

Université de Montréal

**Importance des deux domaines de liaison à l'ADN pour l'action
pionnière du facteur Pax7**

par
Audrey Pelletier

Département de Biochimie et médecine moléculaire
Faculté de Médecine

Thèse présentée à la Faculté de Médecine
en vue de l'obtention du grade de Doctorat
en Biochimie et médecine moléculaire
option Génétique moléculaire

Avril, 2021

© Audrey Pelletier, 2021

Cette thèse intitulée

**Importance des deux domaines de liaison à l'ADN pour l'action
pionnière du facteur Pax7**

Présentée par
Audrey Pelletier

A été évaluée par un jury composé des personnes suivantes

Sylvie Mader
Président-rapporteur

Jacques Drouin
Directeur de recherche

Thierry Alquier
Membre du jury

Vincent Giguère
Examineur externe

Résumé

Durant le développement embryonnaire, une seule cellule, le zygote, est à l'origine de la formation de tous les types de cellules de l'organisme. L'information génétique nécessaire au destin d'une cellule y est gardé dans la chromatine condensée, d'abord inaccessible, et se déploie progressivement au moment opportun de la différenciation cellulaire grâce à des facteurs pionniers. Les facteurs de transcription pionniers sont des acteurs clés des cascades génétiques et épigénétiques déterminant le destin cellulaire. Les facteurs pionniers ont la propriété unique de reconnaître leurs cibles dans la chromatine fermée. Leur action permet d'augmenter l'accessibilité à un répertoire d'*enhancers* spécifique à un destin cellulaire, ouvrant la voie aux autres facteurs de transcription. Les aspects entourant la reconnaissance initiale des facteurs pionniers à leurs cibles dans la chromatine fermée sont mal compris.

Dans l'hypophyse, le facteur pionnier Pax7 est spécifique aux cellules du lobe intermédiaire et met en place le programme génétique mélanotrope. Pour se faire, Pax7 repère les régions régulatrices clé de l'identité mélanotrope dans la chromatine fermée, afin d'accroître l'accessibilité à l'ADN permettant la liaison d'autres facteurs non-pionniers. Pax7 possède deux domaines de liaison à l'ADN, le domaine *paired* (PD) et l'homéodomaine (HD), chacun liant un motif de séquence caractéristique. De plus, une séquence cible composite, formé de la juxtaposition des motifs reconnus par PD et HD, est enrichie aux sites pionniers de Pax7.

Dans le but de définir les propriétés de liaison à l'ADN de Pax7, nous avons caractérisé les interactions de Pax7 avec ses différentes séquences cibles. Nous avons démontré par des expériences de liaison à l'ADN *in vitro* (retard sur gel) que l'interaction de Pax7 avec le motif composite est d'affinité supérieure et sous forme de monomère où le PD est principalement impliqué. Comme les sites cibles de Pax7 dans l'hétérochromatine sont marqués par la méthylation de l'ADN, nous avons montré que la méthylcytosine dans le motif de liaison n'interfère pas avec la liaison à l'ADN par Pax7. De plus, des mutations ponctuelles aux domaines de liaison à l'ADN de Pax7 ont montré que les deux domaines de liaison à l'ADN sont nécessaires pour le recrutement aux sites dans la chromatine fermée, a fortiori pour l'ouverture de la chromatine.

Des études antérieures ayant identifié des transcrits alternatifs de Pax7, nous avons étudié l'impact fonctionnel des variants Pax7. Les quatre isoformes Pax7 comportent des résidus d'acides

aminés alternatifs au niveau du PD : elles ont des propriétés de liaison à l'ADN *in vitro* ainsi qu'un potentiel de transactivation très similaires. Bien que les isoformes de Pax7 aient des activités pionnières variables, leur action différentielle n'est pas dû à leurs propriétés intrinsèques de liaison à l'ADN ou de transactivation.

Nos travaux ont identifié les particularités de Pax7 par rapport à d'autres facteurs pionniers dans les mécanismes de reconnaissance des sites dans l'hétérochromatine. Ainsi, les enhancers sujet à l'action pionnière de Pax7 contiennent typiquement plus de motifs cibles que les cibles d'action transcriptionnelle et l'action pionnière de Pax7 un recrutement fort aux sites pionniers ; de plus, les deux domaines de liaison à l'ADN intacts sont essentiels pour ce recrutement et l'action pionnière. Nos travaux ont défini les paramètres requis pour le reprogrammage cellulaire par un pionnier, Pax7, tout particulièrement en regard des sites de recrutement dans la chromatine fermée. La reprogrammation de cellules souches pour la production de cellules différenciées présente un potentiel thérapeutique unique en médecine régénérative et nos travaux supportent le rôle critique des pionniers dans ce contexte.

Mots-clés : Expression génique, transcription, liaison à l'ADN, facteur de transcription, hypophyse, pro-opiomélanocortine, différenciation, facteur pionnier, Pax7

Abstract

During embryonic development, a single cell, the zygote, is responsible for the formation of all cell types in the body. The genetic information necessary for cell fates is stored there in condensed chromatin, initially inaccessible, and is gradually deployed at the opportune moment of cell differentiation by the pioneer factors. Pioneer transcription factors are key determinants in the genetic and epigenetic cascades leading to cell fate. Pioneer factors have the unique property of recognizing their DNA targets in closed chromatin. Their action provides accessibility to a repertoire of enhancers specific to a cell fate, opening the way for other transcription factors. We poorly understand initial recognition of pioneer factors at their targets in closed chromatin.

In the pituitary gland, the pioneer factor Pax7 is specific to the intermediate lobe and implements the melanotrope genetic program. To do so, Pax7 identifies key regulatory regions of melanotrope identity in closed chromatin to provide DNA accessibility allowing the binding of other non-pioneer factors. Pax7 has two DNA binding domains (DBD), the paired domain (PD) and the homeodomain (HD), each binding a characteristic sequence motif. In addition, a composite target sequence, formed from the juxtaposition of PD and HD motifs, is enriched at the pioneer sites of Pax7.

In order to define the DNA binding properties of the pioneer factor Pax7, we characterized the interactions of Pax7 with its different target sequences. We have demonstrated by *in vitro* DNA binding experiments (gel shift) that the interaction of Pax7 with the composite motif is of higher affinity and as a monomeric form in which the PD is primarily involved. As Pax7 target sites in heterochromatin are marked by DNA methylation, we showed that methylcytosine within the binding motif does not interfere with Pax7 DNA binding. Using point mutations in the Pax7 DBDs, we showed that both DBDs are required for the recruitment to sites in closed chromatin, furthermore for chromatin opening.

Since previous studies identified alternative Pax7 transcripts, we investigated the functional impact of Pax7 variants. The four Pax7 isoforms contain alternative amino acid residues in the PD: they have similar *in vitro* DNA binding properties and transactivation potential. Although Pax7 isoforms have varying pioneering activities, their differential action is not due to their intrinsic DNA-binding or transactivation properties.

Our work has identified the particularities of Pax7 compared to other pioneering factors in the mechanisms of site recognition in heterochromatin. Thus, the enhancers subject to the pioneer action of Pax7 typically contain more target motifs than the targets of transcriptional action and the pioneer action of Pax7 a strong recruitment to the pioneer sites; moreover, the two intact DNA-binding domains are essential for this recruitment and pioneering action. Our work has defined the parameters required for cellular reprogramming by a pioneer, Pax7, particularly with regard to recruitment to sites in closed chromatin. The reprogramming of stem cells for the production of differentiated cells has unique therapeutic potential in regenerative medicine and our work supports the critical role of pioneers in this context.

Keywords : Gene expression, transcription, DNA binding, transcription factor, pituitary, pro-opiomelanocortin, differentiation, pioneer factor, Pax7

Table des matières

Résumé.....	1
Abstract.....	3
Table des matières.....	5
Liste des tableaux.....	7
Liste des figures.....	9
Liste des abréviations.....	11
Remerciements.....	15
Chapitre 1 : Introduction	17
1.1 Développement : de la génétique et de l'épigénétique.....	17
1.1.1 Les facteurs de transcription et l'identité cellulaire.....	17
1.1.2 La structure de la chromatine et l'accessibilité à l'ADN.....	21
1.1.3 La régulation génique par l'épigénétique.....	25
1.2 Le développement de l'hypophyse et le rôle du pionnier Pax7.....	31
1.2.1 Anatomie et organogenèse de l'hypophyse.....	31
1.2.2 Les types de cellules endocrines de l'hypophyse.....	33
1.2.3 Le gène sélecteur du lobe intermédiaire : Pax7.....	40
1.3 Les facteurs de transcription de la famille Pax.....	43
Abstract.....	45
1. Introduction.....	46
2. Multiple DNA binding modalities.....	47
3. Transactivation domains and coactivators.....	55
4. Transcriptional repression by Pax factors and corepressors.....	57
5. Co-regulatory DNA binding transcription factors.....	59
6. Regulation of Pax factor activity.....	60
7. Pax factors in chromatin remodelling.....	62
8. Pax factors as pioneers for chromatin remodelling.....	65
9. Conclusions and perspectives.....	67
References.....	68
1.4 Les facteurs pionniers : des propriétés uniques.....	77
1.4.1 Les modalités de liaison à l'ADN.....	77
1.4.2 Les facteurs de pluripotence.....	79
1.4.3 Les facteurs pionniers spécifiques.....	83
1.4.4 Caractéristiques communes des facteurs pionniers.....	86
1.5 Le cas du facteur pionnier Pax7.....	89
1.5.1 L'interaction avec l'ADN par les domaines PD et HD.....	89
1.5.2 L'action pionnière de Pax7.....	91
1.5.3 Les isoformes Pax7.....	93
1.5.3 Mutations du gène <i>PAX7</i> et maladies.....	95
Chapitre 2 : L'action pionnière de Pax7 nécessite ses deux domaines de liaison à l'ADN : paired et homeo	97
Abstract.....	98

Introduction.....	100
Materials and methods	102
Results.....	107
Strongest <i>in vitro</i> Pax7-binding motif is enriched at pioneered target sites	107
Prevalent role of Pax7 paired (PD) domain for DNA binding.....	109
Pax7 binds the composite site as monomer	110
Cytosine methylation does not interfere with Pax7 DNA binding	112
Both Pax7 DNA binding domains are required for pioneer activity	113
Discussion.....	119
Figures supplémentaires.....	123
References.....	138
Chapitre 3 : Les propriétés de liaison à l'ADN et transcriptionnelles des isoformes de Pax7	141
Abstract.....	142
Introduction.....	144
Results and discussion	147
Functional dissection of Pax7 DBD.....	147
DNA binding affinity of Pax7 isoforms.....	150
Transcriptional activity of Pax7 isoforms.....	152
Materials and methods	156
References.....	157
Figure supplémentaire.....	160
Chapitre 4 : Discussion	161
4.1 Les modalités de liaison à l'ADN du facteur pionnier Pax7	161
4.1.1 L'affinité différentielle de Pax7 à ses séquences cibles.....	161
4.1.2 Les deux domaines de liaison à l'ADN intact sont essentiels à l'action pionnière de Pax7.....	162
4.2 Les conditions et obstacles à l'action pionnière de Pax7.....	165
4.2.1 Les sites pionniers sont enrichis de motifs de forte affinité.....	165
4.2.2 L'environnement chromatinien aux sites résistants.....	166
4.2.3 Coopération entre facteurs pionniers et non-pionniers	168
4.3 Les isoformes Pax7 ont des propriétés intrinsèques de liaison à l'ADN et de transactivation similaires	170
Chapitre 5 : Conclusion et perspectives	173
Bibliographie.....	175

Liste des tableaux

Chapitre 2

Supplementary Table 1. Sequences of oligonucleotide primers	135
Supplementary Table 2. Summary of high throughput sequencing datasets and experimental conditions	136
Supplementary Table 3. ChIP and ATAC-seq thresholds to categorize Pax7 peaks.....	137

Liste des figures

Chapitre 1

Figure 1.1 La différenciation cellulaire et la hiérarchie des cellules souches	18
Figure 1.2 L'expression des gènes Pax durant le développement du tube neural (plan transversal).....	19
Figure 1.3 La différenciation et plasticité des cellules	20
Figure 1.4 La structure du nucléosome.....	22
Figure 1.5 Structure du noyau et caractéristiques de l'hétérochromatine et de l'euchromatine.....	23
Figure 1.6 Fonction des complexes de remodelage de la chromatine	26
Figure 1.7 Profil des marques d'histone H3 associé à l'activation transcriptionnelle.....	28
Figure 1.8 La différenciation cellulaire des lignées endocrines de l'hypophyse.....	32
Figure 1.9 Clivage différentiel du précurseur hormonal POMC	36
Figure 1. Structure of Pax transcription factors.....	47
Figure 2. DNA target sites of Pax factor binding.	50
Figure 3. Gene targets of Pax factor action.....	51
Figure 4. ChIPseq analyses of Pax genomic binding sites.....	53
Figure 5. Co-regulator proteins interacting with Pax factors.....	55
Figure 1.10 Exemple d'interactions entre les facteurs de transcription (Pax5 et Ets1) et la double hélice d'ADN.....	78
Figure 1.11 Structure des domaines de liaison à l'ADN de OCT4 et SOX2 lié à l'ADN.....	81
Figure 1. 12 Structure des domaines de liaison à l'ADN de OCT4 et SOX2 lié à l'ADN nucléosomal	82
Figure 1.13 Ressemblance du domaine de liaison à l'ADN de FoxA à l'histone de liaison	84
Figure 1. 14 Structure du domaine de liaison à l'ADN et capacité d'un facteur de transcription à reconnaître l'ADN nucléosomal	87
Figure 1.15 Structure de la protéine Pax7 et des domaines de liaison à l'ADN PD et HD	90
Figure 1.16 Les étapes de l'action pionnière	92
Figure 1.17 Épissage alternatif du transcrit de Pax7 et acides aminés variants des isoformes.	93
Chapitre 2	
Figure 1. Contribution of paired (PD) and homeodomain (HD) DNA binding domains to Pax7 binding.	108
Figure 2. Relative roles of Pax7 PD and HD domains in DNA binding.	111
Figure 3. Pax7 binding is independent of DNA methylation.....	112
Figure 4. Intact and functional PD and HD domains are both required for recruitment and chromatin opening at Pax7 pioneer sites.	114
Figure 5. Mutations of either the PD or HD prevent gene activation at targets that require pioneer action whereas they have variable effects on transcriptional target genes.	117
Supplementary Figure 1. Characterization of Pax7 EMSA complexes with various probes.	124
Supplementary Figure 2. Nucleosome positioning relative to Pax7 Pioneer sites	126
was determined by MNaseSeq analyses in AtT-20 Neo and AtT-20 Pax7 cells.....	126

Supplementary Figure 3. Characterization of Pax7 mutants within the PD (G43A and R56L) and HD (S264E) domains.....	128
Supplementary Figure 4. Alignment of mouse Pax7, human PAX3 and human PAX6 sequences showing structural features of PD (green) and HD (blue) DNA binding domains.....	130
Supplementary Figure 5. Representative loci for genomic recruitment (Flag-Pax7 ChIPseq) and ATACseq for wildtype Pax7 and three DBD mutants as indicated.....	132
Supplementary Figure 6. Pax7 target genes.....	134
Chapitre 3	
Figure 1. Pax7 and Pax3 amino acid sequences and protein features.....	147
Figure 2. Contribution of HTH modules of Pax7 DBDs to DNA binding.	149
Figure 3. Relative DNA binding affinity of Pax7 isoforms.....	151
Figure 4. Transcriptional activity of Pax7 isoforms.	154
Supplementary Figure 1. Pax7 target probes and natural PC2 enhancer.....	160

Liste des abréviations

α MSH	alpha-Melanocyte Stimulating Hormone
ACTH	Adreno Corticotropic Hormone
ADN	Acide Désoxyribonucléique
AL	Anterior Lobe
ARMS	Alveolar Rhabdomyosarcoma
ARN	Acide Ribonucléique
ATAC-seq	Assay for Transposase of Accessible Chromatin followed by high throughput sequencing
ATP	Adenosine Triphosphate
bHLH	basic Helix–Loop–Helix
ChIP-seq	Chromatin Immunoprecipitation followed by high throughput sequencing
CRH	Corticotropin-Releasing Hormone
DBD	DNA Binding Domain
DNA	Deoxyribonucleic Acid
ER	Estrogen Receptor
FSH	Follicle-Stimulating Hormone
FT	Facteur de transcription
GH	Growth Hormone
GnRH	Gonadotropin Releasing Hormone
GR	Glucocorticoid Receptor
HD	Homeodomain
HTH	Helix-Turn-Helix
IL	Intermediate Lobe
LH	Luteinizing Hormone
MNase-seq	Micrococcal Nuclease followed by high throughput sequencing
mRNA	messenger Ribonucleic Acid
ncRNA	non-coding RNA
OAR	Otp/ aristaless/ Rax
OP	Octapeptide
Pax	Paired box
PD	Paired domain
Pitx	Pituitary Homeobox
POMC	Pro-Opiomelanocortin
Prl	Prolactin
RNA	Ribonucleic Acid
SHL	Super Helical Location
TF	Transcription Factor
Tpit	Pituitary Tbox factor
TSH	Thyroid-Stimulating Hormone
TSS	Transcription Start Site
WGBS	Whole Genome Bisulfite Sequencing

*Je dédie cette thèse à ma famille
avec tous mes sentiments de reconnaissance pour leur soutien*

Remerciements

Je tiens d'abord à remercier les Drs Sylvie Mader, Thierry Alquier, Vincent Giguère et Julie Lessard, de l'honneur qu'ils m'ont fait en acceptant d'être rapporteur pour cette thèse.

Merci au Dr Jacques Drouin qui m'a encadré tout au long de mes études graduées. Tes qualités scientifiques sont impressionnantes et n'ont d'égal que tes qualités humaines qui font de toi un être inspirant pour toute personne ayant la chance de te côtoyer. Merci d'avoir été patient et de m'avoir donné la chance de faire mon chemin et de développer mon autonomie.

Merci à tous les membres du laboratoire du Dr Jacques Drouin, anciens et présents, avec qui j'ai évolué depuis la maîtrise et qui ont collaboré, de près ou de loin aux travaux de recherches pour cette thèse. Merci à Konstantin pour l'aide quotidienne, et pour avoir relu et corrigé cet ouvrage. Merci à Aurélio qui m'a transmis des informations essentielles tout au long de ma formation. Merci à Alexandre dont les travaux et réflexions ont permis d'améliorer le contenu de cette thèse. Merci à Arthur, Virginie, Juliette, Kevin, Ryhem, Justine, Vincent, Yves, Amandine et Valérie pour l'aide, les analyses et les discussions. Aux membres passés Maëva, Steve, Deepak, Panojot, Shinobu, Lionel, Catherine, Taby et Évelyne, merci pour ces belles années.

Sur une note plus personnelle, je tiens à remercier toutes ces personnes que j'ai rencontré pendant mon parcours et qui sont devenu des amis très chers. Konstantin et Alexandre, on nous a appelé les inséparables et ce sera toujours vrai. Maëva, une amie qui me sera toujours chère malgré la distance. Yacine, Marine, Rushi, Ines, Rachel ainsi qu'aux amis qui s'y sont joint et avec qui on forme maintenant une grande famille qui s'agrandit toujours : Mallika, Ankush, Prabhas merci pour tous ces moments qui ont fait que la route était belle. Merci à mes amis, Jessica, Myrèla, Akram et Vincent qui ont pris d'autres voies dans leurs études, mais qui ont toujours su me comprendre malgré tout.

Merci à ma famille et surtout à mes parents, qui m'ont soutenu pendant toutes ces années d'études et qui, dès mon plus jeune âge, m'ont appris à être curieuse et à travailler pour repousser mes limites. Enfin, merci à Jerry et Olivia qui ont su tenir le fort et me supporter malgré les sacrifices que mes études graduées ont demandés.

Chapitre 1 : Introduction

1.1 Développement : de la génétique et de l'épigénétique

Dans le domaine de la biologie du développement, un objectif est de comprendre comment une cellule particulière se développe en un type de cellule final. À partir d'un simple zygote, ou cellule-œuf, on trouve chez l'humain adulte environ 37 mille milliards de cellules comprenant des centaines de types cellulaires (Bianconi et al., 2013). La complexité cellulaire de l'organisme s'observe déjà au stade embryonnaire et s'amplifie pour définir les différents tissus et organes comptant une multitude de cellules spécialisées. Le destin des cellules peut être déterminé par des facteurs de développement endogènes, l'interaction avec des cellules adjacentes ou des signaux externes, tels que les morphogènes et les hormones. Par conséquent, chaque cellule d'un embryon reçoit des signaux moléculaires des cellules voisines sous forme de protéines, d'ARN et même d'interactions de surface (Hagolani, Zimm, Marin-Riera, & Salazar-Ciudad, 2019). Les processus développementaux sont ainsi dirigés par le microenvironnement cellulaire tout en maintenant l'homéostasie tissulaire. Ces processus comprennent la prolifération cellulaire, la différenciation, le mouvement cellulaire et la mort cellulaire programmée (Fuchs & Steller, 2011; Solnica-Krezel, 2005). Le déploiement des différents types cellulaires avec une régulation spatio-temporelle précise est essentiel au cours du développement pour établir les différentes structures de l'organisme.

1.1.1 Les facteurs de transcription et l'identité cellulaire

Au front de ce déploiement cellulaire, ce sont les facteurs de transcription (FT) : une classe de protéine qui lie l'ADN pour activer ou réprimer la transcription des gènes. Il y a une progression de la régulation menée par l'ensemble des FT présents dans le noyau cellulaire ; c'est la cascade de régulation transcriptionnelle (Bolouri & Davidson, 2003). Celle-ci se déroule par étape : d'abord, un FT se lie à des régions régulatrices et active l'expression d'autres FTs, ces nouveaux facteurs s'accumulent dans le noyau et interagissent avec de nouvelles régions du génome permettant l'expression d'un réseau complexe de gènes. La nature séquentielle de la différenciation cellulaire est due à la contribution de gènes régulateurs et des signaux externes, tel que l'acide rétinoïque, les facteurs de croissance et des cytokines. L'expression simple de

ces facteurs ne suffit pas à cerner leur rôle puisqu’eux-mêmes sont modulés à plusieurs niveaux incluant, évidemment, l’efficacité de transcription du gène, mais aussi l’épissage alternatif de l’ARNm, la demi-vie de celui-ci, la traduction et l’accumulation de la protéine puis des modifications post-traductionnelles (Baralle & Giudice, 2017; Burow et al., 2015). Il y a aussi un aspect combinatoire des différents FTs présents dans une cellule qui contrôlent et coordonnent la différenciation cellulaire. Le parcours menant au destin cellulaire est parsemé de « décisions » qui poussent la cellule à s’engager vers une voie de différenciation particulière (Fig. 1.1). À chaque étape de cet engagement, des contrôles internes sont opérés pour maintenir l’état de différenciation cellulaire et éviter qu’il y ait un retour en arrière ou une permutation de l’identité de la cellule (Natoli, 2010).

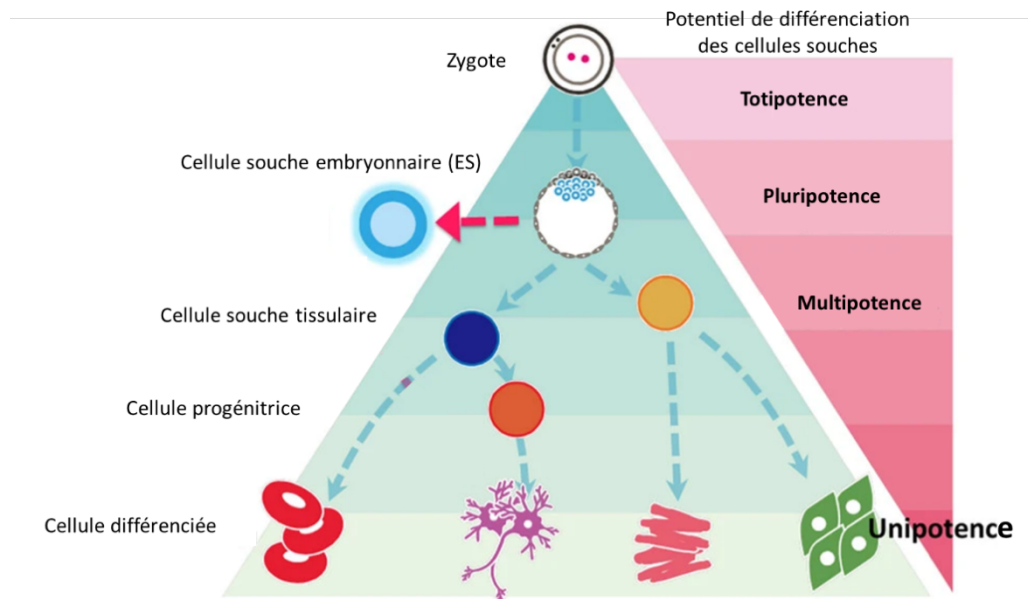


Figure 1.1 La différenciation cellulaire et la hiérarchie des cellules souches

Adapté de (Sugawara, Nishino, Umezawa, & Akutsu, 2012)

Au cours du développement embryonnaire, l’identité positionnelle des cellules est déterminée par les gènes sélecteurs, des FTs dont l’expression est restreinte à des compartiments particuliers et qui fournissent une structuration précoce. L’analyse génétique et développementale des mutants morphogénétiques chez la drosophile a permis de définir le rôle des gènes sélecteurs comme des gènes agissant de manière autonome afin d’informer les cellules

de leur position dans l'embryon et de préciser leur positionnement définitif au cours de la formation des organes (García-Bellido, 1975). Ainsi, l'identité positionnelle précoce fournie par les gènes sélecteurs est la base d'un contexte cellulaire permettant ultérieurement les mécanismes de régulation et de différenciation terminale pour produire des destins alternatifs.

