

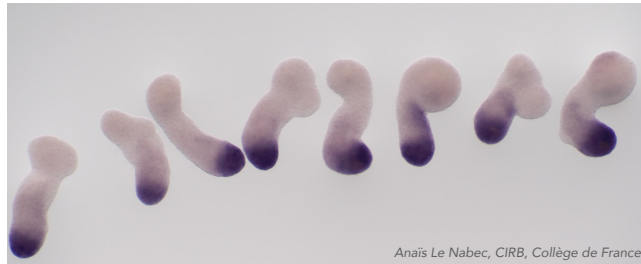


COLLÈGE
DE FRANCE
—1530—

Collège de France; Chaire:
Evolution du Développement et des Génomes
Denis.Duboule@college-de-france.fr

Cours 2025: Régulation des Gènes du Développement: Les Séquences 'Enhancers'

Leçon 3, vendredi 7 mars 2025



Anais Le Nabec, CIRB, Collège de France

@denisduboule.blsk.social

1

Denis Duboule/2025



COLLÈGE
DE FRANCE
—1530—

Régulation des Gènes du Développement

Approches pour identifier des enhancers

*Comment trouver et valider fonctionnellement des enhancers?

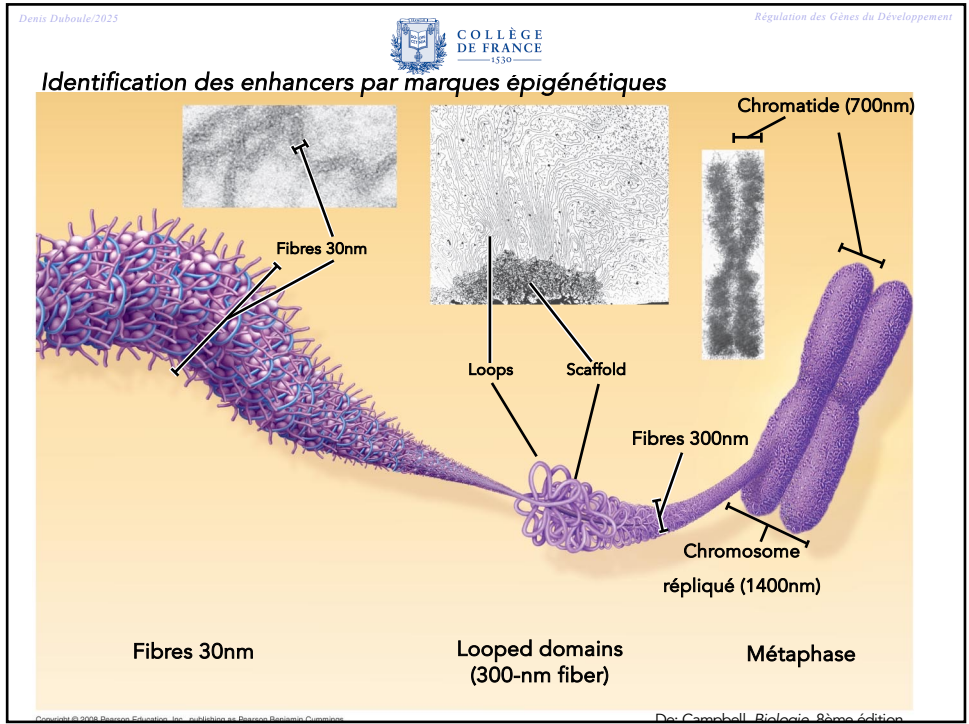
Historiquement:

- 1) Comparaisons de séquences
- 2) Pièges à enhancers, STARR-seq, trajectoires de régulations...
- 3) Marques épigénétiques
- 4) Accessibilité de la chromatine
- 5) Contacts inter-chromatine
- 6)

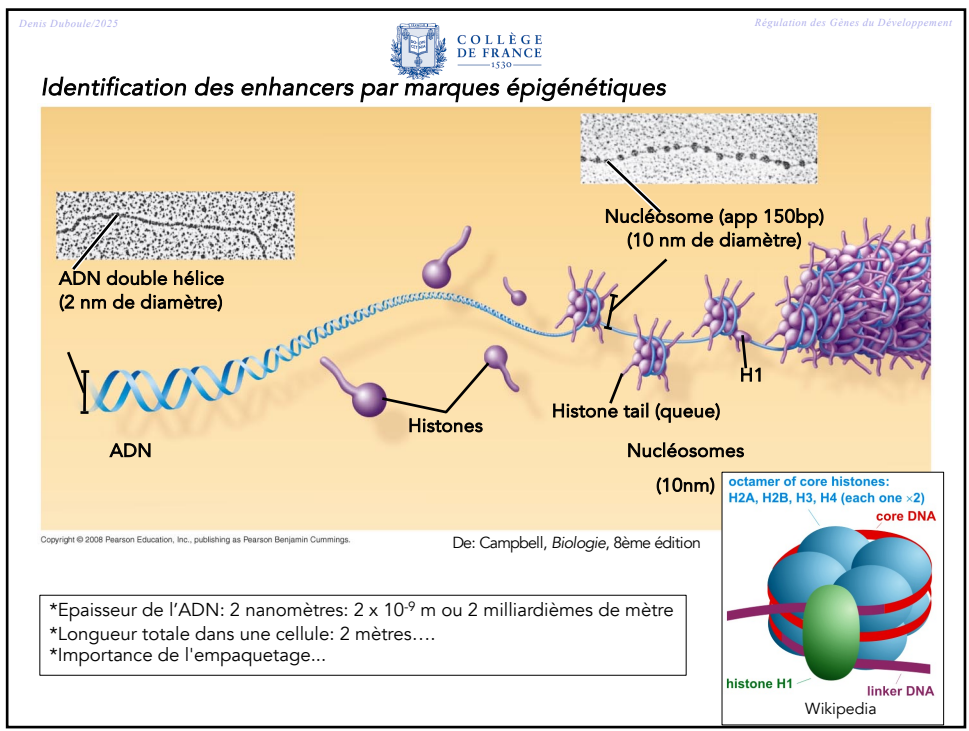


Méthodologies à très hauts débits
(dont des combinaisons variées font
souvent office de tests fonctionnels...)

2



3



4

Modifications épigénétiques de la chromatine (Histones)

- *Modifications post-traduction, faites par des enzymes (réversibles)
- *Modifications associées à des états différents de la chromatine (active, inactive..)
- *Modifications utilisées pour évaluer l'état fonctionnel de séquence d'ADN

Lowe et al., 2019 Cancers (11) 660



H3K27ac → Chromatine active (gènes transcripts, enhancers..)

