system encompasses 3 to-express-elements, all driven under the Tet-On system, which makes the optimization procedure more complicated than traditional BioID. Three factors can affect the final biotinylated protein output, including labeling time, TurboID-csGBP expression level (transfected plasmids and dox concentration), and exogenous biotin concentration. Univariate experiments should be designed for each of four possible variables.

During the next proteomics assay, transfected or stably gene-expressing cells will be lysed and biotinylated proteins enriched with streptavidin beads. After on-bead digestion of proteins to peptides, the peptides will be labeled with TMT (tandem mass tag) labels, enabling quantify relative abundance of each protein across samples. LC-MS/MS analysis of pooled peptides will generate the final true positive and false positive protein list for final validation. Our local PSF (Proteomics on Protein Science Facility), provides a large panel of proteomics analysis as the detection of low abundant proteins, relative and targeted quantitative approaches, using Q Exactive HF and Ion Mobility Spectrometry device, will further collaborate with our lab for BibID-based proteome test.
However, robust elucidation of interacting proteins, including not only strong direct protein-protein interactions, but also weak, transient or indirect interactions is challenging. Either BiFC-screening or BibID method could be a response to this issue. Single approaches can supply only limited information about the key parameters of interactions or complexes. Therefore, one day, it will be promising to combine these two alternative approaches, to generate a context-specific and comprehensive PPI interaction data.
Chapter III. General Discussion and Perspectives
3.1 BiFC approach: direct or indirect interaction interrogation?

Pairs of proteins can form either direct contacts or indirect interactions (Luck et al., 2017). Contrary to Y2H based on direct binary PPIs, the BiFC was considered that fluorescence complementation occurs simply when the two FP fragments are reconstituted, and two POIs are part of a protein complex, not necessary to directly interact with each other (Kerppola, 2006b). Reflecting on the target protein overexpression and the FP fragment-tag introduction, we have several points to argue against this explanation of BiFC, giving rise to the question “In BiFC assay, the binary PPIs are direct or indirect interactions?” or “Which type of binary PPI is dominant in BiFC assay, direct or indirect?”

Prior to answering this question, it is important to clarify the different layers of macromolecular complex characteristics, which can be investigated at 5 levels (Figure III-1): (A) composition, (B) stoichiometry, (C) copy number, (D) topology, and (E) dynamics.

**Figure III-1. Different layers of macromolecular complex characteristics.** Complex composition (A) encompasses the list of complex constituents. Complex stoichiometry (B) defines the abundance of constituents relative to each other, while the copy number (C) quantifies the absolute number of constituents per complex. Complex topology (D) describes the spatial setup of the constituents in the complex. Complex dynamics (E) represent the alterations in composition, stoichiometry, and topology over time. Reprinted from figure 1 in (Wohlgemuth et al., 2015).

In BiFC, given the widely used transfection, the overexpression of target proteins is still a key concern of background signal, despite that optimization of plasmid amount and microscopy parameter can abate this side effect, minimizing the non-specific background signal. As previously mentioned (Kerppola, 2006b), many interactions can be visualized when the fusion proteins are expressed at concentrations comparable to their endogenous...
counterparts. Therefore, this problem can be alleviated by expressing the fusion proteins at concentrations approximating their endogenous counterparts.

However, because of the simplicity and convenience of BiFC, many experiments were performed by routine transfection, with strong enough expression regardless of their endogenous expression level. This is not surprising due to the following reasons. First, limited information of protein isoform abundance in given cellular context, thus hindering the antibody selection for the most representative protein blotting. Second, the unavailability of antibodies for certain special proteins, resulting from the binding specificity and technical problem. The last is that not all tested proteins exist in a given cell line. Sometimes, we want to only confirm a new interaction with a candidate protein regardless of the endogenous status. Therefore, in most cases, people prefer to validate only the expression and colocalization of fusion proteins instead of their endogenous counterparts, potentially resulting in artificial/non physiologically relevant expression levels during the BiFC assay.

Besides overexpression, fusion proteins, generally, possess different physiological properties from wildtype proteins, such as the non-native folding and increased protein size, possibly influencing intrinsic dynamics and topology of related protein complex. For example, the binding kinetics of manipulated fusion proteins should be much slower than that of endogenous proteins when interacting with additional cofactors or components that belong to the same protein complex. Consequently, the exogenous overexpressed proteins are prone to interact with each other, forming direct binary contacts instead of indirect PPIs. The overexpressed manipulated proteins, in part, are redundant to well-organized inner microenvironments. As competitors of endogenous proteins for restricted and relatively low-expressed complex components, which have potential to form multimeric protein complex with, the exogenous fusion protein is at a binding disadvantage in this competition and hard to interact with additional limited factors, even of small part, would not enough overrule the main binary direct PPI signals.

Additionally, thinking about the stoichiometry of macromolecular complexes, in BiFC and other PPI individual tests, such as Co-IP, the equal genetic expression of two target proteins
(1:1) were traditionally applied in practice. This arbitrary decision conferred the same mole fraction on all resulting target proteins in the complex’s constituents, which is regarded as an extremely rare and ideal simulated situation in a given native context.

In our experiments, we occasionally noticed that stoichiometry could regulate HOX/cofactor subcellular localization in living cells. Indeed, it is infeasible to optimize the mole ratio prior to manipulation, due to the lack of information of the given protein complex in the specific biological context. As such, manipulation of two proteins with comparable expression is a reasonable compromise in protein functional and interaction study.

In our retrospective analysis, fortuitously, it was observed that dose-dependent effect on subcellular localization of HOXA9/PBX1 interaction in living cells, demonstrating the importance of the stoichiometry of the protein complex in PPI detection (Figure III-2).
Figure III-2. Hypothesis model and preliminary results of dose-dependent effect in binary PPI. (A) Model of Dose effect in binary PPI. (Upper) Protein A and B as putative interaction partners. Depending on the stoichiometry of A/B complex, the status of A/B abundance ratio was defined as “Balance”, “Moderate imbalance” and “Severe imbalance”. The resulting phenotypes was assigned to different numbers: “1” indicates that the A/B complex was localized in nucleus, upon their interaction. “2” refers to a transitional change of A/B complex localization from nucleus to cytoplasm. “3” means a cytoplasmic localization of A/B complex. The letter size represents the relative abundance comparing to A or B. The bigger size, the higher relative expression in cells. (Lower) For example, in “Balance” range, HOX/cofactor interaction occurs in nucleus, as phenotype “1”. Along with the A/B abundance ratio changing, the corresponding phenotypes will turn: “2” to “3”. (B) The BiFC test in VN-HOXA9 stable cell line. CC-PBX1 was transiently expressed in cells by transfection, under control of Dox concentration. The expression of CC-PBX1 increased in parallel with Dox increasing, resulting in “Moderate imbalance” between HOXA9 and PBX1. A represents VN-HOXA9, B refers to CC-PBX1. (C) BiFC assay for interaction of VN-HOXA9 and CC-PBX1. Expression of VN-HOXA9 was controlled by Dox concentration. CC-PBX1 and VN-HOXA9 was coexpressed in HEK cells by transfection. The BiFC under different Dox concentrations generated three distinct phenotypes, according with dose-dependent model.
3.2 Limitation and future directions of current large-scale BiFC screening

3.2.1 Lack of rigorous negative control

The random collisions of two halves of reporter protein fragments are the main source of BiFC background signals, as well as other PCA-based methods. Thus, in a high-throughput BiFC scan, the negative control is important to non-specific signal removal and subsequent candidate validation. Unlike the loss-of-function screens that the paired negative controls are well defined and integrated in the whole genetic library, such as gRNA or shRNA library, the BiFC negative control is more flexible and very study-dependent. Even for individual BiFC, a rigorous control is not always available, not to speak of a genome-wide BiFC screening. Though sometimes the gain-of-function or overexpression screening shares the same hORFeome library with BiFC screening, the control using function-free and fluorescent proteins are not compatible with BiFC. To address this point, four types of negative controls are widely recruited in the BiFC screening: (a) N- and C-terminal FP fragments without fusion to the POI; (b) N-terminal FP fragment fused to POI with the unfused C-terminal FP fragment; (c) C-terminal FP fragment fused to POI with the unfused N-terminal FP fragment. (d) One FP fragment fused to POI alone. As reported, at least one, or up to 3 of the above mentioned controls are used in BiFC screenings (Xia et al., 2018; Yue et al., 2017). Unfortunately, all of these controls fall into the inappropriate controls, as described in Section 3.3.4, Chapter I.

The splitFAST approach could constitute a promising alternative to FP-based BiFC (as described in Figure 19, Section 3.3.5, Chapter I). Indeed, splitFAST is reversible and can therefore be used for selecting interactions that will be specifically affected by a given inhibitor to test. In contrast to FP-based BiFC, where the inhibitor has to be mixed with the fusion proteins for a long time incubation before BiFC signal detection, splitFAST allows testing the inhibitor directly on the fluorescent complexes in real time (Figure III-3). This strategy could be suitable for a conventional genome-wide PPI screening (e.g. to know the potential range of action of a therapeutic molecule), and could also be amenable to conduct a TF functional domain screening with a known corresponding inhibitor, as well as drug screening for fitness genes within potential PPIs in human cell lines. Finally, given the
reversibility of splitFAST, it allows having access to the PPI dynamics, a key molecular parameter that conventional FP-based BiFC could not approach.

Figure III-3. Schematic of splitFAST-based BiFC screening. (A) Using the Cell-PCA strategy, C-terminal FAST (CFAST)-tagged prey cell library was coexpressed with N-terminal FAST (NFAST) bait protein after HMBR treatment (not shown in the figure). The GFP+ cells were sorted and collected after first round FACS. (B) The GFP+ positive cells will be treated with a given PPI inhibitor to reverse the positive protein interactions, leading to the loss of fluorescence. Only the PPI-independent non-specific signals exist in the cell population. Through the second cell sorting, the true interactor-containing cells with GFP- were sorted and collected. FAST, the fluorescence-activating and absorption shifting tag (Tebo and Gautier, 2019). HMBR, 4-hydroxy-3-methylbenzylidene rhodanine, which provides green-yellow fluorescence.

3.2.2 Multicolor BiFC screening

Growing concern regarding the rise in protein complex-specific interactome, the split-BioID was reported as a promising and powerful method to investigate the proximal binding partners of a given binary protein complex, as well as our new developed BibID approach. Likewise, multicolor BiFC analysis provides an effective assay to compare the subcellular distributions of protein complexes formed with different binding partners.
Furthermore, this method can be regarded as one single-scale method to verify the binary protein complex-specific cofactors. Although its utility in high-throughput manner is theoretically feasible, multicolor BiFC assay applied for high-throughput screen is as yet unpublished. Owing to the availability of Bi-PTRE, two baits can be easily cloned into the same plasmid, keeping the same system complexity as single bait screening. Here I propose a strategy of high-throughput multicolor BiFC screening, which will break the limit of conventional single target BiFC screening, empowering it with a protein complex-compatible screening, in addition to single bait protein screening (Figure III-4).

In principle, split-FP fragments will be used and derived from two different fluorophores, mVenus and mCerulean, with distinct spectra. Either VC155 or CC155 could achieve mVenus-like BiFC with VN173 fragment (see 4.3, Part 1, Chapter II). In addition, the CC155 fragment can complement with the CN173 for making mCerulean-like BiFC. This property allows visualizing two different PPIs simultaneously by doing mVenus- and mCerulean-like BiFC with three fusion proteins. The CN173/CC155/VN173 has proven to be the best combination for multicolor BiFC (Shyu et al., 2006). Based on these versatile FP fragments, VN-HOXA9/CN-PBX1 bait proteins were used in plasmid construction (Figure III-4A), and the rationale of high-throughput multicolor BiFC was detailed in Figure III-4B. Depending on the needs, the double baits are not necessary to be coexpressed. This system can easily be used as conventional single-bait BiFC screening, if keeping one of Bi-PTRE flanks empty. Collectively, proposed high-throughput multicolor BiFC screening enables two interactions to be examined simultaneously, facilitating the detection of binary protein complex-specific partners, as a cost-effective and time-saving method.
Figure III-4. The strategy of high-throughput multicolor BiFC, exemplified by HOXA9 and PBX1. (A) Schema of plasmid construction, illustrated by bait proteins, VN-HOXA9 and CN-PBX1. All constructs are controlled by the Tet-On system, driving the gene expression in presence of Dox. (Upper) Bi-P_{TRE} promotor conducts the tagged bait proteins (VN-HOXA9 and CN-PBX1) at the same time. mCherry element is used as an expression marker of target bait proteins, controlled by an independent hPGK promotor. (Lower) CC-tagged ORFeome (prey) plasmid library was made by Gateway cloning, as described previously. Ptight-TRE, Tet-responsive tight promoter as 2nd generation doxycycline-inducible version, consisting of seven tet operator sequences followed by the minimal CMV promotor. VN, the N-terminal fragments of Venus, 1-173aa. CN, N-terminal fragment of Cerulean, 1-173aa. CC, C-terminal fragment of mCerulean, 155-238aa. T2A, self-cleaving 2A peptides. (B) The red fluorescent cells represent the bait/prey co-expressed cell population, which can be further gated in mCherry+. Upon these mCherry+ cells, multiple gating can be used, dividing total mCherry+ cells into 4 different populations in 4 different quadrants. CC/VN, as GFP+/Cerulean- cell population, represents HOXA9 interacting CC-ORF candidates (residing in cells). VN/CC/VN, as GFP+/Cerulean+ cell population, indicates HOXA9/PBX1 complex interacting CC-ORF candidates. CC/CN, as GFP-/Cerulean+ cell population, refers to PBX1 interacting CC-ORF candidates. CC, as double negative cell population, no interaction between baits and prey library.
3.2.3 The tag effect

There are two main concerns, when applying a tag-based method. One is the biotechniques used for tagged fusion protein introduced into a live context, such as the classical methods, including transfection, transformation or virus-mediated transduction, and advanced approaches, for example, CRISPR-Cas9-mediated gene tagging (Lackner et al., 2015). Though the side effect of tag was largely improved, especially the steric hindrance of interactions caused by overexpression of tagged bait protein, on the other hand, the tag per se, can result in non-native folding of fusion proteins and further influencing the stoichiometry of the complex.

However, the split-FP fragment as tags of tested proteins, is the basis of BiFC principle. The tag is an integral part of the BiFC system, unlike the antibody-based methods such AP-MS or Co-IP, in which alternative tag-free procedure can be performed via endogenous protein-specific antibody. Given that large and bulky fragments impair BiFC fragment solubility and folding, consequently leading to high background signals (Y and Cd, 2010), several efforts have been addressed on the tag minimization, for instance, the micro-tagging system based on tripartite split-GFP (Cabantous et al., 2013) (See Section 3.4.1, Chapter I).

Concerning our BiFC system, generic and large fragments were used in a large-scale screen, which is potentially aggregation-prone and of high backgrounds from self-assembly. I should still emphasize here that (i) only 1-2% of cells were fluorescent in the HOX screen and (ii) different partners were found with the different HOX proteins, thus making the BiFC background of low incidence for capturing specific signals. In any case, to know whether other BiFC systems could be performed with a minimum background, I tested a tripartite split-GFP system, also known as TriFC (Tripartite Fluorescence Complementation) assay (Figure III-5A). As showcase, I generated sfGFP10-HOXA9 and sfGFP11-PBX1 functional fusions in a single pcDNA3 plasmid, with a ready-to-use detector plasmid containing sfGFP1-9 fragment fused to mCherry reporter (Figure III-5B). The test was performed in HEK cells by transfection. Each plasmid was individually transfected to check the background signals. Preliminary results showed that the TriFC

- 193 -
worked properly in a two-plasmid transfection system, comparable to our routine 3-plasmid system. The HOXA9/PBX1 interaction localized in the nucleus as expected (Figure III-5C). Only requiring attention is that the reconstituted GFP signal is seemingly much weaker than CC/VN refolded fluorescence, which needs a further evaluation.

Figure III-5. Schematic of TriFC and preliminary test. (A) Principle of TriFC. beta-strand 10 (GFP10) and beta-strand 10 (GFP11) are fused to bait A and prey B proteins, respectively and the detector fragment GFP1–9 is added separately. When protein interaction occurs, GFP10 and GFP11 are tethered and then
spontaneously associate with GFP1–9 fragment to form a full-length GFP. GFP10, 194–212aa. GFP11, 213–233aa. GFP1-9, 1–193aa. OPT, optimized. Adapted from figure 1 in (Cabantous et al., 2013). (B) Illustration of constructs for TriFC. The HOXA9 was fused to GFP11 by a linker, and labeled by HA tag used for further immunostaining test. GFP10 was fused to PBX1 at its N-terminus, coupled with V5 tag for next immunostaining. The functional cassettes were outlined by grey dashed boxes. Detector fragment GFP1-9 was fused to mCherry reporter. All constructs were cloned in pcdna3 plasmid and conducted by CMV promoter. (C) Preliminary test for TriFC. All plasmids were transiently expressed in HEK cells using Jetprime, as described previously. Each plasmid used in different conditions is 0.5ug. (Up panel) GFP11 and GFP10 were coexpressed in HEK cells, to check the background, when GFP1-9 is absent. pcDNA3-mCherry was cotransfected, using for transfection system control. (Middle panel) Single GFP1-9 was transfected to check the background in absence of GFP10 and GFP11. Only the GFP channel needs to be checked. (Bottom panel) complete TriFC system was tested by coexpressing 3 parts of split-GFP. Finally, the moderate GFP-positive signals were generated. Scale bar = 50 μm.