Chez les vertébrés, l'implication des gènes sélecteurs des familles Hox et Pax a été démontrée dans l'embryogenèse pour le positionnement cellulaire le long des axes antéropostérieur et dorsoventral des organes (Mallo, 2018; Mansouri, Stoykova, & Gruss, 1994; Monsoro-Burq, 2015). Ces gènes restreignent le potentiel de différenciation des domaines où ils sont exprimés. Par exemple, les FTs à homéodomaine (HD) Pax3 et Pax7 confèrent une identité dorsale au tube neural qui est la structure embryonnaire à l'origine du système nerveux central (Mansouri & Gruss, 1998), tandis que Pax6 établit des populations distinctes de cellules progénitrices ventrales (Fig. 1.2) (Ericson et al., 1997). D'ailleurs, l'expression aberrante de gènes régulateurs ou la perturbation d'éléments de régulation causent un développement pathologique et des malformations congénitales (Chalepakis, Stoykova, Wijnholds, Tremblay, & Gruss, 1993; Quinonez & Innis, 2014; M. Smith & Flodman, 2018).



Figure 1.2 L'expression des gènes Pax durant le développement du tube neural (plan transversal)
Selon (Tanabe & Jessell, 1996) et (Chalepakis et al., 1993), créé à l'aide de BioRender

Différents concepts ont été proposés afin d'élucider les mécanismes moléculaires à la base de l'action des gènes sélecteurs. Certains modèles montrent l'action combinatoire de plusieurs FTs spécifiques pour la régulation de gènes dépendant de la différenciation cellulaire (Heinz et al., 2010). Bien que la coopérativité de liaison puisse être suffisante pour de nombreux types de FTs afin d'activer l'expression génique, les mécanismes sous-jacents à l'initiation de nouveaux programmes génétiques restent peu compris. La mise en place du programme génétique durant la différenciation cellulaire (la programmation cellulaire, Fig. 1.3) doit

surmonter de nombreux obstacles d'accessibilité à l'ADN aux régions clés dans une chromatine condensée qui sera discuté dans le prochain chapitre. Seulement un sous-groupe de FTs possède la capacité remarquable de reprogrammer un type cellulaire en un autre.

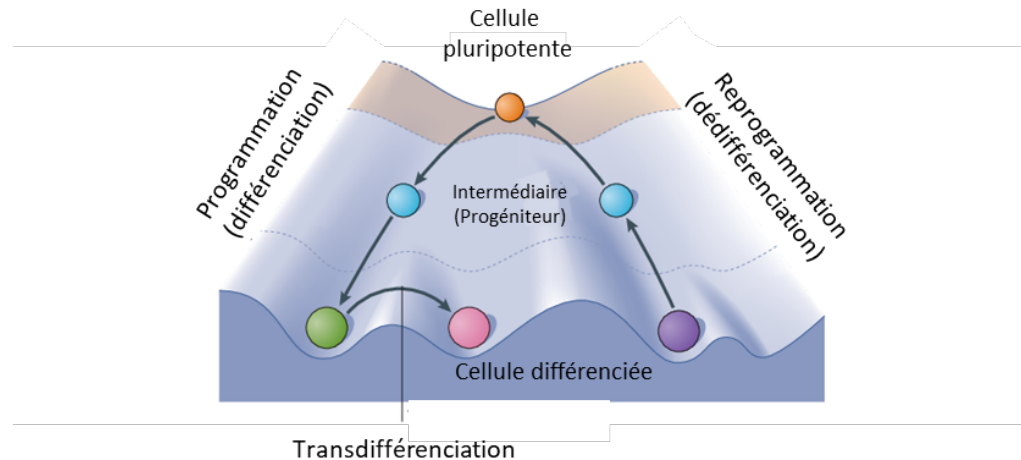


Figure 1.3 La différenciation et plasticité des cellules

Adapté de (Takahashi & Yamanaka, 2016)

Les FTs qui reprogramment le destin cellulaire sont invariablement ceux qui sont cruciaux à la programmation cellulaire initiale dans le développement embryonnaire. Pour entraîner la programmation ou la reprogrammation cellulaire, ces FTs doivent être capables d'engager des gènes qui sont silencieux développementalement ou dont l'expression est inappropriée dans la cellule originale. Ces gènes silencieux sont typiquement dans des régions de chromatine fermée qui sont couvertes de nucléosomes. Les gènes sélecteurs ou les facteurs impliqués dans la programmation pourraient correspondre au concept de facteurs pionniers qui ont la propriété unique de sonder la chromatine fermée et engager des sites cibles dans l'ADN nucléosomal (Drouin, 2014; Zaret & Carroll, 2011). La liaison initiale de facteurs pionniers dans des régions de chromatine moins accessibles permettrait d'ouvrir localement la chromatine et la rendre directement compétente pour la liaison d'autres FTs.

1.1.2 La structure de la chromatine et l'accessibilité à l'ADN

L'organisation du génome dans le noyau et la structure de la chromatine jouent un rôle clé dans la détermination du statut transcriptionnel des gènes. De fait, l'accessibilité à l'ADN est un prérequis à la liaison de la majorité des FTs. Toutefois, les différents niveaux de compaction de l'ADN limitent l'accessibilité au double brin pour la machinerie protéique qui régule les différentes fonctions tel que la transcription, mais aussi la réplication, la réparation et la recombinaison.

Le nucléosome

Chez les eucaryotes, l'unité de base qui forme la chromatine est le nucléosome constitué de 146 paires de base (pb) d'ADN enroulées sur un octamère protéique comprenant deux molécules de chacune des histones H2A, H2B, H3 et H4 (Fig.1.4). La courbure de l'ADN sur la surface de la protéine rapproche le squelette phosphate des deux brins à l'intérieur de la spirale (Bowman & Poirier, 2015).

L'enroulement de l'ADN autour du noyau de l'histone a pour conséquence d'occlure stériquement l'ADN à la liaison d'autres protéines. Cette conformation de nature inhibitrice est observée chez pratiquement tous les eucaryotes pour réguler l'accès à l'ADN (Kornberg & Lorch, 2020). L'assemblage du nucléosome est possible sur n'importe quelle séquence d'ADN, quoi que la séquence nucléotidique puisse en favoriser ou défavoriser la formation. Par exemple, les protéines du noyau des histones ont une plus grande affinité pour certaines séquences de nucléotides tel que les séquences d'ADN à haute teneur en GC ainsi que des intervalles d'environ 10 pb des dinucléotides AA/TT/TA, ce qui favorise la formation de nucléosomes (Eslami-Mossallam, Schiessel, & van Noort, 2016; Ioshikhes, Albert, Zanton, & Pugh, 2006; Segal et al., 2006). À l'opposé, les séquences avec des étendues poly (dAdT) sont connues pour déstabiliser les nucléosomes (Anderson & Widom, 2001). Par conséquent, la séquence d'ADN influence la stabilité et la mobilité des nucléosomes qui, par ailleurs, ne sont pas des structures statiques, mais fluctuent dynamiquement (Huertas, MacCarthy, Schöler, & Cojocar, 2020).

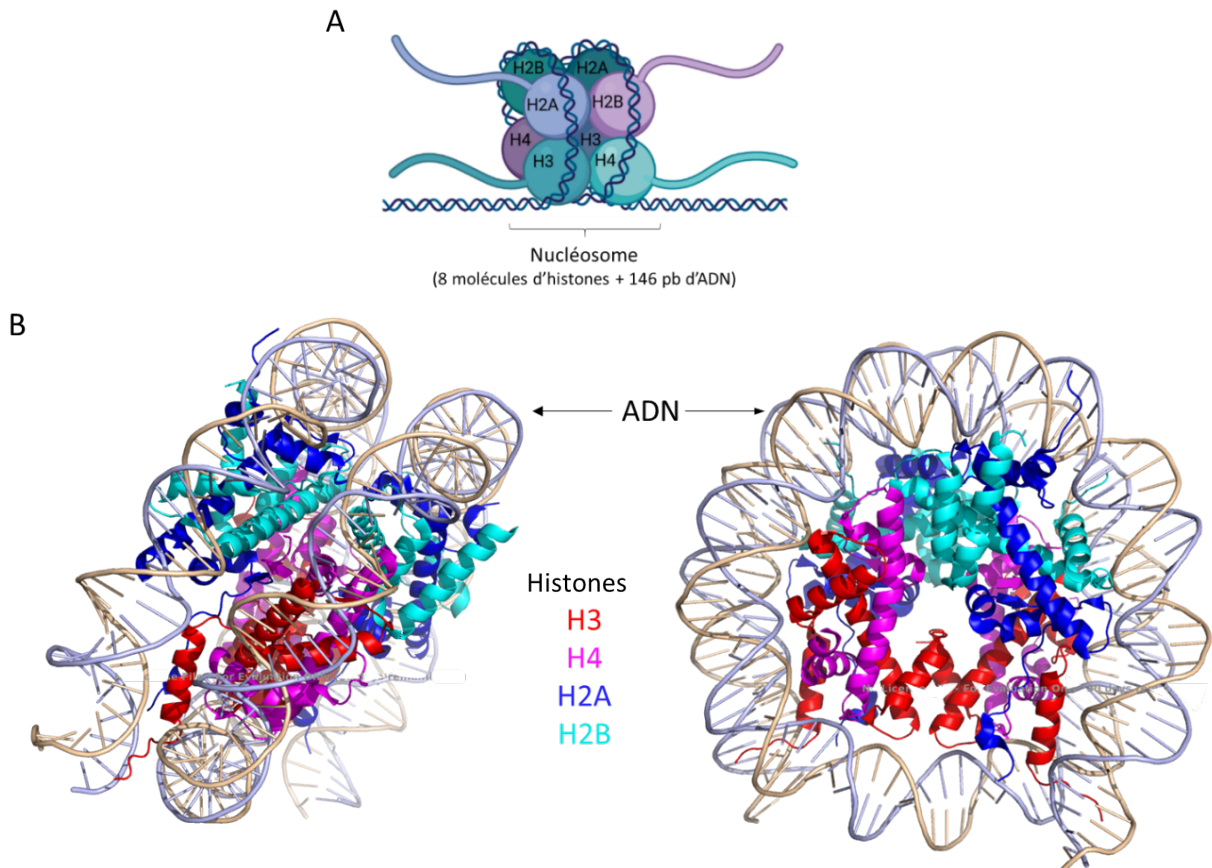


Figure 1.4 La structure du nucléosome

(A) Représentation schématique créée à l'aide de BioRender (B) Structure de la particule cœur du nucléosome humain tiré de (Tsunaka, Kajimura, Tate, & Morikawa, 2005), modifiée à l'aide de PyMOL

Il existe aussi des formes variantes d'histones dont l'incorporation est essentielle dans des domaines spécifiques du génome. Par exemple, un variant de l'histone H3, CENP-A, est associé aux régions centromériques (Palmer, O'Day, Wener, Andrews, & Margolis, 1987). GammaH2A-X est essentiel à la réparation dans les régions où des cassures d'ADN double brin ont eu lieu (Fillingham, Keogh, & Krogan, 2006). H2A-Z modifie la structure de la chromatine afin de réguler la transcription. Ce dernier variant se situe principalement aux promoteurs des gènes (H. Zhang, Roberts, & Cairns, 2005) et occupe souvent les deux nucléosomes entourant une région libre de nucléosome (Raisner et al., 2005).

Un niveau supérieur de compaction de la chromatine fait intervenir l'ADN internucléosomal et l'histone de liaison (soit H1 ou H5) dont l'interaction donnent un motif structural qui accompagne le repliement du nucléofilament en fibre de 30 nm (Bednar et al., 1998). La longueur de l'ADN internucléosomal varie entre 20 et 90 pb selon les espèces et les

tissus, et peut fluctuer à l'intérieur du génome d'une cellule donnée (van Holde & Zlatanova, 1995). Le nucléofilament s'organise en boucles de 250 nm pendant l'interphase du cycle cellulaire et le niveau maximal de compaction est le chromosome métaphasique de 850 nm.

La transcription dans le paysage chromatinien

La structure de la chromatine est étroitement liée à l'expression génique et l'identité de la cellule. Les deux grandes catégories de structure d'ordre supérieur de la chromatine durant l'interphase sont : l'euchromatine et l'hétérochromatine (Fig. 1.5). L'euchromatine est une forme décondensée localisée au centre du nucléoplasme et correspond à la chromatine transcriptionnellement active ou compétente de la cellule. L'hétérochromatine est une forme compacte et transcriptionnellement silencieuse, généralement située en périphérie du noyau. Il y a l'hétérochromatine constitutive incluant les régions structurales comme les centromères et les télomères, et l'hétérochromatine facultative impliquée dans le silençage de gènes (*gene silencing*) spécifiques dans un type cellulaire donné. L'hétérochromatine constitutive est marquée par la triméthylation de l'histone H3 au niveau de la lysine 9 (H3K9me3), tandis que H3K27me2 est enrichie dans l'hétérochromatine facultative (Allshire & Madhani, 2018). L'hétérochromatine est globalement désacétylée et l'ADN est méthylé, contrairement à l'euchromatine.

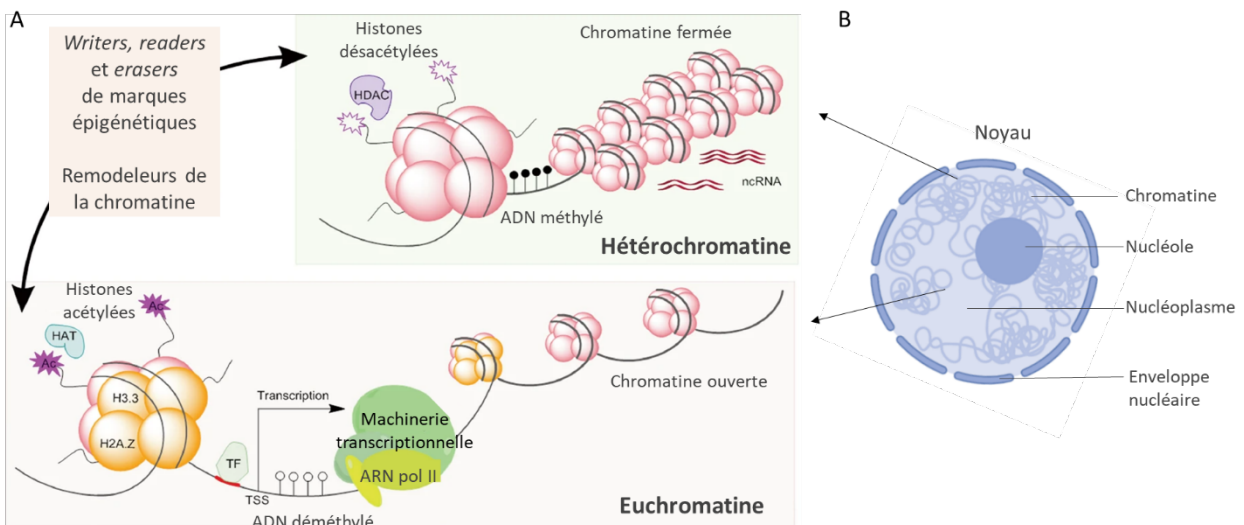


Figure 1.5 Structure du noyau et caractéristiques de l'hétérochromatine et de l'euchromatine

(A) Adapté de (Cazaly, Charlesworth, Dickinson, & Holloway, 2015) (B) Représentation schématique créée à l'aide de BioRender

Les protéines HP1 (*Heterochromatin Protein 1*) sont une composante majeure des domaines d'hétérochromatine. Il a été démontré chez la levure *S. pombe* que HP1 (Swi6) reconnaît la modification H3K9me3 et forme un réseau de complexes protéine HP1-nucléosome à la base de l'assemblage et de la propagation de l'hétérochromatine constitutive (Canzio et al., 2011). Cette conformation favorise la formation de compartiments d'hétérochromatine distincts et multi-chromosomiques causé par un effet de séparation de phase (Larson et al., 2017; Strom et al., 2017). Ces domaines en « gouttelettes » ont des propriétés physicochimiques particulières et seraient moins perméables notamment à la machinerie transcriptionnelle (Larson & Narlikar, 2018). Il y a aussi des évidences du rôle d'ARN non-codants (ncRNA) dans l'assemblage de l'hétérochromatine (Quinodoz et al., 2020).

L'accès des FTs aux éléments de régulation dans le contexte chromatinien peut être fortement limité par l'architecture local de la chromatine (Wang et al., 2012). Il est bien établi que le nucléosome réprime la transcription puisque sa présence peut altérer la liaison à l'ADN de nombreux FTs, de même que des enzymes de réparation de l'ADN (Blomquist, Li, & Wrangé, 1996; Schild-Poulter, Sassone-Corsi, Granger-Schnarr, & Schnarr, 1996; Taylor, Workman, Schuetz, & Kingston, 1991). Toutefois, la chromatine n'est pas un environnement stable. Au contraire, c'est dans un état dynamique qu'évoluent les histones, les protéines régulatrices de la transcription et les enzymes qui modifient et déplacent les nucléosomes. Les facteurs associés à une région coopèrent ou entrent en compétition pour modifier l'équilibre entre les états de chromatine permissifs et non permissifs, menant à l'activation ou à la répression de la transcription (Mellor, 2006; Perino & Veenstra, 2016). La transcription implique typiquement des régions *enhancers* de l'ADN réunissant des sites de liaison de FTs régulateurs qui peuvent parfois être éloignés du promoteur. Pour être active, une région régulatrice doit être dépourvue de nucléosome, c'est-à-dire que l'ADN soit accessible, pour permettre la liaison de FTs et de la machinerie transcriptionnelle de base.

Toute l'architecture du génome est réorganisée lors de la différenciation cellulaire et cette dynamique se produit de concert avec les changements dans la régulation de la transcription génique menant à la transition vers le destin cellulaire (Bonev et al., 2017; Dixon et al., 2015; Phillips-Cremins et al., 2013). Pour maintenir l'identité cellulaire, les domaines de la chromatine et leurs marques épigénétiques doivent être fidèlement hérités durant la division

cellulaire. C'est un défi pour une cellule dupliquant son génome de transférer les nucléosomes parentaux préexistants avec leurs modifications post-traductionnelles (Escobar et al., 2019).

1.1.3 La régulation génique par l'épigénétique

La stabilité de l'identité cellulaire est assurée par des mécanismes robustes qui maintiennent le statut de l'épigénome. L'organisation et la maintenance de l'état de la chromatine sont assurées par deux types de facteurs : les facteurs de remodelage de la chromatine et les enzymes de modifications post-traductionnelles des histones. Ils peuvent induire des changements conformationnels au niveau du nucléosome de même qu'au niveau de larges domaines de chromatine.

Le remodelage de la chromatine et les modifications des histones

Les facteurs de remodelage de la chromatine participent à l'activation ou à la répression de certains gènes (Elgin, Elgin, & Workman, 2000). Pour activer la transcription, l'octamère d'histones doit être déplacé par les facteurs de remodelage de la chromatine afin que les régions régulatrices de l'ADN soient accessibles à la machinerie transcriptionnelle (Steger & Workman, 1996). Les complexes de remodelage de la chromatine comprennent les familles SWI/SNF (*switch/sucrose non-fermentable*), ISWI (*imitation SWI*) et CHD (*chromodomain helicase DNA-binding*) qui modifient la structure nucléosomale grâce à l'énergie libérée par l'hydrolyse de l'ATP afin de rompre les interactions entre les acides nucléiques et les protéines. Les complexes de remodelage de la chromatine dissocient l'ADN de la surface de l'octamère d'histone, permettant aux nucléosomes soit de glisser le long de l'ADN ou d'être dissociés ou déplétés et ainsi réguler l'accès des FTs à leurs sites de liaison à l'ADN (Figure 1.6) (Clapier, Iwasa, Cairns, & Peterson, 2017).

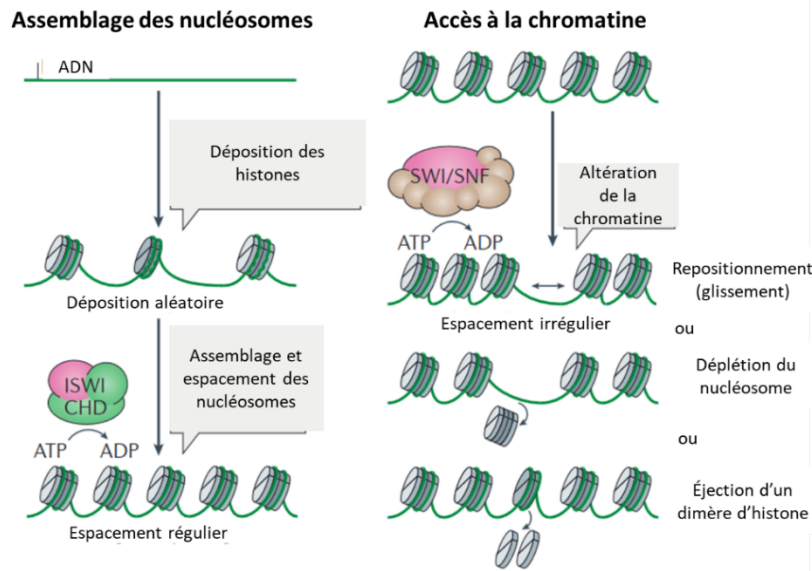


Figure 1.6 Fonction des complexes de remodelage de la chromatine

Adapté de (Clapier et al., 2017)

Les protéines d'histones sont conservées au cours de l'évolution et les extrémités N-terminales des histones, aussi appelées les queues d'histone, pointant vers l'extérieur de l'unité nucléosomale sont la cible de modifications post-traductionnelles. Ces modifications peuvent non-seulement affecter leur charge, mais aussi l'accessibilité à l'ADN et les interactions protéine:protéine servant à réguler l'expression génique. Il existe plusieurs marques d'histones et la multitude de combinaisons possibles de celles-ci pourrait expliquer la diversité des activités de la chromatine dans le noyau (Strahl & Allis, 2000). Les modifications post-traductionnelles qui peuvent être ajoutées ou enlevées dynamiquement par voie enzymatique comprennent l'acétylation, la méthylation, la phosphorylation, l'ubiquitylation, la sumoylation et l'ADP-ribosylation. Ces marques ont le potentiel de modifier les interactions histone-ADN et histone-histone et fournissent ainsi un moyen de cibler de manière transitoire les changements dans la dynamique des nucléosomes. Des modifications post-traductionnelles spécifiques peuvent être reconnues via des domaines protéiques particuliers que sont les modules *readers* et les enzymes qui modifient les histones sont soit les *writers* qui ajoutent des marques, soit les *erasers* qui en suppriment. Le profil des marques épigénétiques associées aux régions actives sont résumées à la figure 1.7.

L'acétylation des histones

Plusieurs protéines clés impliquées dans l'activation de la transcription possèdent une activité histone-acétyltransférase (HAT), tandis que les histone-désacétylases (HDAC) sont associées à des complexes multiprotéiques répresseurs (Eberharter & Becker, 2002; Y. Shen, Wei, & Zhou, 2015). L'acétylation des queues d'histone est reconnue par des protéines ayant un bromodomaine. Ce sont des composantes de complexes protéiques ayant pour fonction de faciliter la transcription génique. Par exemple, les complexes SWI/SNIF contenant les ATPases apparentées Brg1 et Brm se lient aux histones acétylées afin de remodeler la chromatine.

Les enzymes HAT ajoutent un groupement acétyle aux protéines d'histone, ce qui réduit l'affinité de celles-ci à l'ADN, facilitant l'accès aux FTs (Eberharter & Becker, 2002). Les HAT les plus étudiées sont CBP (*CREB-binding protein*) et p300 (*adenoviral E1A-associated protein of 300kDa*) qui marquent les enhanceurs actifs et déposent la modification H3K27ac (Visel et al., 2009). CBP/p300 sont des coactivateurs fondamentaux puisqu'ils peuvent interagir avec une multitude de facteurs liant l'ADN et des composantes de la machinerie transcriptionnelle basale incluant l'ARN polymérase II. Le rôle des enzymes HDAC est de retirer un groupement acétyle aux histones. Il existe des complexes NuRD (*Nucleosome Remodeling Deacetylase complex*) qui combinent l'activité histone-désacétylase des HDAC1/2 et les facteurs de remodelage de la famille Mi-2 dans la réorganisation répressive de la chromatine (Xue et al., 1998; Y. Zhang, LeRoy, Seelig, Lane, & Reinberg, 1998).

La méthylation des histones

La méthylation des histones est très versatile puisque d'un à trois groupements méthyles peuvent être ajoutés aux résidus lysines ou arginines, et ces marques peuvent être soit permissives ou répressives (Cui, Fu, & Dempsey, 2019). Les groupements méthyles ajoutés aux histones sont reconnus par des protéines pourvues d'un chromodomaine.

Les complexes activateurs sont liés au complexe de protéines Trithorax (Trx) chez la drosophile et les complexes MLL (*mixed-lineage leukemia*) chez les mammifères. Les complexes Trx et MLL créent et maintiennent un environnement de chromatine actif. Le groupe Trx médie notamment les marques d'histone H3K4me3 et H3K36me2/3. Il existe quatre complexes apparentés MLL (MLL1-4) qui sont peuvnt méthyler H3K4. Ces quatre gènes

remplissent des fonctions non redondantes car ils ciblent différents sites génomiques. Les complexes MLL sont particulièrement importants dans les régions intergéniques pour la monométhylation de H3K4 (H3K4me) au niveau des *enhancers* (Herz et al., 2012). Les complexes répressifs Polycomb (PRC1 et PRC2) sont impliqués dans l'organisation de la chromatine associés au silençage génique. Le complexe PRC2 a une activité méthylase qui entraîne la méthylation de H3K27 et contribue à la formation de chromatine répressive (Di Croce & Helin, 2013; Margueron et al., 2008; X. Shen et al., 2008).