H3K27me1

H3K27me2

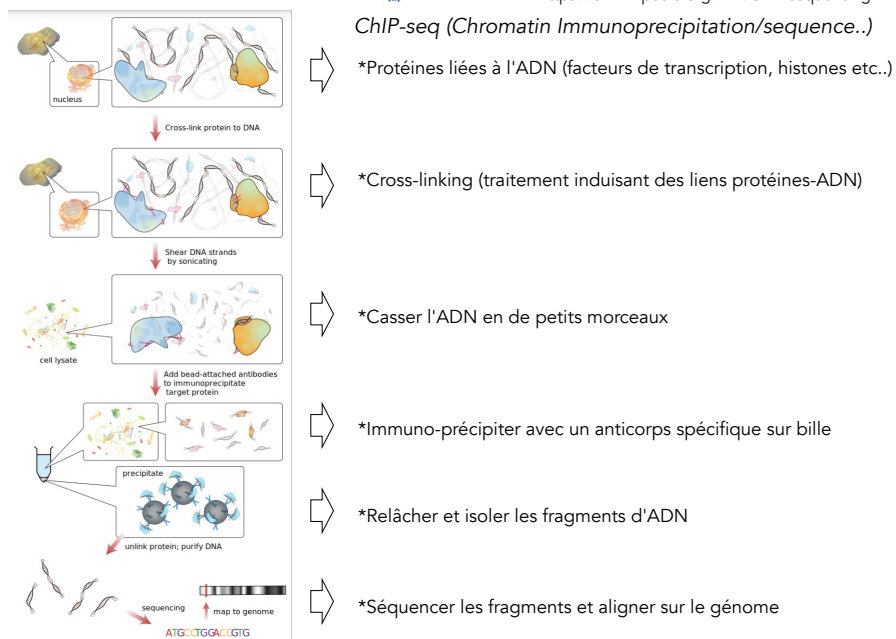
H3K27me3 → Chromatine inactive (gènes non-transcripts)

- *Il existe des anticorps qui reconnaissent ces modifications de façon spécifique
- *L'utilisation de ces anticorps associée au séquençage ADN à haut débit permet de localiser les endroits où ces modifications se trouvent (ChIP-seq)

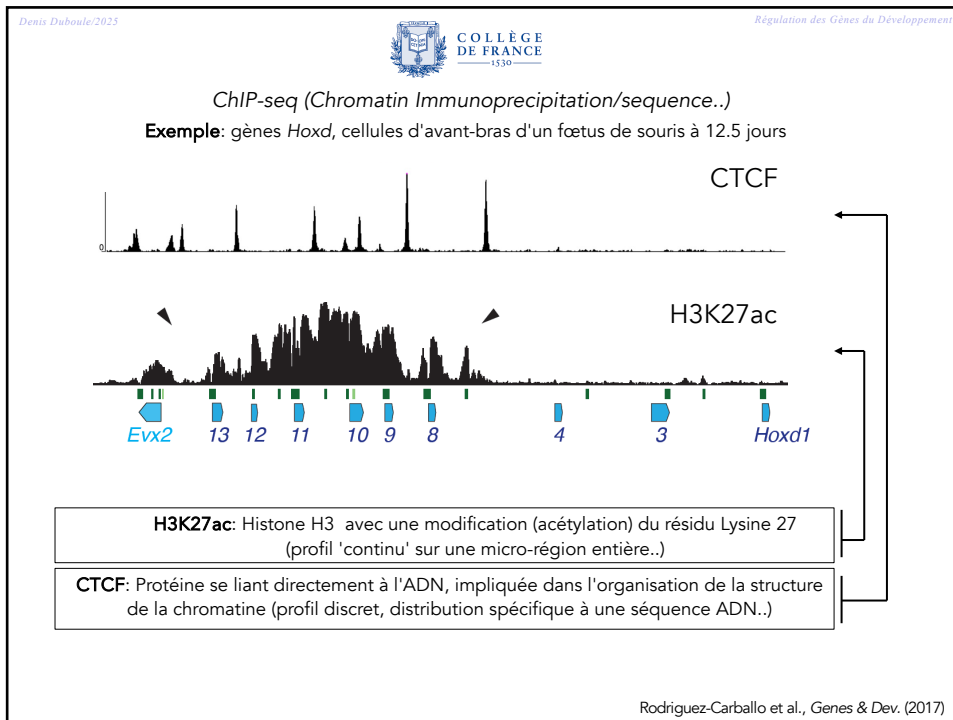
5

<https://en.wikipedia.org/wiki/ChIP-seq>

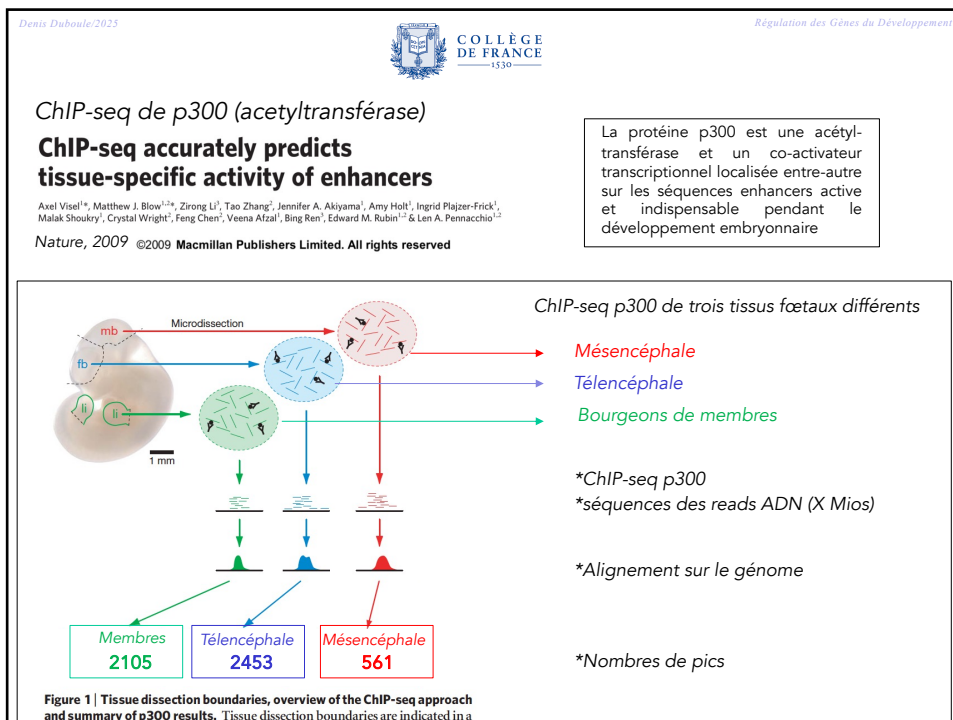
ChIP-seq (Chromatin Immunoprecipitation/sequence..)



6



7



8

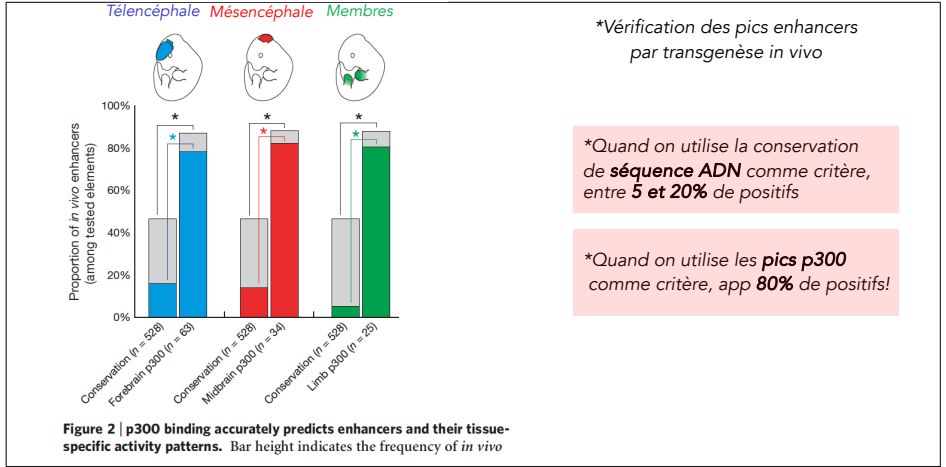
ChIP-seq de p300 (acétyltransférase)

ChIP-seq accurately predicts tissue-specific activity of enhancers

Axel Visel^{1*}, Matthew J. Blow^{1,2*}, Zirong Li¹, Tao Zhang², Jennifer A. Akiyama¹, Amy Holt¹, Ingrid Plajzer-Frick¹, Malak Shoukry¹, Crystal Wright², Feng Chen², Veena Afzal¹, Bing Ren¹, Edward M. Rubin^{1,2} & Len A. Pennacchio^{1,2}