Furthermore, upon this verified advanced TriFC method, I present here its applicability in a high-throughput system. There is no difference for the tagged bait protein and prey library preparation (with a small GFP tag in both cases). One more step should be considered is how to introduce the big GFP1-9 detector fragment into the cells. On this point, making a GFP1-9 stable cell line will be best to subsequently be used to generate GFP10 or GFP11-tagged prey cell libraries. The resulting endogenous GFP1-9-expressed prey cell library will function similarly as that used in our previous Cell-PCA. Hopefully, this micro-tagged TriFC screening could minimize the unexpected protein interference and aggregation.

### 3.2.4 Limited quantitative information

Several screening methods can provide both a final candidate list and quantitative information according to the different selection conditions. For example, the detection of gene-independent cell fitness effects in CRISPR-CAS9 screening (Joung et al., 2017). This quantitative information can help to know the candidate relative abundance compared with the control, as well as their functional potency in a specific phenotype. In terms of PPI study, AP-MS method characterizes the target protein interactome, and the relative abundance of interaction candidates can also be obtained (which requires more materials),
giving the quantitative information. Through label-free quantification, the differentially enriched proteins can be defined under the tested conditions (Smaczniak et al., 2012).

The final readout of BiFC screening has also the quantitative information that represents the candidate ORF abundance. More exactly, this abundance deciphered the BiFC-positive cell number during the FACS procedure for a given ORF. It is restricted only for deciding the positive or negative status compared to the fixed BiFC-positive gate. Consequently, this ORF abundance can neither depict the bait/prey interaction intensity, nor the BiFC-positive signal change. As shown in Figure III-6, a microplate-based high-throughput BiFC screening was performed coupled with FACS and two ORF interaction profiles were illustrated. They have the same number of positive cells in red, defined by the same gate. The ORF abundance can be calculated as positive ratio (PR). Despite that Intensity of interactions between two different ORFs are quite different, they have the same abundance (PR) as 0.1. Besides, we can assume that two profiles belong to the same ORF. Left one is drug-treated, resulting in a great positive signal loss, but still bigger than the threshold as positive cells. Thus the quantitative information of BiFC screening is limited in that we can know whether a given ORF is positive or not.

In turn, the sequential gates can solve this problem to some degree. Several gates were set along with the signal strength, and count the cell number in different gates. The final calculated weighted positive ratio (WPR) revealed the interaction difference between ORFs (Figure III-6). This strategy was described only for individual BiFC signal evaluation. In BiFC screening, five gates represent five positive cell populations, assigning with PPI intensity degree from 1 to 5. The resulting ORFs will be intensity degree-specific candidates, which will be useful for further weak or strong PPI analysis. As example, this strategy has been used in a genome-scale ORFeome screen (Mengwasser, 2018). The whole-genome ORFeome library under TRE promoter control. The screen for PD-L1 responsiveness to IFNγ performed in the presence of 100ng/mL IFNγ. ORF expression was induced for 48 hours with doxycycline, stained the cells with a monoclonal antibody recognizing PD-L1, and FACS sorted the cells into PD-L1\textsuperscript{high}, PD-L1\textsuperscript{med}, and PD-L1\textsuperscript{low}-staining populations. Finally, the validated ORFs were linked to different PD-L1 expression to explain their potency.
Figure III-6. Calculation of the Weighted Positive Ratio (WPR) and positive ratio (PR). Comparison of WPR and the Positive Ratio (PR) for two hypothetical profiles. (Left) PR = 0.1. (Right) PR = 0.29. Adapted from figure S3 in (Lee et al., 2011).

3.2.5 Convoluted polyclonal prey cell library

The heterogeneity of cells is mostly resulting from genome instability and cellular division during culturing and passaging, especially in the case of immortalized model cell lines. As such, the cell-to-cell variability is of wide concern for proteomics. However, the cell-based PPI studies have commonly been performed in bulk, with substantial materials that can largely compensate this intrinsic variance, obtaining a global and acceptable proteome profile. Since the emergence of high-throughput screening, the pooled cell library was frequently used in different screening-based approaches. The lentiviruses are frequently used to make the well-known one-ORF-per-Cell library. Due to the random lentiviral insertion of ORF, genetic interruption and insertional mutagenesis were often observed in stable cells. Moreover, the expression of integrated genes will depend on the transcriptional activity of the surrounding sequences at the integration site. Taken altogether, the random insertion will aggravate the library inner variance, which will further influence the screen performance.
Therefore, knock-in at the identical target genomic locus is highly demanded. Generally, there are two well-documented methods to introduce the target DNA sequence to a predefined genome site. First is the Flp-In system that involves introduction of a Flp Recombination Target (FRT) site into the genome of the mammalian cell line of choice (O’Gorman et al., 1991). Once the biotic-resistant Flp-In cell line is established, the subsequent generation of isogenic stable line is rapid and efficient. Consequently, this method was frequently used in function gene stable cell line generation. As the probability of obtaining stable integrants containing a single FRT site or multiple FRT sites, with subsequent chromosomal position effect, this method is as yet not reported to be used in high-throughput study. Second, clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9)-based techniques have transformed our ability to genetically manipulate mammalian genomes. Knock-in method with Cas9 RNPs (Cas9-gRNA ribonucleoproteins) mediated homology directed repair template (HDRT) permits alteration of the endogenous sequence or insertion of an exogenous sequence at the target locus. More recently, a method for pooled knockin screens was developed which combines simultaneous delivery of a library of dsDNA HDRTs and Cas9 RNPs by electroporation. The HDRTs are integrated into a single locus (endogenous TCR-α locus), specified by the sgRNA to generate a T cell population expressing on average one insert per cell. The delivery of the library is followed by assays to evaluate the impact of each construct on T cells (Roth et al., 2020). This impactful attempt sheds light on the generation of a pooled and BiFC-compatible tagged prey cell library in means of a unique single integration site. The resulting pooled tagged ORF knock-in involves the same locus being targeted in every cell, however the identity of the genetic modification at this locus differs between the cells in the edited population.

After all, the future pooled prey cell library with identical ORF insert sites will largely decrease the side effects caused by the transcriptional activity of the surrounding sequences for the cell population harboring the same ORF. Given the same ORF flanking sequence, it will further improve the PCR efficiency and bias of target region amplification during the NGS library preparation process. Moreover, the less convoluted cell library will benefit the ORF representation throughout the BiFC screening. In addition, during the sequential
gating procedure, the cells with less variance will generate more concentrated cell populations, which could further increase the precision of cell sorting.

3.2.6 *A bulk interactome investigation instead of single functional complex*

Proteins participate in most cellular processes and fulfil most of their biological tasks in complexes through interaction with other proteins (*Alberts, 1998*). Protein complexes that are composed of more than one component are found in many different classes of proteins. Like that cell is the fundamental structural and functional unit of all living organisms, the protein complex is the functional unit of cellular biological processes in which proteins involve (*Figure III-7, center*). However, the PPI detection methods are generally performed in bulk manner, which measures only the global interactome of target proteins across a large population of input cells, resulting in a pool of potential interacting candidates regardless of the protein complex integrity (*Figure III-7, left*). For example, BioID-like AP-MS investigates the whole complexome of target protein. In turn, our Cell-PCA generates a collection of possible binary PPIs. These methods do not allow the analysis of specific complexes but rather give an overview of all possible PPIs of a given protein, resulting in one of the most representative interactomes by smoothie-like PPI analysis.

![Figure III-7. Protein complexome in cell, illustrated by protein A.](image)

The protein complexome refers to the entire set of protein complexes produced in a cell. Each protein has its specific complexome composed of
many individual protein complexes. For example, (Centre) protein A complexome is a collection of protein A complexes, potentially, localized in various cellular compartments, including nucleus, cytoplasm, membranes and membrane-bound organelles (not shown). (Left) Protein A interactome includes all protein A interacting proteins in a specific biological context, which can be detected by either co-complex manner (e.g. AP-MS or BioID) or binary PPI screens (e.g. Cell-PCA). The individual protein A complex can perform biological functions in a DNA-binding or -free fashion. (Right) The former, as genome locus-specific protein factors can be deciphered by reverse ChIP method.

Regarding the growing concern about the intra-tissue heterogeneity as well as the cell-to-cell variability in bulk analysis, single-cell omics gained widespread popularity since 2014 (Kharchenko et al., 2014; Picelli et al., 2014), along with more accessible protocols and lower sequencing costs. Following the first whole-transcriptome analysis of a single cell (Tang et al., 2009), more than 100 different single cell sequencing methods have been published (Wikipedia, 2021). These substantial advances have led to the transition from initial scRNA-seq to single-cell multi-omics, allowing multimodal measurements and integration of transcriptome, proteome, and spatial localization from the same cell. For example, the commercial 10X Genomics Visium solution combines whole transcriptome spatial analysis with immunofluorescence protein detection in the same tissue section, which empowers a deeper, more holistic understanding of tissue organization. Moreover, the classical large-scale genetic perturbation screens stand to benefit from single-cell sequencing. Recently, screens combining genetic perturbations with scRNA-seq readouts have emerged as promising and scalable alternatives over traditional screens, enabling direct readout of transcriptomic changes from the final fitness-responded cell population. As example, Maehr and colleagues combined single-cell RNA-seq with parallel CRISPR perturbations to comprehensively define the loss-of-function phenotype of those factors in definitive endoderm development (Genga et al., 2019). Innovatively, barcoded genome-scale ORF expression libraries were used by Mali lab, to systematically overexpress a pooled library of TFs in hPSCs, coupling scRNA-seq and fitness screen (Parekh et al., 2018). While other groups have demonstrated different scRNA-seq-based screens, notably, scRNA-seq based PPI screens have yet to be demonstrated.

Thanks to the well-performed barcoded ORFeome library (Parekh et al., 2018; Sack et al., 2018) and scRNA-seq (Chromium Single Cell 3’ library) (Genga et al., 2019), herein I
present a parallel scRNA-seq BiFC screening method, enabling pooled BiFC screens with single-cell transcriptome resolution (Figure III-8). In contrast to the genetic perturbation screens that pooled shRNAs and gRNAs themselves can serve as specific barcodes because they consist of uniquely identifiable DNA sequences, ORF sequences vary substantially in length, introducing bias during PCR recovery, as longer templates are recovered less efficiently by PCR. Thus, pairing ORFs uniquely with DNA BCs of uniform length will provide a marked improvement in screen fidelity, in which the BCs serve as the surrogate reporter to monitor ORF abundance. Moreover, this scRNA-seq-based BiFC screen method, at the time of deconvolution, is based on RNA-seq using a single cell 3’ gene expression library. The resulting CC-ORF was paired with a unique length barcode sequence located 200 bp upstream of the PolyA region (Figure III-8C). This yields a polyadenylated transcript bearing the barcode proximal to the 3’ end, facilitating efficient detection in scRNA-seq. Consequently, scRNA-seq based BiFC screening simultaneously assays both bait interacting ORF candidates and PPI-coupled cell-specific changes in transcriptome, more significantly, making a place for high-throughput PPI screening method in this new single-cell era.

As a long-term issue that pooled BiFC screening is restricted to apply in only cell lines or single-cell organisms, the multi-cell self-organized organoid is a potential model to be used in high-throughput BiFC screening. Recent progress in stem cell biology led to a strong revival of the organoid field. Organoid technology can therefore be used to model human organ development and various human pathologies ‘in a dish’, reflecting key structural and functional properties of organs (Lancaster and Knoblich, 2014). To date, many genetic manipulations have already performed in organoids, such as transfection (Laperrousaz et al., 2018), transduction (Maru et al., 2016) and even CRISPR/Cas9 precision genome editing (Artegiani et al., 2020). In addition, single-cell analyses of matched organoids by FACS was also widely used in different studies (Fujimichi et al., 2019; Rosenbluth et al., 2020). Accordingly, the current achievements have shown that an organoid-based pooled BiFC screening is very promising to be conceived and carried out in the near future. This feasibility opens new perspectives for pooled BiFC screening, as well as high-throughput
genetic screening and functional genomic applications, further giving precise and valuable insight into gene function and PPIs in a more human-like context.

Figure III-8. Schematic of barcoded CC-ORF library generation and capture of ORF candidates during scRNA-seq. (A) Construction of Gateway-compatible pooled barcode library (CC tagging). Prior to barcode sequence integration, the plasmid was inserted with a CC155 fragment (between TRE promotor and Gateway cassette) and UTR-PolyA sequence (between Gateway cassette and PGK promotor). 24 or 30mer barcodes of random sequence were PCR amplified and cloned into plix-CC-Gateway-UTR-PGK-Puro-DEST vector, using rare unique restriction sites I-CeuI and I-SceI. LTR, long terminal repeat; TRE, tetracycline responsive element; UTR, 3′ untranslated region; PolyA, an SV40 polyadenylation signal; PGK, phosphoglycerate kinase 1 promoter; Puro, puromycin resistance gene. (B) Construction of barcoded CC-ORF library expression vector. ORF collections were cloned into Gateway DEST site by LR recombination. The CC-ORF-BC plasmid DNA was then sheared and size-selected to recover 350-550 bp fragments (not...
shown). Through Illumina paired-end sequencing, allowing identification of the BC sequences uniquely paired to each ORF. (C) Schematic of lentiviral CC-ORF-BC vector and capture of CC-ORF-BC transcript during scRNA-seq. The final pooled CC-ORF-BC library will be used to produce lentivirus, for generating barcoded prey cell libraries. After co-expressed with corresponding tagged bait (VN tagging), the cell library will be sorted by FACS. Fluorescence-positive cells, with paired control, will be used for 10X single cell 3’ gene expression library preparation. The final ORF hits will be deconvolved by BCs. UMI, Unique Molecular Identifier. Adapted from figure 1 in (Parekh et al., 2018).

Given the applicability of single-cell-based or/and organoid-based pooled PPI screening, it will permit measuring expression levels for each interacting candidate across a population of cells and allow studying new biological questions in which PPI-affected cell-specific changes in transcriptome. However, this single cell-based binary protein complex analysis is still far from a real functional protein complex. To fill this gap, a reverse-ChIP method, named CLASP (Cas9 locus-associated proteome), was reported to capture functionally relevant gene-specific regulators targeted to the gene locus of interest (Tsui et al., 2018). By using purified recombinant catalytically inactive Cas9 (dCas9)–guide RNA ribonucleoprotein (RNP) complexes, CLASP does not require specialized cell lines and can be easily prepared with different guide RNAs to target multiple loci in any cell line or tissue. By fusing dCAS9 to PL enzyme, a new attempt was to fuse dCAS9 with BirA* to create a novel technology CASID, which was applied to analyze binding proteins in the direct vicinity of specific loci (Schmidtmann et al., 2016). Whereas dCAS9-based method enables a single genome locus-proximal proteome analysis, in which may include dozens of functional protein complexes instead of a single functional complex, it still provides insight into the real-time binding activities of these proteins at a specific DNA locus and uncovers the identities of these proteins simultaneously. However, one mentionable caveat is that the dCAS9-based method is designed to focus on only the nuclear protein activities depending on the DNA binding and its highly nucleus-restricted. Referring to the huge number of non-nucleus-localized protein complexes, a global non-compartment-specific detection method is needed, which enables a whole protein complexome analysis, like single transcript-based transcriptome analysis via RNA-seq.
To close this gap, co-elution or co-fractionation (CoFrac) approaches are collectively a global approach used to simultaneously study the whole interactome \cite{Havugimana2012, Kristensen2012}. They all rely on separation of protein complexes under native conditions, with the fact that proteins belonging to the same complex co-elute or migrate together during separation, showing the same migration profile (Figure III-9). As such, hundreds to thousands of protein complexes can be simultaneously and rapidly analyzed by co-elution in a single experiment, enabling the all-to-all protein analysis at single-protein-complex resolution. An added attraction of co-elution is, to date, that generated interactome does not rely on the genetic manipulation of cells or organisms, co-elution has thus been able to predict endogenous and unmanipulated protein complexes on a considerably large scale and in more physiologically relevant manner, as opposed to the results involving the tagged or overexpressed bait proteins. Nonetheless, one main drawback of co-elution that enslaves its popularity, is requirement of sophisticated bioinformatics analyses, facing million pairs of proteins quantified in a sample. However, co-elution is a powerful tool for next-generation interactomics, and it provides higher dimensional data information over existing high-throughput PPI screen methods. Looking forward, co-elution methods will progress toward increasing separation resolution and maximizing quantitation accuracy, along with miniaturization of sensitive MS measurement, and guide future single-cell interactomics.