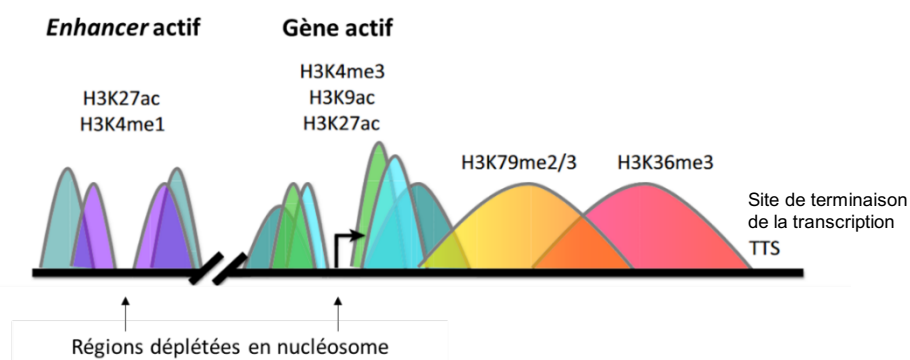


Figure 1.7 Profil des marques d'histone H3 associé à l'activation transcriptionnelle

Adapté de (Gates, Foulds, & O'Malley, 2017)

Paradoxalement, dans les cellules souches embryonnaires (*embryonic stem* - ES), on trouve à la fois des marques actives (H3K4me3) et répressives (H3K27me3) au niveau des gènes développementaux (Bernstein et al., 2006; Fakhouri, Stevenson, Chisholm, & Mango, 2010). Ces *enhancers* bivalents, aussi appelés *poised enhancers*, ne présentent toutefois pas de marque activatrice H3K27ac. Ces domaines bivalents formeraient un équilibre pour les gènes dont une activation ultérieure influencerait la décision du destin cellulaire (Hoffmann, Zimmermann, & Spengler, 2015). D'ailleurs, les protéines des groupes Trx (MLL) et Polycomb à l'origine de ces marques sont impliquées dans le maintien de l'identité cellulaire et sont nécessaires à une différenciation cellulaire normale (Aloia, Di Stefano, & Di Croce, 2013; Sauvageau & Sauvageau, 2010). Les mécanismes d'hérités des modifications post-traductionnelles sont surtout associées aux régions réprimées de la chromatine marquées par H3K9me3 et H3K27me3. Le complexe PRC2 joue un rôle dans la propagation de la marque H3K27me3 pour couvrir un locus entier, assure l'établissement et le maintien de ces marques et préserve ainsi l'identité cellulaire dans les cellules filles (Margueron et al., 2009; Reinberg & Vales, 2018).

La méthylation de l'ADN : la méthylcytosine

La méthylation de l'ADN joue un rôle critique dans la régulation des gènes et l'architecture du noyau de la cellule (Bernstein, Meissner, & Lander, 2007). C'est une marque épigénétique stable qui reflète la mémoire cellulaire afin de préserver l'information sur le programme génétique. Les patrons de méthylation de l'ADN sont établis très tôt durant le développement de l'embryon par des vagues de méthylation et de déméthylation (Zeng & Chen, 2019). Des profils uniques de méthylation de l'ADN sont observés dans les cellules souches et entre différentes lignées cellulaires (Shipony et al., 2014).

Les protéines responsables d'apposer la méthylation sur l'ADN sont Dnmt3a, Dnmt3b et dans une moindre mesure Dnmt2 (Hermann, Gowher, & Jeltsch, 2004). Ces enzymes catalysent l'addition d'un groupement méthyle au cinquième carbone d'une base cytosine qui précède une guanine de l'ADN. La méthylation de l'ADN est maintenue lors de la division cellulaire par une autre enzyme, Dnmt1 (Robert et al., 2003). La méthylcytosine n'est toutefois pas immuable et peut être activement inversée en sa forme non modifiée par catalyse enzymatique impliquant les enzymes TET (Koivunen & Laukka, 2018). Les méthylcytosines sont reconnues spécifiquement par les protéines MeCP 1/2 (*methyl-CpG-binding protein*) et MBD 1-4 (*methyl-CpG-binding domain*) qui, en médiant l'interaction avec d'autres modifications épigénétiques, participent à la formation de structures répressives de la chromatine (Bommarito & Fry, 2019).

Les régions riches en dinucléotides CpG sont fréquentes à proximité des régions régulatrices des gènes et la méthylation de l'ADN à ces endroits marque des régions inactives. La méthylation de l'ADN peut altérer l'expression génique de façon directe ou indirecte. Par exemple, lorsque les CpGs sont méthylés au niveau du promoteur d'un gène, cela bloque l'accès à la machinerie transcriptionnelle qui ne peut initier la transcription. Toutefois, la régulation de l'expression génique par la méthylation de l'ADN n'est pas seulement opérée par un mécanisme de blocage direct puisque ce ne sont pas tous les FTs qui sont sensibles à la méthylation de l'ADN au niveau de leur site de liaison.

Le portrait de l'ensemble des marques épigénétiques

Les récentes avancées techniques permettent de cartographier la myriade de marques épigénétiques ainsi que l'environnement chromatinien sur tout le génome grâce, entre autres, au séquençage à haut débit. L'épigénome étudié par ChIP-seq (immunoprécipitation de la chromatine suivie de séquençage massif) pour décrire l'état chromatinien est typiquement déterminé par les modifications d'histone H3K27ac et H3K4me, associées avec un état actif, et H3K27me3 représentant les régions inactives. Les mesures d'accessibilité à l'ADN sont évaluées par ATAC-seq (*Assay for Transposon-Accessible Chromatin*) et la cartographie des nucléosomes par MNase-seq (*Micrococcal Nuclease*) (Ninova, Fejes Tóth, & Aravin, 2019; Stricker, Köferle, & Beck, 2017). Pour l'étude de la méthylation de l'ADN, le traitement au bisulfite de sodium suivi de séquençage massif (BS-seq ou *Whole Genome Bisulfite Sequencing* – WGBS) reflète le méthylome des cellules (Bibikova et al., 2011). Ce sont sur ces profils que reposent la compréhension et l'action des gènes régulateurs, dont les FTs pionniers, dans le contexte du développement et de la différenciation cellulaire.

1.2 Le développement de l'hypophyse et le rôle du pionnier Pax7

L'hypophyse, aussi appelée glande pituitaire, constitue un modèle pour l'étude de l'établissement de l'identité cellulaire. Longtemps considérée comme la glande maîtresse, l'hypophyse, située à la base du cerveau, est le chef d'orchestre de plusieurs fonctions endocriniennes. Les hormones hypophysaires peuvent agir directement sur des organes cibles ou sur d'autres glandes (les hormones trophiques), régulant des fonctions corporelles telles que la croissance, le métabolisme et la fertilité (Drouin, 2011).

1.2.1 Anatomie et organogenèse de l'hypophyse

L'hypophyse compte 3 lobes : le lobe antérieur, le lobe intermédiaire et le lobe postérieur. Si le lobe intermédiaire est bien développé chez certains animaux tel que les rongeurs et les chiens, l'humain voit cette structure seulement au stade fœtal, puis celle-ci régresse à partir de la quinzième semaine de gestation (Dubois, el Amraoui, & Héritier, 1997). Ce que l'on connaît des mécanismes développementaux de l'hypophyse provient d'études sur des modèles murins. Chez l'embryon, le lobe antérieur se développe par une excroissance épithéliale du plafond du larynx, tandis que le lobe postérieur provient d'une excroissance ventrale du diencephale (Drouin, 2011; Rizzoti, 2015). La structure antérieure est une invagination de l'ectoderme oral, aussi appelée poche de Rathke, dont l'expansion se produit sous les contacts soutenus avec l'épithélium neural (Fig. 1.8). L'hypophyse est reliée à l'hypothalamus par la tige pituitaire composée principalement de vaisseaux sanguins et de prolongations de cellules nerveuses. Les lobes antérieur et intermédiaire sont modulés par les facteurs de contrôle de nature hormonale passant par les vaisseaux sanguins qui proviennent de l'hypothalamus, tandis que le lobe postérieur est sujet à ses influx nerveux.

Plusieurs facteurs sont critiques pour le développement de l'hypophyse. Des étapes fondamentales à la formation de la poche de Rathke, il y a au jour embryonnaire e9,5, l'expression de *Bmp4* (*Bone morphogenetic protein 4*) dans la partie ventrale du diencephale et qui est remplacée par *Fgf8* (*Fibroblast growth factor 8*) au jour e10 (Drouin, 2011). Aussi essentiel au développement de l'hypophyse, le facteur *Shh* (*Sonic hedgehog*) qui est exprimé dans tout l'ectoderme oral, excluant la poche de Rathke. Cette dernière voit l'expression transitoire de *Hesx1* entre e9,5 et e11,5 pendant que la structure forme rapidement une glande

fermée par la séparation du lien avec l'ectoderme oral et transit ainsi de l'épithélium au mésenchyme. Très tôt dans la formation de la poche de Rathke, la structure est marquée par l'expression des FTs Pitx1 et Pitx2 (*Pituitary homeobox 1* et 2), puis des facteurs à homeodomaine *Lim-homeo* Lhx3, Lhx4 et Is11 (*Islet1*). Entre les jours e12 et e15, l'hypophyse subit une expansion importante dû à la prolifération cellulaire, toutefois l'expression de marqueurs de différenciation cellulaire est plus tardive. La différenciation des cellules hypophysaires est progressive et séquentielle entre les jours e11,5 et e16,5. C'est seulement au jour e17,5 que toutes les cellules spécialisées de l'hypophyse expriment leurs hormones respectives, et ont ainsi acquis leur identité cellulaire unique (Davis, Mortensen, & Camper, 2011).

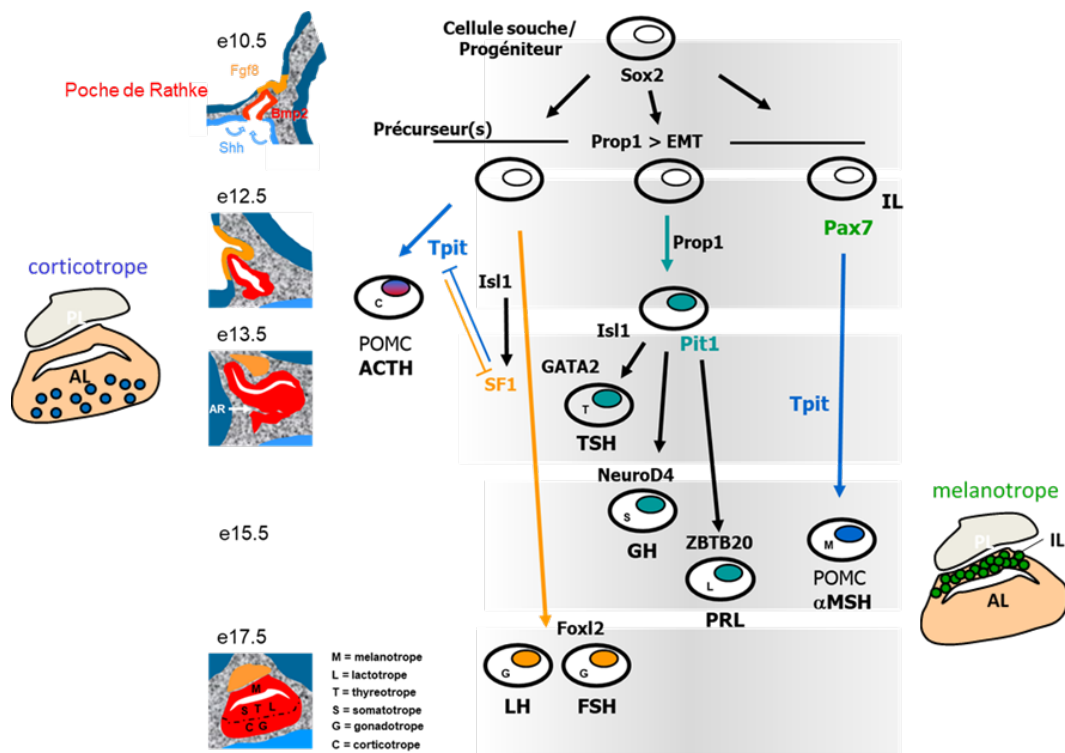


Figure 1.8 La différenciation cellulaire des lignées endocrines de l'hypophyse

Adapté de (Drouin, 2011)

Le contrôle de la prolifération cellulaire est critique lors de l'organogenèse afin de déterminer la taille des organes et la composition cellulaire des tissus. Les cellules progénitrices de l'hypophyse, indifférenciées et marquées par l'expression de Sox2, prolifèrent activement

autour de la lumière (entre les lobes antérieur et intermédiaire putatifs) (Ikeda & Yoshimoto, 1991). La prolifération des cellules progénitrices diminue au fur et à mesure que l'organogenèse progresse de telle sorte que, chez l'adulte, les cellules Sox2-positives quiescentes se trouvent principalement tapissant la fente hypophysaire (Fauquier, Rizzoti, Dattani, Lovell-Badge, & Robinson, 2008; Gremeaux, Fu, Chen, & Vankelecom, 2012). Cette fente qui sépare le lobe intermédiaire du lobe antérieur constitue le reste de la lumière d'origine du primordium hypophysaire. Les cellules Sox2-positives ont la capacité de former des « pituisphères » en culture, qui peuvent se développer et former des sphères secondaires (J. Chen et al., 2005; Fauquier et al., 2008). Les pituisphères peuvent être induites pour se différencier vers tous les types de cellules hypophysaires.

1.2.2 Les types de cellules endocrines de l'hypophyse

On distingue 5 types de cellules endocrines dans le lobe antérieur : les somatotropes qui produisent l'hormone de croissance (GH), les lactotropes qui produisent la prolactine (PRL), les gonadotropes qui produisent l'hormone folliculo-stimulante (FSH) et l'hormone lutéinisante (LH), les thyrotropes qui produisent la thyroestimuline (TSH) et les corticotropes qui produisent l'hormone adrénocorticotrope (ACTH). Le lobe intermédiaire est un tissu homogène ne comptant que les mélanotropes qui produisent la mélano-stimuline (MSH). Quant au lobe postérieur, aussi appelé lobe neural, il sécrète la vasopressine (AVP) et l'ocytocine (OT).

Les avancés en imagerie tridimensionnelle des tissus ont permis de mettre en lumière le caractère interrelié du déploiement des cellules hypophysaires. En effet, les cellules de chaque lignée forment des réseaux homotypiques étendus et structurés (Budry et al., 2011; Mollard, Hodson, Lafont, Rizzoti, & Drouin, 2012). C'est grâce à ces réseaux que la réponse cellulaire aux sécrétagogues est opérée de façon robuste et coordonnée (Mollard et al., 2012). Les gènes codant pour les hormones caractéristiques de chaque lignée ont servi à identifier les FTs qui sont impliqués dans l'identité des cellules hypophysaires. Par conséquent, ce sont surtout les facteurs responsables de la différenciation terminale qui sont les mieux connus, ceux-ci sont généralement exprimés de 12 à 24 heures avant le gène de l'hormone lui-même (Drouin, 2011). Les premières étapes d'engagement et de spécification de la différenciation cellulaire des lignées hypophysaires sont ainsi peu connues.

Les lignées dépendantes de Pit1

Un marqueur précoce de l'identité hypophysaire est Prop1 (*Prophet of Pit1*), un FT à homéodomaine de type *paired*, dont l'expression est observable au jour e12,5 (Raetzman, Ward, & Camper, 2002). Prop1 est requis pour l'activation de l'expression du FT Pit1 (*POU homeodomain protein*), lui-même essentiel à la différenciation des lignées thyrotrope, lactotrope et somatotrope (Cohen & Radovick, 2002; S. Li et al., 1990; Potok et al., 2008; Ward et al., 2005).

Les thyrotropes

Les thyrotropes expriment α GSU (*glycoprotein hormone alpha-subunit*), une sous-unité partagée avec les gonadotropes pour la production de TSH (*thyroid-stimulating hormone*) et des gonadotropines respectivement. Déjà au jour e11,5, l'expression d' α GSU est détectable dans la poche de Rathke (Japón, Rubinstein, & Low, 1994). D'ailleurs, les régulateurs du gène *Cga*, codant pour α GSU, sont communs parmi ces deux lignées. Entre autres, Foxl2, un FT de type *forkhead* dont l'expression persiste chez l'adulte participe à l'activation de *Cga* (Ellsworth et al., 2006; Parks, Adess, & Brown, 1997). Foxl2 est exprimé au jour e11,5 au niveau de la partie ventrale de la poche de Rathke et constitue un marqueur de l'engagement des lignées thyrotrope et gonadotrope (Ellsworth et al., 2006). La différenciation de ces deux populations de cellules est aussi dépendante du FT Gata2, dont l'expression au jour e12,5 est aussi maintenue chez l'adulte (Charles et al., 2006). En plus de participer à l'activation du gène *Cga*, Gata2 active l'expression du gène *TSH β* , spécifique aux thyrotropes (Gordon et al., 1997). Dans les thyrotropes, Pit1 participe aussi à l'activation des gènes *Cga* et *TSH β* . Les thyrotropes représentent 1-5% des cellules du lobe antérieur et la thyrostimuline qu'elles sécrètent (TSH, comprenant les sous-unités α GSU et TSH β) ont pour action principale de stimuler la glande thyroïde pour le développement et la croissance de la glande, ainsi que la production d'hormones thyroïdiennes nécessaires notamment au métabolisme.

Les somatotropes

Les somatotropes apparaissent vers le jour e13,5, lorsque Pit1 active l'expression de *NeuroD* dans la partie médiane de l'hypophyse en développement (Davis et al., 2011; Rizzoti, 2015). Elles stimulent la croissance par la sécrétion de somatotrophine (*Growth Hormone* - GH)

dont l'expression est détectable au jour e 15.5 (Japón et al., 1994). De plus, Foxo1 contribuerait à la différenciation des somatotropes par la régulation de gènes importants pour l'identité de ces cellules tel que *NeuroD*, mais aussi *Ghl* et *Pou1f1* (Kapali et al., 2016). La population de somatotropes augmente considérablement au niveau post-natal et représente les cellules les plus abondantes de l'hypophyse formant environ 50% du lobe antérieur adulte.

Les lactotropes

L'émergence des lactotropes est détectable par l'expression de l'hormone prolactine (Prl) vers le jour e15.5, mais l'expression de Prl est augmentée de manière post-natale (Drouin, 2011). Faisant partie de la population exprimant Pit1 relativement tôt dans le développement de la glande, ces cellules semblent être engagées bien avant l'expression de leur hormone. Jusqu'ici, l'identification des FTs requis pour l'expression de Prl n'ont pas réussi à déterminer les facteurs spécifiques à la différenciation des lactotropes. Malgré tout, l'œstrogène est un signal critique pour la fonction de ces cellules puisque le récepteur d'œstrogène (ER) synergise avec Pit1 au niveau de l'*enhancer* de Prl (Day, Koike, Sakai, Muramatsu, & Maurer, 1990; Peers et al., 1990). L'abondance de lactotropes dépend aussi de ER et varie selon la période de lactation chez les femelles et chez les mâles cette lignée hypophysaire représente entre 15 et 20% des cellules du lobe antérieur.

La lignée SF1

Les gonadotropes

Le marqueur spécifique le plus précoce des gonadotropes est le récepteur de la GnRH (*Gonadotropin Releasing Hormone*), détectable au jour e12,5 (Wen, Ai, Alim, & Boehm, 2010). Toutefois, c'est seulement au jour e16,5 que ces cellules expriment leurs hormones caractéristiques, les gonadostimulines LH (hormone lutéinisante) et FSH (hormone folliculo-stimulante) (Japón et al., 1994). La signalisation GnRH contrôle la biosynthèse et la libération de la LH et de la FSH qui à leur tour régulent le développement et l'activité des ovaires et des testicules. Tel que mentionné précédemment, les gonadotropes sont marqués par l'expression des FTs Foxl2 et Gata2, génétiquement en aval des Lhx3/4 (Charles et al., 2006; Ellsworth et al., 2006; Gordon et al., 1997; Hashimoto et al., 2005). Le facteur de différenciation terminale qui est exprimé de façon restreinte dans les gonadotropes est le récepteur nucléaire SF1

(*Steroidogenic Factor 1*), détectable au jour e13,5 (Barnhart & Mellon, 1994; Zhao et al., 2001). SF1 intervient à plusieurs niveaux de la fonction reproductive de l'axe hypothalamo-hypophysogonadien afin de réguler le développement des glandes surrénales, des gonades et la production de stéroïdes sexuels.

Les lignées POMC

L'expression de *Tpit* (*Pituitary Tbox factor*) est restreinte aux deux lignées hypophysaires exprimant POMC chez la souris, les corticotropes du lobe antérieur et les mélanotropes du lobe intermédiaire (Lamolet et al., 2001). *Tpit* est un FT à boîte T qui a d'abord été identifié comme un partenaire transcriptionnel du facteur *Pitx1* agissant sur le promoteur du gène *POMC*. La différenciation terminale des deux lignées POMC est menée par *Tpit*, et les souris déficientes en *Tpit* (*Tpit*^{-/-}) ne parviennent pas à exprimer POMC et la différenciation des mélanotropes et des corticotropes n'aboutit pas (Pulichino, Vallette-Kasic, Tsai, et al., 2003).

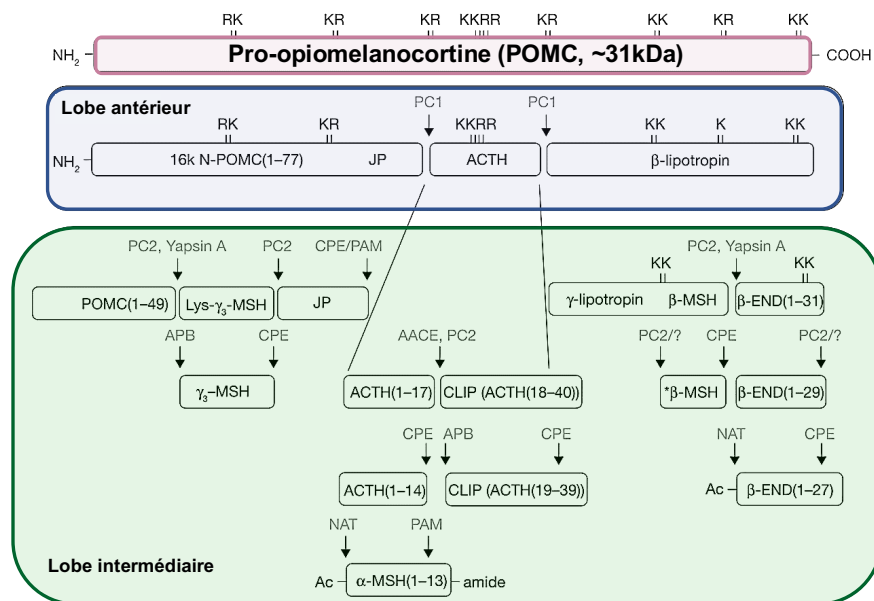


Figure 1.9 Clivage différentiel du précurseur hormonal POMC

Adapté de (Cawley, Li, & Loh, 2016)

Les corticotropes sont les premières cellules endocrines de l'hypophyse à atteindre leur différenciation terminale au jour e12,5 (Japón et al., 1994). Dans le primordium hypophysaire, elles sont détectées par l'expression d'ACTH (hormone adrénocorticotrope), une hormone qui

provient d'un clivage enzymatique de la POMC (proopiomélanocortine). Le produit d'un seul gène codant pour la POMC est un précurseur hormonal qui est traité de manière différentielle pour produire une multitude de peptides dont l'ACTH et l' α MSH, collectivement connues sous le nom de mélanocortines. La POMC est transformée en ACTH par la proprotéine-convertase 1 (PC1, codée par le gène *Pcsk1*) dans les corticotropes et davantage clivé en α MSH par la PC2 (*Pcsk2*) dans les mélanotropes (Figure 1.9) (Seidah et al., 1999). Le contrôle hormonal de la transcription du gène *POMC* et de la libération de ces hormones est également tissu-spécifique. En plus de ces différences, l'expression restreinte de certains récepteurs membranaires marque ces deux lignées ; les corticotropes expriment les récepteurs des glucocorticoïdes (GR), de l'hormone de libération de la corticotropine (*corticotropin-releasing-hormone receptor 1 - Crhr1*) et de la vasopressine (*vasopressin receptor 1B - Avpr1b*) (Papadimitriou & Priftis, 2009), tandis que les mélanotropes expriment le récepteur de la dopamine D2 (*Drd2*) (Cote et al., 1986).

Les corticotropes

La fonction des corticotropes est au centre de l'axe hypothalamo-hypophyso-surrénalien (*Hypothalamic–Pituitary–Adrenal – HPA*) (Drouin, 2011). Le signal de l'hypothalamus pour la sécrétion d'ACTH par l'hypophyse est médié par la sécrétion d'un peptide, le CRH (*Corticotropin-Releasing Hormone – CRH*). L'ACTH libérée dans le sang permet la production de glucocorticoïdes par les glandes surrénales. Les glucocorticoïdes interagissent avec le récepteur cellulaire GR exprimé dans de nombreux organes afin de réguler notamment le métabolisme énergétique. Il y a aussi un phénomène de rétroaction négatif de l'axe HPA puisque les glucocorticoïdes qui se fixent aux récepteurs GR de l'hypothalamus et de l'hypophyse vont occasionner la répression de la libération du CRH et de l'ACTH, et ainsi bloquer la sécrétion de glucocorticoïde par le cortex surrénalien. La modulation hormonale de l'expression de POMC par les corticotropes devient possible avec l'expression des récepteurs *Crhr1* (récepteur de la CRH) et GR lors de la différenciation des corticotropes (Lugo & Pintar, 1996). Les corticotropes deviennent alors sensibles à ces signaux et commencent à sécréter l'ACTH en cours de développement fœtal et cette régulation est maintenue jusque chez l'adulte (G. W. Smith et al., 1998; Tremblay et al., 1988).