Nature, 2009©2009 Macmillan Publishers Limited. All rights reserved

La protéine p300 est une acétyltransférase et un co-activateur transcriptionnel localisée entre-autre sur les séquences enhancers active et indispensable pendant le développement embryonnaire



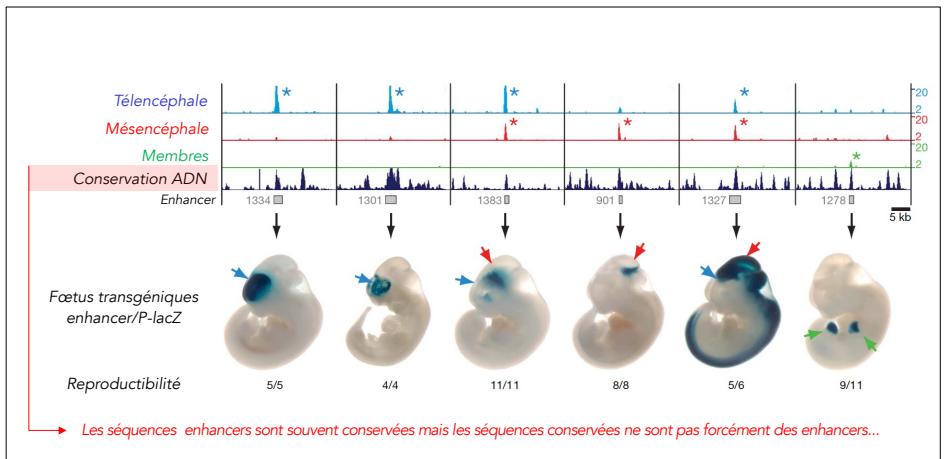
ChIP-seq de p300 (acétyltransférase)

ChIP-seq accurately predicts tissue-specific activity of enhancers

Axel Visel^{1*}, Matthew J. Blow^{1,2*}, Zirong Li¹, Tao Zhang², Jennifer A. Akiyama¹, Amy Holt¹, Ingrid Plajzer-Frick¹, Malak Shoukry¹, Crystal Wright², Feng Chen², Veena Afzal¹, Bing Ren¹, Edward M. Rubin^{1,2} & Len A. Pennacchio^{1,2}

Nature, 2009©2009 Macmillan Publishers Limited. All rights reserved

La protéine p300 est une acétyltransférase et un co-activateur transcriptionnel localisée entre-autre sur les séquences enhancers active et indispensable pendant le développement embryonnaire



ChIP-seq de p300 (acétyltransférase)

ChIP-seq accurately predicts tissue-specific activity of enhancers

Axel Visel^{1*}, Matthew J. Blow^{1,2*}, Zirong Li³, Tao Zhang³, Jennifer A. Akiyama⁴, Amy Holt¹, Ingrid Plajzer-Frick¹, Malak Shoukry¹, Crystal Wright⁵, Feng Chen⁶, Veena Afzal⁷, Bing Ren⁸, Edward M. Rubin^{1,2} & Len A. Pennacchio^{1,2}

Nature, 2009 ©2009 Macmillan Publishers Limited. All rights reserved

Conclusions

A major yet unresolved quest in decoding the human genome is the identification of the regulatory sequences that control the spatial and temporal expression of genes. Distant-acting transcriptional enhancers are particularly challenging to uncover because they are scattered among the vast non-coding portion of the genome. Evolutionary sequence constraint can facilitate the discovery of enhancers, but fails to predict when and where they are active *in vivo*. Here we present the results of chromatin immunoprecipitation with the enhancer-associated protein p300 followed by massively parallel sequencing, and map several thousand *in vivo* binding sites of p300 in mouse embryonic forebrain, midbrain and limb tissue. We tested 86 of these sequences in a transgenic mouse assay, which in nearly all cases demonstrated reproducible enhancer activity in the tissues that were predicted by p300 binding. Our results indicate that *in vivo* mapping of p300 binding is a highly accurate means for identifying enhancers and their associated activities, and suggest that such data sets will be useful to study the role of tissue-specific enhancers in human biology and disease on a genome-wide scale.



...nos résultats indiquent que la cartographie *in vivo* des sites de liaison de p300 est un moyen précis pour identifier les enhancers et leurs activités. Ces résultats suggèrent que de tels datasets seront utiles pour étudier le rôle des enhancers tissu-spécifiques dans la biologie humaine normale et dans certaines pathologies, à l'échelle du génome entier.

11

La chromatine présente des 'états d'accessibilité' différents

- *Chromatine fermée: Les facteurs n'accèdent pas aux séquences d'ADN
- *Chromatine permissive: Des facteurs peuvent accéder et réorganiser la chromatine
- *Chromatine ouverte: la transcription peut s'accomplir

Ces états peuvent être révélés par l'accessibilité d'un transposon Tn5

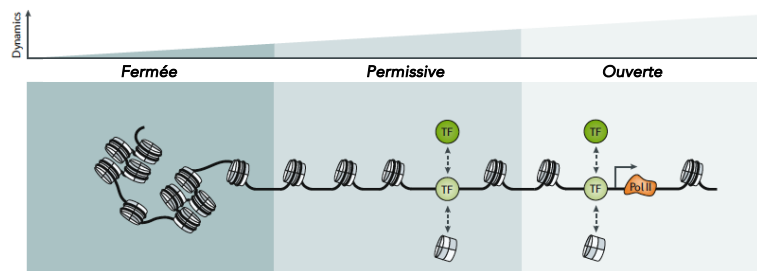
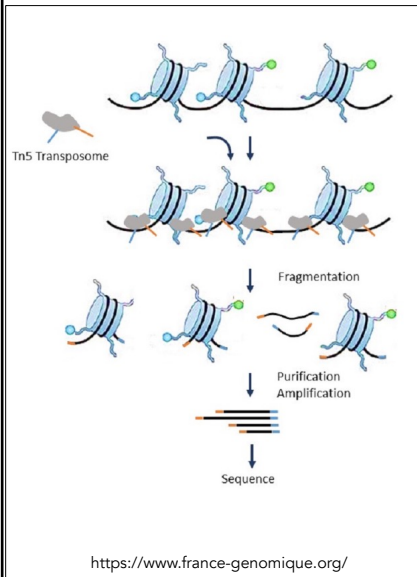


Fig. 1 | A continuum of accessibility states broadly reflects the distribution of chromatin dynamics across the genome. In contrast to closed chromatin, permissive chromatin is sufficiently dynamic for transcription factors to initiate sequence-specific accessibility remodeling and establish an open chromatin conformation (illustrated here for an active gene locus). Pol II, RNA polymerase II; TF, transcription factor.

Klemm, Shipony and Greenleaf, 2019 Nature Review Genetics

12

ATAC-seq ('Assay for transposase-accessible chromatin using sequencing')



Buenrostro et al., (2013) Nature Methods 10: 1213-1218

Transposase du Tn5 mutante hyperactive

La transposase peut aller dans les régions accessibles, ouvertes de la chromatine

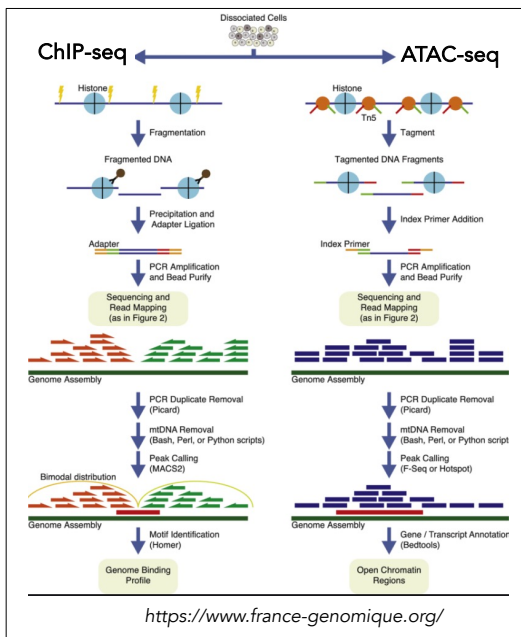
La transposase coupe et 'tague' les régions accessibles, ('tagmentation')

L'ADN coupé et tagué est purifié, séquencé et aligné sur le génome

*Pas d'anticorps, pas d'extraction, pas de sonication...
*Possibilité d'aller chercher des sites de liaison à des facteurs de transcription 'à l'intérieur' des signaux ATAC

13

Combinaison d'approches (ATAC-seq; CHIP-seq, CUT&RUN...)