Figure III-9. General workflow of a co-elution experiment. The lysed sample containing protein complexes under native conditions is separated in a set number of fractions. Proteins from the same complex show the same co-elution profile after a bioinformatic analysis to extract an interactome map, including single protein complex info. Adapted from figure 1 in \cite{Salas2020}.
Chapter IV. Annex

A Systematic Survey of HOX and TALE Expression Profiling in Human Cancers (article 4, review)

Yunlong Jia¹, Francois Bleicher¹ and Samir Merabet¹,#

¹ Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, Université Lyon 1, CNRS, Ecole Normale Supérieure de Lyon, 46, allée d’Italie 69364 Lyon cedex 07 France

# Correspondence: samir.merabet@ens-lyon.fr

Published in Int. J. Dev. Biol., 2018, doi: 10.1387/ijdb.180286fb
A systematic survey of HOX and TALE expression profiling in human cancers

YUNLONG JIA, FRANÇOISE BLEICHER and SAMIR MERABET*
Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, Université Lyon 1, CNRS,
Ecole Normale Supérieure de Lyon, Lyon, France

ABSTRACT HOX and TALE genes encode homeodomain (HD)-containing transcription factors that act in concert in different tissues to coordinate cell fates and morphogenesis throughout embryonic development. These two evolutionary conserved families contain several members that form different types of protein complexes on DNA. Mutations affecting the expression of HOX or TALE genes have been reported in a number of cancers, but whether and how the two gene families could be perturbed together has never been explored systematically. As a consequence, the putative collaborative role between HOX and TALE members for promoting or inhibiting oncogenesis remains to be established in most cancer contexts. Here, we address this issue by considering HOX and TALE expression profiling in normal and cancer adult tissues, using normalized RNA-sequencing expression data derived from The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) research projects. Information was extracted from 28 cancer types originating from 21 different tissues, constituting a unique comparative analysis of HOX and TALE expression profiles between normal and cancer contexts in human. We present the general and specific rules that could be deduced from this large-scale comparative analysis. Overall this work provides a precious annotated support to better understand the role of specific HOX/TALE combinatorial codes in human cancers.

KEY WORDS: HOX, TALE, homeodomain, cancer

Introduction

HOX proteins are homeodomain (HD)-containing transcription factors (TFs) that control various developmental processes during embryogenesis, including axis patterning (Pearson et al., 2005), limb formation (Zakany and Duboule, 2007) or organ differentiation (Chojnowski et al., 2014; Gligorov et al., 2013; Wessiek, 2011). HOX proteins are also required in the adult, in particular for the homeostasis of stem cell lineages in different tissues (Lebert-Gial et al., 2016, Rux et al., 2016, Sugimura et al., 2017; Xin et al., 2017). These various and specific functions are thought to rely on the partnership with diverse types of cofactors that remain to be identified for most of them (Merabet and Dard, 2014).

The best-characterized class of HOX cofactors are the PBC (Pre-B cell complex) proteins, which belong to the TALE (Three Amino acids Loop Extension) family of HD-containing TFs (Borglin, 1997). PBC cofactors interact with the large majority of HOX proteins on DNA, forming protein complexes with higher DNA-binding

*Address correspondence to: Samir Merabet, ENSL, IGFL, 46 Allée d’Italie, 69346 LYON cedex 07, France. Tel.: +33 (0)4 26731317. E-mail: samir.merabet@ens-lyon.fr

Supplementary Material (three tables) for this paper is available at: https://doi.org/10.13677/jdb.1802686

Submitted: 30 August, 2018; Accepted: 13 September, 2018.

ISSN: Online 1696-3547, Print 0214-6282
© 2018 UPV/EHU Press
Printed in Spain
specificity and affinity (Mann et al., 2009). Two other classes of the TALE family, the MEIS (Myeloid Ecotropic viral Integration Site) and PREP (Pbx-regulating protein, also called PKNOX) classes, also participate in HOX functions. MEIS and PREP originate from a common ancestor called MEINOX (Bürghel, 1998) and display intricate regulatory relationships with HOX and PBC proteins. For example, MEIS and PREP compete to interact with PBC and this interaction is necessary for the nuclear translocation hence activity of each TALE component. Competition between MEIS and PREP also occurs at the level of DNA-binding for target gene regulation on the genome (Dardel et al., 2014). In addition, the role of MEIS and PREP as direct HOX-binding partners on DNA without PBC proteins remains to be unequivocally determined. Along the same line, genome-wide binding analysis of PBC, MEIS and PREP in mice showed preferential and distinct combinations with HOX proteins. In particular, PBC/PREP DNA-binding sites appear to be enriched in promoters and nearby regions, while HOX/PBC/MEIS DNA-binding sites are preferentially found in intronic and intergenic regions (Perkov et al., 2013). Finally, MEIS but not PREP induces strong HOX-PBC interaction remodelling, revealing the role of various and specific HOX protein motifs for trimeric complex formation (Dardel et al., 2018). This interaction plasticity has been proposed to be important for diversifying and specifying HOX-TALE functions during development and evolution (Merabet and Mann, 2016).

What about HOX and TALE in cancer? In fact, the role of these families in cancer was historically identified from mutations affecting the protein function. PBC and MEIS were first identified in vertebrates from chromosomal translocations and viral insertions that respectively led to Pre-B acute lymphoblastic leukemia in children and acute myeloid leukemia in mouse (Kamps et al., 1990; Moskow et al., 1995; Nourse et al., 1990; Steelman et al., 1997). HOXA9 was described in 1996 as being responsible for acute and chronic leukemia, due to a chromosomal translocation leading to a fusion with the nucleoprotein NUP98 (Borrow et al., 1996; Nakamura et al., 1996). More recently, mutations in HOX513 coding sequences have also been associated with familial or high-risk prostate cancers (Karsson et al., 2014; Saunders et al., 2014). The large majority of HOX and TALE mutations in cancer are however not affecting the protein function but the expression profile. A number of studies have reported aberrant expression profiles of HOX, and to lesser extent, TALE genes, in solid cancers and leukemia. Moreover, the correlation with poor prognosis along with functional validations upon artificial expression in cancer-derived cell lines has been demonstrated in several instances. Interestingly, HOX genes are described to be up- or down-regulated depending on the cancer type, acting as pro- or anti-oncogenes. The same HOX gene can even have opposite functions in different cancer types (see Angioloupolous and Humphries, 2007; Bhath kar et al., 2014; Ekund, 2011; Sivakula et al., 2008) for excellent compiling reviews on this subject). With regard to TALE proteins, PBC and MEIS members are generally described as oncoproteins, while PREP is more frequently associated with a tumor suppressor function, due to its competitive role against MEIS (see Blasi et al., 2017) for review). The cooperative role of HOX and TALE in cancer is best established during leukemogenesis, where it has been demonstrated that PBX3 and MEIS1 are important cofactors for the transformation/immortalization activity of HOX proteins (in particular HOXA9) in hematopoietic stem cells (Li et al., 2013; Rozovskaya et al., 2001). Such cooperative role has rarely been shown in solid cancers (see for example (Fernandez et al., 2008)) and is principally deduced from indirect studies using a HOX-PBC interaction inhibitory peptide in cancer cell lines (see for example (Morgan et al., 2012). Moreover, the role of PBC, MEIS and PREP was analyzed with one member of each subfamily in most cancer studies (PBX1, MEIS1 and PREP1), which asks for the role of the other TALE members (PBX2-4, MEIS2-3 and PREP2).

Here we present a systematic analysis of the expression profile of the 39 human HOX members and their associated TALE cofactors (PBX1-4, MEIS1-3 and PREP1-2) in 28 cancer types deriving from 21 different tissues, comparing the normal and oncogenic context in each case. Raw RNA sequencing expression was extracted from the TCGA (The Cancer Genome Atlas; https://cancergenome.nih.gov) and GTEx (Genotype Tissue Expression: https://www.gtexportal.org/home/) projects and encompasses 5526 normal and 9018 cancer patient samples in total (Table 1, see also methods). Our analysis provides a global picture of enriched HOX and TALE expression profiles in normal and cancer tissues, allowing identifying the most significant regulatory changes associated with tumor progression. This information could serve as a molecular support for future therapeutic strategies aiming at targeting specific HOX/TALE complexes for developing anti-cancer agents or biomarkers.

Results

Expression of HOX and TALE genes in normal tissues

The 39 human HOX genes are distributed in four genomic clusters (A, B, C and D) that are located on different chromosomes. Each genomic cluster has a different number of HOX genes, due to gene loss events during evolution. Overall, human HOX genes are organized into 13 paralogous groups (PGs) that are defined as anterior (PG13-2), central (PG4-9) or posterior (PG9-13), based on the expression profile along the anterior-posterior axis in the early embryo. HOX genes from anterior PGs are, for example, expressed earlier and in more anterior parts of the vertebrate embryo than HOX genes from posterior PGs. In addition, HOX genes from the same PG usually display a highly similar if not identical expression profile along the AP axis, although the underlying cis-regulatory logic could be different (Kmita and Duboule, 2003). The nine TALE cofactor-encoding genes under study (PBX1-4, MEIS1-3 and PREP1-2) are also distributed on different chromosomes and display overlapping as well as distinct expression domains throughout development. Importantly, genetic and expression analyses are consistent with the role of TALE proteins as generic HOX cofactors in vertebrates (Moens and Sellier, 2006).

The role of TALE proteins as HOX cofactors is generally less documented in the adult, except in the context of the hematopoietic stem cell lineage (Altharbi et al., 2013). Given that the role of HOX and TALE genes in cancer could directly be linked to a change in the adult expression profile, we have considered the TCGA and GTEx database to annotate their expression profiles in 28 cancer types derived from 21 different tissues (Table 1).

Practically, normal HOX and TALE expression profiles in the adult were represented in heatmap with a specific color code (Fig. 1). Because HOX and TALE expression levels could strongly fluctuate, not only between different HOX or TALE members within the same tissue, but also for a same HOX or TALE member from one tissue to another, we arbitrarily decided to consider the mean expression level of all HOX genes in all normal tissues as the reference value
of a significant enriched expression level (hereafter defined as the global HOX mean value, TPM=4.51, Tables S1 and S2). This choice allowed us to normalize each HOX or TALE expression level in the different tissues (see also Methods). Of note, the same rational has previously been applied for establishing the tissue-specific atlas of the human proteome, although in this study authors arbitrarily decided to consider a fold change of at least 5 as a significant enrichment (Uhién et al., 2015). Here, we voluntarily did not apply the same threshold to consider weakly expressed HOX and TALE genes that could be of functional relevance between the normal and cancer contexts.

Each individual HOX or TALE expression level was reported as a ratio over this global HOX mean value and annotated as a log2 fold change. Expression levels below the global HOX mean correspond to negative log2 values and were not considered (light-grey cases in the heatmap of the Fig. 1). Expression levels that were equal or superior to the global HOX mean correspond to positive log2 values and the fold change (log2FC≥0) was illustrated by a light-to-dark green gradient color code in the heatmap (Fig. 1).

Below we summarized the main conclusions that could be deduced from the analysis of normal and tissue-specific HOX and TALE expression profiles in the adult (Fig. 1):

(1) - HOX genes are widely expressed in adult tissues, with only five tissues that show a significant HOX-poor expression profile (blood, brain, liver, pancreas and HINT (internal head and neck tissue)). Four tissues have few enriched HOX genes (less than six: esophagus, stomach, testis, thyroid) while five tissues harbor enriched HOX genes from the four HOX genomic clusters (Adrenal gland, breast, kidney, skin and uterus). The other tissues have enriched HOX genes belonging to one or two different genomic clusters. In all those cases, tissues express HOX genes that cover anterior, central and posterior PGs.

(2) - Several HOX genes show no or very low frequent expression profiles in adult tissues, as noticed for HOXA1-2, HOXA6, HOXB1, HOXB3, HOXC5, HOXC8, HOXC11-13, HOXD1 and HOXD12-13 (bottom graph). HOX genes of the HOXC genomic cluster are also less widely expressed in general than HOX genes of the three other clusters. In contrast, HOXA9-11, HOXB2-3 and HOXD8-10 show a global high enrichment of expression when compared to the other HOX genes. Together with the point mentioned in (1), these observations highlight that the HOX expression profile is quite diverse but also specific in adult tissues, which is also illustrated by a weak global mean frequency of expression (value=9.46, which accounts for 1/4 of all HOX genes).

(3) - The five tissues with enriched HOX (blood, brain, liver, pancreas and head neck) express at least one TALE member, illustrating potential HOX-independent functions of TALE members in those tissues. Such independent roles have previously been characterized for craniofacial and spinal cord motor neuron development (Feretti et al., 2011; Hanley et al., 2016). In the case of blood and liver samples, only PBX members are significantly enriched, which could thus represent non-functional contexts given

### TABLE 1

<table>
<thead>
<tr>
<th>TCGA</th>
<th>Full name of the tumor</th>
<th>Number of tumor tissue samples (TCGA)</th>
<th>Normal tissue name</th>
<th>Number of normal tissue samples (TCGA+GTEX)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGC</td>
<td>Adrenocortical carcinoma</td>
<td>77</td>
<td>Adrenal Gland [1]</td>
<td>0+128</td>
</tr>
<tr>
<td>BLCA</td>
<td>Bladder Urothelial Carcinoma</td>
<td>404</td>
<td>Bladder [2]</td>
<td>19+9</td>
</tr>
<tr>
<td>DLBC</td>
<td>Lymphoid Neoplasm Diffuse Large B-cell Lymphoma</td>
<td>47</td>
<td>Blood [3.1]</td>
<td>0+337</td>
</tr>
<tr>
<td>THYM</td>
<td>Thymoma</td>
<td>116</td>
<td>Blood [3.2]</td>
<td>0+337</td>
</tr>
<tr>
<td>LAML</td>
<td>Acute Myeloid Leukemia</td>
<td>173</td>
<td>Bone Marrow [4]</td>
<td>0+70</td>
</tr>
<tr>
<td>GBLM</td>
<td>Globlastoma multiforme</td>
<td>163</td>
<td>Brain [5.1]</td>
<td>0+267</td>
</tr>
<tr>
<td>LGG</td>
<td>Brain Lower Grade Glioma</td>
<td>518</td>
<td>Brain [5.2]</td>
<td>0+207</td>
</tr>
<tr>
<td>CESC</td>
<td>Cervical squamous cell carcinoma and endocervical adenocarcinoma</td>
<td>306</td>
<td>Cervix Uteri [7]</td>
<td>3+10</td>
</tr>
<tr>
<td>COAD</td>
<td>Colon adenocarcinoma</td>
<td>275</td>
<td>Colon [8.1]</td>
<td>41+308</td>
</tr>
<tr>
<td>READ</td>
<td>Rectum adenocarcinoma</td>
<td>92</td>
<td>Colon [8.2]</td>
<td>10+368</td>
</tr>
<tr>
<td>ESCA</td>
<td>Esophageal carcinoma</td>
<td>182</td>
<td>Esophagus [9]</td>
<td>13+273</td>
</tr>
<tr>
<td>KICH</td>
<td>Kidney Chromophobe</td>
<td>66</td>
<td>Kidney [10.1]</td>
<td>25+28</td>
</tr>
<tr>
<td>KIRC</td>
<td>Kidney Renal Clear cell carcinoma</td>
<td>523</td>
<td>Kidney [10.2]</td>
<td>72+28</td>
</tr>
<tr>
<td>KRP</td>
<td>Kidney Renal Papillary cell carcinoma</td>
<td>286</td>
<td>Kidney [10.3]</td>
<td>32+28</td>
</tr>
<tr>
<td>LUAD</td>
<td>Lung adenocarcinoma</td>
<td>483</td>
<td>Lung [12.1]</td>
<td>59+268</td>
</tr>
<tr>
<td>LUSC</td>
<td>Lung squamous cell carcinoma</td>
<td>486</td>
<td>Lung [12.2]</td>
<td>50+268</td>
</tr>
<tr>
<td>OV</td>
<td>Ovarian serous cystadenocarcinoma</td>
<td>426</td>
<td>Ovary [13]</td>
<td>0+88</td>
</tr>
<tr>
<td>PAAO</td>
<td>Pancreatic adenocarcinoma</td>
<td>179</td>
<td>Pancreas [14]</td>
<td>4+167</td>
</tr>
<tr>
<td>PRAO</td>
<td>Prostate adenocarcinoma</td>
<td>492</td>
<td>Prostate [15]</td>
<td>52+100</td>
</tr>
<tr>
<td>SKCM</td>
<td>Skin Cutaneous Melanoma</td>
<td>461</td>
<td>Skin [16]</td>
<td>1+557</td>
</tr>
<tr>
<td>STAD</td>
<td>Stomach adenocarcinoma</td>
<td>408</td>
<td>Stomach [17]</td>
<td>36+175</td>
</tr>
<tr>
<td>TGCT</td>
<td>Testicular Germ Cell Tumors</td>
<td>137</td>
<td>Testis [18]</td>
<td>0+165</td>
</tr>
<tr>
<td>THCA</td>
<td>Thyroid carcinoma</td>
<td>512</td>
<td>Thyroid [19]</td>
<td>59+278</td>
</tr>
<tr>
<td>UCEC</td>
<td>Uterine Corpus Endometrial Carcinoma</td>
<td>174</td>
<td>Uterus [20.1]</td>
<td>13+78</td>
</tr>
<tr>
<td>UCS</td>
<td>Uterine Carcinoma</td>
<td>57</td>
<td>Uterus [20.2]</td>
<td>0+78</td>
</tr>
<tr>
<td>HNSC</td>
<td>Head and Neck squamous cell carcinoma</td>
<td>519</td>
<td>Head and Neck [21]</td>
<td>44+40</td>
</tr>
</tbody>
</table>

Samples from normal tissues (last column) were from TCGA (blue) and GTEX (gene) portals. “O” indicates absence of available samples in the TCGA or GTEX database. Samples from cancer tissues (third column) are from TCGA (blue). The full name of each cancer type abbreviation used for the Figures 3-5 is provided (second column). Numbers in brackets (1 to 21) refer to the tissue type for each cancer.
that PBX proteins normally need MEIS or PREP partners to translocate into the nucleus.