Dans la lignée corticotrope, l'interaction entre *Ptx1*, *NeuroD1/β2* et *Tpit* est essentielle pour l'histospécificité de la transcription du gène *POMC* (Lamolet et al., 2001; Poulin, Lebel,

Chamberland, Paradis, & Drouin, 2000; Poulin, Turgeon, & Drouin, 1997). Les corticotropes expriment Tpit à partir du jour e11,5 et celui-ci est critique pour la différenciation terminale et la maintenance des corticotropes (Pulichino, Vallette-Kasic, Tsai, et al., 2003). Un antagonisme existe entre Tpit et SF1 qui joue un rôle dans l'établissement des lignées POMC et gonadotropes. Cela suggère que ces lignées pourraient provenir de précurseurs communs (Pulichino, Vallette-Kasic, Couture, Brue, & Drouin, 2004). C'est en partie le résultat d'un mécanisme de trans-répression entre les deux FTs spécifiques à leur lignée qui s'exerce sur leurs cibles génétiques respectives (Pulichino, Vallette-Kasic, Tsai, et al., 2003). Ainsi, Tpit est un régulateur négatif du destin gonadotrope, autant que SF1 est un régulateur négatif du destin corticotrope. En conséquence, l'hypophyse antérieure des souris Tpit^{-/-} présente une augmentation du nombre de gonadotropes. Les mutations du gène *TPIT* chez l'humain entraînent un déficit isolé en ACTH (*isolated ACTH deficiency* – IAD), attestant du caractère restreint de sa distribution cellulaire (Pulichino, Vallette-Kasic, Couture, et al., 2003). Un déficit congénital de type IAD est très rare, mais peut mener à la mort chez les nouveau-nés s'il n'y a pas de traitement adéquat (Vallette-Kasic et al., 2005). Cette maladie héréditaire récessive est causée par la carence en ACTH, ayant pour effet secondaire sur les surrénales d'entraîner une carence en glucocorticoïdes, ce qui provoque une hypoglycémie brutale et sévère (Vallette-Kasic et al., 2005). Le déficit hormonal peut toutefois être corrigé s'il est pris en charge assez tôt par une thérapie de remplacement de glucocorticoïde. Différentes mutations au gène TPIT ont été identifiées, y compris des arrêts, défauts d'épissage, délétions génomiques et mutations ponctuelles (Vallette-Kasic et al., 2005).

NeuroD1, un FT de type bHLH, a été identifié sur la base de son action sur la transcription cellule-spécifique du promoteur POMC (Poulin et al., 1997). NeuroD1, de concert avec Ptx1, contribue à la transcription de POMC de façon spécifique aux corticotropes et participe à la mise en place du programme génique de cette lignée. Le développement de l'hypophyse fœtale est marqué par l'expression transitoire et à des niveaux élevés de NeuroD1 dans les corticotropes, toutefois il est exclu des mélanotropes (Poulin et al., 1997). En absence de NeuroD1, le lobe antérieur de l'hypophyse accuse un retard de l'expression de POMC, qui est récupéré vers le jour e15,5, lorsque l'expression de NeuroD1 normalement diminue (Poulin et al., 2000).

Les mélanotropes

La deuxième lignée de cellules hypophysaires exprimant la POMC sont les mélanotropes formant la totalité des cellules endocrines du lobe intermédiaire. Cependant, la régulation de la fonction mélanotrope diffère de celle des corticotropes (Proulx-Ferland et al., 1983). Des études chez le rat ont démontré que la libération d'hormones du lobe intermédiaire est modulée par des neurones hypothalamiques dopaminergiques et GABA-ergiques qui sécrètent des peptides MIF (*MSH-release inhibiting factor*) ou MRH (*MSH-releasing hormone*) (Kastin, Plotnikoff, Viosca, Anderson, & Schally, 1973; Saland, 2001). Ceux-ci agissent à travers les récepteurs *Drd2* et les récepteurs GABA qui sont exprimés sur les mélanotropes (Demeneix, Desaulles, Feltz, & Loeffler, 1984). Chez les amphibiens, l' α MSH joue un rôle crucial dans le processus d'adaptation de la couleur de la peau, en stimulant la synthèse de la mélanine. Chez les mammifères, l' α MSH favorise la pigmentation de la peau, des cheveux et des muqueuses. Chez l'homme, où le lobe intermédiaire est absent, la pigmentation est régulée localement au niveau de la peau en réponse à l'exposition aux UV et à la production locale d' α MSH.

Durant le développement de l'hypophyse, l'expression de la POMC dans le lobe intermédiaire commence vers le jour e15,5 et est précédée par l'expression de *Tpit* (Lamolet et al., 2001). L'inactivation du gène *Tpit* conduit à une perte presque complète de l'expression de POMC dans les mélanotropes et les corticotropes putatifs car les cellules destinées à devenir des cellules POMC n'atteignent pas la différenciation terminale (Pulichino, Vallette-Kasic, Tsai, et al., 2003). En l'absence de *Tpit*, environ 10% des cellules du lobe intermédiaire changent de destin et deviennent gonadotropes SF1-positives et n'expriment aucun marqueur mélanotrope, en accord avec l'antagonisme observé entre *Tpit* et SF1 (Pulichino, Vallette-Kasic, Tsai, et al., 2003). D'autres cellules du lobe intermédiaire restent bloquées dans un état non prolifératif et non différencié exprimant p57 (du gène *Cdkn1c* codant pour l'inhibiteur de kinase dépendant de la cycline 1C) qui est normalement exprimé par des cellules qui viennent de sortir de la prolifération, mais qui n'expriment pas encore de marqueur connu (Bilodeau et al. 2009).

Des travaux du laboratoire du Dr Sally Camper ont montré que la perte de *Hes1*, un transducteur de la voie de signalisation Notch, provoque une altération de l'identité mélanotrope (Raetzman, Cai, & Camper, 2007). Les expériences de perte et de gain de fonction de *Hes1* dans l'organogenèse hypophysaire ont montré que la signalisation Notch via *Hes1* d'une part, inhibe

la différenciation des cellules hypophysaires et d'autre part, favorise le destin des cellules mélanotropes (Raetzman et al., 2007). Plus récemment, notre laboratoire a fait la démonstration que le lobe intermédiaire hypophysaire est spécifié à un destin développemental mélanotrope par le facteur pionnier Pax7.

1.2.3 Le gène sélecteur du lobe intermédiaire : Pax7

Les cellules mélanotropes sont marquées par l'expression de Pax7, un FT de type *Paired-box* (Budry et al., 2012). Pax7 est essentiel pour l'activation des gènes spécifiques des mélanotropes et la répression des gènes corticotropes. Ainsi, Pax7 est un gène sélecteur du lobe intermédiaire et détermine le destin mélanotrope dans l'une des deux lignées POMC qui sont autrement toutes deux dépendantes de Tpit pour la différenciation terminale. Pax7 est exprimé dans le lobe intermédiaire juste avant Tpit et son expression chevauche de façon transitoire avec celle du marqueur de cellules souches hypophysaires Sox2. Pax7 amène les progéniteurs hypophysaires à se différencier notamment grâce à son action pionnière. Pax7 cible les *enhancers* spécifiques aux mélanotropes et augmente l'accessibilité permettant le recrutement d'autres facteurs tel que Tpit (Mayran et al., 2018). En absence de Pax7 (Pax7^{-/-}), les cellules du lobe intermédiaire changent d'identité vers un destin corticotrope.

Toutefois, Pax7 n'est pas suffisant pour la différenciation des mélanotropes car celle-ci est menée par Tpit. En ce sens, les cellules Pax7-positives du lobe intermédiaire de souris Tpit^{-/-} sont engagées dans la voie de différenciation, mais nécessitent Tpit pour terminer le processus (Budry et al., 2012). Tpit exerce un effet positif sur l'expression de Pax7 et la coopération de ces deux facteurs permet de mettre en place le programme d'expression génique spécifique aux mélanotropes (Mayran et al., 2019). Dans les cellules AtT-20, une lignée cellulaire dérivée de corticotropes, l'expression ectopique de Pax7 conduit à un changement de l'identité cellulaire vers une signature génétique mélanotrope (Budry et al., 2012). Pax7 peut reprogrammer ces cellules pour exprimer des marqueurs mélanotropes et ce changement d'identité implique un remodelage de la chromatine. Les régions régulatrices mélanotrope-spécifiques arborent de nouvelles marques de chromatine actives telles que l'histone acétyltransférase p300 et des marques bimodales de H3K4me qui flanquent les *enhancers* actifs (Budry et al., 2012). L'action pionnière de Pax7 aux *enhancers* mélanotrope-spécifiques donne

lieu à de nombreux nouveaux sites de liaison pour le FT Tpit, notamment aux locus PC2 et DrD2 caractéristiques des mélanotropes. Ces nouveaux sites Tpit contiennent des motifs de liaison à l'ADN canoniques, TpitREpal, qui ne peuvent pas être distingués des autres sites de liaison de Tpit actifs dans les cellules AtT-20 (Budry et al., 2012). L'action pionnière de Pax7 a un effet majeur sur l'accessibilité au niveau d'un répertoire d'*enhancers* spécifiques aux mélanotropes et crée donc un environnement de chromatine permissif pour la liaison de Tpit (Mayran et al., 2018).

Pax7 fait partie de la famille conservée de gènes *Pax* qui ont des rôles importants dans le développement et l'organogenèse. Plusieurs études accumulant des données de liaison à l'ADN indiquent que les protéines de la famille Pax peuvent utiliser différentes modalités de liaison à l'ADN pour élargir leur répertoire de reconnaissance de séquence et leur éventail de fonctions (Czerny, Schaffner, & Busslinger, 1993; Epstein, Cai, Glaser, Jepeal, & Maas, 1994; J. A. Epstein et al., 1994; Kozmik, Czerny, & Busslinger, 1997).

1.3 Les facteurs de transcription de la famille Pax

Ce sous-chapitre est constitué d'un article de revue écrit avec mon collègue Alexandre Mayran (Mayran, Pelletier, & Drouin, 2015). Il se trouve dans sa langue de publication originale, soit l'anglais.

Pax factors in transcription and epigenetic remodelling

Alexandre Mayran¹, Audrey Pelletier¹ and Jacques Drouin²

Laboratoire de génétique moléculaire, Institut de recherches cliniques de Montréal (IRCM),
Montréal QC, H2W 1R7 Canada

Keywords: DNA binding, gene expression, chromatin, pioneer, repression, coactivator

¹ These authors contributed equally.

² Corresponding author: Jacques Drouin, Laboratoire de génétique moléculaire, Institut de recherches cliniques de Montréal (IRCM), jacques.drouin@ircm.qc.ca

Abstract

The nine Pax transcription factors that constitute the mammalian family of paired domain (PD) factors play key roles in many developmental processes. As DNA binding transcription factors, they exhibit tremendous variability and complexity in their DNA recognition patterns. This is ascribed to the presence of multiple DNA binding structural domains, namely helix-turn-helix (HTH) domains. The PD contains two HTH subdomains and four of the nine Pax factors have an additional HTH domain, the homeodomain (HD). We now review these diverse DNA binding modalities together with their properties as transcriptional activators and repressors. The action of Pax factors on gene expression is also exerted through recruitment of chromatin remodeling complexes that introduce either activating or repressive chromatin marks. Interestingly, the recent demonstration that Pax7 has pioneer activity, the unique property to “open” chromatin, further underlines the mechanistic versatility and the developmental importance of these factors.

1. Introduction

Pax factors exert critical roles in early development as revealed by mouse and human mutations (as discussed in other papers of this series). One unique aspect of the structure of Pax factors is the presence of two DNA binding domains (DBD), the paired (PD) and homeodomain (HD) in a subset of factors and their differential use for target sequence/gene recognition (reviewed in [1]).

These dual modalities of DNA interactions provide versatility for Pax gene action in the control of diverse developmental programs. Targeted mutagenesis experiments of one or the other DBD have suggested that each DBD may be more critical for specific developmental roles. This review paper will primarily focus on the structure and function of Pax proteins as transcription factors (TF) and particularly, on molecular mechanisms that may provide versatility and specificity of action towards subsets of gene regulatory networks. Beyond the action of Pax factors as classical DNA binding TFs, the recent demonstration that Pax7 can function as a pioneer TF adds a new dimension to the action of these factors [2]. Pioneer factors can access their target sequence in compacted chromatin and lead to chromatin “opening” [3,4]. The pioneer activity is particularly well suited for the action of master regulatory genes because it can modify entire developmental programs through remodeling of the chromatin landscape.

2. Multiple DNA binding modalities

The Pax family of TFs is quite conserved in evolution (reviewed in [1]) and the hallmark of these DNA binding proteins is the presence of a PD DBD. This review will focus on mouse and human genes that have been the subject of numerous studies. In mammals, there are nine Pax genes (Fig. 1). These genes fall into four subgroups depending on the presence or absence of two domains: these are in addition to PD, a complete or truncated HD and an octapeptide (OP) motif in the linker region between PD and HD. The OP is related to the engrailed-homology motif that was associated with strong transcriptional repression activity [5]. Pax7 is unique in the family for the presence of a C-terminal 14 amino acids sequence, the OAR (Otp/aristaless/ Rax), that is found in over 30 paired-type HD proteins but is still of poorly defined activity [6].

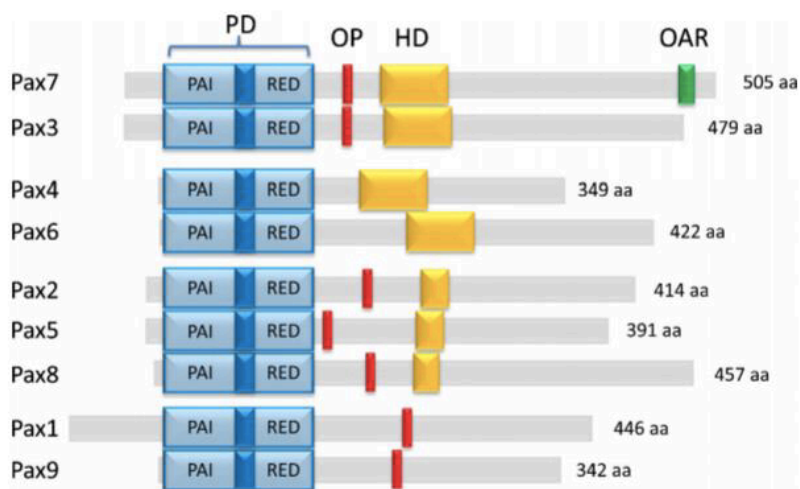


Figure 1. Structure of Pax transcription factors.

Schematic drawings represent the conserved features of the mouse Pax factors. These include the paired (PD) DNA binding domain (blue) present in all Pax factors and subdivided into two helix-turn-helix (HTH) motif (PAI and RED). The Pax family is classified in four groups depending on the presence of other domains such as the homeodomain (HD, yellow, complete in Pax3/7/4/6, partial in Pax2/5/8) and the octapeptide (OP, red, absent in Pax4/6); this domain is homologous to the engrailed homology 1 (En1) domains that recruit co-repressors of the Groucho/Tle family. Pax7 is the only member of the family that contains an OAR domain (green) that is also found in many paired-type HD factors.

DNA binding by Pax factors can primarily involve either PD or HD but the simplistic model of either DBD being responsible for protein:DNA interaction may not be prevalent. Indeed, DNA binding activity is clearly dependent on interactions between the two DBDs. In addition, the C-terminal region of some Pax proteins interacts with the DBDs to influence DNA binding activity. In accordance with the involvement of PD and/or HD in DNA binding, some

target DNA sequences contain either a core PD or a HD recognition motif; also, both motifs are often found in the same target regulatory sequences [2,7–9]. Although the specificity of binding of each Pax protein for individual DNA sequences is not well resolved, the available data indicates that different Pax proteins have distinct binding specificity, in particular when comparing the family subgroups. A large-scale analysis of DNA binding properties for many over-expressed human TFs revealed the core PD and HD motifs [10] but not the subtlety of factor-specific binding provided by ChIPseq as discussed below.

The implication of various Pax protein domains in determining DNA binding specificity, and ultimately activity, is further complicated by the existence of splicing variant isoforms for many Pax proteins (variant list reviewed in reference [11]). Some of these variants differ from the canonical form by only a few amino acids or represent truncations: they have been associated with either varying strengths of DNA binding or altered specificity/activity.

2.1 The PD domain and DNA recognition

The PD is a bipartite DNA binding structure composed of two helix-turn-helix (HTH) motifs that resemble bacterial DNA binding proteins [7,12]. These two motifs are separated by a flexible linker. The two DNA interacting subdomains of PD are thus structurally related to the HD. Crystal structures of *Drosophila* Prd [13], PAX6 [14], and PAX5 [15] together with nuclear magnetic resonance analysis of PAX8 [16] have supported the model of the third helix of these HTH motifs being implicated in sequence-specific interactions. In addition, the N-terminal HTH motif, named PAI subdomain, is preceded by a β -hairpin that acts as a clamp interacting with the phosphate backbone and minor groove. The flexible linker between the N-terminal (PAI) and C-terminal (RED) subdomains interacts with bases in the minor groove. The N-terminal PAI subdomain, more specifically its 3rd helix, fits directly in the major groove [13] and recognizes the essential bases for PD binding specificity, i.e. the core motif GTCACGC (Fig. 2). For some Pax proteins, DNA interactions extend beyond this core motif, mostly in 3', to include sequences that interact with the C-terminal RED subdomain. For example, the Pax6 RED subdomain contacts sequences 6bp downstream of the divergent core TTCACGC to interact with the sequence TG/TA/CN [7,13]. An elegant *in vivo* study of Pax6 point mutations affecting either PAI or RED domains showed that they each have tissue-specific effects on development and gene expression [17]. Much remains to be clarified about the extent to which

each (or both) PD subdomain is implicated in DNA interactions and at what subset of genomic targets.

It is noteworthy that splicing variants within the Pax3, Pax6 and Pax8 PDs were shown to switch the binding modality from interactions involving both PD subdomains to variants interacting primarily through the N-terminal PAI subdomain [18–22]. For Pax6, these splicing variants with altered DNA specificity were associated with distinct roles in brain and retinal development [22,23]. For Pax3 and Pax7, alternate use of splicing acceptor sites leads to the inclusion of an additional glutamine residue within the PD with significant effect (about 5-fold) on DNA binding affinity toward the PD target sequence but not toward the HD target sequence, thus enhancing discrimination between the two classes of DNA targets [18,24].

On its own, the PD domain is thus a complex structure that has capability for many different modalities of DNA interactions. This complexity is further enhanced for the subgroups of Pax proteins that also include a separate HD domain.

2.2 The HD domain of Pax factors.

Four of nine Pax factors, namely the Pax4/Pax6 and Pax3/Pax7 pairs, contain a separate DNA-binding HD (Fig. 1). The HD is also a DBD of the HTH class like the PD subdomains. In addition, the subgroup of Pax2, Pax5 and Pax8 contain part of a HD corresponding primarily to the first helix. In Pax5, this helix was shown to interact with TBP and underphosphorylated Rb [25]; similar interactions were also reported for Pax6 [26]. The paired-type HD is characterised by a serine residue at position 50 of the 60-amino acids HD. This residue of the third helix is critical for DNA sequence recognition and defines the paired subgroup within the large family of HD factors [27]. The paired-type HD binds sequences related to the target sequences of other HD factors, namely sequences containing the TAAT motif (Fig. 2).

The paired-type HD has the unique property of forming cooperative dimers that bind to palindromic sites containing two TAAT motifs separated by 2 or 3 base pairs (Fig. 2) [28,29]. While the HD target of Pax proteins may appear, *in vitro* at least, to suffice for DNA interaction, *in vivo* assessment of Pax protein mutants deleted of their HD has shown that some biological functions do not require the HD: for example, a HD-deficient form of Pax6 is sufficient for pancreas development in zebrafish but not for lens development [21].

Pax	Domain	Motif	Consensus motif (PCR)	Reference
2,5,8,4,3	Isolated PD	P	GTCACGC	[7]
1	FL	P	GTTCCGC	[101]
2	FL	P	GTCA ^T _C -GC ^A _G T	[102]
5	FL	P	GT ^C _T -A ^A _C CGC	[12,49]
8	FL	P	GTCAC ^G _C	[7]
4	Isolated HD	H	TAATT	[103]
4	Isolated PD	P	^G _T - ^C _G - ^A _A - ^T _G GC GTG	[47]
6	FL	P	TTCACGC	[102,104]
6	Isolated PD	P.12	TTCACGCATGAGTGCC	[102,104]
6-5a	Isolated PD	P.16	CTCAGTGAATGTTTCATTGAC	[19]
6	Isolated HD	H.13 H.12	TAAT ^G _T CGATTA TAATCGATTA	[49]
3	Isolated HD	H.12	TAAT ^T _C GATTA	
3	FL	P	gTCACG ^G _A	[104]

Figure 2. DNA target sites of Pax factor binding.

Pax factors are regrouped according to their homology and the domain used in each DNA binding study is indicated, FL full-length. Consensus binding sequences are shown together with a systematic nomenclature where P refers to the paired consensus sequence GTCACGC and H to the homeodomain consensus binding site TAAT. Complex binding motifs are labelled with the letter symbol of the motif(s) present followed by D for direct or same orientation repeats, I for inverted, E for everted repeats and a number indicating the spacing between the repeats.

The investigation of natural Pax target sites revealed a wide spectrum of sequence arrangements that include combinations of PD and HD target sequences in various topologies, including single sites, direct repeats, inverted or everted sites. A sample of documented (i.e. investigated by either gel retardation assay and/or mutagenesis) target sites/genes is presented in Figure 3 together with a systematic nomenclature of target site topology. For each target site, the presence of PD (GTCACGC) or HD (TAAT) target sequence is indicated by P or H,

respectively, followed (when needed) by a letter indicating the relation between two motifs (D for direct or same orientation, I for inverted and E for everted orientations) and a number indicating the number of intervening nucleotides. Thus, the e5 (also called PRS-1) target sequence of the drosophila even skipped gene that was used for many DNA binding studies in the past is labelled PH.E5 because it includes HD and PD target sequences in everted orientation relative to each other separated by 5 bp (Fig. 3).

Pax	Motif	Natural sites	Reference
1 / 9	P	<i>Bapx1</i> enhancer TCCAATCCAATCCAAT CAAACG TAAAAT	[105]
2	P	<i>Pax5</i> enhancer GCTACCTAATTT CATGCATCAGAG	[106]
2	P P.D0	<i>Engrailed-2</i> enhancer BS-I GGCACGG BS-II GTCAAGGCTCAGTG	[107]
5	P.D0	<i>CD19</i> promoter GCGGTGG TACGCCTCAGGCCCA	[108]
8	P	<i>Tg</i> promoter CACTGCC AGTCAAGT GTTCTTGA	[109]
4	P.I-2	Glucagon promoter TTCACGCCTGAC	[110]
6	P.I10	Rat insulin I, C2 element AAATGAGGTGGAAAATGCTCAGCC	[111]
6	P.I2	Glucagon, G3 element TTCACGCCTGACTGAG	[111]
3	PH.E0 (PH.I4, PH.D6)	<i>Mitf</i> promoter TTTAGTCCAGTAGTATTAAT	[8]
	P	<i>Trp-1</i> promoter GTCACACTT	[8]
	PH.E0	<i>Myf5</i> enhancer GCATGACTAAT	[31]
	PH.E0	<i>Dmrt2</i> enhancer, site 2 TGTGTGACTAAT	[32]
	PH.E5	<i>Evenskipped</i> ATTAGCACCGTCCGC	[42]
7	PH.E0	<i>Pcsk2</i> (pioneer enhancer site) GCGTGACTAAT	[2]

Figure 3. Gene targets of Pax factor action.

A selection of natural binding sites is shown for the indicated Pax factors. Bold characters indicate the position of consensus P or H sequences and motifs are labelled with the same nomenclature as in Figure 2.

The advent of ChIPseq technology now offers the possibility of assessing genome-wide occupancy of recognition motifs for the different Pax proteins. There are a few reports of ChIPseq for Pax factors. Three groups did ChIPseq for the related Pax3 and/or Pax7 and identified by *de novo* motif searches the PD binding motif GTCACGGT at a subset of Pax3 and

Pax7 peaks [2,9,30]; the analyses were performed in rhabdomyosarcoma, myoblasts and pituitary cells. The myoblasts and pituitary Pax7 ChIPseq (which have 20% peaks in common) also revealed the presence of HD motifs (TAAT) at very high frequency. Interestingly, myoblasts show the frequent occurrence of a HD palindrome separated by 2bp (H.I2) whereas the Pax7 pituitary sites mostly have isolated TAAT motifs rather than the H.I2 palindrome (Fig. 4).

The parallel analyses of Pax3 and Pax7 recruitment in myoblasts [9] supported the idea of extensive redundancy in the targets of these factors but also documented specific binding sites for each factor; interestingly, the Pax7 ChIPseq revealed far more sites than the Pax3 data. This was unexpected in view of their relative roles in myoblasts and it highlights the importance of context for DNA binding. At this point, the binding site analyses only showed slightly different target sequence bias, with Pax3 showing preference for the PD motif whereas Pax7 appears to prefer the H.I2 palindrome [9]. In contrast, the Pax7 peaks in pituitary cells (Fig. 4b) have a much higher frequency of PD target sequences (44% versus 11%) and far less of the HD palindrome H.I2 (2% versus 78%). These striking differences likely reflect the importance of cellular context to influence recruitment at specific target sites. This context may result from interactions with co-regulatory DNA binding factors that are uniquely expressed in different cell types and/or with transcriptional co-regulators (activators or repressors). The role of splicing variants in target site selection also needs to be assessed at genome level.

A unique feature of the pituitary cell study is the identification of the composite Pax7 binding motif PH.E0 ATTAGTCACGG (Fig.2 and 4) that is preferentially associated with the pioneer action of Pax7 [2]. Similar sequence motifs are present at well-characterised myogenic transcription targets of Pax3/7 such as at the Myf5 and Dmrt2 enhancers (Fig. 3) but their putative roles as pioneering sites has not been investigated [31,32]. Interestingly, detailed analysis of Pax3 binding to a target sequence in the Mitf promoter that includes PH.I4, PH.D6 and PH.E0 motifs (Fig. 3) indicated that the PH.E0 motif is contributing the most to binding affinity by Pax3 although the downstream TAAT motifs also contribute [8]; these studies were conducted with truncated proteins and thus, the motif dependence for *in vivo* binding of full-length Pax3 remains to be defined.