La distribution des séquences d'ADN sera différente selon la technique, ce qui apporte une confirmation supplémentaire à la spécificité de la liaison

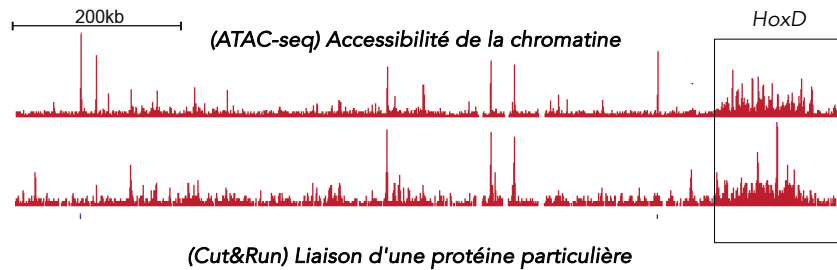
! Les enhancers détectés sont utilisés de façons différentes dans chaque type cellulaire, dans chaque organe. Par conséquent, les profils obtenus seront différents selon le type cellulaire, selon la spécificité de tel ou tel enhancer. Ce sont des profils de nature épigénétique, qui ne peuvent pas être établis une fois pour toute, comme une séquence génomique. Ces profils varient, avec l'activité des gènes.

14



Combinaison d'approches (ATAC-seq; CHIP-seq, CUT&RUN...)

Exemple: Région régulatrice (paysage) du complexe de gènes *HoxD*



La Convergence des deux analyses permet d'identifier des sites potentiels de liaisons de facteurs nécessaires au contrôle de l'expression des gènes cibles..

MAIS: Comment être certain de la coexistence des signaux dans les mêmes cellules...? Echantillonnages hétérogènes...(cellules mélangées..)

15



Combinaison d'approches (ATAC-seq; CHIP-seq, CUT&RUN...)

MAIS: Comment être certain de la coexistence des signaux dans les mêmes cellules...? Echantillonnages hétérogènes...(cellules mélangées..).

*Triage des cellules (marquages avec une fluorescence...). Quantité, types cellulaires nombreux...



*Approches par analyse de cellules uniques, en combinant deux 'read-out' différents; le scATAC-seq et le scRNA-seq. Dans cette approche 'multiomes', deux librairies sont faites à partir d'une seule cellule, avec le même barcode de façon à pouvoir associer un transcriptome à un profile ATAC.

16

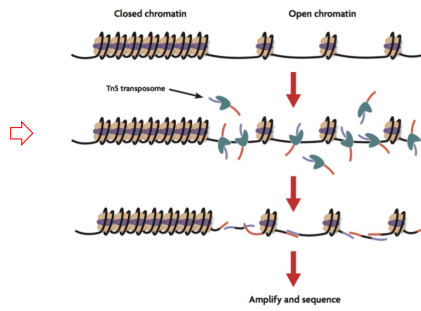
Analyse par **scATAC-seq**.

L'approche ATAC globale (X milliers-ions de cellules)

CURRENT PROTOCOLS in Molecular Biology 2015

UNIT
ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide
 Jason D. Buenrostro, Beijing Wu, Howard Y. Chang, William J. Greenleaf
 First published: 05 January 2015 | <https://doi.org/10.1002/0471142727.mb2129c109> | Citations: 1,512

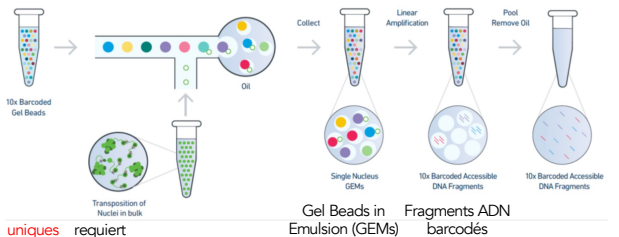
L'approche **ATACseq** mesure l'accessibilité à la chromatine, donc les endroits 'actifs' (là où des facteurs sont déposés, ou où la chromatine est modifiée etc..



17

Analyse par **scATAC-seq**.

L'approche **ATAC en cellules (noyaux) uniques**



L'approche **ATACseq en cellules uniques** requiert l'association de noyaux uniques avec des billes barcodées (micro-fluidique)



18

La fabrication d'un ADN (double brin) à partir d'un ARN (simple brin)

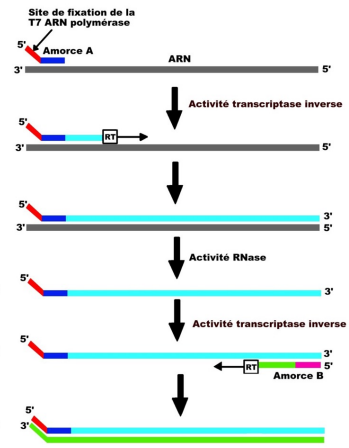
La clé: une enzyme, la transcriptase inverse:

La **transcriptase inverse** ou rétrotranscriptase (en anglais reverse transcriptase ou encore RT) est une enzyme utilisée par les rétrovirus et les rétrotransposons qui transcrivent l'information génétique des virus ou rétrotransposons de l'ARN en ADN, qui peut s'intégrer dans le génome de l'hôte.

Wikipedia
https://fr.wikipedia.org/wiki/Transcriptase_inverse

microbiologiemedicale.fr

- ARN simple brin, avec une amorce (multiples séquences..)
- Transcription inverse, ADN simple brin
- Élimination de l'ARN par digestion enzymatique
- Fabrication du brin d'ADN complémentaire



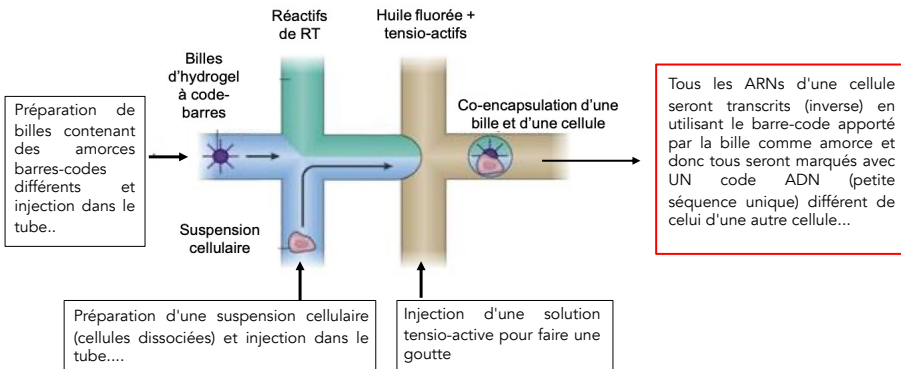
19

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Sophie FOULON
Le 10 Octobre 2019

Préparée à l'École supérieure de physique et de chimie industrielles de la ville de Paris

Stratégie générale de l'approche par microfluidique en goutte (avec un système de codes barres pour amorcer la synthèse ...).