(4) - PBX (with the exception of PBX4) and MEIS members are generally more widely and strongly expressed than HOX members in adult tissues. PBX4 is particularly enriched in testis and thyroid tissues. Specific enrichment of PBX4 in testis was also previously reported in mouse (Wagner et al., 2001). PREP members are less strongly and less frequently expressed than PBX or MEIS, which recapitulates previous observations (Longobardi et al., 2014). Overall, the global mean expression level of TALE genes in adult tissues is five times higher than the global HOX mean expression level, with a global high mean frequency value (corresponding to 3/5 of all TALE genes).

We next performed a clustering analysis to assess whether HOX and TALE genes could be preferentially organized in specific ensembles based on their normal expression profile in adult tissues (Fig. 2). This analysis confirms the distinct expression properties of PBX (except PBX4) and MEIS members, which form a unique ensemble of highly and widely expressed genes. In contrast, PREP1 and PBX4/PREP2 form two independent and isolated branches within ensembles of highly- or poorly-expressed HOX genes, respectively. This observation suggests that PREP1, but not PBX4 and PREP2, could participate to HOX functions in normal tissues.

HOX members form three main ensembles that basically regroup high, medium and low expression levels. Interestingly, ensembles corresponding to high and medium expression levels can be divided

Fig. 1. HOX and TALE expression profiles in normal tissues. Expression profile is obtained from RNA-sequencing data performed in 21 different tissue types (see Table 1 for the full nomenclature), using TCGA and GTEx portals. Each value (see Table S1) was reported to the global HOX mean (TPM = 4.51), followed by log2 fold change (FC) conversion. Heatmap colors represent HOX and TALE enrichment as indicated in the color key. Varying shades of green indicate enrichment level (from 0 to 5), while light gray represents non-significant expression levels, lacking the global HOX mean expression level in all tissues as the reference value (TPM=4.51). Histograms around the heatmap indicate mean expression levels (dark gray bars) and frequency (red connected dots). Gray and red dotted lines note the global mean of expression level and frequency, respectively (A) HOX mean expression and enriched HOX frequency in each normal tissue; (B) TALE mean expression and enriched TALE frequency in each normal tissue; (C) all normal tissues’ mean expression and enrichment frequency for each HOX member; (D) all normal tissues’ mean expression and enrichment frequency for each TALE member.
into several sub-ensembles that are quite homogenous in terms of the genomic cluster (sub-ensembles a, c, g or k), grouped into consistent anterior/central/posterior PG identity (sub-ensembles b, f, 1 and j). The ensemble regrouping weakly and non-significantly expressed HOX genes is more disparate although continuous HOX genes are present in small homogenous groups.

Tissue clustering confirms that different samples providing from the same tissue are quite homogenous, with highly similar HOX and TALE expression profiles. The clustering also revealed similarities between different tissues, as noticed for the breast and adrenal gland, thyroid and testis, or cluster composed by HNIT, pancreas and stomach. Other tissues appear more distinct, like the ovary, skin and bone marrow.

Overall the clustering analysis showed that the majority of HOX genes formed homogenous ensembles of two to five members based on their expression profile in adult tissues. This organization follows the PG affiliation (anterior, central or posterior) and/or the genomic cluster identity, highlighting that various cis-regulatory rules could be responsible for the expression of specific combinations of HOX genes in different tissues. Moreover, tissue-specific HOX combinatorial codes are systematically associated with a high expression level of several PBX and MEIS members. This observation suggests that the information provided by each specific combination of HOX genes in normal adult tissues is dependent on a general partnership with PBX and MEIS members.

**Expression of HOX and TALE genes in oncogenic tissues**

The expression profile of HOX and TALE genes in cancer was analyzed in 28 cancer types deriving from the same 21 different tissues, using data samples from TCGA (Tables 1 and S2). For

---

**Fig. 2. Unsupervised clustering of HOX and TALE expression profiles in normal tissues.** Hierarchical clustering of HOX and TALE members was classified based on the non-defined cluster (ordered by a set of letters), using complete linkage method and log2 fold change (FC) values as the Euclidean distance metric. All “non-significant” genes’ values are set as -1 for facilitating clustering. The color code highlights HOX paralog groups (graded pink), HOX genomic clusters (graded gray) and TALE members (yellow), as indicated.
better comparison with the normal context, each \( \log_2 \) value was calculated by considering individual HOX and TALE cancer level over the same global HOX mean value previously defined in normal tissues. The deduced \( \log_2 \) values were given in a heatmap with the same color code as in the Fig. 1 to represent non-enriched (grey) or enriched (green) expression levels (Fig. 3).

The heatmap shows that HOX genes are globally less frequently and less strongly expressed in cancer tissues (see the global mean values and graphs in the Fig. 3). The decrease of expression level is most apparent for tissues or HOX genes that displayed high enrichments in the normal condition (compare for example kidney, uterus or the HOXD cluster between Fig. 1 and 2). Still, there are also novel HOX expression profiles, particularly in cancers derived from tissues that were negative in normal condition (lymphoid neoplasia, thymoma, glioblastoma, pancreatic adenocarcinoma and HNIT: Fig. 3). In contrast, only one HOX gene among the nine that were classified as non-enriched in the normal condition becomes positive in one cancer tissue (HOX2 in thymoma: Fig. 3). This observation highlights that most HOX genes that were not enriched in normal tissues remain refractory to up-regulatory mechanisms in cancer. The expression level of TALE factors in cancer tissues is also significantly diminished, with a two folds decrease on average when compared to normal tissues (Fig. 3). Still, the TALE frequency pattern is comparable between cancer and normal tissues, and the expression level remains more than three times higher than the global HOX mean level. In conclusion, the cancer heatmap shows that a number of HOX and TALE genes have a modified expression profile that principally results from a reduced expression level.

We next performed a clustering analysis to assess whether expression changes in cancer could modify the overall distribution of HOX and TALE members when compared to the normal tissues.
Results show that cancer types remain clustered in function of their tissue origin, showing that the same logic of deregulation applies for different cancers deriving from the same tissue (Fig. 4). The overall organization of HOX and TALE members is however drastically remodeled when compared to normal tissues, with basically two main ensembles of strongly or weakly expressed genes, and an independent branch corresponding to the unique wide and strong expression pattern of PBX2 (as already noticed in normal tissues). The two main ensembles are disparate in their HOX formula, although they contain groups of two or three continuous HOX genes of the same genomic cluster several times. Among the TALE members, PREP1 became closer to PBX1 and PBX3 than MEIS1, showing a potentially interesting role of PREP1 in place of MEIS1 in cancer tissues. Interestingly, MEIS1 becomes closer to HOXB2 and HOXB3 in the same ensemble, suggesting that a specific and general relationship could exist between these three factors in cancer. All other HOX genes are quite distantly related to the TALE factors, with HOXB7 and HOXB13 lying apart in independent branches. Finally, two tandems belong to the ensemble of weakly expressed HOX genes: PBX4/PREP2 and MEIS2/MEIS3. The novel distribution of the MEIS2/MEIS3 tandem highlights that these two TALE members are more generally affected in cancers than the other TALE members. Overall the clustering analysis shows that the HOX expression profile is more sensitive to deregulatory changes in cancer than the TALE expression profile, which is explained by their respective moderate and high expression levels in general in normal tissues.

To get more insights into the HOX and TALE expression changes in cancer contexts, we generated a third cancer heatmap based on the fold change enrichment or loss of individual HOX and TALE members when compared to their respective expression level in the normal tissue (Table S3 and Fig. 5). Fold changes were calculated as a log2 value and considered as significant when they were at least two times superior (red cases in the Fig. 5) or inferior (blue

---

Fig. 4. Unsupervised clustering of HOX and TALE expression profiles in cancer tissues. Data processing and representation were performed as mentioned in Fig. 2.
cases in the Fig. 5) to the normal value. The magnitude of positive or negative change was represented in each case with a graded red or blue color code, respectively. Values corresponding to a non-significant change are represented by light-grey cases. Of note, contexts with a significant change but with values that were below the threshold defined for an enriched expression level in the normal tissue are considered apart (corresponding to boxes surrounded by a dotted line in the Fig. 5).

This novel representation confirmed that the large majority of TALE, and to some extent HOX genes, are down-regulated in the cancer context when compared to the normal tissue, as illustrated with the number of blue cases (Fig. 5). For example, 15 different cancer contexts show a decreased expression level of TALE members while the reverse is only observed in three cases (thyroidoma (THYM), brain lower grade glioma (LGG) and pancreatic adenocarcinoma (PAAD)). Along the same line, members of the HOXA and HOXD clusters are most often down-regulated, while members of the HOXB and HOXC clusters show a more balanced pattern, with up-regulation events that are slightly more frequent than down-regulation events (Fig. 5). In total, all TALE members and 33/39 HOX genes show a significant change in the expression level when comparing cancer and normal tissues. Accordingly, only one (liver hepatocellular carcinoma) or two (brain lower grade glioma and liver hepatocellular carcinoma) cancer types show no significant changes in TALE or HOX expression profiles, respectively (Fig. 5).

In several cases, HOX expression changes in cancer do not systematically correspond to a down-regulation but also to up-regulations in few instances, showing that the same HOX gene could be repressed or activated depending on the cancer type. This fluctuated pattern is observed for most members of the HOXA-C clusters (exceptions are for HOX genes that are mono-deregulated in a few cancers, as noticed for HOXA6, A11, A13 and B13). In contrast, all but two members of the HOXD cluster show a more uniform deregulation, with a large majority of down regulations (bottom graph in the Fig. 5), especially in breast (BRCA), cervical (CESC), colon (COAD), rectum (READ), and kidney (KICH, KIRC, KIRP) and uterus (UCEC, UCS) cancers. This observation suggests that HOXD members could have a more general antitumorigenic action.

The majority of cancers have a homogenous distribution of up- or down-regulation in the HOX gene expression profile (graph on the right in the Fig. 5). For example, large B-cell lymphoma (DLBC), thyroma (THYM), glioblastoma (GBM), esophageal carcinoma

---

**Fig. 5. HOX and TALE differential expression profiles between normal and cancer tissues.** The heatmap represents significant HOX and TALE expression changes between normal and cancer cell contexts. Values (Table S3) are calculated as a log2 ratio between the normal and the corresponding cancer tissue. Non-significantly modified HOX and TALE expression are set in light-grey cases. Heatmap colors represent significantly up-regulated (red) or down-regulated (blue) HOX and TALE expression levels between cancer and normal cell types, as shown in the color key. Blue and red boxes surrounded by a dashed line indicate a significant change of expression for genes that are not significantly enriched in the normal or cancer tissue. They correspond to weakly expressed HOX and TALE genes (see table S3). Histograms around the heatmap margins indicate the frequency of significant changes for enriched expression levels across members (bottom graphs) and in different tissues (left and right graphs).
(ESCA), pancreatic/stomach adenocarcinoma (PAAD/STAD) and head/neck squamous carcinoma (HNSC) show only up-regulation of HOX genes, while the reverse is observed for adenocortical carcinoma (ACC), bladder urothelial carcinoma (BLCA), kidney cancers (KICH, KIRC, KIRP), lung adenocarcinoma (LUAD), testicular germ cell tumors (TGCT) and thyroid carcinoma (THCA). Notably, cancers with a balanced pattern of both strong up- and down-regulation of HOX genes are less frequent, and relate to acute myeloid leukemia (LAML), colon (COAD), rectum (READ) and uterine (UCEC/UCS) adenocarcinoma. Interestingly, these versatile distributions can be HOX cluster specific (LAML, UCEC and UCS), with antagonistic distribution between HOX members of the same PG (like HOX8 and HOX8b in LAML for example), or occur between members of the same HOX cluster (as observed for the HOXB complex in colon (COAD) and rectum adenocarcinoma (READ)). Cancer tissues have also a uniform distribution of TALE down-regulations, except for thymoma (THYM) and pancreatic adenocarcinoma (PAAD), where TALE factors are systematically up-regulated (right graph in the Fig. 5). Acute myeloid leukemia (LAML) and glioblastoma (GBM) are the only cancer contexts with a more balanced pattern of up- and down-regulations (Fig. 5).

This heatmap also shows that HOX and TALE expression profiles follow the same type of remodeling in the majority of cancers. Cancers with a major decrease of HOX expression levels have an associated decrease of TALE enrichment. Interestingly, the two cancer contexts with a strong and homogenous enrichment of TALE expression coincide with a uniform up-regulation of HOX genes (thymoma and pancreatic adenocarcinoma: PAAD). Finally, one of the rare contexts with a strong opposite enrichment between two different TALE members (MEIS2 and PBX3) corresponds to a cancer context with the same kind of strong opposite enrichment between HOXA and HOXB members (LAML: Fig. 5).

Altogether, these observations highlight that HOX and TALE factors are deregulated in a coordinated manner in the majority of cancers, suggesting that the two families are not acting independently of each other for cancer progression or arrest.

A comparison with the literature

Our systematic analysis of the 99 human HOX members and their associated TALE cofactors (PBX1-4, MEIS1-3 and PREP1-2) provides a global picture of HOX and TALE expression profiles in normal and cancer tissues.

To our knowledge, this analysis is the first attempt to couple the aberrant expression profile of HOX and TALE genes in a vast number of different cancer cell types. Data were extracted from RNA-sequencing experiments deposited in the TGCA and GTEx portals and we arbitrarily decided to assign the global HOX mean expression level in the 21 normal tissues under study as the minimal reference value for a significant enriched expression level. Under this condition, weakly expressed HOX genes could not be considered. In any case, given that all measures were reported to this minimal reference value, this choice has no influence on the log2 fold change that could be find between the normal and the corresponding cancer tissue. In addition, we considered significant fold changes from values that were below the global HOX mean value, considering that these modifications could still be not neutral in the cancer context (see below).

A large number of samples were considered for each cancer type (from 47 to 1085 cancer samples and 13 to 558 normal tissue samples), making the data collection highly heterogeneous. This collection of cancers likely comprises tumors of various natures, with different aggressive behaviors, at different stages, with or without metastases. By comparison, most of the analyses of HOX expression in normal (Takahashi et al., 2004) or cancer (Abdel-Fattah et al., 2006; Buccoliero et al., 2009; Hur et al., 2014; Kanai et al., 2010; Kelly et al., 2016; Makiyama et al., 2005; Miller et al., 2003; Plowright et al., 2009; Yamashita et al., 2006) tissues are based on RT-PCR or QPCR in cell lines or from more restricted and more homogenous tumor samples (using approximately 4 to 20 tumor samples). These differences might explain that we did not systematically reproduce observations from previous work.

Comparison with the published literature is symbolized in the Fig. 5 by up and down arrows for each studied HOX or TALE gene in the different cancer types (corresponding to activated or repressed states, respectively). This comparison shows that several HOX genes that we considered as significantly enriched in the normal tissue but not significantly affected between the normal and cancer condition (grey cases) were described as being up- or down regulated in previous studies. For example, HOX1A1 is reported as being overexpressed in breast cancer patient datasets (Taminiau et al., 2016) and its forced expression was shown to increase the proliferation rate of mammary carcinoma cell lines (Taminiau et al., 2018; Zhang et al., 2003). Still, HOX1A1 was also reported as being silent or down regulated in some primary breast cancer samples (Cantile et al., 2003; Taminiau et al., 2016). Our analysis of HOX1A1 in 361 normal and 1085 tumor breast samples revealed neither an enriched normal expression level nor a significant increase in the cancer context. Along the same line, members of the HOXB and HOXD complexes are described to be up-regulated in a subtype-specific manner in breast cancer cell lines (Hur et al., 2014). Here, we noticed that HOX1A/D and HOX1C members were respectively down- or up-regulated in breast cancer, which is coherent with a previous study of HOX gene expression in invasive ductal breast cancer tissues (Makiyama et al., 2005). Meanwhile, we observed a systematic up-regulation of HOX genes in glioblastoma (GBM), which is in accordance with previous observations (Abdel-Fattah et al., 2006). On the other hand, our analysis did not reveal HOX gene deregulation in brain lower grade glioma (LGG), while another study based on RT-PCR noticed that few HOXD genes were either up- or down-regulated in 14 pediatric low-grade gliomas (Buccoliero et al., 2009). Together these observations highlight that HOX expression profiles can be quite heterogeneous when considering different samples of the same cancer, suggesting that HOX deregulatory changes could be highly subtype-specific. In this context, our large set of heterogeneous samples will principally give access to the most frequent and/or dramatic changes when considering the cancer type as a whole. In addition, we noticed that some of the HOX genes that we considered as not significantly enriched in the normal or cancer tissue but that displayed a significant log2 fold change (boxes surrounded by a dashed line in the Fig. 5) were previously described in the literature as being deregulated in cancer. This is for example the case for HOXA1 up regulation in stomach adenocarcinoma (STAD) and head and neck squamous cell carcinoma (HNSC), or HOXB2 down regulation in colon adenocarcinoma (COAD) and rectum adenocarcinoma (READ, Fig. 5). This observation suggests that changes of weakly expressed HOX genes could also be not neutral in certain cancers.