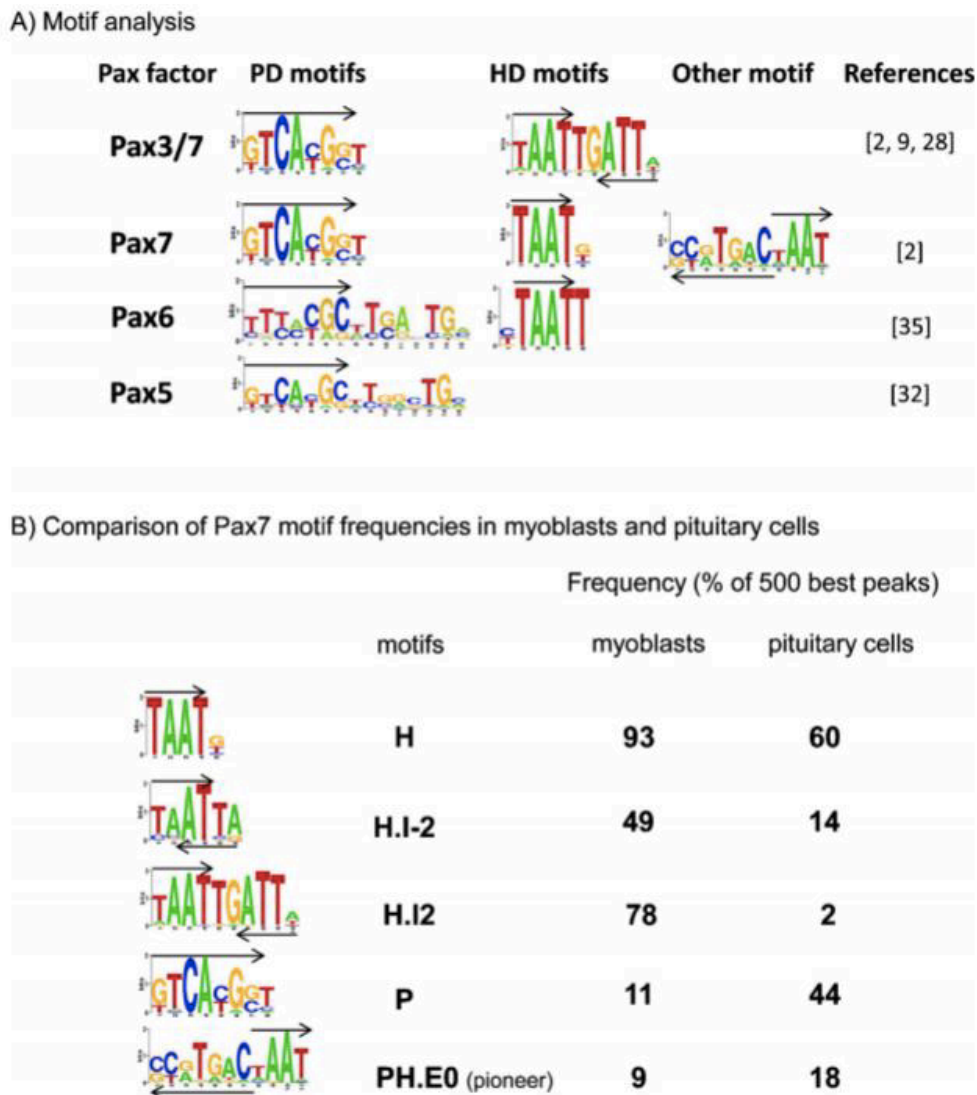


Figure 4. ChIPseq analyses of Pax genomic binding sites.

(A) The results of *de novo* motif analyses performed on ChIPseq datasets are shown.

(B) Comparison of motif frequencies present under Pax7 peaks in ChIPseq data from myoblasts [9] and pituitary [2] cells. Motif occurrence was determined for the 500 stronger peaks of each dataset using the HOMER program.

The analysis of Pax5 ChIPseq in proB and mature B cells identified a dominant motif GTCACGC-TGG-TG [33] that corresponds to previous findings (Figs. 2-4), and a similar conclusion was found for Pax8 in thyroid cells [34]. The Pax6 ChIPseq also revealed a dominant and related PD motif, together with a subset of HD motifs [35].

2.3 Complex interactions between PD and HD domains.

The Pax factors that contain both PD and HD (Pax3, Pax4, Pax6 and Pax7) exhibit an intriguing interdependence between their DBDs for efficient DNA interactions. Indeed, mutations in one DBD often affect DNA interactions by both DBDs; this was first revealed through analyses of human mutations found in patients harbouring PAX3 (Waardenburg syndrome) or PAX6 (aniridia) mutations. When produced individually, both PD and HD bind their cognate sequence *in vitro*, and their combination results in synergistic binding to either cognate sequence [7]. However, *in vivo* function appears to require that both domains be within the same protein as trans complementation was not found to be as efficient [36,37].

Biochemical analysis of this interdependence was carried out in particular for Pax3: in this case, the PD enhances binding by the HD and alters the sequence requirements for this binding [38]. Thus, the PD and HD behave as a single DNA binding module in which the linker sequences are critical for proper synergism and binding efficiency; this conclusion is well supported by numerous human PAX3 mutations that alter DNA binding properties [8,39–41]. For example, mutations in the PD of PAX3 from Waardenburg syndrome patients (such as G43A and P50L) impair DNA binding to sequences that lack a HD recognition motif [42]. Also, deletion of 6 amino acids in a highly conserved region of the PAI subdomain of the PD (aa 29-34) completely abolish binding to either PD or HD target sequences [42,43]. Similar conclusions were also reached for PAX6 [44,45]. The interdependence between PD and HD was mostly studied in relation to binding of cognate DNA sequences by either DBD. However, studies of natural targets for Pax3 and Pax6 have also revealed a subset of targets that include composite sequences containing both PD and HD motifs juxtaposed next to each other [7,36] or separated by varying numbers of nucleotides as summarised in Fig. 2. The modalities and structural constraints for recognition of such composite motifs remain poorly analysed.

3. Transactivation domains and coactivators

In most Pax proteins, transactivation domains were identified in the C-terminal part of the protein, i.e. downstream of the DBD [46–50]. However, these transactivation domains were not mapped very precisely but a number of coactivators recruited by Pax factors have been identified (Fig. 5).

Pax factor	HD	OP	Co-Activator	Co-Repressor	Chromatin remodeling
2	1/2	+	PTIP [50,51] (Grg4 [63,64] PPM1B [64]	MLL3/4 [51] Prmt5 [63]
5	1/2	+	PTIP [52] Daxx [112] CBP [112]	Grg4 [62] NCoR1 [52] Daxx [112]	Brg1 [52]
8	1/2	+	Rb [59]		
3	+	+	Pax3/7BP [56] Taz, Tap65 [57] HDAC10 [70]	HIRA [67] Daxx [68] KAP1, HP1 [69]	Tif1b [97] HP1 [97]
7	+	+	Pax3/7BP [56] Carm1 (Prmt4) [55]		H3K4/HMT [56] MLL1/2:Wdr5 [54,55]
4	+	-			
6	+	-	Tip60 [58]		Brg1 [94] Snf2h [94]

Figure 5. Co-regulator proteins interacting with Pax factors.

The co-activators, co-repressor and chromatin remodeling complexes that are recruited directly or indirectly to Pax factors are shown together with references to the relevant work.

The C-terminal domain of Pax2 was used as bait to identify a co-activator protein that interacts with a variety of Pax factors [50]. This ubiquitous nuclear co-activator, named Pax Transactivation-Domain Interacting Protein (PTIP), is a bridge for recruitment of the MLL3/4 chromatin remodelling complexes that has H3K4 methyltransferase activity. Indeed, PTIP is required for Pax2-driven recruitment of the MLL complex to specific DNA binding sites [51]. PTIP was also shown to interact with Pax5 and both were found to co-localise in developing B-cells, for example at an enhancer of immunoglobulin light-chain genes [52]. Also at Pax5dependent promoters and enhancers, PTIP recruitment is accompanied by recruitment of Brg1, the ATPase of the SWI/SNF chromatin remodelling complex. It is also noteworthy that

PTIP interacts with Pituitary Homeobox 2 (Pitx2) through its HD and that this leads to recruitment of the MLL4 complex [53].

Pax3 and Pax7 appear to use different co-activator proteins to recruit the MLL complex to target genes [54]. Two different co-activators interacting with the PD were identified, CARM1 (PRMT4) and Pax3/7BP [55,56]. The Hippo pathway effectors Taz and Yap65 act as Pax3 coactivators and neural crest inactivation of their genes decreased expression of Pax3-dependent genes [57].

Also, the coactivator Tip60, a histone acetyltransferase, interacts with the Pax6 PD and enhances Pax6-dependent transcription [58]. Finally, Pax8 interacts directly with Rb in thyroid follicular cells for activation of the thyroperoxydase gene promoter [59], Rb thus behaving as a co-activator of Pax8 activity.

4. Transcriptional repression by Pax factors and corepressors

Most Pax factors have one predominant feature for transcriptional repression, the octapeptide (OP) present downstream of the PD, in the linker region between PD and HD for the factors that contain both (Fig.1). This conserved OP contains the motif also known as the engrailed-homology 1 (Eh1) motif that is a binding site [60] for transcriptional repressors of the groucho (Grg) / transducin-like enhancer of zest (Tle) family (Fig. 5). A similar motif (conveying transcriptional repression) is found in many other structural families of TFs, including factors of the Tbox, Forkhead and Homeodomain families [61].

The action of these co-repressors was investigated in greater detail for the Pax2/5/8 subfamily that recruit the Groucho family corepressor Grg4 to exert transcriptional repression [62]. Interestingly, Grg4 is present in a complex that includes a phosphatase that can dephosphorylate the C-terminal activation domain of Pax2 [63,64]. Grg4 thus primarily exerts its repressor activity by antagonising the phosphorylated transactivation domain where phosphorylation enhances interaction with the co-activator PTIP via its phosphoserine binding BRCT domain [65]. In addition, the Grg4 complex contains an arginine methyltransferase, PRMT5, that dimethylates histone arginines, in particular H4R3 that is associated with repressed epigenetic states [63,66]. Pax5-dependent repression was also associated with recruitment of HDAC-containing NCor1 complexes [52].

In addition, repression domains were identified in the C-terminal domain of some Pax factors, such as Pax4 [47]. Pax3 was found to recruit different proteins with co-repressor activity (Fig. 5): HIRA that interacts with the Pax3 HD [67] and Daxx that interacts with the HD and linker domains [68]. Also, the repressive activity of Pax3 is enhanced by recruitment of KAP1 and HP1 [69] and this activity is reversed by co-recruitment of HDAC10 at melanogenic target genes [70]; this reversal was associated with deacetylation of Pax3 itself.

Pax7 is the only mammalian Pax protein that has a PHT/OAR domain at its C-terminus; this domain does not in itself confer transcriptional repression, but it appears to contribute to transcriptional repression, possibly by masking the activity of transactivation domains [6]. Similar OAR sequences are found toward the C-terminus of about 35 paired-type HD proteins such as the Prx1/2 and Pitx1/2/3. It is noteworthy that one Pax7 variant does not include this C-

terminal region and consequently, this variant protein would be subject to quite different regulation compared to the major Pax7 isoforms.

5. Co-regulatory DNA binding transcription factors

Pax factor action was studied for many target genes and biological systems. These studies identified recurrent partners for DNA binding of target regulatory sequences. For example, Pax3 binds the c-RET promoter at a site that is adjacent to a binding site for Sox10 suggesting that they may operate together towards transcriptional regulation [71]. This idea is reinforced by the observation that the same two factors, Pax3 and Sox10, also cooperate for regulation of MITF gene expression in melanocytes [72–74]. While a large number of PAX3 mutations account for Waardenburg syndrome, it is noteworthy that this syndrome has also been associated with mutations in SOX10 [75] and MITF [76].

Another pair of Pax and Sox factors interact for transcriptional control of the δ -crystallin gene enhancer involved in lens-specific transcription, namely Pax6 and Sox2 [77]. These factors bind cooperatively to adjacent binding sites [78] and this results in synergetic activation of transcription. It would be interesting to determine whether the Pax6/Sox2 interaction involves similar protein domains as the interactions between Pax3 and Sox10.

Pax3 was also shown to elicit synergistic activation of transcription at the Myf5 epaxial somite enhancer through interactions with the Shh-dependent activator Gli2 and with Zic1 [79].

In the case of Pax5, protein/protein interactions with Ets-1 have been shown to influence the DNA binding properties of Ets1 [15,80]. Ets1 also interacts with Pax3 to synergistically activate expression of the MET gene promoter [81].

6. Regulation of Pax factor activity

The first level of regulation of Pax factor activity is developmental, namely production of different isoforms depending on tissue or time of development. Most Pax factors are expressed as various isoforms that are set for a developmental context, meaning for a defined tissue or period of development. The bulk of these are produced by differential splicing and this developmental choice produces proteins with different, and sometimes opposite, activities. For example, some variants of Pax6 and Pax8 have altered DNA binding activity as a result of alternate splicing events affecting their DBDs [20,22,37] whereas others have altered transactivation potential and developmental activity because of the absence of their C-terminal domain [21]. Pax3 is a good case in point because it has numerous variants with altered properties, e.g. [18], that have either positive or negative effects on proliferation or differentiation in specific Pax3-expressing tissues such as melanocytes [82]. Interestingly, the Pax3 C-terminus contains lysine residues that can be acetylated with differential effects on gene expression [83]; and the deacetylases SIRT1 and HDAC10 may reverse these effects [70,83]. The differential activities of Pax3 variants are mirrored by PAX3 mutations found in Waardenburg patients [18]. An extensive review of Pax factor variants is provided in a Table of a recent review [11].

SUMOylation was shown to play critical roles for the activity of two Pax factors, namely short splicing variants of Pax6 and Pax7. The p32 isoform of Pax6 is deleted of its PD and thus can only bind DNA through its HD; interestingly, SUMOylation of its K91 residue significantly increases DNA binding and transcriptional activation [84]. Similarly, SUMOylation of K85 of Pax7 appears critical for its activity in neural crest and in myogenic differentiation [85].

As for many other TFs, the activity of Pax factors is regulated by cytoplasmic signalling pathways that exert their action through activation of kinases. The transcriptional activity of many Pax factors is potentiated by phosphorylation, and this most often involves the C-terminal transactivation domain such as in Pax3 [86] and Pax6 [87]. Studies of Pax2 regulation by phosphorylation of its transactivation domain is particularly detailed as it was shown that this phosphorylation enhances recruitment of the PTIP co-activator [65] which is itself required for recruitment of the chromatin remodelling MLL3/4 complexes [63]. Further, this

phosphorylation is itself the target of regulation by the Grg4 complex that is recruited to the octapeptide and includes the PPM1B phosphatase that dephosphorylates the C-terminal transactivation domain and thus reverses interaction with the co-activator PTIP [64].

7. Pax factors in chromatin remodelling

As for other TFs, the recruitment of co-activators and co-repressors contributes to the recruitment of chromatin remodelling complexes (Fig. 5) in addition to direct transcriptional regulation. These complexes either introduce permissive marks onto chromatin that favor gene expression or they introduce negative marks that impede chromatin access and transcription. The activating complexes are related to the drosophila Trithorax (Trx) complex of proteins and in mammals, there are four related complexes that take part in this process: they are the MLL (mixed-lineage leukemia) 1-4 complexes and each MLL complex contains a set of common proteins, such as Wdr5, Rbbp5 and Ash2 [88]. The chromatin repressor complexes are related to the drosophila Polycomb (Pc) complex and in mammals, two complexes containing either the protein PRC1 or PRC2 fulfill complementary roles, one as reader and the other as writer of repressive marks, respectively [89,90].

7.1. Chromatin activation by Pax factors:

As discussed above, both Pax2 and Pax5 recruit the co-activator PTIP to their transactivation domains. PTIP associates with the MLL3/4 complexes that include H3K4 methyltransferases [51,91,92]. Thus, Pax2 binding leads to PTIP and MLL3/4 recruitment and to increased levels of H3K4me3. This mark is associated with activated promoters and is permissive for transcription. The importance of PTIP in this paradigm was clearly established in PTIP-deficient cells where Pax2 target genes exhibit reduced levels of H3K4me1/2/3, AshL2 and Rbbp5 without affecting Pax2 binding [51].

A similar paradigm was suggested for the action of Pax3 and Pax7 in muscle cells where the MLL complex is recruited to muscle-specific genes at differentiation [93]. This recruitment can be exerted either through Pax7 or Pax3 [55,56]. Two different co-activators may facilitate this recruitment. The co-activator PRMT4 (Carm1) methylates many arginine residues in the N-terminus of Pax7 and methylated Pax7 then recruits the MLL1/2 complex via interactions with Wdr5 [55] for induction of the myogenic regulatory gene Myf5. Also, the protein Pax3/7BP is an adaptor for Pax7 and Pax3 recruitment of H3K4 methyltransferases to these Pax factors [56]. Thus, recruitment of the MLL and histone methyltransferase complexes to target promoters leads to enhanced H3K4me3 that promotes gene transcription.

Pax6 was suggested to recruit Brg1 and Snf2h, components of the SWI/SNF and ISWI ATP-dependent chromatin remodeling complexes, respectively. At the α A-crystallin locus, this is associated with H3K9 acetylation and H3K4 trimethylation [94].

In B cells, Pax5 action leads to removal of repressive chromatin marks. Indeed, during early B-cell development, Pax5 is required for VH-to-DJH recombination that normally occurs in committed B cells. Non-committed B cells show high level of H3K9 methylation at the VH locus consistent with inactive chromatin, and Pax5 is necessary and sufficient for removal of H3K9me3 at the VH locus through a replication-independent mechanism involving histone exchange [95].

7.2. Pax factors and repressive chromatin.

The epigenetic contribution to specific gene repression by Pax factors was investigated in particular for the Pax2-dependent recruitment of Grg4 [63]. Grg4 is within a protein complex that also includes the arginine methyltransferase Prmt5 that introduces symmetric dimethylation at H4R3 [66]. This dimethylated arginine is associated with repressive chromatin and contributes to the recruitment of the PRC2 repressor complex that includes the KMT6 (Ezh2) and Suz12 proteins; this promotes trimethylation of H3K27, a mark of repressed or paused chromatin states [96]. This is one mechanism used by Pax factors to exert gene-specific repression. In addition, Prmt5-mediated methylation of H4R3 recruits Dnmt3a, a DNA cytosine methylase that is linked to *de novo* methylation and gene silencing [66]. This suggest that Pax factors may have a far more profound effect on the epigenome than histone modification as *de novo* DNA methylation may have long term effects on gene expression and cell fate.

It is interesting and intriguing that two different Pax factors (Pax3 and Pax9) have been implicated in silencing of mouse heterochromatin [97]. This action is presumably initiated through extensive binding of these factors to repeated satellite DNA leading to recruitment of the heterochromatin proteins HP1 and Kap1 (Trim28). The recruitment of these proteins to heterochromatin is associated with the presence of the repressive epigenetic marks H3K9me3 and H4K20me3 and indeed, the abundance of both these marks at pericentric heterochromatin is reduced in Pax3-deficient cells [97]. This heterochromatinization function of Pax factors may

represent a specialised function that is distinct from its gene-specific action but it may also draw on the activity of similar coregulatory protein partner.

8. Pax factors as pioneers for chromatin remodelling

The biology associated with the developmental roles of many Pax factors is suggestive of critical actions in early development. A prevalent mechanism for cell specific control of gene expression is the combinatorial use of a subset of TFs; this model has been validated in numerous systems and cells and the combination of factors required for identity of specific lineages often includes Pax factors. However, major developmental decisions, or bottlenecks, involve significant remodelling of the epigenetic landscape in order to radically change permissive versus non-permissive chromatin environments. In the mid-seventies, this type of function was ascribed to so-called selector genes that define early developmental fields and determine the ultimate fates of lineages that will be specified later in development [98]. At chromatin level, this idea was borne out in the form of so-called “pioneer factors” that would have the ability to remodel (“open”) the chromatin environment and thus unravel a whole new array of possibilities for gene regulation through combinatorial use of TFs. The last five years have provided some experimental demonstration of this pioneer activity and one remarkable example is provided by the pluripotency factors, Sox2, Oct4 and Klf4, that can reprogram differentiated cells into induced pluripotent stem (iPS) cells [4].

One Pax factor was shown to have pioneer activity: this is Pax7 that specifies the intermediate pituitary identity and in so doing, selects the fate of cells that are driven into terminal differentiation by the Tbox factor Tpit [2]. In absence of Pax7 as occurs in the remaining of the pituitary gland, the so-called anterior lobe, Tpit-driven differentiation yields the corticotrope fate. Whole-genome assessment of epigenetic marks associated with active regulatory elements in a reconstituted system showed that Pax7 has the ability to target a subset of ≈ 2500 putative enhancers and trigger remodelling of their chromatin into a permissive state. These pioneered target enhancers preferentially contain the Pax7 target sequence PH.E0 (Figs. 3, 4) that is an everted composite of paired and homeodomain recognition motifs.

This action unmaskes Tpit binding sites that were previously inaccessible and leads to Tpit-driven differentiation into the melanotrope fate. In this simple binary alternate cell differentiation scheme, Pax7 is thus the pioneer that alters the epigenome in order to change the differentiation potential of precursor cells. In both lineages, terminal differentiation is specified

by another factor, Tpit. At enhancers targeted by Pax7 for determination of the melanotrope fate, Pax7 binding results in appearance of an active chromatin mark, H3K4me1, and increases DNA accessibility as revealed by FAIREseq (formaldehyde-assisted isolation of regulatory elements) [99,100].

Notwithstanding this pioneer action, Pax7 acts on a very large subset of enhancers and promoters in the same cells to participate in the combinatorial control of cell-specific transcription. At present, the similarities and differences in those two actions of Pax7 (pioneer and classical TF actions) remain to be defined.

9. Conclusions and perspectives

Looking back on this review provides an interesting snapshot on the evolution of research. While investigations of the biological roles of Pax factors progressed steadily, mechanistic studies of Pax interactions with DNA and transcriptional partners shifted more towards recently uncovered aspects of chromatin remodeling. Notwithstanding this novelty, it is also clear that DNA interaction studies were hampered by technical limitations. The tremendous power of genome-wide techniques such as ChIPseq now opens the way to tackle the great versatility of Pax factor interactions with diverse DNA target sequences and its implications for regulation of gene networks. Taken together with the actions of Pax factor as pioneers of chromatin remodeling and with genome-wide assessment of human PAX gene mutations, these approaches will radically deepen our understanding of the critical roles of Pax factors in development and disease.

Acknowledgements

We thank the many colleagues who have contributed to our work over the years, both lab members and collaborators from around the world. The critical feedback provided by Alan Underhill on this review is greatly appreciated. We are grateful to Tabasum Abdul-Rasul for her expert assistance. Work in our lab was supported by grants from the Canadian Institutes of Health Research (CIHR).

References

1. Underhill DA (2012) PAX proteins and fables of their reconstruction. *Crit Rev Eukaryot Gene Expr* 22: 161-177.
2. Budry L, Balsalobre A, Gauthier Y, Khetchoumian K, L'honore A, Vallette S, Brue T, Figarella-Branger D, Meij B, Drouin J (2012) The selector gene Pax7 dictates alternate pituitary cell fates through its pioneer action on chromatin remodeling. *Genes Dev* 26: 2299-2310.
3. Drouin J (2014) Minireview: pioneer transcription factors in cell fate specification. *Mol Endocrinol* 28: 989-998.
4. Iwafuchi-Doi M, Zaret KS (2014) Pioneer transcription factors in cell reprogramming. *Genes Dev* 28: 2679-2692.
5. Smith ST, Jaynes JB (1996) A conserved region of engrailed, shared among all en-, gsc-, Nk1-, Nk2- and msh-class homeoproteins, mediates active transcriptional repression *in vivo*. *Development* 122: 3141-3150.
6. Norris RA, Kern MJ (2001) The identification of Prx1 transcription regulatory domains provides a mechanism for unequal compensation by the Prx1 and Prx2 loci. *J Biol Chem* 276: 26829-26837.
7. Jun S, Desplan C (1996) Cooperative interactions between paired domain and homeodomain. *Development* 122: 2639-2650.
8. Corry GN, Underhill DA (2005) Pax3 target gene recognition occurs through distinct modes that are differentially affected by disease-associated mutations. *Pigment Cell Res* 18: 427-438.
9. Soleimani VD, Punch VG, Kawabe Y, Jones AE, Palidwor GA, Porter CJ, Cross JW, Carvajal JJ, Kockx CE, van Ijcken WF, Perkins TJ, Rigby PW, Grosveld F, Rudnicki MA (2012) Transcriptional dominance of pax7 in adult myogenesis is due to high-affinity recognition of homeodomain motifs. *Dev Cell* 22: 1208-1220.
10. Jolma A, Yan J, Whittington T, Toivonen J, Nitta KR, Rastas P, Morgunova E, Enge M, Taipale M, Wei G, Palin K, Vaquerizas JM, Vincentelli R, Luscombe NM, Hughes TR, Lemaire P, Ukkonen E, Kivioja T, Taipale J (2013) DNA-binding specificities of human transcription factors. *Cell* 152: 327-339.
11. Blake JA, Ziman MR (2014) Pax genes: regulators of lineage specification and progenitor cell maintenance. *Development* 141: 737-751.
12. Czerny T, Schaffner G, Busslinger M (1993) DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site. *Genes Dev* 7: 2048-2061.
13. Xu W, Rould MA, Jun S, Desplan C, Pabo CO (1995) Crystal structure of a paired domain-DNA complex at 2.5 Å resolution reveals structural basis for Pax developmental mutations. *Cell* 80: 639-650.
14. Xu HE, Rould MA, Xu W, Epstein JA, Maas RL, Pabo CO (1999) Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev* 13: 1263-1275.
15. Garvie CW, Hagman J, Wolberger C (2001) Structural studies of Ets-1/Pax5 complex formation on DNA. *Mol Cell* 8: 1267-1276.