20

L'approche 'multiomics' permet d'associer un transcriptome à un profil ATAC (donc à l'accessibilité de séquences d'ADN liées par des facteurs connus)

Une autre librairie est préparée avec l'approche ATAC-seq (deux librairies séquencées séparément mais avec les mêmes barcodes par cellule..

Après séquençage de X milliers de cellules différentes, les ADNs (ARNs ainsi que les produits ATAC...) sont triés par un logiciel en fonction de leurs barres-codes..

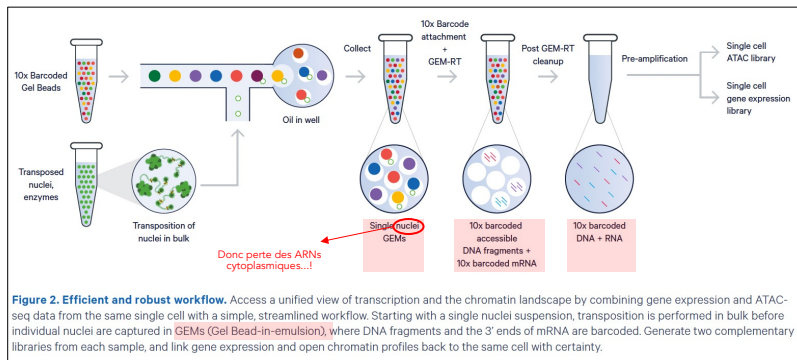


Figure 2. Efficient and robust workflow. Access a unified view of transcription and the chromatin landscape by combining gene expression and ATAC-seq data from the same single cell with a simple, streamlined workflow. Starting with a single nuclei suspension, transposition is performed in bulk before individual nuclei are captured in GEMs (Gel Bead-in-emulsion), where DNA fragments and the 3' ends of mRNA are barcoded. Generate two complementary libraries from each sample, and link gene expression and open chromatin profiles back to the same cell with certainty.



10xgenomics.com/products/epi-multiome

21

L'approche 'multiomics' permet d'associer un transcriptome à un profil ATAC (donc à l'accessibilité de séquences d'ADN liées par des facteurs connus)

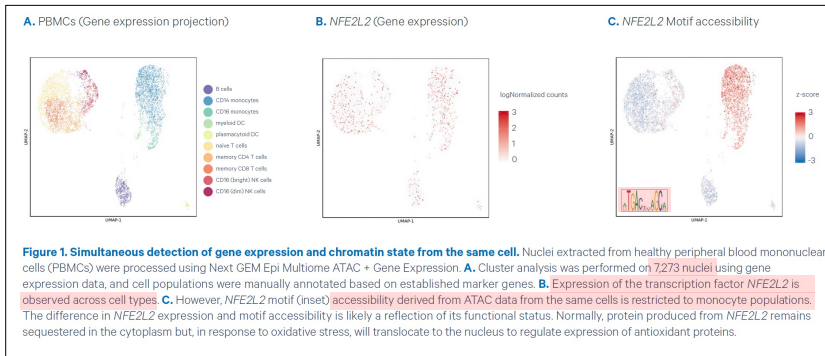
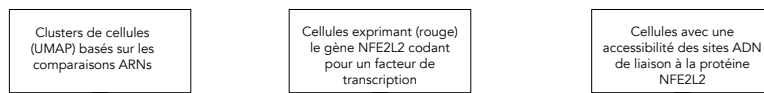


Figure 1. Simultaneous detection of gene expression and chromatin state from the same cell. Nuclei extracted from healthy peripheral blood mononuclear cells (PBMCs) were processed using Next GEM Epi Multiome ATAC + Gene Expression. **A.** Cluster analysis was performed on 7273 nuclei using gene expression data, and cell populations were manually annotated based on established marker genes. **B.** Expression of the transcription factor *NFE2L2* is observed across cell types. **C.** However, *NFE2L2* motif (inset) accessibility derived from ATAC data from the same cells is restricted to monocyte populations. The difference in *NFE2L2* expression and motif accessibility is likely a reflection of its functional status. Normally, protein produced from *NFE2L2* remains sequestered in the cytoplasm but, in response to oxidative stress, will translocate to the nucleus to regulate expression of antioxidant proteins.



10xgenomics.com/products/epi-multiome

22



Zebrahub – Multimodal Zebrafish Developmental Atlas Reveals the State Transition Dynamics of Late Vertebrate Pluripotent Axial Progenitors

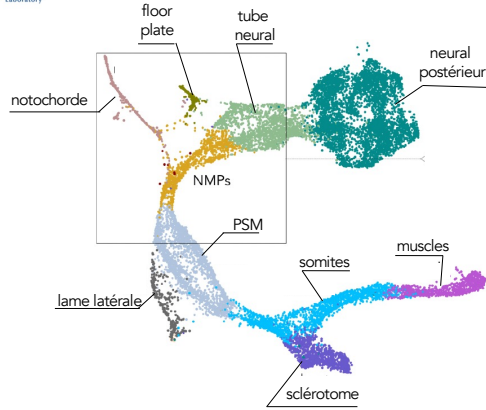


Clustering par comparaisons des ARNs

Schéma général de distribution des types cellulaires (UMAP, clustering) après séquençage d'ARNs de cellules uniques d'embryons. Pour chaque cluster (dans une expérience multiomes) on peut en principe aller rechercher les profils ATAC-seq

Merlin Lange,^{1,2} Alejandro Granados,^{1,2} Sheethi Vipin Kumar,^{1,2} Jordao Bragantini,^{1,2} Sarah Ancheba,^{1,2} Sreejith Santhosh,¹ Michael Borja,¹ Hirofumi Kobayashi,¹ Erin McGeever,¹ Ahmet Can Solak,¹ Bin Yang,¹ Xiang Zhao,¹ Yang Liu,¹ Angela Detweiler,¹ Sheryl Paul,¹ Honey Makonen,¹ Tiger Lao,¹ Rachel Banks,¹ Adrien Jacobo,¹ Koiri Balis,¹ Kyle Aweyem,¹ Samuel D'Souza,¹ Robert Hasse,¹ Alexandre Dizeux,¹ Olivier Pourquie,¹ Rafael Gómez-Sjöberg,¹ Greg Huber,¹ Mattia Serra,² Norma Neff,¹ Angela Oliveira Pisco,¹ & Loïc A. Royer^{1,2}

bioRxiv
THE PREPRINT SERVER FOR BIOLOGY



23



Zebrahub – Multimodal Zebrafish Developmental Atlas Reveals the State Transition Dynamics of Late Vertebrate Pluripotent Axial Progenitors

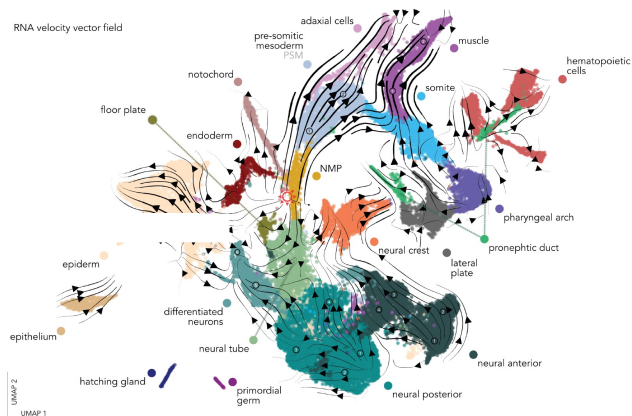


Dynamique possible des expériences multiomes

Analyse en 'RNA velocity'. Les flèches représentent les moyennes de RNA velocity de cellules uniques. Stades de 10 à 24 hpf. En principe, la dynamique des profils ATAC pourrait être extraite...