A recent review based on the published literature with patient
samples and cell lines highlighted that only two (HOXC10 and HOXC12) of the 39 human HOX genes are reported as being not affected in a solid tumors (Bhatotkar et al., 2014). Here we noticed that HOXC10 is either up- or down-regulated in different cancer contexts, including glioblastoma (GBM), breast BRCA, cervical (CESC), esophageal (ESCA), kidney (KICH), skin (SKCM) and stomach (STAD) cancers (Fig. 5). In fact, more recent studies have also described a role of HOXC10 in gastric (Guo et al., 2017) breast (Sadik et al., 2016), and cervix (Zhai et al., 2007) cancers. More generally, our analysis reveals most significant variations of HOX gene expression in eight novel cancer types (adrenocortical carcinoma (ACC), Lymphoid Neoplasm (DLBC), Thymoma (THYM), Kidney chromophobe (KICH), Kidney renal papillary cell carcinoma (KIRC), Pancreatic adenocarcinoma (PAAD), Testicular germ cell tumor (TGCT) and Uterus corpus endometrial carcinoma (UCEC)). It also validates and expands previous important conclusions, including i) the specific pattern of change in HOX gene expression depending on the cancer type, ii) the more frequent up-regulation of posterior HOX genes in solid tumors, and iii) a cluster-specific enrichment, with different preferential patterns depending on the cancer type.

TALE factors are generally less studied than HOX genes in cancers, and our survey is therefore informative with this regard. Among the few studies on TALE in cancer, one reported that PBX1 was up-regulated in mouse immortalized hepatoblast cells screened for transposon germline insertions that led to mesenchymal liver tumor upon transplantation into nude mice (Kodama et al., 2016). We did not find a particularly high level of PBX1 expression in either wild type or oncogenic liver tissues (covering 160 and 369 different normal or cancer tissue samples, respectively; Table 1). Results obtained from this study are however difficult to compare with our datasets since the cell contexts are quite different. Another study reported that high level of PBX1 correlated with shorter survival in post-chemotherapy ovarian cancer patients, and that silencing PBX1 reduced stem-like properties of ovarian tumor cells (Jung et al., 2016). We noticed that PBX1 was highly expressed in both wild type and tumor ovarian tissues (Figs. 1 and 3), explaining why it was not captured as an enriched gene in ovarian cancer in our analysis. Similarly, high level of PREP1 was reported as being important for triggering epithelial-mesenchymal transition (EMT), invasion and metastasis in lung adenocarcinoma cells (Risolino et al., 2014). Our study revealed that PREP1 is strongly expressed in the normal and cancer lung tissues (Figs. 1 and 3). These observations suggest that the same TALE factor could have opposite functions between the normal and cancer cell context, being strongly expressed in both contexts for promoting differentiation or proliferation, respectively. Such antagonistic functional switch could potentially be linked to a variation of the HOX formula during cancer progression.

Our analysis reveals that TALE members are generally down-regulated in cancer although PBX and MEIS are usually considered as oncoproteins. In fact, the role of PBX and MEIS as oncoproteins results principally from studies in leukemia, while conclusions in solid cancers are deduced from few classical expression analyses (considered as high with no systematic comparison with the normal tissue) and overexpression experiments in few cancer cell lines (see Blasi et al., 2017) for review). Interestingly, we found a systematic enrichment of TALE factors in three cancer types (Thymus (THYM), Brain lower grade glioma (LGG), and Pancreatic adenocarcinoma (PAAD), two of them correlating to increased expression of a specific HOX cluster (THYM and PAAD). Our analysis thus revealed that THYM and PAAD constitute the best examples of a putative pro-oncogenic and collaborative role between HOX and TALE proteins in solid cancers.

Finally, our heatmaps show that TALE members are five times more enriched on average in normal tissues than HOX genes. Although this enrichment is strongly diminished (from five to two times on average, except for PBX2) in cancer tissues, it is maintained, which could explain why PBX and MEIS are considered as enriched factors in previous studies. In addition, PREP members remain systematically much less expressed than MEIS members in cancer, which is in accordance with the general tumor suppressor function of PREP due to its competitive role against MEIS.

**Conclusion**

A striking aspect of our large-scale analysis is the global strong expression level of TALE members, especially PBX and MEIS, when compared to the HOX family members (three times stronger on average). As a consequence, the HOX expression profile is much more sensitive to subtle deregulations in cancers than the TALE expression profile. Accordingly, several cancer contexts express only a significant level of TALE members and no HOX, while the reverse is never observed. Thus, a hallmark of tumor cells is the expression of PBX and MEIS in a HOX-low or free state. Surprisingly, most of the studies have focused on HOX genes in cancer while the role of TALE factors remains to be determined in many different cancers.

Several cancers have a mixture of down- and up-regulation of HOX genes, highlighting the importance of the cell context for HOX function. In those cases, up-regulated HOX members could potentiate the effect of TALE members. This effect is often cluster-specific, as observed for HOXB members in uterus, ovarian and pancreatic cancers, or HOXA members in thymoma and glioblastoma. What could dictate the pro- or anti-oncogenic activity of HOX proteins with PBX and MEIS is clearly a key issue to understand the HOX/TALE molecular code in cancer. The dose of each HOX and TALE molecule is certainly not neutral, as is the role of the different TALE members in association with the HOX family. In any case, given the general high expression level of PBX and MEIS members in cancer, one promising avenue for future therapeutic strategies could be to alter the activity of TALE products specifically in neoplastic cells. This could be achieved by overexpressing PREP, to block the activity of MEIS, or targeting MEIS, to make PBX non-functional (cytoplasmic). Interestingly, a dominant negative form of MEIS is also described in the literature (Jaw et al., 2000) and could be easily tested as a potential and general inhibitory peptide in many different cancer cell lines.

**Materials and Methods**

**Data collection**

The RNA sequencing normalized data used in this study were obtained from GEPIA (Gene Expression Profiling Interactive Analysis; http://geopia.cancer-pku.cn/) (Tang et al., 2017). GEPIA is a web-based tool for gene expression analysis based on the TCGA (the Cancer Genome Atlas) and the GTEx (Genotype-Tissue Expression) databases, using the output of a standard processing pipeline of UCSC Xena project (http://xenia.ucsc.edu)(Goldman et al., 2017). Here we considered 28 cancer types from 21
different tissues that have more than 10 different normal control samples from TCGA and/or GTEx. Table 1).

**Data preprocessing**

A mean TPM (Transcripts Per Kilobase Million) value was calculated for each Hox and TALE member in each tissue type (Tables S1 and S2). Values that were inferior to the global Hox mean value in all tissues (TPM = 4.51) were considered as not significantly enriched (gray cases in Tables S1 and S2). We noticed that Hox genes that have no known functions in a given tissue could have a TPM value closed to the global Hox mean TPM value, as noticed for example for HOXC4 in breast (TPM = 4.12) or HOXA6 in adrenal gland (TPM = 3.73). In contrast, known Hox functions correspond to higher TPM values, from 5.23 (HOXA3 in breast) to 49.19 (HOX10 in cervix uteri). This observation suggests that the global Hox mean TPM value constitutes a good threshold for discriminating enriched expression profiles linked to a putative function in vivo. All HOX and TALE expression levels were then normalized over the global Hox mean TPM value and log2 processed to do the heatmap (Figs. 1 and 3). This normalization allowed comparing Hox and TALE expression levels between different tissues. Hierarchical clustering is according to the Euclidean distance based on log2 fold change (FC) values, using complete linkage method.

**Identification of differentially expressed genes between normal and cancer tissues**

Differentially expressed gene (DEG) analysis was performed using GE-PIA, which was assessed by the R package limma using linear model and empirical Bayes method, with adjusted p-value (Benjamini and Hochberg FDR). Significantly modified Hox and TALE expression profiles meet the following conditions (Tables S3): 1) Enrichment: log2FC > 1, and p-value < 0.01. 2) Depletion: log2FC < -1, and p-value < 0.01. All data are displayed as log2 transformed in heatmap.

**Data visualization**

All heatmaps were performed in the statistical programming environment R (version 3.3.0) using functions available from Bioconductor (Huber et al., 2015), and histograms were performed using GraphPad Prism 7 software.

**Acknowledgments**

We thank LS Shashidhar for comments on the manuscript. Research in the laboratory is supported by the CNRS, ENS-Lyon and grants from Cellpura, FRM, ARC, Ligue Contre le Cancer and China Scholarship Council (grant #201708070003 to Y. Jia).

**References**


Chapter V. Materials and Methods
### 5.1 Plasmids used in different projects

All plasmids made and used during my PhD, as summarized in the following Table V1.