16. Codutti L, van Ingen H, Vascotto C, Fogolari F, Corazza A, Tell G, Quadrifoglio F, Viglino P, Boelens R, Esposito G (2008) The solution structure of DNA-free Pax-8 paired box domain accounts for redox regulation of transcriptional activity in the pax protein family. *J Biol Chem* 283: 33321-33328.
17. Walcher T, Xie Q, Sun J, Irmeler M, Beckers J, Ozturk T, Niessing D, Stoykova A, Cvekl A, Ninkovic J, Gotz M (2013) Functional dissection of the paired domain of Pax6 reveals molecular mechanisms of coordinating neurogenesis and proliferation. *Development* 140: 1123-1136.
18. Vogan KJ, Underhill DA, Gros P (1996) An alternative splicing event in the Pax-3 paired domain identifies the linker region as a key determinant of paired domain DNA-binding activity. *Mol Cell Biol* 16: 6677-6686.
19. Epstein JA, Glaser T, Cai J, Jepeal L, Walton DS, Maas RL (1994) Two independent and interactive DNA-binding subdomains of the Pax6 paired domain are regulated by alternative splicing. *Genes Dev* 8: 2022-2034.
20. Kozmik Z, Czerny T, Busslinger M (1997) Alternatively spliced insertions in the paired domain restrict the DNA sequence specificity of Pax6 and Pax8. *EMBO J* 16: 6793-6803.
21. Verbruggen V, Ek O, Georgette D, Delporte F, Von B, V, Detry N, Biemar F, Coutinho P, Martial JA, Voz ML, Manfroid I, Peers B (2010) The Pax6b homeodomain is dispensable for pancreatic endocrine cell differentiation in zebrafish. *J Biol Chem* 285: 13863-13873.
22. Azuma N, Tadokoro K, Asaka A, Yamada M, Yamaguchi Y, Handa H, Matsushima S, Watanabe T, Kohsaka S, Kida Y, Shiraishi T, Ogura T, Shimamura K, Nakafuku M (2005) The Pax6 isoform bearing an alternative spliced exon promotes the development of the neural retinal structure. *Hum Mol Genet* 14: 735-745.
23. Haubst N, Berger J, Radjendirane V, Graw J, Favor J, Saunders GF, Stoykova A, Gotz M (2004) Molecular dissection of Pax6 function: the specific roles of the paired domain and homeodomain in brain development. *Development* 131: 6131-6140.
24. Vogan KJ, Gros P (1997) The C-terminal subdomain makes an important contribution to the DNA binding activity of the Pax-3 paired domain. *J Biol Chem* 272: 28289-28295.
25. Eberhard D, Busslinger M (1999) The partial homeodomain of the transcription factor Pax-5 (BSAP) is an interaction motif for the retinoblastoma and TATA-binding proteins. *Cancer Res* 59: 1716s-1724s.
26. Cvekl A, Kashanchi F, Brady JN, Piatigorsky J (1999) Pax-6 interactions with TATA-box-binding protein and retinoblastoma protein. *Invest Ophthalmol Vis Sci* 40: 1343-1350.
27. Duboule, D. (1994) *Guidebook to the homeobox genes*. Oxford, U.K.: Sambrook & Tooze Publication. 284 p.
28. Wilson D, Sheng G, Lecuit T, Dostatni N, Desplan C (1993) Cooperative dimerization of paired class homeo domains on DNA. *Genes Dev* 7: 2120-2134.
29. Birrane G, Soni A, Ladias JA (2009) Structural basis for DNA recognition by the human PAX3 homeodomain. *Biochemistry* 48: 1148-1155.
30. Cao L, Yu Y, Bilke S, Walker RL, Mayeenuddin LH, Azorsa DO, Yang F, Pineda M, Helman LJ, Meltzer PS (2010) Genome-wide identification of PAX3-FKHR binding sites in rhabdomyosarcoma reveals candidate target genes important for development and cancer. *Cancer Res* 70: 6497-6508.

31. Buchberger A, Freitag D, Arnold HH (2007) A homeo-paired domain-binding motif directs Myf5 expression in progenitor cells of limb muscle. *Development* 134: 1171-1180.
32. Sato T, Rocancourt D, Marques L, Thorsteinsdottir S, Buckingham M (2010) A Pax3/Dmrt2/Myf5 regulatory cascade functions at the onset of myogenesis. *PLoS Genet* 6: e1000897.
33. Revilla ID, Bilic I, Vilagos B, Tagoh H, Ebert A, Tamir IM, Smeenk L, Trupke J, Sommer A, Jaritz M, Busslinger M (2012) The B-cell identity factor Pax5 regulates distinct transcriptional programmes in early and late B lymphopoiesis. *EMBO J* 31: 3130-3146.
34. Ruiz-Llorente S, Carrillo Santa dP, Sastre-Perona A, Montero-Conde C, Gomez-Lopez G, Fagin JA, Valencia A, Pisano DG, Santisteban P (2012) Genome-wide analysis of Pax8 binding provides new insights into thyroid functions. *BMC Genomics* 13: 147.
35. Sun J, Rockowitz S, Xie Q, Ashery-Padan R, Zheng D, Cvekl A (2015) Identification of *in vivo* DNA-binding mechanisms of Pax6 and reconstruction of Pax6-dependent gene regulatory networks during forebrain and lens development. *Nucleic Acids Res* .
36. Bertuccioli C, Fasano L, Jun S, Wang S, Sheng G, Desplan C (1996) *In vivo* requirement for the paired domain and homeodomain of the paired segmentation gene product. *Development* 122: 2673-2685.
37. Clements J, Hens K, Merugu S, Dichtl B, de Couet HG, Callaerts P (2009) Mutational analysis of the eyeless gene and phenotypic rescue reveal that an intact Eyeless protein is necessary for normal eye and brain development in *Drosophila*. *Dev Biol* 334: 503-512.
38. Underhill DA, Gros P (1997) The paired-domain regulates DNA binding by the homeodomain within the intact Pax-3 protein. *J Biol Chem* 272: 14175-14182.
39. Underhill DA, Vogan KJ, Gros P (1995) Analysis of the mouse Splotch-delayed mutation indicates that the Pax-3 paired domain can influence homeodomain DNA-binding activity. *Proc Natl Acad Sci USA* 92: 3692-3696.
40. Fortin AS, Underhill DA, Gros P (1997) Reciprocal effect of Waardenburg syndrome mutations on DNA binding by the Pax-3 paired domain and homeodomain. *Hum Mol Genet* 6: 1781-1790.
41. Apuzzo S, Gros P (2002) Site-specific modification of single cysteine Pax3 mutants reveals reciprocal regulation of DNA binding activity of the paired and homeo domain. *Biochemistry* 41: 12076-12085.
42. Chalepakis G, Goulding M, Read A, Strachan T, Gruss P (1994) Molecular basis of splotch and Waardenburg Pax-3 mutations. *Proc Natl Acad Sci U S A* 91: 3685-3689.
43. Tassabehji M, Read AP, Newton VE, Harris R, Balling R, Gruss P, Strachan T (1992) Waardenburg's syndrome patients have mutations in the human homologue of the Pax-3 paired box gene [see comments]. *Nature* 355: 635-636.
44. Singh S, Stellrecht CM, Tang HK, Saunders GF (2000) Modulation of PAX6 homeodomain function by the paired domain. *J Biol Chem* 275: 17306-17313.
45. Mishra R, Gorlov IP, Chao LY, Singh S, Saunders GF (2002) PAX6, paired domain influences sequence recognition by the homeodomain. *J Biol Chem* 277: 49488-49494.
46. Chalepakis G, Jones FS, Edelman GM, Gruss P (1994) Pax-3 contains domains for transcription activation and transcription inhibition. *Proc Natl Acad Sci U S A* 91: 12745-12749.
47. Fujitani Y, Kajimoto Y, Yasuda T, Matsuoka TA, Kaneto H, Umayahara Y, Fujita N, Watada H, Miyazaki JI, Yamasaki Y, Hori M (1999) Identification of a portable repression domain and an

- E1A-responsive activation domain in Pax4: a possible role of Pax4 as a transcriptional repressor in the pancreas. *Mol Cell Biol* 19: 8281-8291.
48. Kalousova A, Benes V, Paces J, Paces V, Kozmik Z (1999) DNA binding and transactivating properties of the paired and homeobox protein Pax4. *Biochem Biophys Res Commun* 259: 510-518.
 49. Czerny T, Busslinger M (1995) DNA-binding and transactivation properties of Pax-6: three amino acids in the paired domain are responsible for the different sequence recognition of Pax-6 and BSAP (Pax-5). *Mol Cell Biol* 15: 2858-2871.
 50. Lechner MS, Levitan I, Dressler GR (2000) PTIP, a novel BRCT domain-containing protein interacts with Pax2 and is associated with active chromatin. *Nucleic Acids Res* 28: 2741-2751.
 51. Patel SR, Kim D, Levitan I, Dressler GR (2007) The BRCT-domain containing protein PTIP links PAX2 to a histone H3, lysine 4 methyltransferase complex. *Dev Cell* 13: 580-592.
 52. McManus S, Ebert A, Salvagiotto G, Medvedovic J, Sun Q, Tamir I, Jaritz M, Tagoh H, Busslinger M (2011) The transcription factor Pax5 regulates its target genes by recruiting chromatin-modifying proteins in committed B cells. *EMBO J* 30: 2388-2404.
 53. Liu Y, Huang Y, Fan J, Zhu GZ (2014) PITX2 associates with PTIP-containing histone H3 lysine 4 methyltransferase complex. *Biochem Biophys Res Commun* 444: 634-637.
 54. McKinnell IW, Ishibashi J, Le Grand F, Punch VG, Addicks GC, Greenblatt JF, Dilworth FJ, Rudnicki MA (2008) Pax7 activates myogenic genes by recruitment of a histone methyltransferase complex. *Nat Cell Biol* 10: 77-84.
 55. Kawabe YI, Wang YX, McKinnell IW, Bedford MT, Rudnicki MA (2012) *Carm1* Regulates Pax7 Transcriptional Activity through MLL1/2 Recruitment during Asymmetric Satellite Stem Cell Divisions. *Cell Stem Cell* 2012;11:333-45.
 56. Diao Y, Guo X, Li Y, Sun K, Lu L, Jiang L, Fu X, Zhu H, Sun H, Wang H, Wu Z (2012) Pax3/7BP Is a Pax7- and Pax3-Binding Protein that Regulates the Proliferation of Muscle Precursor Cells by an Epigenetic Mechanism. *Cell Stem Cell* 11: 231-241.
 57. Manderfield LJ, Engleka KA, Aghajanian H, Gupta M, Yang S, Li L, Baggs JE, Hogenesch JB, Olson EN, Epstein JA (2014) Pax3 and hippo signaling coordinate melanocyte gene expression in neural crest. *Cell Rep* 9: 1885-1895.
 58. Kim CH, Kim JW, Jang SM, An JH, Song KH, Choi KH (2012) Transcriptional activity of paired homeobox Pax6 is enhanced by histone acetyltransferase Tip60 during mouse retina development. *Biochem Biophys Res Commun* 424: 427-432.
 59. Miccadei S, Provenzano C, Mojzisek M, Natali PG, Civitareale D (2005) Retinoblastoma protein acts as Pax 8 transcriptional coactivator. *Oncogene* 24: 6993-7001.
 60. Tolkunova EN, Fujioka M, Kobayashi M, Deka D, Baynes JB (1998) Two distinct types of repression domain in engrailed: one interacts with the Groucho corepressor and is preferentially active on integrated target genes. *Mol Cell Biol* 18: 2804-2814.
 61. Jennings BH, Pickles LM, Wainwright SM, Roe SM, Pearl LH, Ish-Horowicz D (2006) Molecular recognition of transcriptional repressor motifs by the WD domain of the Groucho/TLE corepressor. *Mol Cell* 22: 645-655.
 62. Eberhard D, Jimenez G, Heavey B, Busslinger M (2000) Transcriptional repression by Pax5 (BSAP) through interaction with corepressors of the Groucho family. *EMBO J* 19: 2292-2303.

63. Patel SR, Bhumbra SS, Paknikar RS, Dressler GR (2012) Epigenetic mechanisms of Groucho/Grg/TLE mediated transcriptional repression. *Mol Cell* 45: 185-195.
64. Abraham S, Paknikar R, Bhumbra S, Luan D, Garg R, Dressler GR, Patel SR (2015) The Groucho Associated Phosphatase PPM1B Displaces Pax Transactivation Domain Interacting Protein (PTIP) to Switch the transcription factor Pax2 from a Transcriptional Activator to a Repressor. *J Biol Chem* 2015;290: 7185–94.
65. Manke IA, Lowery DM, Nguyen A, Yaffe MB (2003) BRCT repeats as phosphopeptide-binding modules involved in protein targeting. *Science* 302: 636-639.
66. Zhao Q, Rank G, Tan YT, Li H, Moritz RL, Simpson RJ, Cerruti L, Curtis DJ, Patel DJ, Allis CD, Cunningham JM, Jane SM (2009) PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat Struct Mol Biol* 16: 304-311.
67. Magnaghi P, Roberts C, Lorain S, Lipinski M, Scambler PJ (1998) HIRA, a mammalian homologue of *Saccharomyces cerevisiae* transcriptional co-repressors, interacts with Pax3. *Nat Genet* 20: 74-77.
68. Hollenbach AD, Sublett JE, McPherson CJ, Grosveld G (1999) The Pax3-FKHR oncoprotein is unresponsive to the Pax3-associated repressor hDaxx. *EMBO J* 18: 3702-3711.
69. Hsieh MJ, Yao YL, Lai IL, Yang WM (2006) Transcriptional repression activity of PAX3 is modulated by competition between corepressor KAP1 and heterochromatin protein 1. *Biochem Biophys Res Commun* 349: 573-581.
70. Lai IL, Lin TP, Yao YL, Lin CY, Hsieh MJ, Yang WM (2010) Histone deacetylase 10 relieves repression on the melanogenic program by maintaining the deacetylation status of repressors. *J Biol Chem* 285: 7187-7196.
71. Lang D, Chen F, Milewski R, Li J, Lu MM, Epstein JA (2000) Pax3 is required for enteric ganglia formation and functions with Sox10 to modulate expression of c-ret. *J Clin Invest* 106: 963-971.
72. Bondurand N, Pingault V, Goerich DE, Lemort N, Sock E, Le Caignec C, Wegner M, Goossens M (2000) Interaction among SOX10, PAX3 and MITF, three genes altered in Waardenburg syndrome. *Hum Mol Genet* 9: 1907-1917.
73. Potterf SB, Furumura M, Dunn KJ, Arnheiter H, Pavan WJ (2000) Transcription factor hierarchy in Waardenburg syndrome: regulation of MITF expression by SOX10 and PAX3. *Hum Genet* 107: 1-6.
74. Verastegui C, Bille K, Ortonne JP, Ballotti R (2000) Regulation of the microphthalmia-associated transcription factor gene by the Waardenburg syndrome type 4 gene, SOX10. *J Biol Chem* 275: 30757-30760.
75. Pingault V, Bondurand N, Kuhlbrodt K, Goerich DE, Prehu MO, Puliti A, Herbarth B, Hermans-Borgmeyer I, Legius E, Matthijs G, Amiel J, Lyonnet S, Ceccherini I, Romeo G, Smith JC, Read AP, Wegner M, Goossens M (1998) SOX10 mutations in patients with Waardenburg-Hirschsprung disease. *Nat Genet* 18: 171-173.
76. Tassabehji M, Newton VE, Read AP (1994) Waardenburg syndrome type 2 caused by mutations in the human microphthalmia (MITF) gene. *Nat Genet* 8: 251-255.

77. Kamachi Y, Uchikawa M, Tanouchi A, Sekido R, Kondoh H (2001) Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes Dev* 15: 1272-1286.
78. Narasimhan K, Pillay S, Huang YH, Jayabal S, Udayasuryan B, Veerapandian V, Kolatkar P, Cojocaru V, Pervushin K, Jauch R (2015) DNA-mediated cooperativity facilitates the co-selection of cryptic enhancer sequences by SOX2 and PAX6 transcription factors. *Nucleic Acids Res* 43: 1513-1528.
79. Himeda CL, Barro MV, Emerson CP, Jr. (2013) Pax3 synergizes with Gli2 and Zic1 in transactivating the Myf5 epaxial somite enhancer. *Dev Biol* 383: 7-14.
80. Fitzsimmons D, Hodsdon W, Wheat W, Maira SM, Wasylyk B, Hagman J (1996) Pax-5 (BSAP) recruits Ets proto-oncogene family proteins to form functional ternary complexes on a B-cell-specific promoter. *Genes Dev* 10: 2198-2211.
81. Kubic JD, Little EC, Lui JW, Iizuka T, Lang D (2014) PAX3 and ETS1 synergistically activate MET expression in melanoma cells. *Oncogene* 2014.
82. Wang Q, Kumar S, Slevin M, Kumar P (2006) Functional analysis of alternative isoforms of the transcription factor PAX3 in melanocytes *in vitro*. *Cancer Res* 66: 8574-8580.
83. Ichi S, Boshnjaku V, Shen YW, Mania-Farnell B, Ahlgren S, Sapru S, Mansukhani N, McLone DG, Tomita T, Mayanil CS (2011) Role of Pax3 acetylation in the regulation of Hes1 and Neurog2. *Mol Biol Cell* 22: 503-512.
84. Yan Q, Gong L, Deng M, Zhang L, Sun S, Liu J, Ma H, Yuan D, Chen PC, Hu X, Liu J, Qin J, Xiao L, Huang XQ, Zhang J, Li DW (2010) Sumoylation activates the transcriptional activity of Pax-6, an important transcription factor for eye and brain development. *Proc Natl Acad Sci U S A* 107: 21034-21039.
85. Luan Z, Liu Y, Stuhlmiller TJ, Marquez J, Garcia-Castro MI (2013) SUMOylation of Pax7 is essential for neural crest and muscle development. *Cell Mol Life Sci* 70: 1793-1806.
86. Iyengar AS, Loupe JM, Miller PJ, Hollenbach AD (2012) Identification of CK2 as the kinase that phosphorylates Pax3 at Ser209 in early myogenic differentiation. *Biochem Biophys Res Commun* 428: 24-30.
87. Mikkola I, Bruun JA, Bjorkoy G, Holm T, Johansen T (1999) Phosphorylation of the transactivation domain of Pax6 by extracellular signal-regulated kinase and p38 mitogen-activated protein kinase. *J Biol Chem* 274: 15115-15126.
88. Schuettengruber B, Martinez AM, Iovino N, Cavalli G (2011) Trithorax group proteins: switching genes on and keeping them active. *Nat Rev Mol Cell Biol* 12: 799-814.
89. Di Croce L, Helin K (2013) Transcriptional regulation by Polycomb group proteins. *Nat Struct Mol Biol* 20: 1147-1155.
90. Simon JA, Kingston RE (2013) Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol Cell* 49: 808-824.
91. Daniel JA, Santos MA, Wang Z, Zang C, Schwab KR, Jankovic M, Filsuf D, Chen HT, Gazumyan A, Yamane A, Cho YW, Sun HW, Ge K, Peng W, Nussenzweig MC, Casellas R, Dressler GR, Zhao K, Nussenzweig A (2010) PTIP promotes chromatin changes critical for immunoglobulin class switch recombination. *Science* 329: 917-923.

92. Cho YW, Hong T, Hong S, Guo H, Yu H, Kim D, Guszczynski T, Dressler GR, Copeland TD, Kalkum M, Ge K (2007) PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *J Biol Chem* 282: 20395-20406.
93. Rampalli S, Li L, Mak E, Ge K, Brand M, Tapscott SJ, Dilworth FJ (2007) p38 MAPK signaling regulates recruitment of Ash2L-containing methyltransferase complexes to specific genes during differentiation. *Nat Struct Mol Biol* 14: 1150-1156.
94. Yang Y, Stopka T, Golestaneh N, Wang Y, Wu K, Li A, Chauhan BK, Gao CY, Cveklova K, Duncan MK, Pestell RG, Chepelinsky AB, Skoultchi AI, Cvekl A (2006) Regulation of alphaA-crystallin via Pax6, c-Maf, CREB and a broad domain of lens-specific chromatin. *EMBO J* 25: 2107-2118.
95. Johnson K, Pflugh DL, Yu D, Hesslein DG, Lin KI, Bothwell AL, Thomas-Tikhonenko A, Schatz DG, Calame K (2004) B cell-specific loss of histone 3 lysine 9 methylation in the V(H) locus depends on Pax5. *Nat Immunol* 5: 853-861.
96. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448: 553-560.
97. Bulut-Karslioglu A, Perrera V, Scaranaro M, de la Rosa-Velazquez IA, van de NS, Shukeir N, Popow J, Gerle B, Opravil S, Pagani M, Meidhof S, Brabletz T, Manke T, Lachner M, Jenuwein T (2012) A transcription factor-based mechanism for mouse heterochromatin formation. *Nat Struct Mol Biol* 19: 1023-1030.
98. Garcia-Bellido A (1975) Genetic control of wing disc development in *Drosophila*. *Ciba Found Symp* 0: 161-182.
99. Giresi PG, Lieb JD (2009) Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). *Methods* 48: 233-239.
100. Song L, Zhang Z, Grasfeder LL, Boyle AP, Giresi PG, Lee BK, Sheffield NC, Graf S, Huss M, Keefe D, Liu Z, London D, McDaniel RM, Shibata Y, Showers KA, Simon JM, Vales T, Wang T, Winter D, Zhang Z, Clarke ND, Birney E, Iyer VR, Crawford GE, Lieb JD, Furey TS (2011) Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res* 21: 1757-1767.
101. Chalepakis G, Fritsch R, Fickenscher H, Deutsch U, Goulding M, Gruss P (1991) The molecular basis of the undulated/Pax-1 mutation. *Cell* 66: 873-884.
102. Epstein J, Cai J, Glaser T, Jepeal L, Maas R (1994) Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes. *J Biol Chem* 269: 8355-8361.
103. Smith SB, Ee HC, Connors JR, German MS (1999) Paired-homeodomain transcription factor PAX4 acts as a transcriptional repressor in early pancreatic development. *Mol Cell Biol* 19: 8272-8280.
104. Chalepakis G, Gruss P (1995) Identification of DNA recognition sequences for the Pax3 paired domain. *Gene* 162: 267-270.
105. Rodrigo I, Hill RE, Balling R, Munsterberg A, Imai K (2003) Pax1 and Pax9 activate Bapx1 to induce chondrogenic differentiation in the sclerotome. *Development* 130: 473-482.

106. Pfeffer PL, Bouchard M, Busslinger M (2000) Pax2 and homeodomain proteins cooperatively regulate a 435 bp enhancer of the mouse Pax5 gene at the midbrain-hindbrain boundary. *Development* 127: 1017-1028.
107. Song DL, Chalepakis G, Gruss P, Joyner AL (1996) Two Pax-binding sites are required for early embryonic brain expression of an Engrailed-2 transgene. *Development* 122: 627-635.
108. Kozmik Z, Wang S, Dorfler P, Adams B, Busslinger M (1992) The promoter of the CD19 gene is a target for the B-cell-specific transcription factor BSAP. *Mol Cell Biol* 12: 2662-2672.
109. Zannini M, Francis-Lang H, Plachov D, Di Lauro R (1992) Pax-8, a paired domain-containing protein, binds to a sequence overlapping the recognition site of a homeodomain and activates transcription from two thyroid-specific promoters. *Mol Cell Biol* 12: 4230-4241.
110. Ritz-Laser B, Estreicher A, Gauthier BR, Mamin A, Edlund H, Philippe J (2002) The pancreatic beta-cell-specific transcription factor Pax-4 inhibits glucagon gene expression through Pax-6. *Diabetologia* 45: 97-107.
111. Sander M, Neubuser A, Kalamaras J, Ee HC, Martin GR, German MS (1997) Genetic analysis reveals that PAX6 is required for normal transcription of pancreatic hormone genes and islet development. *Genes Dev* 11: 1662-1673.
112. Emelyanov AV, Kovac CR, Sepulveda MA, Birshstein BK (2002) The interaction of Pax5 (BSAP) with Daxx can result in transcriptional activation in B cells. *J Biol Chem* 277: 11156-11164.

1.4 Les facteurs pionniers : des propriétés uniques

L'organogenèse est soumise aux processus de spécification du destin cellulaire par les gènes sélecteurs menant à la différenciation cellulaire. Le modèle d'une combinaison distincte de FTs, comprenant des facteurs généraux, histo- ou cyto- spécifiques, qui conduit à une spécification cellulaire et à un programme génétique unique n'explique pourtant pas comment les régions inaccessibles du génome peuvent être activées. Les années 1990 ont vu l'émergence du concept de FT interagissant avec la chromatine pour en altérer sa structure et permettre l'activation de la transcription génique (Steger & Workman, 1996; Zaret, 1999). L'idée avancée était que l'affinité d'un FT pour l'ADN nucléosomal est déterminée par la localisation de la séquence de reconnaissance sur le nucléosome, et par l'interaction coopérative de plusieurs protéines ciblant leur site de liaison sur ce nucléosome. Plus récemment, il y a eu la découverte de FT qui ont la propriété unique de lier l'ADN nucléosomal et remodeler la chromatine pour modifier l'épigénome; ce sont les facteurs pionniers (Drouin, 2014; Zaret & Carroll, 2011; Zaret & Mango, 2016).

1.4.1 Les modalités de liaison à l'ADN

Pour la régulation de l'expression génique, les interactions protéine-ADN jouent un rôle clé dans la fonction des protéines régulatrices. La spécificité des protéines de liaison à l'ADN repose sur la reconnaissance d'une séquence de 8 à 20 pb. Les interactions protéine-ADN dans le contexte de sites accessibles ont été bien documentés. Les protéines qui reconnaissent des séquences d'ADN spécifiques présentent différents motifs structuraux, constitués typiquement d'hélices alpha ou de brins beta qui se replient pour former le domaine de liaison à l'ADN dont il existe plusieurs types ; motif hélice-tour-hélice (HTH), hélice-boucle-hélice (HLH), doigt de zinc et domaine bZIP (région basique *leucine-zipper*) (Angrand, 1993).

Une séquence d'ADN est privilégiée par une protéine lorsqu'elle peut former un nombre optimal de contacts avec les paires de bases et le squelette d'ADN. Tandis que la conformation du squelette phosphate dépend peu de la séquence d'ADN, chaque paire de base offre une signature chimique caractéristique dans les sillons de la double hélice d'ADN (Figure 1.10). C'est majoritairement dans le sillon majeur que se retrouve la distinction unique pour chaque groupe fonctionnel de l'ADN, par rapport au sillon mineur où les profils des

donneurs/accepteurs de liaison hydrogène sont moins variés (Garvie & Wolberger, 2001). D'ailleurs, ce sont surtout des réseaux complexes de liaisons hydrogène qui sont impliqués dans la spécificité de liaison, mais il y a aussi des interactions de type van der Waals qui stabilisent les complexes protéine-ADN (Luscombe, Laskowski, & Thornton, 2001). Des structures en hélice alpha ou brins beta adjacentes au motif de reconnaissance principal de la protéine peuvent aider à orienter celui-ci par des contacts avec les chaînes latérales et le squelette phosphate (Pabo & Sauer, 1984).