Merlin Lange,^{1,2} Alejandro Granados,^{1,2} Sheethi Vipin Kumar,^{1,2} Jordao Bragantini,^{1,2} Sarah Ancheba,^{1,2} Sreejith Santhosh,¹ Michael Borja,¹ Hirofumi Kobayashi,¹ Erin McGeever,¹ Ahmet Can Solak,¹ Bin Yang,¹ Xiang Zhao,¹ Yang Liu,¹ Angela Detweiler,¹ Sheryl Paul,¹ Honey Makonen,¹ Tiger Lao,¹ Rachel Banks,¹ Adrien Jacobo,¹ Koiri Balis,¹ Kyle Aweyem,¹ Samuel D'Souza,¹ Robert Hasse,¹ Alexandre Dizeux,¹ Olivier Pourquie,¹ Rafael Gómez-Sjöberg,¹ Greg Huber,¹ Mattia Serra,² Norma Neff,¹ Angela Oliveira Pisco,¹ & Loïc A. Royer^{1,2}

bioRxiv
THE PREPRINT SERVER FOR BIOLOGY



24



Approches pour identifier des enhancers

*Comment trouver et valider fonctionnellement des enhancers?

Historiquement:

- 1) Comparaisons de séquences
- 2) Pièges à enhancers, STARR-seq, trajectoires de régulations...
- 3) Marques épigénétiques
- 4) Accessibilité de la chromatine
- 5) Contacts inter-chromatine
- 6)



Méthodologies à très hauts débits
(dont des combinaisons variées font
souvent office de tests fonctionnels...)

25

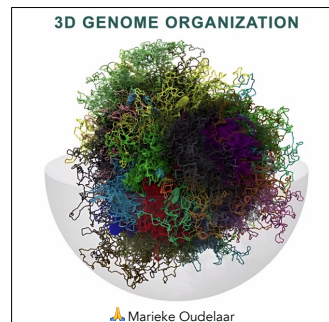


Approches pour identifier des enhancers

*Comment trouver et valider fonctionnellement des enhancers?

Historiquement:

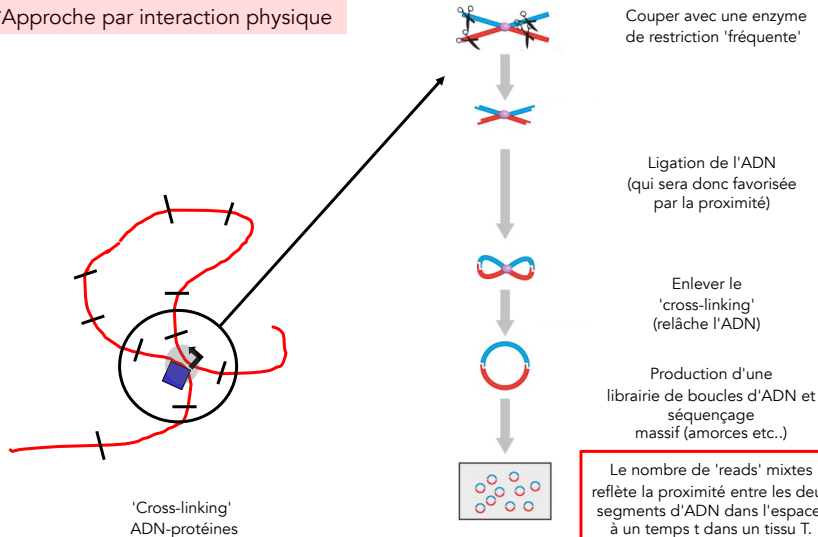
- 1) Comparaisons de séquences
- 2) Pièges à enhancers, STARR-seq, trajectoires de régulations...
- 3) Marques épigénétiques
- 4) Accessibilité de la chromatine
- 5) Contacts inter-chromatine
- 6)



26

Capture de conformation chromosomique

*Approche par interaction physique



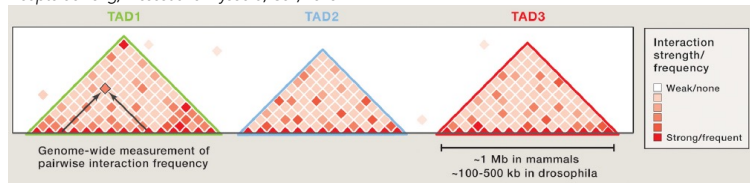
Adapté de Dekker and Misteli, Cold Spring Harb Perspect Biol, 2015

27

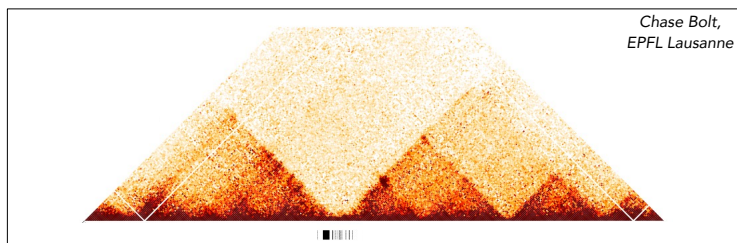
Capture de conformation chromosomique (3C, 4C, 5C, HiC, Capture HiC, Micro-C...)

3C: A contre B
4C: A contre tout
HiC: Tout contre tout

Adapté de Long, Prescott and Wysocka, Cell, 2016

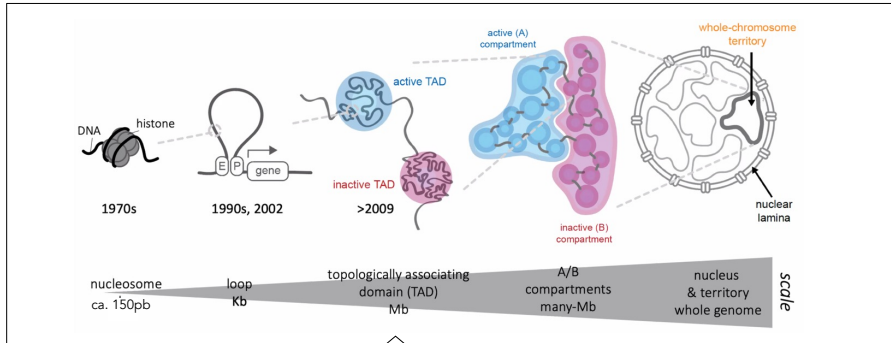


*Le nombre de reads est représenté par l'intensité de la couleur
*Une carte des interactions au niveau du génome entier (ici sur 2-3 mégabases)



28

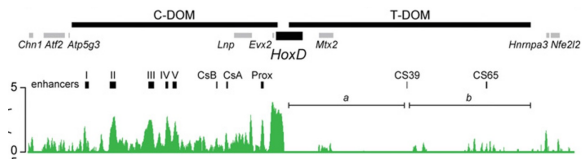
Organisation (repléments multiples) de l'ADN (chromatine) dans le noyau



Une grande partie de la restriction des enhancers pour leurs gènes cibles semble se jouer à l'intérieur des domaines topologiques TADs