<table>
<thead>
<tr>
<th>Barcode</th>
<th>Symbol</th>
<th>Construction</th>
<th>Insert ORF size</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>J4-C8</td>
<td>ACP1</td>
<td>plix-CC155-ORF</td>
<td>477</td>
<td></td>
</tr>
<tr>
<td>J4-C3</td>
<td>ACT16A</td>
<td>plix-CC155-ORF</td>
<td>1290</td>
<td></td>
</tr>
<tr>
<td>J3-F8</td>
<td>AFF4</td>
<td>plix-CC155-ORF</td>
<td>1062</td>
<td>HOXA9 BifC test</td>
</tr>
<tr>
<td>B18-F4</td>
<td>AGO2</td>
<td>plix-CC155-ORF</td>
<td>1134</td>
<td></td>
</tr>
<tr>
<td>J4-D4</td>
<td>ANKRD2</td>
<td>plix-CC155-ORF</td>
<td>1083</td>
<td></td>
</tr>
<tr>
<td>B18-F3</td>
<td>APP</td>
<td>plix-CC155-ORF</td>
<td>918</td>
<td></td>
</tr>
<tr>
<td>J3-D2</td>
<td>ASB8</td>
<td>plix-CC155-ORF</td>
<td>867</td>
<td></td>
</tr>
<tr>
<td>J2-D7</td>
<td>ATG5</td>
<td>plix-CC155-ORF</td>
<td>828</td>
<td></td>
</tr>
<tr>
<td>B18-F5</td>
<td>BACH1</td>
<td>plix-CC155-ORF</td>
<td>2211</td>
<td></td>
</tr>
<tr>
<td>J3-F2</td>
<td>BARX1</td>
<td>plix-CC155-ORF</td>
<td>303</td>
<td></td>
</tr>
<tr>
<td>J2-D3</td>
<td>C6orf201</td>
<td>plix-CC155-ORF</td>
<td>396</td>
<td></td>
</tr>
<tr>
<td>B18-F6</td>
<td>CAMK2D</td>
<td>plix-CC155-ORF</td>
<td>1437</td>
<td>low-concentration</td>
</tr>
<tr>
<td>J2-D1</td>
<td>CAPN3</td>
<td>plix-CC155-ORF</td>
<td>930</td>
<td></td>
</tr>
<tr>
<td>J3-F6</td>
<td>CBFB</td>
<td>plix-CC155-ORF</td>
<td>549</td>
<td></td>
</tr>
<tr>
<td>B18-H5</td>
<td>CRY1</td>
<td>plix-CC155-ORF</td>
<td>1761</td>
<td>HOXA9 BifC test</td>
</tr>
<tr>
<td>J3-D8</td>
<td>CUL3</td>
<td>plix-CC155-ORF</td>
<td>2307</td>
<td></td>
</tr>
<tr>
<td>J2-C5</td>
<td>DDX49</td>
<td>plix-CC155-ORF</td>
<td>1452</td>
<td></td>
</tr>
<tr>
<td>J3-D3</td>
<td>DLX3</td>
<td>plix-CC155-ORF</td>
<td>864</td>
<td></td>
</tr>
<tr>
<td>B18-19</td>
<td>DLX4</td>
<td>plix-CC155-ORF</td>
<td>723</td>
<td></td>
</tr>
<tr>
<td>J4-D8</td>
<td>DLX5</td>
<td>plix-CC155-ORF</td>
<td>870</td>
<td></td>
</tr>
<tr>
<td>J4-C2</td>
<td>EZF6</td>
<td>plix-CC155-ORF</td>
<td>846</td>
<td></td>
</tr>
<tr>
<td>J4-F7</td>
<td>EDF1</td>
<td>plix-CC155-ORF</td>
<td>447</td>
<td>For Kiran project</td>
</tr>
<tr>
<td>J3-F5</td>
<td>EHMT2</td>
<td>plix-CC155-ORF</td>
<td>3006</td>
<td></td>
</tr>
<tr>
<td>J4-C1</td>
<td>ETV5</td>
<td>plix-CC155-ORF</td>
<td>1533</td>
<td></td>
</tr>
<tr>
<td>J2-D6</td>
<td>FAM60A</td>
<td>plix-CC155-ORF</td>
<td>222</td>
<td></td>
</tr>
<tr>
<td>B18-F7</td>
<td>FHL5</td>
<td>plix-CC155-ORF</td>
<td>855</td>
<td></td>
</tr>
<tr>
<td>J2-E8</td>
<td>FOXO3</td>
<td>plix-CC155-ORF</td>
<td>2022</td>
<td></td>
</tr>
<tr>
<td>J3-G3</td>
<td>GATA2</td>
<td>plix-CC155-ORF</td>
<td>1401</td>
<td></td>
</tr>
<tr>
<td>J2-D9</td>
<td>GRB2</td>
<td>plix-CC155-ORF</td>
<td>654</td>
<td></td>
</tr>
<tr>
<td>J4-D5</td>
<td>GRHL3</td>
<td>plix-CC155-ORF</td>
<td>1668</td>
<td>HOXA9 BifC test</td>
</tr>
<tr>
<td>B18-F8</td>
<td>HCF2</td>
<td>plix-CC155-ORF</td>
<td>2379</td>
<td></td>
</tr>
<tr>
<td>B18-F9</td>
<td>HESX1</td>
<td>plix-CC155-ORF</td>
<td>558</td>
<td></td>
</tr>
<tr>
<td>J3-E1</td>
<td>HIF1AN</td>
<td>plix-CC155-ORF</td>
<td>1050</td>
<td></td>
</tr>
<tr>
<td>B18-F1</td>
<td>HLF</td>
<td>plix-CC155-ORF</td>
<td>888</td>
<td>HOXA9 BifC test</td>
</tr>
<tr>
<td>B18-I5</td>
<td>HMGN1</td>
<td>plix-CC155-ORF</td>
<td>303</td>
<td></td>
</tr>
<tr>
<td>J2-C6</td>
<td>HMGN3</td>
<td>plix-CC155-ORF</td>
<td>234</td>
<td></td>
</tr>
<tr>
<td>J4-C4</td>
<td>HOPX</td>
<td>plix-CC155-ORF</td>
<td>222</td>
<td></td>
</tr>
<tr>
<td>B18-G1</td>
<td>IRF2</td>
<td>plix-CC155-ORF</td>
<td>1050</td>
<td></td>
</tr>
<tr>
<td>B18-G2</td>
<td>IRF3</td>
<td>plix-CC155-ORF</td>
<td>1359</td>
<td></td>
</tr>
<tr>
<td>Barcode</td>
<td>Symbol</td>
<td>Construction</td>
<td>Insert ORF size</td>
<td>Comment</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>---------------</td>
<td>----------------</td>
<td>--------------</td>
</tr>
<tr>
<td>B18-G3</td>
<td>IRF5</td>
<td>plix-CC155-ORF</td>
<td>1497</td>
<td></td>
</tr>
<tr>
<td>B18-G4</td>
<td>IRF9</td>
<td>plix-CC155-ORF</td>
<td>1182</td>
<td></td>
</tr>
<tr>
<td>J3-E2</td>
<td>KLF6</td>
<td>plix-CC155-ORF</td>
<td>783</td>
<td></td>
</tr>
<tr>
<td>B19-A1</td>
<td>LIN28A</td>
<td>plix-CC155-ORF</td>
<td>630</td>
<td>HOXA9 BiFC test</td>
</tr>
<tr>
<td>B18-I6</td>
<td>LIN28B</td>
<td>plix-CC155-ORF</td>
<td>753</td>
<td></td>
</tr>
<tr>
<td>J4-C9</td>
<td>LRRC29</td>
<td>plix-CC155-ORF</td>
<td>672</td>
<td></td>
</tr>
<tr>
<td>J4-D6</td>
<td>LUC7L</td>
<td>plix-CC155-ORF</td>
<td>978</td>
<td></td>
</tr>
<tr>
<td>J2-E9</td>
<td>MAFB</td>
<td>plix-CC155-ORF</td>
<td>972</td>
<td></td>
</tr>
<tr>
<td>J3-G1</td>
<td>MAX</td>
<td>plix-CC155-ORF</td>
<td>483</td>
<td></td>
</tr>
<tr>
<td>J4-D2</td>
<td>MED20</td>
<td>plix-CC155-ORF</td>
<td>639</td>
<td></td>
</tr>
<tr>
<td>B18-G6</td>
<td>MED31</td>
<td>plix-CC155-ORF</td>
<td>396</td>
<td></td>
</tr>
<tr>
<td>J3-G9</td>
<td>MEIS2</td>
<td>plix-CC155-ORF</td>
<td>1413</td>
<td></td>
</tr>
<tr>
<td>J4-E4</td>
<td>MGC52282</td>
<td>plix-CC155-ORF</td>
<td>1431</td>
<td>HOXA9 BiFC test</td>
</tr>
<tr>
<td>J4-D3</td>
<td>NARFL</td>
<td>plix-CC155-ORF</td>
<td>1023</td>
<td></td>
</tr>
<tr>
<td>J4-D1</td>
<td>NCAPD3</td>
<td>plix-CC155-ORF</td>
<td>1023</td>
<td></td>
</tr>
<tr>
<td>J3-H2</td>
<td>NCOA4</td>
<td>plix-CC155-ORF</td>
<td>1845</td>
<td></td>
</tr>
<tr>
<td>B18-H4</td>
<td>NEK6</td>
<td>plix-CC155-ORF</td>
<td>942</td>
<td></td>
</tr>
<tr>
<td>J3-F9</td>
<td>NFE2</td>
<td>plix-CC155-ORF</td>
<td>1122</td>
<td></td>
</tr>
<tr>
<td>J4-E2</td>
<td>NRC1</td>
<td>plix-CC155-ORF</td>
<td>1569</td>
<td>HOXA9 BiFC test</td>
</tr>
<tr>
<td>J2-D2</td>
<td>NUP210P1</td>
<td>plix-CC155-ORF</td>
<td>1569</td>
<td>HOXA9 BiFC test</td>
</tr>
<tr>
<td>J3-H1</td>
<td>NUP62</td>
<td>plix-CC155-ORF</td>
<td>1569</td>
<td></td>
</tr>
<tr>
<td>J3-A6</td>
<td>PBX1</td>
<td>plix-CC155-ORF</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>J2-C7</td>
<td>PBX3</td>
<td>plix-CC155-ORF</td>
<td>825</td>
<td></td>
</tr>
<tr>
<td>J2-C3</td>
<td>PCP4</td>
<td>plix-CC155-ORF</td>
<td>189</td>
<td></td>
</tr>
<tr>
<td>J3-D4</td>
<td>PITX1</td>
<td>plix-CC155-ORF</td>
<td>945</td>
<td></td>
</tr>
<tr>
<td>B18-E8</td>
<td>PLEKH1A</td>
<td>plix-CC155-ORF</td>
<td>1215</td>
<td></td>
</tr>
<tr>
<td>J3-D6</td>
<td>POU2F2</td>
<td>plix-CC155-ORF</td>
<td>1203</td>
<td></td>
</tr>
<tr>
<td>J4-E1</td>
<td>POUSF1</td>
<td>plix-CC155-ORF</td>
<td>1083</td>
<td></td>
</tr>
<tr>
<td>J4-C6</td>
<td>PPACA</td>
<td>plix-CC155-ORF</td>
<td>777</td>
<td></td>
</tr>
<tr>
<td>J2-C8</td>
<td>PPP2R1A</td>
<td>plix-CC155-ORF</td>
<td>1770</td>
<td></td>
</tr>
<tr>
<td>J4-E9</td>
<td>PTPRO</td>
<td>plix-CC155-ORF</td>
<td>3651</td>
<td>HOXA9 BiFC test</td>
</tr>
<tr>
<td>J2-C9</td>
<td>RAN</td>
<td>plix-CC155-ORF</td>
<td>651</td>
<td></td>
</tr>
<tr>
<td>J3-G2</td>
<td>RARA</td>
<td>plix-CC155-ORF</td>
<td>1389</td>
<td></td>
</tr>
<tr>
<td>J3-F4</td>
<td>RBBP8</td>
<td>plix-CC155-ORF</td>
<td>2709</td>
<td></td>
</tr>
<tr>
<td>J3-G1</td>
<td>RBFOX2</td>
<td>plix-CC155-ORF</td>
<td>1143</td>
<td></td>
</tr>
<tr>
<td>B18-G7</td>
<td>RELA</td>
<td>plix-CC155-ORF</td>
<td>663</td>
<td></td>
</tr>
<tr>
<td>J4-E8</td>
<td>RUFY1</td>
<td>plix-CC155-ORF</td>
<td>1803</td>
<td>HOXA9 BiFC test</td>
</tr>
<tr>
<td>B18-G8</td>
<td>RUVB12</td>
<td>plix-CC155-ORF</td>
<td>780</td>
<td></td>
</tr>
<tr>
<td>J3-F3</td>
<td>SATB1</td>
<td>plix-CC155-ORF</td>
<td>2292</td>
<td></td>
</tr>
<tr>
<td>J4-E5</td>
<td>SCLY</td>
<td>plix-CC155-ORF</td>
<td>1338</td>
<td>HOXA9 BiFC test</td>
</tr>
<tr>
<td>J3-F7</td>
<td>SIRT2</td>
<td>plix-CC155-ORF</td>
<td>1059</td>
<td></td>
</tr>
<tr>
<td>B18-E7</td>
<td>SMAD4</td>
<td>plix-CC155-ORF</td>
<td>1659</td>
<td></td>
</tr>
<tr>
<td>B18-G9</td>
<td>SMAD9</td>
<td>plix-CC155-ORF</td>
<td>1293</td>
<td></td>
</tr>
<tr>
<td>B18-H1</td>
<td>SMARCE1</td>
<td>plix-CC155-ORF</td>
<td>1236</td>
<td></td>
</tr>
<tr>
<td>B18-I8</td>
<td>SMYD1</td>
<td>plix-CC155-ORF</td>
<td>1473</td>
<td></td>
</tr>
<tr>
<td>J4-E6</td>
<td>SOX2</td>
<td>plix-CC155-ORF</td>
<td>954</td>
<td>HOXA9 BiFC test</td>
</tr>
<tr>
<td>B18-E9</td>
<td>SPN</td>
<td>plix-CC155-ORF</td>
<td>1203</td>
<td></td>
</tr>
<tr>
<td>Barcode</td>
<td>Symbol</td>
<td>Construction</td>
<td>Insert ORF size</td>
<td>Comment</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>--------------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td>B18-H2</td>
<td>TAF6L</td>
<td>plix-CC155-ORF</td>
<td>1380</td>
<td></td>
</tr>
<tr>
<td>J2-C1</td>
<td>TCF4</td>
<td>plix-CC155-ORF</td>
<td>1098</td>
<td></td>
</tr>
<tr>
<td>J4-E3</td>
<td>TFAP2A</td>
<td>plix-CC155-ORF</td>
<td>1296</td>
<td>HOXA9 BifC test</td>
</tr>
<tr>
<td>B18-F2</td>
<td>TFG</td>
<td>plix-CC155-ORF</td>
<td>1203</td>
<td></td>
</tr>
<tr>
<td>J3-D5</td>
<td>TLX3</td>
<td>plix-CC155-ORF</td>
<td>876</td>
<td></td>
</tr>
<tr>
<td>J3-D7</td>
<td>TP53</td>
<td>plix-CC155-ORF</td>
<td>1182</td>
<td></td>
</tr>
<tr>
<td>J3-H4</td>
<td>TP63</td>
<td>plix-CC155-ORF</td>
<td>2043</td>
<td></td>
</tr>
<tr>
<td>B18-H3</td>
<td>TRIB3</td>
<td>plix-CC155-ORF</td>
<td>1077</td>
<td></td>
</tr>
<tr>
<td>J3-G7</td>
<td>TRIM27</td>
<td>plix-CC155-ORF</td>
<td>1542</td>
<td></td>
</tr>
<tr>
<td>J2-C4</td>
<td>UBE2T</td>
<td>plix-CC155-ORF</td>
<td>594</td>
<td></td>
</tr>
<tr>
<td>J3-G6</td>
<td>VDR</td>
<td>plix-CC155-ORF</td>
<td>1284</td>
<td></td>
</tr>
<tr>
<td>J4-F8</td>
<td>ZBTB10</td>
<td>plix-CC155-ORF</td>
<td>2097</td>
<td>For Kiran project</td>
</tr>
<tr>
<td>J2-C2</td>
<td>ZMYM6</td>
<td>plix-CC155-ORF</td>
<td>471</td>
<td></td>
</tr>
<tr>
<td>J3-G5</td>
<td>ZNF148</td>
<td>plix-CC155-ORF</td>
<td>396</td>
<td></td>
</tr>
<tr>
<td>J3-H3</td>
<td>ZNF16</td>
<td>plix-CC155-ORF</td>
<td>2013</td>
<td></td>
</tr>
<tr>
<td>J3-G4</td>
<td>ZNF212</td>
<td>plix-CC155-ORF</td>
<td>1488</td>
<td></td>
</tr>
<tr>
<td>J4-E7</td>
<td>ZNF213</td>
<td>plix-CC155-ORF</td>
<td>1380</td>
<td>HOXA9 BifC test</td>
</tr>
<tr>
<td>J4-D7</td>
<td>ZNF253</td>
<td>plix-CC155-ORF</td>
<td>1272</td>
<td></td>
</tr>
<tr>
<td>J4-C5</td>
<td>ZNF396</td>
<td>plix-CC155-ORF</td>
<td>1002</td>
<td></td>
</tr>
<tr>
<td>B18-I7</td>
<td>ZNF428</td>
<td>plix-CC155-ORF</td>
<td>567</td>
<td></td>
</tr>
<tr>
<td>J4-F6</td>
<td>ZNRD1</td>
<td>plix-CC155-ORF</td>
<td>318</td>
<td>For Kiran project</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Barcode</th>
<th>Name</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>J3-A3</td>
<td>pcDNA3-Kozak-VN173-stop</td>
<td>Invitrogen [Gateway]</td>
</tr>
<tr>
<td>J3-A4</td>
<td>pcDNA3-Kozak-CC155-stop</td>
<td></td>
</tr>
<tr>
<td>J3-A5</td>
<td>pDONR223</td>
<td></td>
</tr>
<tr>
<td>J3-A6</td>
<td>pLix-Kozak-CC155-LR linker-PBX1</td>
<td>BifC positive control, by Thien</td>
</tr>
<tr>
<td>J3-A7</td>
<td>pLix-kozak-VN173-GGGG-GHOXA9HX [D29A-M56W]</td>
<td>BifC negative control</td>
</tr>
<tr>
<td>J3-A8</td>
<td>pLix Kozak-VN173-stop</td>
<td>BifC negative control</td>
</tr>
<tr>
<td>J3-B1</td>
<td>pLix-Kozak-CC155-stop</td>
<td>BifC negative control</td>
</tr>
<tr>
<td>J3-B2</td>
<td>V137</td>
<td>Making lentivirus</td>
</tr>
<tr>
<td>J3-B3</td>
<td>pcDNA3-VN-HOXA1</td>
<td></td>
</tr>
<tr>
<td>J3-C4</td>
<td>plx403-VN173-GGGG-GHOXA1-BleoR</td>
<td>Prepare for double selection</td>
</tr>
<tr>
<td>J3-C5</td>
<td>plx403-VN173-GGGG-GHOXA9-BleoR</td>
<td>Prepare for double selection</td>
</tr>
<tr>
<td>J3-C6</td>
<td>plx403-HA-HOXA1-BleoR</td>
<td>Prepare for double selection</td>
</tr>
<tr>
<td>J3-C7</td>
<td>plx403-HA-HOXA9-BleoR</td>
<td>Prepare for double selection</td>
</tr>
<tr>
<td>J3-E3</td>
<td>pLix-VN173-MeIS2_19a [for Fred]</td>
<td>plx403; AmpR</td>
</tr>
<tr>
<td>J3-E4</td>
<td>pDONR223-LacZ [for control]</td>
<td>pDONR223-LacZ</td>
</tr>
<tr>
<td>J3-E5</td>
<td>pCDH-CMV-mCherry-T2A-PURO</td>
<td>addgene [Plasmid #72264]</td>
</tr>
<tr>
<td>J3-E6</td>
<td>plix-MCS</td>
<td></td>
</tr>
<tr>
<td>J3-E7</td>
<td>pcDNA3-mCherry</td>
<td></td>
</tr>
<tr>
<td>J3-F1</td>
<td>pLenti-smI/R/F</td>
<td>addgene [plasmid-80349]</td>
</tr>
<tr>
<td>J3-H6</td>
<td>plix-VN-HOXA1-PGK-BleoR-mCherry</td>
<td>For virus and cell line</td>
</tr>
<tr>
<td>J3-H7</td>
<td>plix-VN-HOXA9-PGK-BleoR-mCherry</td>
<td>For virus and cell line</td>
</tr>
<tr>
<td>J3-H8</td>
<td>plix-30aa TurboID-HA-GGBP</td>
<td></td>
</tr>
<tr>
<td>Barcode</td>
<td>Name</td>
<td>Comment</td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>J4-A1</td>
<td>pcDNA3-VN-FLAG-HOXA9</td>
<td></td>
</tr>
<tr>
<td>J4-A2</td>
<td>pcDNA3-VN-FLAG-HoxA9HX</td>
<td></td>
</tr>
<tr>
<td>J4-A3</td>
<td>pcDNA3-kozak VN-HOXB13</td>
<td></td>
</tr>
<tr>
<td>J4-A5</td>
<td>pcDNA3-VN-HOXD10</td>
<td></td>
</tr>
<tr>
<td>J4-A6</td>
<td>pcDNA3-VN-HOXD10HX</td>
<td></td>
</tr>
<tr>
<td>J4-A7</td>
<td>pcDNA3-VN-HOXB3</td>
<td></td>
</tr>
<tr>
<td>J4-A8</td>
<td>pcDNA3-VN-HOXB3HX</td>
<td></td>
</tr>
<tr>
<td>J4-B1</td>
<td>pcDNA3-VN-HOXC8-iso2</td>
<td></td>
</tr>
<tr>
<td>J4-B2</td>
<td>pcDNA3-VN-HOXC8HX-iso2</td>
<td></td>
</tr>
<tr>
<td>J4-B4</td>
<td>pcDNA3-VN-FLAG-HoxA7HX</td>
<td></td>
</tr>
<tr>
<td>J4-B5</td>
<td>pcDNA3-VN-HoxC5</td>
<td></td>
</tr>
<tr>
<td>J4-B6</td>
<td>pcDNA3-VN-HoxC8HX</td>
<td></td>
</tr>
<tr>
<td>J4-B7</td>
<td>pcDNA3-VN-[mouse]-HOXD4</td>
<td></td>
</tr>
<tr>
<td>J4-B8</td>
<td>pcDNA3-VN-[mouse]-HOXD4HX</td>
<td></td>
</tr>
<tr>
<td>J4-F1</td>
<td>pTRE-Tight-Bi-Gl NORM-LacZA-TER-LacZB</td>
<td>seq OK</td>
</tr>
<tr>
<td>J4-F2</td>
<td>plix-TAZ-gssg-TurboID-HA</td>
<td>seq OK</td>
</tr>
<tr>
<td>J4-F3</td>
<td>plix_403-TurboID-HA</td>
<td>seq OK</td>
</tr>
<tr>
<td>J4-F4</td>
<td>Plix_403-EcoRI-TurboID-HA-N-ssGBP-BamHI</td>
<td>seq OK</td>
</tr>
<tr>
<td>J4-F5</td>
<td>plix-EGFP</td>
<td>seq OK</td>
</tr>
<tr>
<td>J4-F9</td>
<td>plix-HA-HOXA9 [new]</td>
<td>seq OK</td>
</tr>
<tr>
<td>J4-G1</td>
<td>plix_403-HOXA9-GGSGG-TurboID-HA</td>
<td>seq OK</td>
</tr>
<tr>
<td>J4-G2</td>
<td>pX330- AAVS1_12-CAS9</td>
<td></td>
</tr>
<tr>
<td>J4-G3</td>
<td>pMK243-Tet3G-001R1-PURO</td>
<td></td>
</tr>
<tr>
<td>J4-G4</td>
<td>pcDNA3-A-sfGFP11-T2A-sfGFP10-B-T2A-mCherry</td>
<td></td>
</tr>
<tr>
<td>J4-G5</td>
<td>pcDNA3-sfGFP1-9 M1-plasmid</td>
<td></td>
</tr>
<tr>
<td>J4-G6</td>
<td>pcDNA3-sfGFP1-9 M1-T2A-mcherry_324430</td>
<td></td>
</tr>
<tr>
<td>J4-G7</td>
<td>pTRE-Tight-Bi-ProteinA_VN-FLAG-HOXA9</td>
<td></td>
</tr>
<tr>
<td>J4-G8</td>
<td>plix-NLS-TurboID-HA</td>
<td></td>
</tr>
<tr>
<td>J4-G9</td>
<td>plix-NLS-TurboID-HA-ssGBP</td>
<td></td>
</tr>
<tr>
<td>J4-H1</td>
<td>pTRE-Tight-Bi-CC-PBX1-V5 ProteinB</td>
<td></td>
</tr>
<tr>
<td>J4-H2</td>
<td>pTRE-Tight-Bi-CC-PBX1-V5_VN-FLAG-HOXA9</td>
<td></td>
</tr>
<tr>
<td>J4-H3</td>
<td>NLS-mCherry_HAIRMUS Juliette[unknown plasmid]</td>
<td></td>
</tr>
<tr>
<td>J4-H4</td>
<td>[19] pTRE-Tight-Bi-CC-PBX1-V5</td>
<td></td>
</tr>
<tr>
<td>J4-H5</td>
<td>[20] pTRE-Tight-Bi-VN-FLAG-HOXA9</td>
<td></td>
</tr>
<tr>
<td>J4-H6</td>
<td>[18] plix-TurboID-HA-ssGBP-PGK-PuroT2A-mCherry</td>
<td></td>
</tr>
<tr>
<td>J4-H7</td>
<td>[CF1] plix_403-kozak-CC155-GGSGG-MCS</td>
<td></td>
</tr>
<tr>
<td>J4-H8</td>
<td>[CF2] plix_403-kozak-VN173-gssg-PBX1-V5</td>
<td></td>
</tr>
<tr>
<td>J4-H9</td>
<td>[CF4] plix_403-kozak-CN173-GGSGG-HA-MEIS1</td>
<td></td>
</tr>
<tr>
<td>J4-H11</td>
<td>[AF1] pcDNA3-HA-HOXA9-GFP11-T2A-GFP10-B-T2A-mCherry</td>
<td></td>
</tr>
<tr>
<td>J4-H12</td>
<td>[CF8] plix_403-kozak-CN173-FALG-HOXA9</td>
<td></td>
</tr>
<tr>
<td>J4-H13</td>
<td>[CF3] plix_403-kozak-CC155-FALG-HOXA9</td>
<td></td>
</tr>
<tr>
<td>J4-H14</td>
<td>[CF5] plix_403-kozak-VN173-FALG-HOA9</td>
<td></td>
</tr>
<tr>
<td>J4-H15</td>
<td>[CF6] plix_403-kozak-CC155-PBX1-V5</td>
<td></td>
</tr>
<tr>
<td>J4-H16</td>
<td>[AF2] pcDNA3-HA-HOXA9-GFP11-T2A-GFP10-PBX1-V5</td>
<td></td>
</tr>
<tr>
<td>J4-H17</td>
<td>[CF7] plix-kozak-CC-HA-MEIS1</td>
<td></td>
</tr>
<tr>
<td>J4-H18</td>
<td>plix-3HA-HOXA9</td>
<td></td>
</tr>
</tbody>
</table>
### Plasmids in Box J1