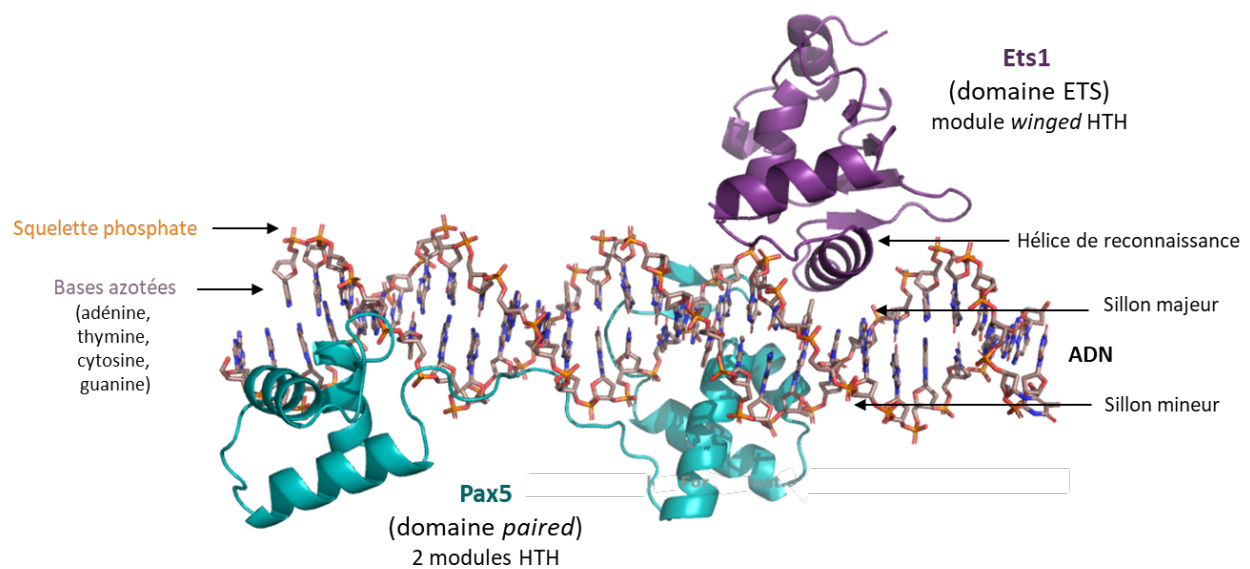


Figure 1.10 Exemple d'interactions entre les facteurs de transcription (Pax5 et Ets1) et la double hélice d'ADN

Structure tirée de (Garvie, Hagman, & Wolberger, 2001), modifiée à l'aide de PyMOL

L'interaction entre les protéines de liaison à l'ADN et leur ADN cible entraîne souvent des changements conformationnels dans l'un et l'autre (Andrabi, Mizuguchi, & Ahmad, 2014; Masse et al., 2002; Paull, Haykinson, & Johnson, 1993; Thompson & Landy, 1988). Ces changements conformationnels ont des implications directes sur la stabilité et la spécificité du complexe, permettant de former des complexes protéine-ADN fonctionnels et également de faciliter les interactions avec d'autres molécules. Par exemple, plusieurs FTs ont la capacité de plier l'ADN, ce qui aurait pour résultat de faciliter les interactions protéines-protéines (en

rapprochant physiquement les sites de liaison qui sont éloignés sur l'ADN) et pourrait inciter la dissociation des complexes nucléoprotéiques (Schultz, Shields, & Steitz, 1991; van der Vliet & Verrijzer, 1993). Ce phénomène de flexion de la double hélice survient notamment durant l'assemblage de très grands complexes protéines-ADN au niveau des régions *enhancers* et promoteurs qui sous-tendent le contrôle de l'expression génique (Tjian & Maniatis, 1994).

Les FTs pionniers à l'origine d'un changement de destin cellulaire ont la capacité de reconnaître leur cible dans des régions de chromatine fermée afin de mettre en place un nouveau programme génétique, notamment durant la différenciation cellulaire et la reprogrammation. Ceux-ci possèdent diverses structures de domaine de liaison à l'ADN et semblent cibler leur site d'interaction sur le nucléosome selon divers modes.

1.4.2 Les facteurs de pluripotence

Découverts par le laboratoire du Dr Shinya Yamanaka, les facteurs de pluripotence ont la capacité de reprogrammer les cellules somatiques en cellules souches pluripotentes (*induced pluripotent stem* - iPS) qui ont un potentiel d'auto-renouvellement et une capacité de différenciation (Takahashi & Yamanaka, 2006). Trois des quatre facteurs initialement identifiés, soit Oct4, Sox2 et Klf4 (OSK), sont suffisants pour la dé-différenciation, tandis que Myc augmente l'efficacité et la vitesse de reprogrammation en stabilisant le recrutement des trois autres facteurs aux loci critiques (Takahashi & Yamanaka, 2016). La génération de cellules iPS nécessite des changements moléculaires globaux sur des mégabases d'ADN menés par OSKM qui effacent l'identité somatique et établissent un programme pluripotent stable. La première étape de liaison de OSKM se produit principalement sur des éléments régulateurs nucléosomiques et dont l'ADN est méthylé, dans des régions de chromatine fermée, puis induit l'accessibilité pour d'autres FTs et co-facteurs (J. Chen et al., 2016; Knaupp et al., 2017).

Cette transition de l'identité cellulaire somatique à pluripotente est relativement lente et semble s'effectuer par vagues de liaison des facteurs OSKM et de changements épigénétiques, avec une phase initiale de loci se refermant, puis une étape finale de loci devenant accessibles (D. Li et al., 2017). Les premiers sont en grande partie composés de gènes associés à un destin somatique, tandis que les seconds sont associés à la pluripotence. Alors que les loci devenant accessibles sont enrichis pour les motifs OSK, les loci somatiques ne le sont pas, suggérant des

mécanismes alternatifs pour la fermeture de la chromatine (Chronis et al., 2017). En outre, Sap30, un composant du complexe corépresseur HDAC, est nécessaire au niveau des loci se refermant et limite le dépôt de la marque H3K27ac (D. Li et al., 2017). La liaison de OSK dépend de plusieurs facteurs dont la capacité de reconnaître des motifs partiels sur l'ADN nucléosomal (Soufi et al., 2015), une forte affinité pour les cibles d'ADN méthylés (Yin et al., 2017), et une forte synergie entre eux (Chronis et al., 2017; Soufi, Donahue, & Zaret, 2012).

Il existe toutefois quelques barrières qui ralentissent le travail des facteurs de pluripotence. Parmi celles-ci, les grands domaines d'hétérochromatine répressive enrichis en H3K9me3 se sont révélés globalement réfractaires à la liaison de OSKM aux premiers stades de la reprogrammation (Soufi et al., 2012). Les régulateurs critiques des cellules souches résident dans ces régions réfractaires, soulignant la nécessité d'un remodelage et d'une relaxation de la chromatine. En outre, la liaison précoce des facteurs pionniers OSKM ne garantit pas l'activation immédiate des gènes cibles qui dépend également du contexte (Apostolou & Stadtfeld, 2018). Aucun profil de marques d'histone particulier n'a toutefois pu être associé à cette activité (Soufi et al., 2012). Une autre étape limitante de la reprogrammation est la synergie entre OSKM et d'autres facteurs de transcription liés à la pluripotence, tel que ESRRB et NANOG, dont l'induction est plus tardive dans le processus de transition (Chronis et al., 2017).

L'activité pionnière des facteurs Oct4 (aussi connu comme POU5F1) et Sox2 semble bien documentée tandis que la liaison de Klf4 à ses cibles dans l'hétérochromatine nécessite l'action d'Oct4 pour relâcher l'état de condensation de l'ADN (K. Chen et al., 2020). Ceci est aussi supporté par les études de liaison aux nucléosomes *in vitro*, où Klf4 montre une faible affinité aux nucléosomes comparativement à Oct4 et Sox2 (Soufi et al., 2015). Structurellement, Oct4 possède deux domaines de liaison à l'ADN de type HTH soit le POU (nom dérivé des FTs Pit1, Oct1/Oct2 et Unc-86) spécifique, POU_S et l'homéodomaine POU, POU_{HD}. Selon la structure cristallographique, ces deux domaines de liaison à l'ADN reconnaissent ensemble l'ADN libre comprenant l'élément de réponse de Oct4 s'étendant sur 15 pb (Figure 1.11) (Esch et al., 2013; Reményi et al., 2003). Toutefois, cette configuration semble incompatible avec la liaison à l'ADN nucléosomal puisqu'elle implique la liaison du POU_S et du POU_{HD} sur les côtés opposés du double brin. Le modèle de reconnaissance partielle proposé est une liaison par laquelle un seul domaine de liaison à l'ADN interagit d'abord avec la cible sur le nucléosome.

La possibilité qu'il existe des configurations alternatives du complexe Oct4-nucléosome a aussi été soulevée (par rapport à la structure d'Oct4 sur l'ADN libre) (Huertas et al., 2020).

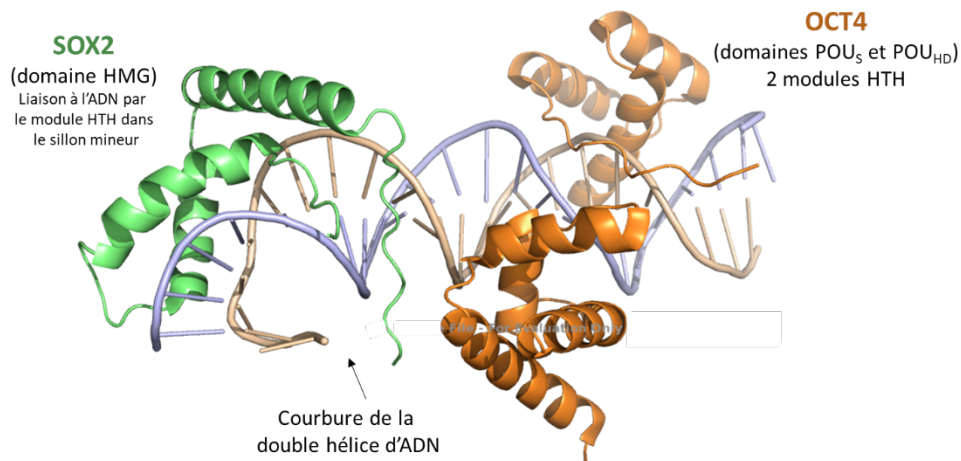


Figure 1.11 Structure des domaines de liaison à l'ADN de OCT4 et SOX2 lié à l'ADN.

Structure tirée de (Reményi et al., 2003), modifiée à l'aide de PyMOL

Des expériences de ChIP-seq dans les cellules ES ont permis de montrer que Oct4 et Sox2 coopèrent pour reconnaître un motif composite, provenant de la juxtaposition de leurs sites de liaison canoniques individuels (X. Chen et al., 2008). La protéine Sox2 interagit avec l'ADN via son domaine HMG (*high-mobility group box*), composé de trois hélices alpha reconnaissant spécifiquement un motif de 6 à 7 pb. Les domaines HMG sont uniques car ils interagissent avec le sillon mineur de l'hélice d'ADN et induisent une courbure drastique de la molécule d'ADN (Figure 1.11) (Reményi et al., 2003). La structure par cryo-microscopie électronique des facteurs SOX (SOX2 et son homologue SOX11) a montré la liaison aux nucléosomes au niveau de la SHL2 de l'ADN (région superhélicale), c'est-à-dire 2 tours d'hélice suivant l'axe dyade du nucléosome. Les facteurs Sox peuvent se lier à l'ADN nucléosomal et déformer localement la double hélice (Dodonova, Zhu, Dienemann, Taipale, & Cramer, 2020). Notons, par ailleurs, que la région superhélicale 2 est la plus assujettie à des déformations et au détachement des histones menant au réarrangement et à la translocation de l'ADN lorsque l'octamère subit des changements conformationnels (Bilokapic, Strauss, & Halic, 2018).

En revanche, l'analyse des sites de liaison de OCT4 et SOX2 *in vivo* a révélé que ces facteurs lient préférentiellement leurs cibles près sites d'entrée et de sortie du nucléosome

(SHL+/-6) (Michael et al., 2020). Les structures générées par cryo-microscopie électronique du nucléosome et des protéines OCT4 et SOX2 liés à ces sites ont mis en lumière l'aspect coopératif de la liaison à l'ADN nucléosomal (Michael et al., 2020). SOX2 crée une distorsion au site cible qui permet un détachement local des histones (Figure 1.12 A) ou a pour effet d'éloigner l'ADN des histones à l'entrée/sortie du nucléosome (Figure 1.12 B). L'interaction allostérique entre les protéines OCT4 et SOX2 est médiée par l'ADN puisque la courbure de la double hélice induite par SOX2 augmenterait l'affinité de liaison à l'ADN de OCT4 (Merino, Bouvier, & Cojocaru, 2015). Les facteurs SOX augmentent ainsi l'accessibilité de l'ADN en facilitant le détachement de la région terminale de l'ADN nucléosomal de l'octamère d'histone (Dodonova et al., 2020). Il a aussi été proposé que la liaison d'un facteur SOX au nucléosome conduisant à un repositionnement de la queue N-terminale de l'histone H4 serait incompatible avec l'empilement de nucléosomes lors de la condensation de l'ADN d'ordre supérieur et faciliterait ainsi l'ouverture de la chromatine et la transcription ultérieure.

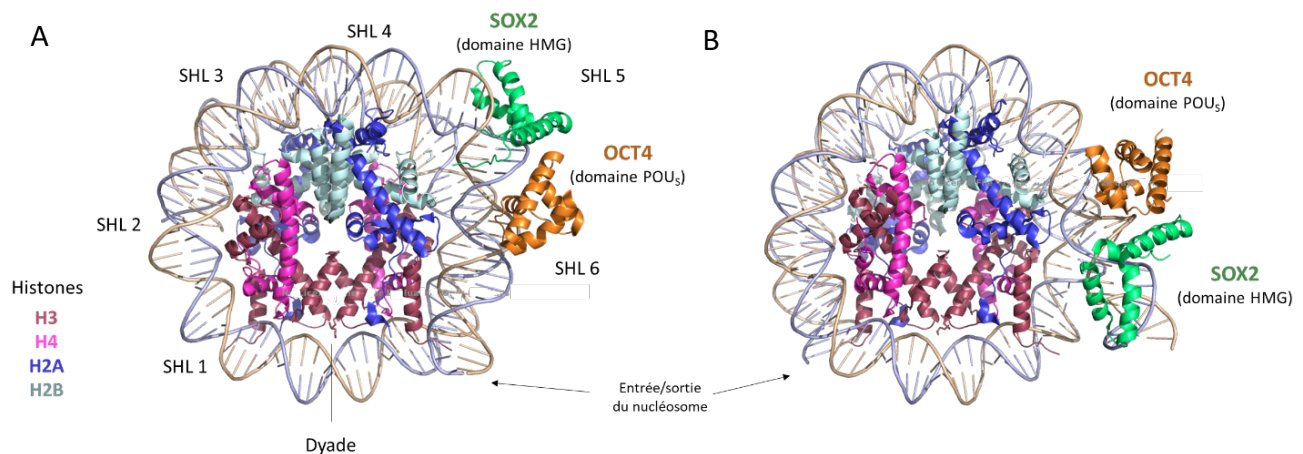


Figure 1. 12 Structure des domaines de liaison à l'ADN de OCT4 et SOX2 lié à l'ADN nucléosomal

OCT4-SOX2 liés à la position A) SHL+6 B) SHL-6. Structures tirées de (Michael et al., 2020), modifiées à l'aide de PyMOL

Des expériences de CHIP-seq effectuées dans les 48h suivant la reprogrammation de fibroblaste en cellule pluripotente, ont montré des interactions marquées par un chevauchement étendu dans la liaison initial des facteurs OSKM, contrairement aux études de liaison dans les

cellules ES où seulement Oct4 et Sox2 corrélaient fortement (Soufi et al., 2012). Les sites reconnus par les facteurs OSK sur les nucléosomes correspondent à des motifs partiels ou dégénérés (Soufi et al., 2015). En somme, la reprogrammation d'une cellule différenciée en cellule souche par les facteurs de pluripotence permet de réactiver le programme génétique associé au stade embryonnaire grâce à l'action pionnière des facteurs Oct4 et Sox2 et la coopération d'autres FT tel que Klf4 et c-Myc.

1.4.3 Les facteurs pionniers spécifiques

Jusqu'à maintenant, peu de FTs spécifiques ont démontré une activité pionnière. Parmi les premiers facteurs pionniers découverts, il y a certains facteurs de la famille *Forkhead* (Fox). Par exemple, FoxA1 et FoxA2 sont des gènes développementaux clés pour l'établissement de la compétence dans l'endoderme de l'intestin antérieur et sont nécessaires au développement du foie, du pancréas, des poumons, de la prostate et neurones dopaminergiques du mésencéphale (Ferri et al., 2007; N. Gao et al., 2008; C. S. Lee, Friedman, Fulmer, & Kaestner, 2005; Wan et al., 2005).

Le rôle de FoxA1 et FoxA2 dans l'initiation de la spécification du foie a d'abord été démontré par leur occupation au niveau de l'enhancer spécifique de l'albumine (Alb1), où Gata4 semble agir comme facteur potentiateur de cette liaison. FoxA1 occupe sa séquence cible activatrice de Alb1 sur un nucléosome *in vitro* et *in vivo* dans les cellules précurseurs hépatiques, puis crée un site hypersensible à la DNaseI, montrant que la protéine ouvre la chromatine localement (Chaya, Hayamizu, Bustin, & Zaret, 2001; Cirillo et al., 2002; McPherson, Shim, Friedman, & Zaret, 1993). En effet, la liaison de FoxA1 à la chromatine *in vivo* crée une région d'ADN accessible et permet la liaison d'autres facteurs tel que le récepteur des œstrogènes (oestrogen receptor - ER) (Carroll et al., 2005; Laganière et al., 2005). D'autres études ont aussi montré que la protéine FoxA1 purifiée se lie à ses sites cibles sur des particules nucléosomiques reconstituées *in vitro* et permet la liaison par Gata4 et NF1 aux sites adjacents (Cirillo et al., 2002; Cirillo et al., 1998; Cirillo & Zaret, 1999).

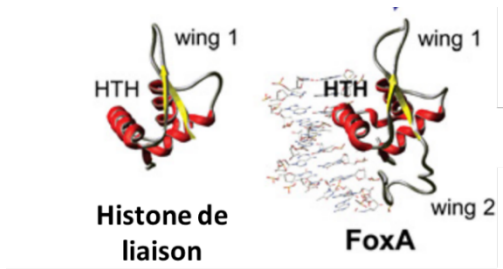


Figure 1.13 Ressemblance du domaine de liaison à l'ADN de FoxA à l'histone de liaison

Adapté de (Zaret & Carroll, 2011)

L'activité pionnière des facteurs FoxA est médiée par le domaine de liaison à l'ADN, qui s'associe spécifiquement à ses cibles, et par le domaine C-terminal de la protéine, qui interagit les histones H3 et H4; cette double interaction est critique pour l'ouverture de la chromatine *in vitro* (Cirillo et al., 2002). L'analyse de la structure cristalline a montré que le domaine de liaison à l'ADN de FoxA se replie en un motif hélice-tour-hélice (HTH) flanqué de chaque côté par des *wings* de chaîne polypeptidique moins ordonné (Clark, Halay, Lai, & Burley, 1993; Ramakrishnan, Finch, Graziano, Lee, & Sweet, 1993). Notamment, ce domaine de liaison à l'ADN en hélice *winged* des protéines FoxA ressemble à celle de l'histone de liaison H1, ces derniers se lient près de l'axe dyade du nucléosome (Figure 1.13) (Goytisolo, Packman, & Thomas, 1996). Par conséquent, il a été suggéré que la liaison de FoxA sur ses cibles dans la chromatine compactée déplace les histones de liaison de la chromatine, expliquant partiellement l'augmentation de l'accessibilité à l'ADN sous-jacent (Iwafuchi-Doi et al., 2016).

Les régions *wings* contribuent à la stabilisation de la liaison du site cible d'ADN puisqu'elles ont le potentiel d'interagir avec le sillon mineur de la double hélice d'ADN, flanquant les interactions du motif HTH avec le sillon majeur (Cirillo & Zaret, 2007). Pour les protéines FoxA, les domaines *wings* interagissent de manière stable avec les sites d'ADN de haute affinité, mais moins au niveau de la cible d'affinité inférieure, où le motif HTH seul semble impliqué (Cirillo & Zaret, 2007). Le facteur FoxA1 reconnaît des motifs dégénérés dans le contexte du nucléosome, suggérant que la liaison au nucléosome peut stabiliser l'interaction à des motifs de faible affinité, tandis que la liaison à des motifs optimaux et de haute affinité serait plutôt opportuniste (Meers, Janssens, & Henikoff, 2019). Ceci est en accord avec le mécanisme de *scanning*, ou balayage de la chromatine, proposé dû à la mobilité faible de FoxA1 conférée

par des interactions spécifiques et non spécifiques avec la chromatine (T. Sekiya, Muthurajan, Luger, Tulin, & Zaret, 2009; Zaret, Lerner, & Iwafuchi-Doi, 2016).

Lorsqu'une cellule se divise, en mitose ou en méiose, l'état extrême de compactage de la chromatine exclut la plupart des protéines régulatrices et le génome devient silencieux (Gottesfeld & Forbes, 1997). Toutefois, un sous-ensemble de facteurs demeure et sert de marque de type *bookmarks* aux régions régulatrices de certains gènes pour une réactivation rapide à mesure que les cellules progressent dans le cycle cellulaire. Il a été démontré que FoxA1 peut se lier à la chromatine mitotique puisqu'environ 15% des sites cibles de FoxA1 dans l'interphase sont liés dans la mitose, notamment au niveau des gènes qui sont importants pour la différenciation hépatique (Caravaca et al., 2013; Zaret et al., 2008). En servant de marque épigénétique, cela constitue un mécanisme par lequel certains FTs aident à conserver l'identité cellulaire pendant la mitose.

Chez le ver *Caenorhabditis elegans*, il a aussi été démontré que le FT de type FoxA, PHA-4, peut induire la décompaction de grands domaines de chromatine et le recrutement du variant d'histone de H2A.Z aux promoteurs (Fakhouri et al., 2010; Updike & Mango, 2006). De plus, l'affinité du PHA-4 pour son site de liaison à l'ADN est déterminant pour la régulation temporelle des gènes: les promoteurs avec des sites de haute affinité sont activés plus tôt dans le développement que ceux avec des sites de moindre affinité (Updike & Mango, 2006). L'affinité de liaison à l'ADN peut donc être un aspect déterminant de la compétence des FTs pionniers. Cependant, cette idée n'est pas ressortie des analyses de l'ensemble des sites cibles de FoxA1 identifiés dans les cellules hépatiques de souris. Cela peut refléter le fait que les FTs pionniers ont des sites de liaison qui présentent des propriétés différentes selon la fonction d'activateur de la transcription classique ou pionnière (Drouin, 2014).

En plus de FoxA, FoxE1 et FoxO1 se sont également avérés ouvrir la chromatine compactée (Cuesta, Zaret, & Santisteban, 2007; Hatta & Cirillo, 2007). L'activité pionnière du facteur spécifique de la thyroïde FoxE1 est d'ailleurs un exemple de différenciation induite par les hormones. En réponse à l'induction hormonale, FoxE1 se lie à la chromatine compactée du promoteur de la thyroperoxydase (TPO) inactif, ce qui coïncide avec l'apparition d'une région d'hypersensibilité à la DNase I. Le rôle de FoxE1 dans la médiation de la régulation hormonale

du gène TPO est de permettre à d'autres facteurs de régulation d'accéder à la chromatine et d'entraîner la différenciation thyroïdienne (Cuesta et al., 2007).

Quelques exemples supplémentaires de FTs ciblant des régions de chromatine fermée ont été répertoriés. Entre autres, la liaison de GR induite par traitement hormonal (dexaméthasone) entraîne un recrutement du remodeleur de la chromatine Brg1 et une augmentation de l'accessibilité localisée au niveau des *enhancers* (Johnson et al., 2018; Voss et al., 2011). La structure du domaine de liaison à l'ADN de GR contient deux modules de type doigt de zinc dont l'un est associé aux interactions spécifiques avec l'ADN et l'autre fournit la surface de dimérisation (Baumann et al., 1993; Härd et al., 1990). Également, il a été proposé que le FT à homéodomaine Isl1 joue un rôle de pionnier dans la différenciation cellulaire des cardiomyocytes (R. Gao et al., 2019). Isl1 se lie à la chromatine fermée et, grâce au recrutement de Brg1-Baf60c du complexe SWI/SNF, entraîne des altérations épigénétiques spécifiques à la lignée cardiaque. Enfin, le FT PU.1, connu pour son rôle crucial dans le développement des lignées myéloïde et lymphoïde, induit le remodelage local du nucléosome suivi du dépôt de marque d'histone active H3K4me1 (Carotta, Wu, & Nutt, 2010; Heinz et al., 2010). PU.1 fait partie de la famille de FTs ETS (*Erythroblast Transformation Specific*) dont le domaine de liaison à l'ADN est formé d'un module *winged* HTH similaire à celui de FoxA (Figure 1.11) (Pio et al., 1995). Dans l'ensemble, ces études montrent que la liaison des facteurs pionniers se produit avant l'engagement de la lignée et déclenche les étapes d'ouverture de la chromatine pour établir la compétence pour l'activation du gène.