Job Dekker, Nancy Kleckner et collègues... source: Susan Mango

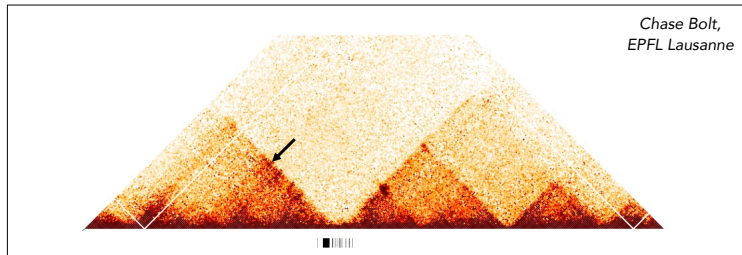
29



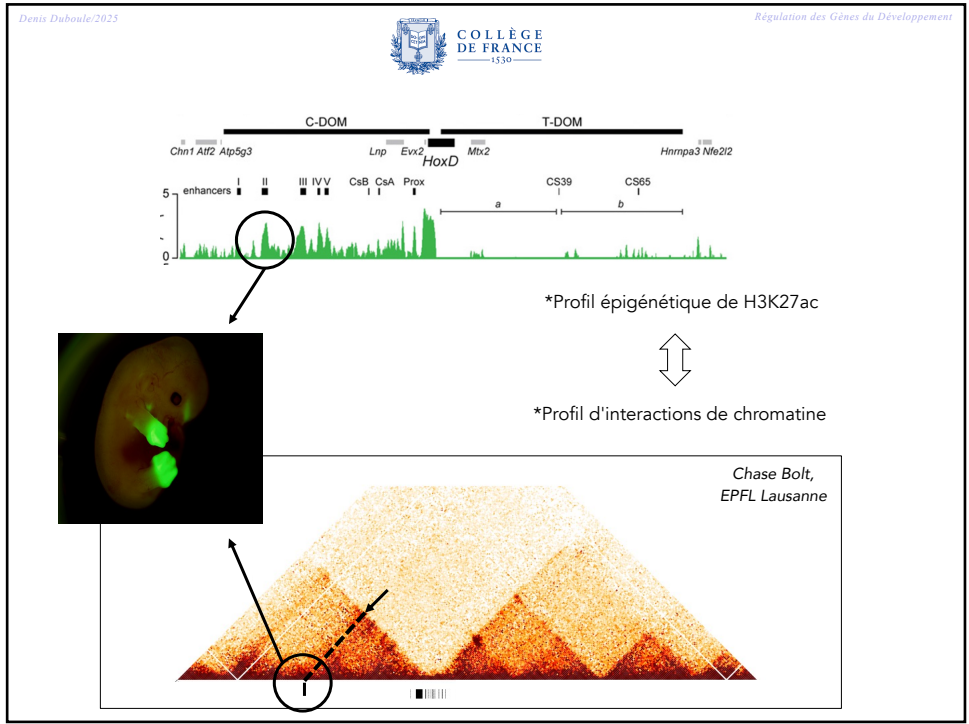
*Profil épigénétique de H3K27ac



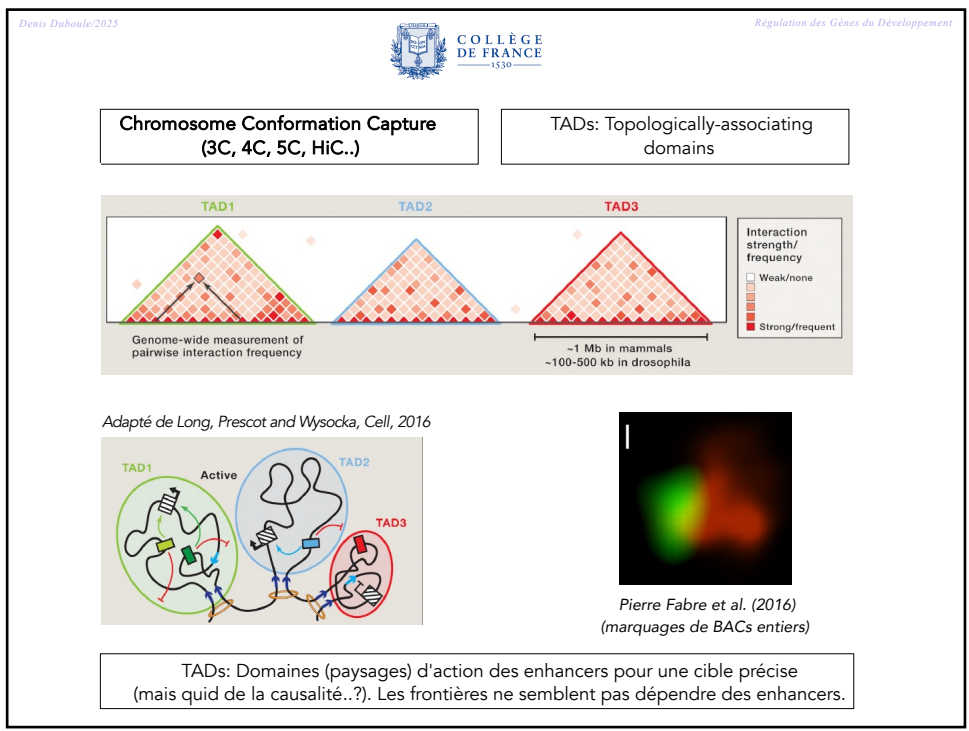
*Profil d'interactions de chromatine



30



31

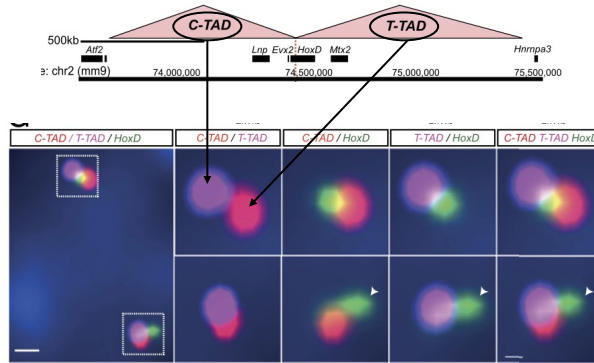


32

Visualisation des TADs par microscopie (DNA FISH)

Nanoscale spatial organization of the *HoxD* gene cluster in distinct transcriptional states

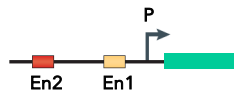
Pierre J. Fabre^a, Alexander Benke^b, Elisabeth Joye^a, Thi Hanh Nguyen Huynh^c, Sullana Manley^{b,1}, and Denis Duboule^{a,c,1}
13964-13969 | PNAS | November 10, 2015 | vol. 112 | no. 45



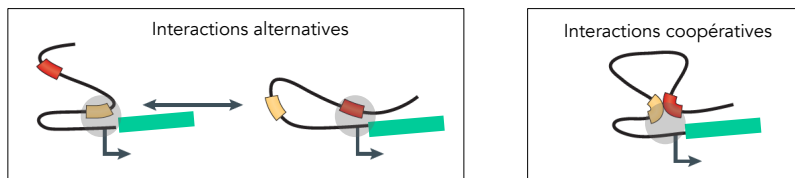
33

Mais comment fonctionne un enhancer à une telle distance du gène cible, à l'intérieur d'un domaine TAD?
Par un rapprochement non-stochastique dans l'espace (formation de 'loops')

P: Promoteur
En1: Enhancer 1
En2: Enhancer 2



Adapté de Spitz and Furlong, Nature Rev. Genetics, 2012

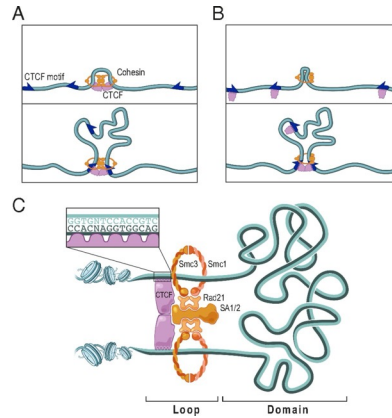


En1 et En2 peuvent avoir des spécificités cellulaires différentes ou similaires, selon les cas (pleiotropie ou redondance fonctionnelle..).

34

Modèle de l'extrusion de loops'

Des loops sont formés en interphase par un mécanisme d'extrusion d'ADN qui est dépendant du complexe protéique de la cohésine. Ces loops sont souvent définis (stabilisés) par la présence des protéines CTCF liées à leurs sites ADN spécifiques



Leonid Mirny laboratory, Adrian L. Sanborn et al. and E. L-Aiden, PNAS (2015)

35

Comment les enhanceurs fonctionnent-ils?