<table>
<thead>
<tr>
<th>Barcode</th>
<th>Name Description</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>J1-A1</td>
<td>plix-Kozak-CC-LacZ</td>
<td>by Thien</td>
</tr>
<tr>
<td>J1-A2</td>
<td>pcDNA3-VN-HOXA1HX[MI-HOXA1HX][D29A]</td>
<td></td>
</tr>
<tr>
<td>J1-A3</td>
<td>pcDNA3-VN-HOXA1HX[MI-HOXA1HX][D29A;M56W]</td>
<td></td>
</tr>
<tr>
<td>J1-A4</td>
<td>pcDNA3-MCS-mCherry(N-ter)</td>
<td></td>
</tr>
<tr>
<td>J1-A5</td>
<td>pcDNA3-MCS-mCherry(C-ter)</td>
<td></td>
</tr>
<tr>
<td>J1-A6</td>
<td>pcDNA3-CC-PBX1</td>
<td></td>
</tr>
<tr>
<td>J1-A7</td>
<td>pcDNA3-VN-HOXA1HX</td>
<td></td>
</tr>
<tr>
<td>J1-A8</td>
<td>pcDNA3-GFP</td>
<td>Control for BIFC</td>
</tr>
<tr>
<td>J1-A9</td>
<td>pcDNA3 VN-HOXA1</td>
<td></td>
</tr>
<tr>
<td>J1-B1</td>
<td>pDest-VC-Mus_Tsh3[Nter]</td>
<td>project-cooperation with Xavier CAUNIT</td>
</tr>
<tr>
<td>J1-B2</td>
<td>pDest-VC-Mus_Tsh3[Cter]</td>
<td>project-cooperation with Xavier CAUNIT</td>
</tr>
<tr>
<td>J1-B3</td>
<td>pBlueScript II SIK(+)-HOXA9[bac]</td>
<td>As template of HOX9A9H, for protein production</td>
</tr>
<tr>
<td>J1-B4</td>
<td>plix-403-mCherry</td>
<td></td>
</tr>
<tr>
<td>J1-C5</td>
<td>pcDNA3-VN-HOXA9[D29A;M56W]</td>
<td>in B13-ES</td>
</tr>
<tr>
<td>J1-B6</td>
<td>Plix-403-RPPH1</td>
<td>FOR QPCR standard plasmid</td>
</tr>
<tr>
<td>J1-B7</td>
<td>pcDNA3-mcherry</td>
<td>Midly by JR, as BIFC reference, instead of GSWIP-mcherry</td>
</tr>
<tr>
<td>J1-B8</td>
<td>pcDNA3-VN-HOXA5</td>
<td></td>
</tr>
<tr>
<td>J1-B9</td>
<td>pcDNA3-VN-HOXA7</td>
<td></td>
</tr>
<tr>
<td>J1-C1</td>
<td>pcDNA3 VN-FLAG-FLAG-HOXA9[H] [D29A]</td>
<td>Modified B13-C1, useless HOXB7 sequence removed</td>
</tr>
<tr>
<td>J1-C2</td>
<td>pcDNA3 VN-FLAG-FLAG-HOXA9[D29A]</td>
<td>Modified B13-G3, useless HOXB7 sequence removed</td>
</tr>
<tr>
<td>J1-C3</td>
<td>plix-403 [Original]</td>
<td>Addgene plasmid # 41395</td>
</tr>
<tr>
<td>J1-C4</td>
<td>pcDNA3-VN-HOXA3</td>
<td></td>
</tr>
<tr>
<td>J1-C5</td>
<td>pET28a(+)-HOXA9H[bac] w/o SD</td>
<td>protein production for HOXA9HX in bac, BUT no SD sequence</td>
</tr>
<tr>
<td>J1-C7</td>
<td>plix VN-HOXA9 [copy 1]</td>
<td>using PlIX-403[Original]</td>
</tr>
<tr>
<td>J1-C8</td>
<td>plix VN-HOXA9 [copy 2]</td>
<td>using PlIX-403[Original]</td>
</tr>
<tr>
<td>J1-C9</td>
<td>pcDNA3-VN155-HOXA9</td>
<td></td>
</tr>
<tr>
<td>J1-D1</td>
<td>pDest-VC-Mus_Tsh3</td>
<td>project-cooperation with Xavier CAUNIT</td>
</tr>
<tr>
<td>J1-D2</td>
<td>pcDNA3-3HA-cGFP</td>
<td>conditionally stable GFP nanobody</td>
</tr>
<tr>
<td>J1-D3</td>
<td>plix-CC-mCherry</td>
<td>For test our plix-CC construction</td>
</tr>
<tr>
<td>J1-D4</td>
<td>pET28a(+)-HOXA9HX[bac] w/o ATG</td>
<td>protein production for HOXA9HX in bac, BUT no ATG [frame shift]</td>
</tr>
<tr>
<td>J1-D5</td>
<td>plix MCS</td>
<td>1st step for future useful construction</td>
</tr>
<tr>
<td>J1-D6</td>
<td>plix-MCS-GGSGG-GV173</td>
<td>for expressing the protein in a c-fusion manner</td>
</tr>
<tr>
<td>J1-D7</td>
<td>pDest-VC156-7mSbz1-Delta_acidic</td>
<td>project-cooperation with Xavier CAUNIT</td>
</tr>
<tr>
<td>J1-D8</td>
<td>pDest-VC156-7mSbz3-Delta_acidic</td>
<td>project-cooperation with Xavier CAUNIT</td>
</tr>
<tr>
<td>J1-D9</td>
<td>pDest-VC156-7mSbz1</td>
<td>project-cooperation with Xavier CAUNIT</td>
</tr>
<tr>
<td>J1-E1</td>
<td>pDest-VN173-Mus_Tsh3</td>
<td>project-cooperation with Xavier CAUNIT</td>
</tr>
<tr>
<td>J1-E2</td>
<td>pDest-VN173-Mus_Tsh3</td>
<td>project-cooperation with Xavier CAUNIT</td>
</tr>
<tr>
<td>J1-E3</td>
<td>plix-kozak-VN173-GGSGG-MCS</td>
<td>OK, tested by JR using BIFC</td>
</tr>
<tr>
<td>J1-E4</td>
<td>plix-kozak-CN171-GGSGG-MCS</td>
<td></td>
</tr>
<tr>
<td>J1-E5</td>
<td>plix kozak-CC-DLX6</td>
<td></td>
</tr>
<tr>
<td>J1-E6</td>
<td>pcDNA3-Kan II-PGK-Bleo-R-T2A</td>
<td></td>
</tr>
<tr>
<td>J1-E7</td>
<td>pcDNA3-2nd and 3rd Lenti-Envelope protein</td>
<td>Making lentivirus</td>
</tr>
<tr>
<td>J1-E8</td>
<td>pcDNA3-FKBP1-Kan II-PGK-Bleo-R-T2A</td>
<td>New BleoR selection</td>
</tr>
<tr>
<td>J1-E7</td>
<td>pcDNA3-Kan II-PGK-Bleo-R-T2A</td>
<td>New BleoR selection</td>
</tr>
<tr>
<td>J1-E2</td>
<td>pcDNA3-CN171-GGSGG-MCS-PGK-BleoR</td>
<td>New BleoR selection</td>
</tr>
<tr>
<td>J1-E3</td>
<td>plix-kozak-VN173-GGSGG-MCS-PGK-BleoR</td>
<td>New BleoR selection for VNN-Myc</td>
</tr>
<tr>
<td>J1-F4</td>
<td>plix - Kozak - CC155 - Gateway</td>
<td>Maybe have a frame shift, always one more &quot;G&quot; after ATG, need further validation. According to the sequencing result from JR, it's the sequencing problem.</td>
</tr>
<tr>
<td>J1-F5</td>
<td>plix-kozak-CN171-GGSGG-cMyc-PGK-BleoR</td>
<td>New BleoR selection for C-Myc</td>
</tr>
<tr>
<td>J1-F6</td>
<td>plix-kozak-VN173-GGSGG-cMyc</td>
<td>done by Agnes</td>
</tr>
<tr>
<td>J1-F7</td>
<td>pcDNA3-3-Myc - HOXA9HX[D29A;M56W]</td>
<td></td>
</tr>
<tr>
<td>J1-F8</td>
<td>pcDNA3-VDE</td>
<td></td>
</tr>
<tr>
<td>J1-G1</td>
<td>plix - Kozak - CC155 - LR linker MAX</td>
<td>The original construction made by Agnes.</td>
</tr>
</tbody>
</table>
5.2 Bacterial glycerol stocks for long-term storage of plasmids

Bacterial glycerol stocks are important for long-term storage of plasmids. If you want to store bacteria for a longer time, you will need to establish glycerol stocks. The addition of glycerol stabilizes the frozen bacteria, preventing damage to the cell membranes and keeping the cells alive. A glycerol stock of bacteria can be stored stably at -80°C for many years. Once this stock is established, the plasmids can be easily reproduced and purified through an overnight culture, followed by a standard midi or maxi prep.

This stock can be simply prepared by adding 500 μL of the overnight culture to 500 μL of 50% glycerol in a 2 mL screw top tube or cryovial and gently mix. To recover bacteria from your glycerol stock, open the tube and use a sterile loop, toothpick or pipette tip to scrape some of the frozen bacteria of the top. Do not let the glycerol stock unthaw. The optimal concentration of long-term glycerol storage is unknown. Most labs store bacteria in 15-25% glycerol. In my condition, I used a 25% glycerol concentration for making the stocks, as summarized below in Table V2.
Table V2. Available bacterial glycerol stocks in the lab.

<table>
<thead>
<tr>
<th>Bac location @ Bac.box</th>
<th>Name</th>
<th>Plasmid ID</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>V137 [pCMVR8.7d]</td>
<td>[J1-E6]</td>
<td>10 octobre 2018</td>
</tr>
<tr>
<td>2</td>
<td>V138 [pMDG2]</td>
<td>[J1-E7]</td>
<td>10 octobre 2018</td>
</tr>
<tr>
<td>3</td>
<td>pcdna3-VN-HOXA9</td>
<td>[J1-A1]</td>
<td>11 novembre 2018</td>
</tr>
<tr>
<td>4</td>
<td>plix-Kozak-CC-Gateway</td>
<td>[J1-F1]</td>
<td>1 janvier 2019</td>
</tr>
<tr>
<td>5</td>
<td>plix-VN-MCS-BleoR</td>
<td>[J1-F1]</td>
<td>2 décembre 2019</td>
</tr>
<tr>
<td>6</td>
<td>plix-CN-MCS-BleoR</td>
<td>[J1-F2]</td>
<td>2 décembre 2019</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Plix-CN-Myc-BleoR</td>
<td>[J1-F5]</td>
<td>3 juillet 2019</td>
</tr>
<tr>
<td>10</td>
<td>plix-Kozak-VN-HOXA1</td>
<td>[J1-G2]</td>
<td>21 mai 2019</td>
</tr>
<tr>
<td>13</td>
<td>plix-HA-HOXA1</td>
<td>[J1-G5]</td>
<td>21 mai 2019</td>
</tr>
<tr>
<td>15</td>
<td>plix-HA-HOXC6_iso2</td>
<td>[J1-G7]</td>
<td>21 mai 2019</td>
</tr>
<tr>
<td>16</td>
<td>pDONR223</td>
<td>[J1-A5]</td>
<td>19 juin 2019</td>
</tr>
</tbody>
</table>
5.3 Digested vectors and fragments

For a quick routine cloning, all digested vectors and DNA fragments were stocked, as summarized in following Table V3.

<table>
<thead>
<tr>
<th>Barcode</th>
<th>Digested Plasmids and DNAs in Box D1</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1-A2</td>
<td>plix-Nhel_AsiSI [7671bp]</td>
<td>Verified, works well.</td>
</tr>
<tr>
<td>D1-A3</td>
<td>pcdna3-BamHI_XhoI [5390bp]</td>
<td>Verified, works well.</td>
</tr>
<tr>
<td>D1-A4</td>
<td>AsiSI-a-HOXA1-BamHI [1023bp]</td>
<td>Sequenced, all right</td>
</tr>
<tr>
<td>D1-A5</td>
<td>AsiSI-a-HOXA0-BamHI [816bp]</td>
<td>Sequenced, all right</td>
</tr>
<tr>
<td>D1-A8</td>
<td>AsiSI-A-HOXA8_BamHI [4812bp]</td>
<td>for plix-HA-HOXA8, Sequenced, all right</td>
</tr>
<tr>
<td>D1-A7</td>
<td>AsiSI-HA-HOXA1-BamHI [1057bp]</td>
<td>for plix-HA-HOXAI, Sequenced, all right</td>
</tr>
<tr>
<td>D1-A1</td>
<td>AsiSI-HA-HOXAG_Sso2-BamHI [511bp]</td>
<td>Sequenced, all right</td>
</tr>
<tr>
<td>D1-B1</td>
<td>plix_AsiSI_BamHI [7671bp]</td>
<td>Verified, works well.</td>
</tr>
<tr>
<td>D1-B3</td>
<td>pcdna3-EcoRI_XhoI [5168bp]</td>
<td>Verified, works well.</td>
</tr>
<tr>
<td>D1-B4</td>
<td>pcdna3-VN-AgiP-PCR_BamHI [816bp]</td>
<td>Verified, works well.</td>
</tr>
<tr>
<td>D1-B7</td>
<td>pcdna3_BamHI_XbaI [5367bp]</td>
<td>Verified, works well.</td>
</tr>
<tr>
<td>D1-B8</td>
<td>pcdna3-VN173-XhoI_XbaI [5367bp]</td>
<td>Verified, works well.</td>
</tr>
<tr>
<td>D1-C1</td>
<td>pcdna3-BamHI_XhoI [5367bp]</td>
<td>Verified, works well.</td>
</tr>
<tr>
<td>D1-C2</td>
<td>plix-Nhel-XhoI [5367bp]</td>
<td>Verified, works well.</td>
</tr>
<tr>
<td>D1-C3</td>
<td>plix_AgiP_BamHI [7671bp]</td>
<td>Verified, works well.</td>
</tr>
<tr>
<td>D1-C4</td>
<td>plix_Nhel_BamHI [7671bp]</td>
<td>Verified, works well.</td>
</tr>
<tr>
<td>D1-C5</td>
<td>Xbal-smURP-stop-Xbal [816bp]</td>
<td>For making the T2A-smURP</td>
</tr>
<tr>
<td>D1-C8</td>
<td>Asp1-T2A-Xbal-smURP-stop-BamHI [816bp]</td>
<td>Verified, works well.</td>
</tr>
<tr>
<td>D1-C7</td>
<td>Nhel-Kozak-VN173-GSGSG-HOXA8 no stop Asp1 [5571bp]</td>
<td>Verified, works well.</td>
</tr>
<tr>
<td>D1-C6</td>
<td>Nhel-Kozak-VN173-GSGSG-HOXA8 no stop Asp1 [2582bp]</td>
<td>Verified, works well.</td>
</tr>
<tr>
<td>D1-C9</td>
<td>AslSI-HA-HOXA1-BamHI [1057bp]</td>
<td>for plix-HA-HOXA1_BamHI</td>
</tr>
<tr>
<td>D1-D1</td>
<td>AslSI-HA-HOXA1-BamHI [816bp]</td>
<td>for plix-HA-HOXA1_BamHI</td>
</tr>
<tr>
<td>D1-D2</td>
<td>Xbal-mCherry-T2A-XbaI [789bp]</td>
<td>for plix-HA-HOXA1_BamHI</td>
</tr>
<tr>
<td>D1-D4</td>
<td>EcOAI-EGFP-stop-XhoI</td>
<td>Verified, works well.</td>
</tr>
<tr>
<td>D1-D5</td>
<td>plix-VN-HOXA1-PCR-Bleor-XbaI [8131bp]</td>
<td>Verified, works well.</td>
</tr>
<tr>
<td>D1-D6</td>
<td>pcdna3-BamHI-XbaI [2582bp]</td>
<td>for plix-HA-HOXA8_BamHI</td>
</tr>
<tr>
<td>D1-D7</td>
<td>XhoI-PBXS1-V5-stop-Xbal [5367bp]</td>
<td>for plix-HA-HOXA8_BamHI</td>
</tr>
<tr>
<td>D1-D8</td>
<td>pcdna3-HA-HOXA9-linker-sGFP-P11-T2A-sGFP-P10-linker-XhoI-XbaI [5367bp]</td>
<td>for plix-HA-HOXA9-linker-sGFP-P11-T2A-sGFP-P10-linker-XhoI-XbaI</td>
</tr>
<tr>
<td>D1-D9</td>
<td>AslSI-HA-HOXA9-GSGSG-EcoRI</td>
<td>for plix-HA-HOXA9-GSGSG-EcoRI</td>
</tr>
<tr>
<td>D1-E1</td>
<td>AslSI-HA-HOXA9-GSGSG-EcoRI [816bp]</td>
<td>for plix-HA-HOXA9-GSGSG-EcoRI</td>
</tr>
<tr>
<td>D1-E2</td>
<td>plix-403-Ashl_EcoOIBlack-HA</td>
<td>for plix-HA-HOXA9-GSGSG-EcoRI</td>
</tr>
<tr>
<td>D1-E4</td>
<td>plix-403-Ashl_BamHI</td>
<td>for plix-HA-HOXA9-GSGSG-EcoRI</td>
</tr>
<tr>
<td>D1-E5</td>
<td>AslSI-HA-HOXA8-stop-BamHI</td>
<td>for plix-HA-HOXA8-stop-BamHI</td>
</tr>
<tr>
<td>D1-E6</td>
<td>pTRE-Tight-M-ProteinA-EcoOIBlack-XbaI</td>
<td>for plix-HA-HOXA8-stop-BamHI</td>
</tr>
<tr>
<td>D1-E7</td>
<td>pTRE-Tight-B-Xmal</td>
<td>for plix-HA-HOXA8-stop-BamHI</td>
</tr>
<tr>
<td>D1-E8</td>
<td>pTRE-Tight-B-Xmal</td>
<td>for plix-HA-HOXA8-stop-BamHI</td>
</tr>
<tr>
<td>D1-E9</td>
<td>plix_AsSI_BamHI-PBXS1-XbaI</td>
<td>for plix-HA-HOXA8-stop-BamHI</td>
</tr>
<tr>
<td>D1-F1</td>
<td>AslSI-BamHI-T2A-smCGBP-BamHI</td>
<td>for plix-HA-HOXA8-stop-BamHI</td>
</tr>
<tr>
<td>D1-F2</td>
<td>pcdna3-Nhel_Pac-linker-sGFP-P11-T2A-sGFP-P10-linker-ProteinB</td>
<td>for plix-HA-HOXA8-stop-BamHI</td>
</tr>
<tr>
<td>D1-F4</td>
<td>plix-PX3-PCR-stop-BamHI [1048bp]</td>
<td>for plix-HA-HOXA8-stop-BamHI</td>
</tr>
<tr>
<td>D1-F5</td>
<td>AslSI-C-T2A-BamHI [822bp]</td>
<td>for plix-HA-HOXA8-stop-BamHI</td>
</tr>
<tr>
<td>D1-F6</td>
<td>plix-Kozak-CC-AsSI_BamHI</td>
<td>for plix-HA-HOXA8-stop-BamHI</td>
</tr>
<tr>
<td>D1-F7</td>
<td>AslSI-HA-MEIS1_BamHI [1210bp]</td>
<td>for plix-HA-HOXA8-stop-BamHI</td>
</tr>
<tr>
<td>D1-F8</td>
<td>plix-Kozak-CN-AsSI_BamHI</td>
<td>for plix-HA-HOXA8-stop-BamHI</td>
</tr>
<tr>
<td>D1-F9</td>
<td>Nhel-HA-HOXA9-no stop-Pac [845bp]</td>
<td>for plix-HA-HOXA9-no stop-Pac</td>
</tr>
<tr>
<td>D1-G1</td>
<td>Nhel-Kozak-CC-GSGSG-AsI [295bp]</td>
<td>for plix-HA-HOXA9-no stop-Pac</td>
</tr>
<tr>
<td>D1-G2</td>
<td>AslSI-NLsTURBID-NA_smCGBP-BamHI</td>
<td>for plix-HA-HOXA9-no stop-Pac</td>
</tr>
<tr>
<td>D1-G3</td>
<td>Nhel-NLsTURBID-NA-smCGBP-BamHI</td>
<td>for plix-HA-HOXA9-no stop-Pac</td>
</tr>
<tr>
<td>D1-G4</td>
<td>plix-Nhel-Ashi_HA-HOXA9</td>
<td>for plix-HA-HOXA9-no stop-Pac</td>
</tr>
<tr>
<td>D1-G5</td>
<td>Nhel-kozak-2DNA-AslSI [annealed oligos]</td>
<td>for plix-HA-HOXA9-no stop-Pac</td>
</tr>
</tbody>
</table>

For plix-HA-HOXA9 for generating plix-Kozak-2DNA-HA-HOXA9
### Table V4. All the cell lines used or made during PhD, as summarized below.

<table>
<thead>
<tr>
<th>Barcode</th>
<th>Cell line</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1_2</td>
<td>HEla</td>
<td>2019/6/27</td>
</tr>
<tr>
<td>R1_3</td>
<td>HEla</td>
<td>2019/6/27</td>
</tr>
<tr>
<td>R1_4</td>
<td>HEla</td>
<td>2019/6/27</td>
</tr>
<tr>
<td>R1_5</td>
<td>lentix-HEK</td>
<td>2020/2/9</td>
</tr>
<tr>
<td>R1_6</td>
<td>lentix-HEK</td>
<td>2020/2/9</td>
</tr>
<tr>
<td>R1_7</td>
<td>lentix-HEK</td>
<td>2020/2/9</td>
</tr>
<tr>
<td>R1_8</td>
<td>lentix-HEK</td>
<td>2020/2/9</td>
</tr>
<tr>
<td>R1_9</td>
<td>lentix-HEK</td>
<td>2020/2/9</td>
</tr>
<tr>
<td>R2_10</td>
<td>lentix-HEK</td>
<td>2020/2/9</td>
</tr>
<tr>
<td>R2_11</td>
<td>lentix-HEK</td>
<td>2020/2/9</td>
</tr>
<tr>
<td>R2_12</td>
<td>lentix-HEK</td>
<td>2020/2/9</td>
</tr>
<tr>
<td>R2_13</td>
<td>lentix-HEK</td>
<td>2020/2/9</td>
</tr>
<tr>
<td>R2_14</td>
<td>lentix-HEK</td>
<td>2020/2/9</td>
</tr>
<tr>
<td>R2_15</td>
<td>HEK [ONET]</td>
<td>2020/6/9</td>
</tr>
<tr>
<td>R2_16</td>
<td>HEK [ONET]</td>
<td>2020/6/9</td>
</tr>
<tr>
<td>R2_17</td>
<td>HEK [ONET]</td>
<td>2020/6/9</td>
</tr>
<tr>
<td>R2_18</td>
<td>HEK [ONET]</td>
<td>2020/6/9</td>
</tr>
<tr>
<td>R3_22</td>
<td>MDA-MB-231 [multiple-plex-VN-HOXA9-BeloR]</td>
<td>2019/12/30</td>
</tr>
<tr>
<td>R3_23</td>
<td>MDA-MB-231 [wt]</td>
<td>2019/2/15</td>
</tr>
<tr>
<td>R3_24</td>
<td>HEK-293T [plex-VN-HOXA9-PGK-BleoR-mCherry][J3-H7]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R3_25</td>
<td>HEK-293T [plex-VN-HOXA9-PGK-BleoR-mCherry][J3-H7]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R3_26</td>
<td>HEK-293T [plex-VN-HOXA9-PGK-BleoR-mCherry][J3-H7]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R3_27</td>
<td>HEK-293T [plex-VN-HOXA9-PGK-BleoR-mCherry][J3-H7]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R4_28</td>
<td>MDA-MB-231 [ONET]</td>
<td>2018/5/5</td>
</tr>
<tr>
<td>R4_29</td>
<td>MDA-MB-231 [ONET]</td>
<td>2018/5/5</td>
</tr>
<tr>
<td>R4_30</td>
<td>MDA-MB-231 [ONET]</td>
<td>2018/5/5</td>
</tr>
<tr>
<td>R4_31</td>
<td>MDA-MB-231 [ONET]</td>
<td>2018/5/5</td>
</tr>
<tr>
<td>R4_32</td>
<td>HEK-293T [plex-TurboID-HA-csGBP][J4-F4]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R4_33</td>
<td>HEK-293T [plex-TurboID-HA-csGBP][J4-F4]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R4_34</td>
<td>HEK-293T [plex-TurboID-HA-csGBP][J4-F4]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R4_35</td>
<td>HEK-293T [plex-TurboID-HA-csGBP][J4-F4]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R4_36</td>
<td>HEK-293T [plex-TurboID-HA-csGBP][J4-F4]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R5_37</td>
<td>HEK-293T [plex-TurboID-HA-csGBP][J4-F4]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R5_38</td>
<td>HEK-293T [plex-TurboID-HA-csGBP][J4-F4]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>Barcode</td>
<td>Cell lines</td>
<td>Comment</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
<td>---------</td>
</tr>
<tr>
<td>R5_39</td>
<td>HEK[wt]</td>
<td>2019/8/30</td>
</tr>
<tr>
<td>R5_40</td>
<td>HEK[wt]</td>
<td>2019/8/30</td>
</tr>
<tr>
<td>R5_41</td>
<td>HEK[wt]</td>
<td>2019/8/30</td>
</tr>
<tr>
<td>R5_42</td>
<td>HEK[wt]</td>
<td>2019/8/30</td>
</tr>
<tr>
<td>R5_43</td>
<td>HEK[wt]</td>
<td>2019/8/30</td>
</tr>
<tr>
<td>R5_44</td>
<td>HEK[wt]</td>
<td>2019/8/30</td>
</tr>
<tr>
<td>R6_46</td>
<td>HEK-293T [plix-VN-HOXA9-PGK-BleoR-mCherry][J3-H7]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R6_47</td>
<td>HEK-293T [plix-VN-HOXA9-PGK-BleoR-mCherry][J3-H7]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R6_48</td>
<td>HEK-293T [plix-VN-HOXA9-PGK-BleoR-mCherry][J3-H7]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R6_49</td>
<td>HEK-293T [plix-VN-HOXA9-PGK-BleoR-mCherry][J3-H7]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R6_50</td>
<td>HEK-293T [plix-VN-HOXA9-PGK-BleoR-mCherry][J3-H7]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R6_51</td>
<td>HEK-293T [plix-VN-HOXA9-PGK-BleoR-mCherry][J3-H7]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R6_52</td>
<td>HEK-293T [plix-VN-HOXA9-PGK-BleoR-mCherry][J3-H7]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R6_53</td>
<td>HEK-293T [plix-VN-HOXA9-PGK-BleoR-mCherry][J3-H7]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R6_54</td>
<td>HEK-293T [plix-VN-HOXA9-PGK-BleoR-mCherry][J3-H7]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R7_55</td>
<td>HEK[wt]</td>
<td>2018/10/26</td>
</tr>
<tr>
<td>R7_56</td>
<td>HEK[wt]</td>
<td>2018/10/25</td>
</tr>
<tr>
<td>R7_57</td>
<td>HEK[wt]</td>
<td>2018/10/25</td>
</tr>
<tr>
<td>R7_58</td>
<td>HEK[wt]</td>
<td>2018/10/25</td>
</tr>
<tr>
<td>R7_59</td>
<td>HEK[wt]</td>
<td>2018/10/25</td>
</tr>
<tr>
<td>R7_60</td>
<td>HEK[wt]</td>
<td>2018/10/25</td>
</tr>
<tr>
<td>R7_61</td>
<td>HEK[wt]</td>
<td>2018/10/25</td>
</tr>
<tr>
<td>R7_62</td>
<td>HEK[wt]</td>
<td>2018/10/25</td>
</tr>
<tr>
<td>R8_63</td>
<td>lentiv-HEK</td>
<td>2018/08/22</td>
</tr>
<tr>
<td>R8_64</td>
<td>lentiv-HEK</td>
<td>2018/08/22</td>
</tr>
<tr>
<td>R8_65</td>
<td>lentiv-HEK</td>
<td>2018/08/22</td>
</tr>
<tr>
<td>R8_66</td>
<td>HEK-293T [plix-VN-HOXA9-PGK-BleoR-mCherry][J3-H7]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R8_67</td>
<td>HEK-293T [plix-VN-HOXA9-PGK-BleoR-mCherry][J3-H7]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R8_68</td>
<td>HEK-293T [plix-VN-HOXA9-PGK-BleoR-mCherry][J3-H7]</td>
<td>2020/12/15</td>
</tr>
<tr>
<td>R8_69</td>
<td>HEK[plix-VN-Myc]</td>
<td>2019/3/18</td>
</tr>
<tr>
<td>R8_70</td>
<td>HEK[plix-VN-Myc]</td>
<td>2019/3/18</td>
</tr>
<tr>
<td>R8_71</td>
<td>HEK[plix-VN-Myc]</td>
<td>2019/3/18</td>
</tr>
<tr>
<td>R9_72</td>
<td>lentiv-HEK</td>
<td>2018/05/09</td>
</tr>
<tr>
<td>R9_73</td>
<td>lentiv-HEK</td>
<td>2018/05/09</td>
</tr>
<tr>
<td>R9_74</td>
<td>lentiv-HEK</td>
<td>2018/05/09</td>
</tr>
<tr>
<td>R9_75</td>
<td>lentiv-HEK</td>
<td>2018/05/09</td>
</tr>
<tr>
<td>R9_76</td>
<td>lentiv-HEK</td>
<td>2018/05/09</td>
</tr>
<tr>
<td>R9_77</td>
<td>lentiv-HEK</td>
<td>2018/05/09</td>
</tr>
<tr>
<td>R9_78</td>
<td>HEK-293T [plix-TurboID-HA-csGBP][J4-F4]</td>
<td>2020/12/16</td>
</tr>
<tr>
<td>R9_79</td>
<td>HEK-293T [plix-TurboID-HA-csGBP][J4-F4]</td>
<td>2020/12/16</td>
</tr>
<tr>
<td>R9_80</td>
<td>HEK-293T [plix-TurboID-HA-csGBP][J4-F4]</td>
<td>2020/12/16</td>
</tr>
<tr>
<td>R9_81</td>
<td>HEK-293T [plix-TurboID-HA-csGBP][J4-F4]</td>
<td>2020/12/16</td>
</tr>
</tbody>
</table>
References


Waldo, G.S., and Cabantous, S. (2010). Protein-protein interaction detection system using fluorescent protein microdomains (Los Alamos National Lab. (LANL), Los Alamos, NM (United States)).