*Au niveau moléculaire, de leur(s) grammaire(s), i.e. des codes de liaisons avec les facteurs de transcription?

*Au niveau du 'choix' des gènes cibles et de l'organisation chromosomique?

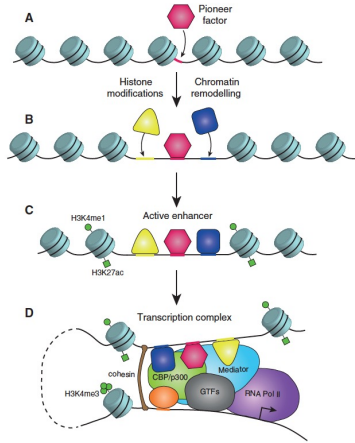
*...

36

Mobilisation des enhancers, facteurs pionniers

DE GRUYTER

W. Schaffner: Transcription enhancers and gene regulation — 321



Reconnaissance par un facteur pionnier



Modifications locales de la chromatine, réorganisation..



Liaisons d'autres facteurs enhancer 'actif'



Contact avec le promoteur activation de la transcription

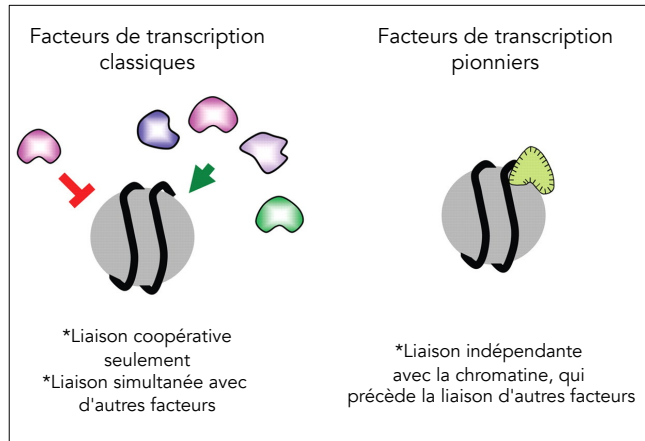
Mobilisation des enhancers, facteurs pionniers

REVIEW

Pioneer transcription factors: establishing competence for gene expression

Kenneth S. Zaret^{1,3} and Jason S. Carroll²

GENES & DEVELOPMENT 25:2227-2241 © 2011 by Cold Spring Harbor Laboratory Press

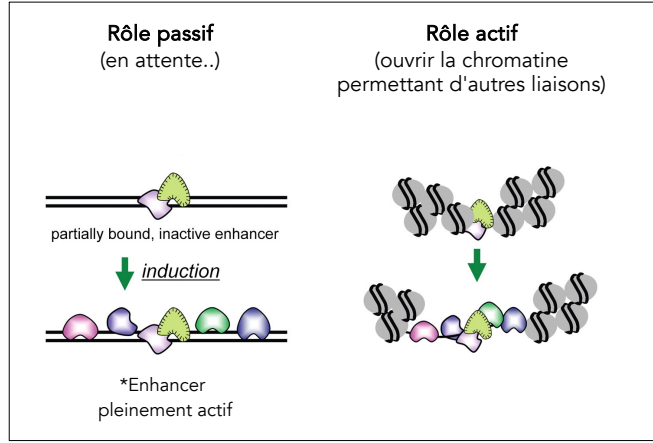


Mobilisation des enhancers, facteurs pionniers

Pioneer transcription factors: establishing competence for gene expression

Kenneth S. Zaret^{1,3} and Jason S. Carroll²

GENES & DEVELOPMENT 26:2227-2241 © 2011 by Cold Spring Harbor Laboratory Press



Copyright © 2011 by Cold Spring Harbor Laboratory Press



Biology
A Global Approach

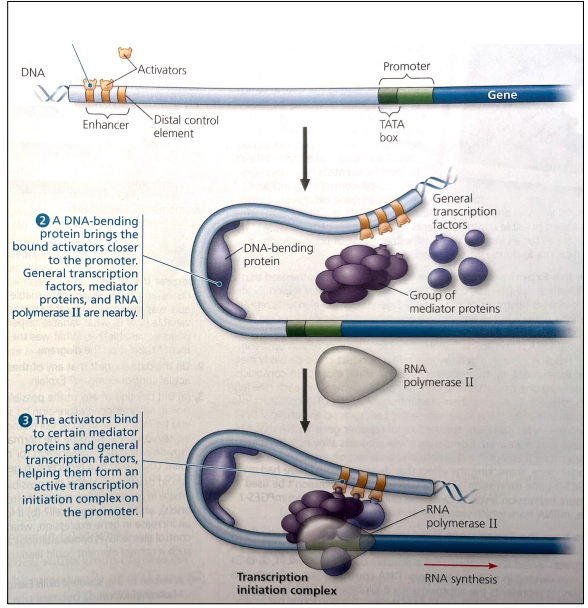
10^{ème} Edition
TENTH EDITION
Campbell • Reece • Urry • Cain • Wasserman • Minorsky • Jackson



La liaison de facteurs de transcription (activateurs) sur les séquences enhancers va déclencher la formation d'un complexe protéique contenant des facteurs généraux (dont le grand complexe 'médiateur', une interface physique et fonctionnelle entre les facteurs liés à l'ADN et la machinerie basale PolII).

Cela va conduire à la formation d'un complexe d'initiation de la transcription

PolII (ARN polymérase II) est un complexe multiprotéique (12 sous unités chez les humains). Activité nucléotidyltransférase..



Biology

A Global Approach

10^{ème} Edition

TENTH EDITION

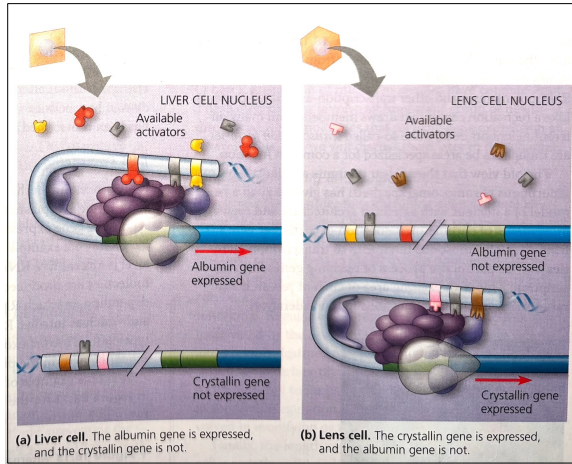
Campbell • Reece • Urry • Cain • Wässerman • Minorsky • Jackson

GLOBAL EDITION

La spécificité de la transcription est fixée par les combinaisons de facteurs se liant sur les séquences enhancers.

Contrairement aux gènes 'housekeeping' qui pour la plupart n'en ont pas besoin.. (moins besoin)

La spécificité est souvent fixée par la présence de facteurs uniques, ou de combinaisons uniques de facteurs, la machinerie transcriptionnelle étant très fortement conservée.. La mise en route de la transcription nécessite également la présence de facteurs spécifiques (souvent les mêmes..)



Feature Review

Enhancer Logic and Mechanics in Development and Disease

Ryan Fickels¹ and Ali Shilatifard^{1*}

Trends in Cell Biology, August 2018, Vol. 28, No. 8 <https://doi.org/10.1016/j.tcb.2018.04.001>
© 2018 Published by Elsevier Ltd.

- *Facteur pionnier, éviction de nucléosome
- *Promoteur (Pol II) en mode pause
- *Looping et contact (cohesin, CTCF..)
- *Recrutement modificateurs histones (p300)
- *Recrutement du complexe Mediateur
- *Démarrage de la Pol II, élongation

NELF: Negative elongation complex
SEC: Super elongation complex
TF: Transcription factors