1.4.4 Caractéristiques communes des facteurs pionniers

En somme, il existe divers modes de liaison à l'ADN nucléosomal par les facteurs pionniers (Figure 1.14). Ce qui est commun et compatible avec la liaison nucléosomique est la reconnaissance de l'ADN via une hélice alpha courte d'un module HTH (Oct4, Sox2, FoxA, Pax7), à doigt de zinc (Klf4) ou de type « *scissor-like* » (Ascl1) (Fernandez Garcia et al., 2019). Ces propriétés de liaison à l'ADN peuvent être importantes pour la capacité des facteurs à reconnaître des séquences d'ADN spécifiques dans la chromatine nucléosomique condensée. En revanche, des FTs de type immunoglobuline dépourvus d'hélice de reconnaissance (hélice alpha) nécessitent une liaison coopérative pour cibler des sites moins accessibles. À titre

d'exemple, le FT à boîte T, Tpit, se lie à des sites cibles situés dans une chromatine fermée grâce au facteur pionnier Pax7 dont il sera discuté plus en détail dans le prochain sous-chapitre. De plus, il a été démontré que TP53, un FT de type immunoglobuline, se lie aux régions flanquantes du nucléosome, c'est-à-dire à l'ADN sans histone (Yu & Buck, 2019). De même, la protéine virale BZLF1, un FT dont le domaine de liaison à l'ADN de type bZIP est formé d'un module « *scissor-like* » prolongé, a montré une liaison au nucléosome minimale à des concentrations élevées de BZLF1 (Schaeffner et al., 2019).

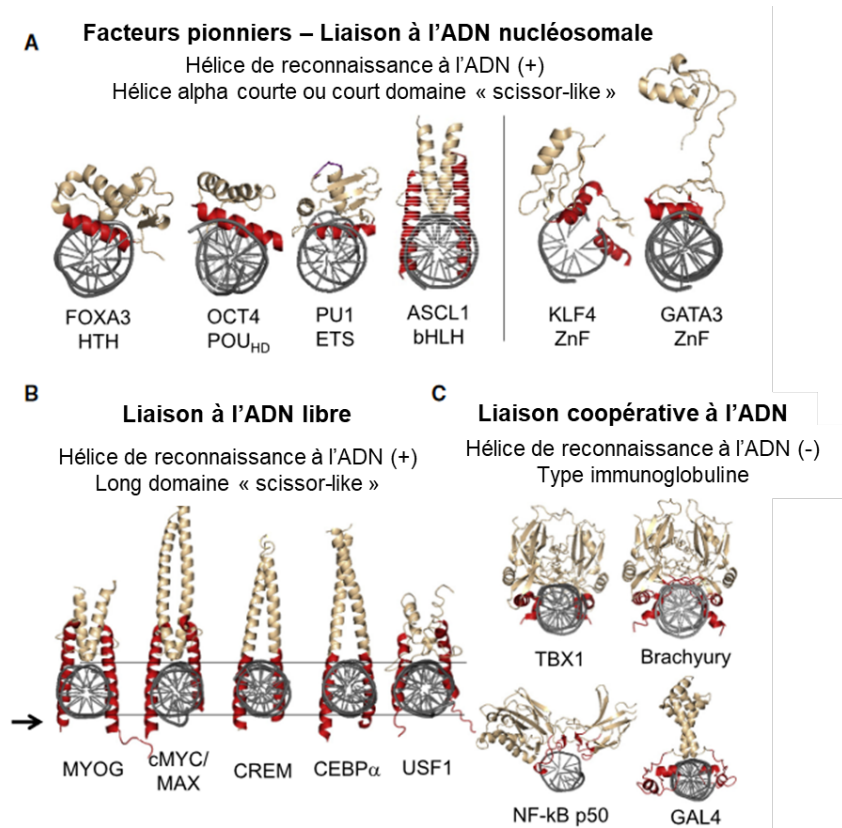


Figure 1. 14 Structure du domaine de liaison à l'ADN et capacité d'un facteur de transcription à reconnaître l'ADN nucléosomal

Les FT ayant une liaison nucléosomale forte possèdent des motifs de reconnaissance en hélices α courtes, figure adaptée de (Fernandez Garcia et al., 2019)

(A) Structures cristallines des domaines de liaison à l'ADN des FTs pionniers.

(B) FTs se liant à l'ADN libre. Structures cristallines des domaines de liaison à l'ADN en forme de ciseaux et des motifs de reconnaissance en hélices α prolongées.

(C) FTs se liant de manière coopérative. Structures cristallines des domaines de liaison à l'ADN à repliement de type immunoglobuline.

Tous les FTs, y compris les facteurs pionniers, se lient uniquement à une partie de tous leurs motifs de liaison putatifs dans le génome, et ce sous-ensemble diffère d'un type cellulaire

à l'autre. Des caractéristiques de la chromatine peuvent permettre ou, au contraire, empêcher la liaison du facteur pionnier. Les facteurs pionniers peuvent initier le ciblage d'une région, or le recrutement d'autres facteurs, stabilisateurs ou potentiateurs, est nécessaire. L'aspect combinatoire de la liaison de multiples FTs étant inhérent à la régulation génique, en particulier en ce qui a trait au contrôle cellule-spécifique, les événements de reprogrammation ou de différenciation cellulaire n'y font pas exception. Les exemples d'action pionnière dans les différents systèmes suggèrent que les facteurs pionniers peuvent se lier de manière coordonnée avec d'autres facteurs. Les facteurs pionniers ont des propriétés intrinsèques de reconnaissance nucléosomique leur permettant de sonder la chromatine fermée à la recherche de sites cibles potentiels, puis de recruter d'autres facteurs, qui peuvent à leur tour stabiliser la liaison du facteur pionnier à la chromatine. Comme mentionné précédemment, pour les protéines Oct4, Sox2, Klf4 et FoxA, l'activité pionnière dépend de la capacité à cibler des motifs partiels exposés sur la surface du nucléosome. Une telle reconnaissance de motif partiel peut se produire par liaison coordonnée entre les facteurs. Bien que la liaison coopérative la plus connue est basée sur l'interaction protéine-protéine entre les FTs, la coopération médiée par le nucléosome entre les FTs a été observée depuis quelques décennies (Adams & Workman, 1995; Chávez & Beato, 1997; Vashee, Melcher, Ding, Johnston, & Kodadek, 1998). La synergie de la liaison entre facteurs peut être obtenue par leur compétition simultanée avec les histones pour l'ADN sous-jacent et a été notée principalement pour des sites situés sur la même moitié du nucléosome (Mirny, 2010; Moyle-Heyrman, Tims, & Widom, 2011).

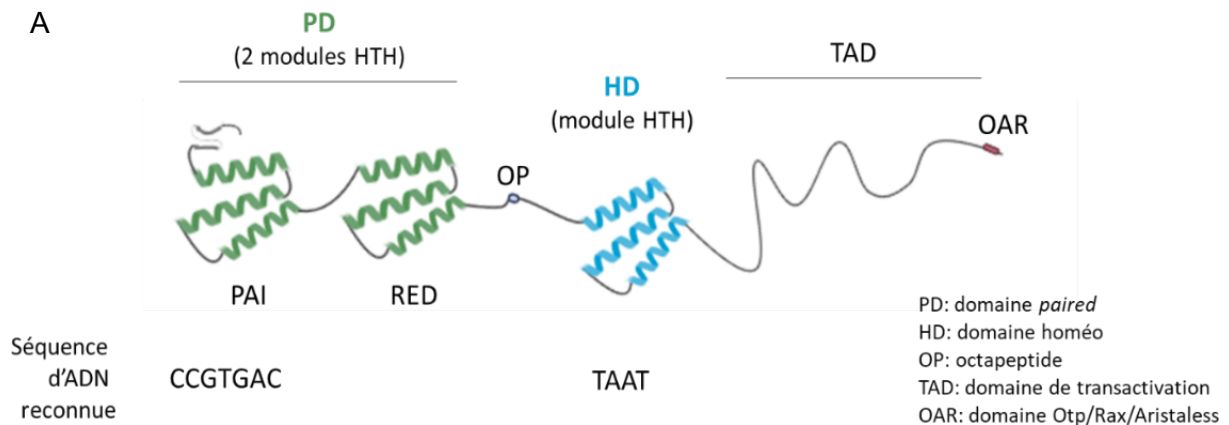
Ainsi, les facteurs pionniers ont la propriété unique de reconnaître à eux seuls les sites cibles nucléosomiques puis recruter d'autres FTs non-pionniers de sorte que ce modèle est compatible avec un mécanisme de liaison spécifique au type de cellule. De plus, la capacité accrue des facteurs pionniers à interagir avec l'ADN nucléosomique fournit un avantage pour les mécanismes de coopération facilitant la liaison de multiples FTs à la chromatine fermée. Il est important de noter que les études sur l'accessibilité de l'ADN indiquent que les promoteurs sont généralement ouverts de manière ubiquitaire tandis que les régions *enhancers* s'ouvrent de manière tissu-spécifique (Schones et al., 2008; Thurman et al., 2012). Un facteur pionnier représente ainsi une clé pour accéder à un répertoire d'*enhancers* associé à un programme génétique.

1.5 Le cas du facteur pionnier Pax7

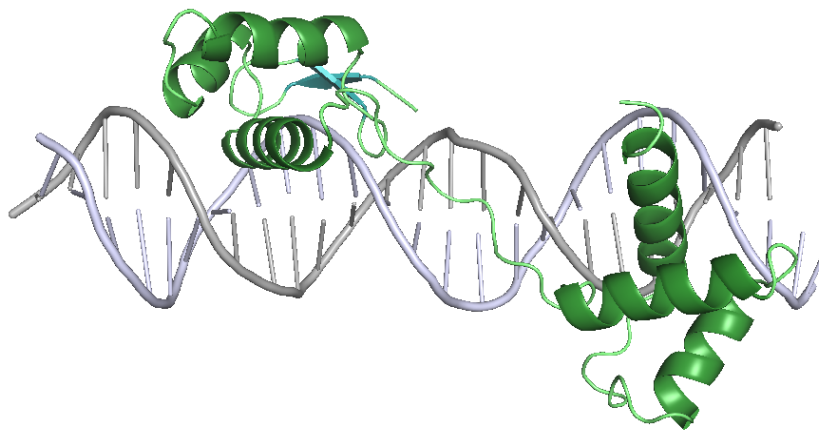
En tant que FT pionnier, Pax7 contrôle la spécification du destin cellulaire mélanotrope (Budry et al., 2012; Mayran & Drouin, 2018) et son expression est nécessaire au développement de la crête neurale, du cerveau et du muscle squelettique (Relaix, Rocancourt, Mansouri, & Buckingham, 2005; Stoykova & Gruss, 1994). Pax7 est exprimé dans les cellules myogéniques en prolifération et les cellules satellites musculaires où il constitue un régulateur clé en amont de la myogenèse chez l'adulte (von Maltzahn, Jones, Parks, & Rudnicki, 2013).

1.5.1 L'interaction avec l'ADN par les domaines PD et HD

Rappelons que Pax7 possède deux domaines de liaison à l'ADN, le *paired* (PD) et l'homéodomaine (HD), chacun reconnaissant un motif canonique (Figure 1.15). Le domaine PD est une structure bipartite comprenant deux modules HTH nommément les sous-domaines PAI (N-terminal) et RED (C-terminal) (W. Xu, Rould, Jun, Desplan, & Pabo, 1995). Le domaine HD est un troisième module HTH, celui-ci ayant la capacité de former des homo- et hétéro-dimères sur des séquences palindromique HD avec son homologue Pax3 (Birrane, Soni, & Ladias, 2009; Schäfer, Czerny, Bernasconi, Genini, & Busslinger, 1994; D. Wilson, Sheng, Lecuit, Dostatni, & Desplan, 1993; D. S. Wilson, Guenther, Desplan, & Kuriyan, 1995). Les protéines Pax peuvent d'ailleurs utiliser plusieurs combinaisons de leurs HTH pour reconnaître une variété de types de sites cibles (Jun & Desplan, 1996). Les travaux sur Pax7 et Pax3, deux membres de la famille Pax partageant un niveau élevé d'identité de séquence, ont montré l'interaction coopérative entre les domaines PD et HD, dont la liaison à l'ADN de l'un ou l'autre des domaines de liaison à l'ADN entraîne des changements conformationnels au niveau de l'autre domaine (Apuzzo, Abdelhakim, Fortin, & Gros, 2004; Jun & Desplan, 1996).



B Complexe du domaine PD de PAX6 humain avec l'ADN



C Complexe du domaine HD de PAX3 humain avec l'ADN

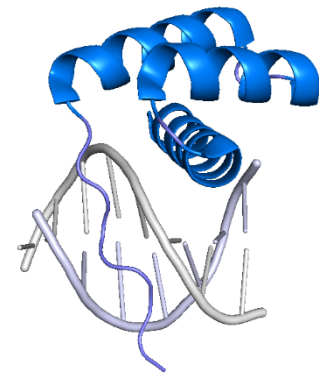


Figure 1.15 Structure de la protéine Pax7 et des domaines de liaison à l'ADN PD et HD

(A) Représentation schématique de la structure de Pax7 et les séquences d'ADN reconnues par ses domaines de liaison à l'ADN, créée à l'aide de BioRender

(B) Structure du domaine PD de PAX6 associé à l'ADN, tirée de (H. E. Xu et al., 1999) et modifiée à l'aide de PyMOL

(C) Structure du domaine HD de PAX3 associé à l'ADN, tirée de (Birrane et al., 2009) et modifiée à l'aide de PyMOL

La découverte de l'action pionnière de Pax7 dans l'hypophyse a permis d'identifier le motif composite, qui comprend les deux sites de liaison de Pax7 reconnus par son domaine PD et HD, enrichi aux cibles pionnières (Budry et al., 2012). La reconnaissance initiale de motif

dans la chromatine fermée par les facteurs pionniers est une propriété unique encore insaisissable. La diversité de mode de liaison à l'ADN possible et la variété de motifs de séquence reconnus par Pax7, dont le motif composite, supportent l'idée que l'action pionnière de Pax7 implique une interaction particulière avec l'ADN. Cette perspective est au cœur des hypothèses pour expliquer les propriétés pionnières uniques de Pax7 qui seront étudiées dans cet ouvrage.

1.5.2 L'action pionnière de Pax7

En permettant l'accessibilité à un répertoire unique d'enhancers, Pax7 est nécessaire et suffisant pour la spécification de la lignée mélanotrope. Les expériences avec une chimère ER-Pax7 inductible par le Tamoxifène a permis de démontrer que Pax7 se lie rapidement à ses cibles aux *enhancers*, mais le remodelage de la chromatine et l'activation des gènes sont plus lents (Mayran et al., 2018). Les *enhancers* ouverts par Pax7 montrent une perte de méthylation de l'ADN et acquièrent une mémoire épigénétique stable, comme en témoigne la liaison de facteurs non-pionniers après le retrait de Pax7. L'état permissif de la chromatine pour l'action pionnière semble être l'hétérochromatine facultative, marquée par H3K9me2 (Figure 1.16). La liaison initiale de Pax7 à l'ADN nucléosomal est d'abord faible, puis il y a stabilisation de la liaison de Pax7 dans les 24 heures. La chromatine du site cible subit une première transition où et un nucléosome central devient plus accessible ou dans un état de *primed enhancer* caractérisé par un signal H3K4me1 faible au centre des enhancers cibles. L'activation complète des *enhancers* implique l'action d'autres FT non-pionniers et est caractérisée par une déplétion du nucléosome, une distribution bimodale de marques d'*enhancers* actif H3K4me1 et H3K27ac, ainsi que par le recrutement du coactivateur général p300. Bien que la liaison de Pax7 ne semble pas affectée par la méthylation des cytosines à ses sites d'ADN cibles, il existe une corrélation entre le remodelage de la chromatine dépendante de l'action pionnière et la perte de méthylation au niveau des cytosines aux *enhancers* nouvellement accessibles.

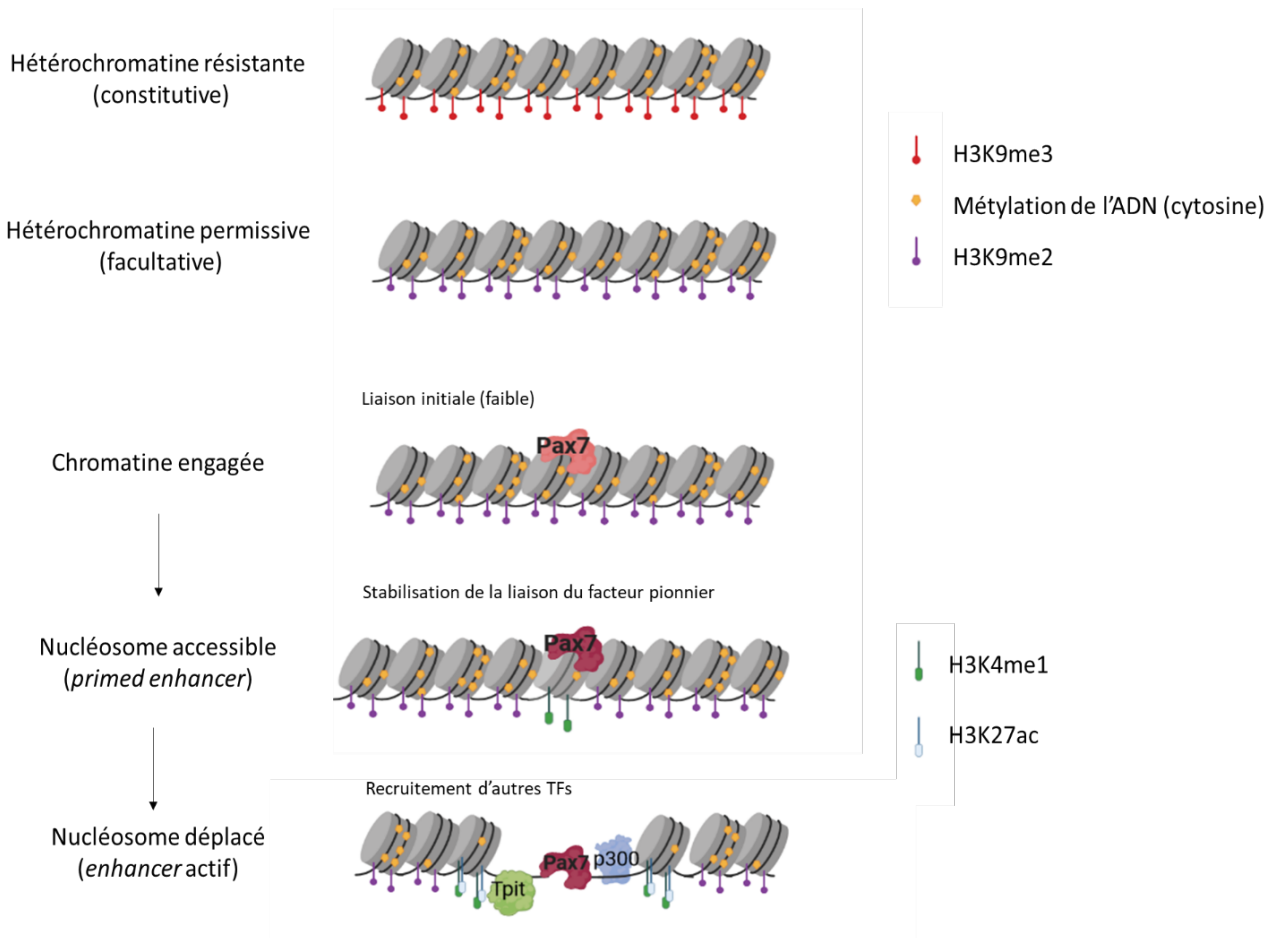


Figure 1.16 Les étapes de l'action pionnière

Inspiré de (Mayran & Drouin, 2018), créé à l'aide de BioRender

Ainsi, l'action pionnière de Pax7 permet la mise en place du programme génétique mélanotrope dans l'hypophyse, mais il n'agit pas seul. Pax7 reconnaît et engage des sites pionniers dans l'hétérochromatine, toutefois l'ouverture de la chromatine nécessite l'action du facteur non-pionnier Tpit (Mayran et al., 2019). Une combinaison de modèles d'inactivation génique de Tpit ou de Pax7 a permis de mettre en lumière l'importance de la coopération entre les facteurs pionniers et non-pionniers pour le déploiement du répertoire d'*enhancers* mélanotropes et l'ouverture de la chromatine (Mayran et al., 2019). De même, l'expression ectopique de Pax7 dans un modèle de cellule gonadotrope en culture (α T3) n'exprimant pas Tpit, montre que la liaison des cibles dans l'hétérochromatine par Pax7 est indépendante de Tpit, mais l'ouverture de la chromatine subséquente nécessite l'action de Tpit (Mayran et al., 2019). Cela démontre que les propriétés intrinsèques des facteurs pionniers sont limitées à la capacité

de reconnaître les cibles dans l'ADN nucléosomal et de faciliter la liaison non-pionnière. L'ouverture de la chromatine en soi peut impliquer la coopération avec des facteurs non-pionniers.

1.5.3 Les isoformes Pax7

Différents isoformes de Pax7 ont été identifiées dans divers tissus murins et humains. L'épissage de transcrits alternatifs se produit soit à la jonction de l'intron 2 et de l'exon 3 (trinuécléotide) ou à la jonction de l'intron 3 et de l'exon 4 (hexanucléotide) (Figure 1.17). Les quatre isoformes en résultant diffèrent par l'inclusion ou l'exclusion de résidus glutamine (Q) et/ou glycine et leucine (GL) dans le domaine PD. Plus précisément, le résidu glutamine est situé dans la région linker des sous-domaines du PD, à proximité du sous-domaine RED, et le dipeptide GL est situé en N-terminal de la troisième hélice alpha du RED. Ces isoformes appelées Pax7 Q-GL+, Pax7 Q+GL+, Pax7 Q-GL- et Pax7 Q+GL- ont des caractéristiques structurelles différentes qui peuvent altérer la liaison à l'ADN.

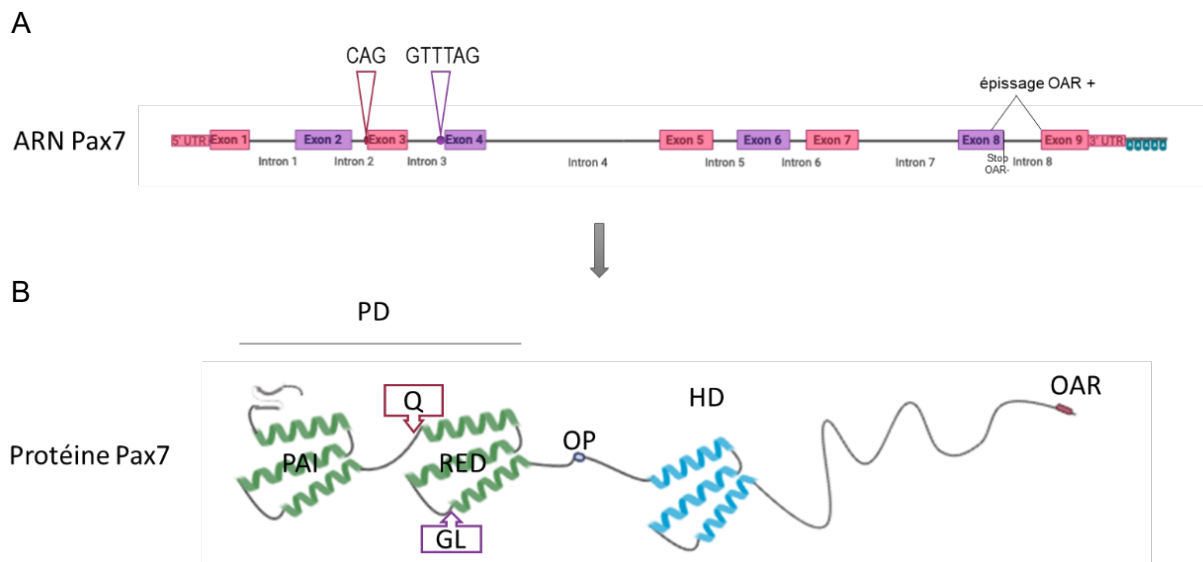


Figure 1.17 Épissage alternatif du transcrit de Pax7 et acides aminés variants des isoformes

Représentation schématique des nucléotides subissant un épissage alternatif du transcrit (A) et des résidus différentiels de la protéine Pax7 (B), créée à l'aide de BioRender.

De fait, un variant d'épissage de Pax3 qui présente la forme Q- a une affinité de liaison plus forte aux sites PD et une activité transcriptionnelle plus élevée que la forme Q+ (Vogan &

Gros, 1997; Vogan, Underhill, & Gros, 1996). Il a été proposé que l'absence de Q qui se trouve dans la région liant les sous-domaines PAI et RED crée une proximité plus étroite des sous-domaines qui augmente l'affinité de liaison. Jusqu'à présent, il n'est pas clair si les variations structurelles dans le domaine PD des isoformes de Pax7 influencent la spécificité ou l'affinité de liaison des domaines PD et HD. Pour ce qui est de la présence du dipeptide GL, des prédictions de structure secondaire de protéine indique qu'il exercerait une influence importante sur la conformation de la protéine puisque la troisième hélice alpha du sous-domaine RED serait allongée (Ziman & Kay, 1998). Étant donné l'interdépendance des domaines de liaison à l'ADN des facteurs Pax et que la variation structurelle dans le domaine PD s'est avéré être responsable des différentes spécificités de liaison à l'ADN, il est probable que les différents transcrits confèrent une diversité fonctionnelle supplémentaire à Pax7 (Du et al., 2005; Vogan & Gros, 1997; Vogan et al., 1996). Cette diversité aurait le potentiel de médier des fonctions distinctes dans l'embryon en développement.

Les quatre transcrits de Pax7 possèdent des profils d'expression différentiels selon les lignées cellulaires et même selon les souches de souris (Ziman, Fletcher, & Kay, 1997). Ziman et al. ont comparé l'abondance de chacun des transcrits dans les tissus musculaires et cérébraux (Ziman et al., 2001). Dans les cellules musculaires, quoique les quatre transcrits soient détectables, Pax7 Q+GL- est enrichi et Pax7 Q+GL+ semble majoritaire. Pour ce qui est du cerveau, les transcrits Pax7 Q+GL+ et Pax7 Q+GL- sont tous deux enrichis et les transcrits Pax7 Q-GL+ et Pax7 Q-GL- sont peu ou pas représentés. Les isoformes présentes dans le lobe intermédiaire de l'hypophyse sont Pax7 Q+GL-, et dans une moindre mesure Pax7 Q+GL+.

Des expériences de transfection dans des cellules pluripotentes P19 ont montré le rôle de Pax7 dans la détermination de l'identité cellulaire puisque l'isoforme Pax7 Q+GL+ a entraîné la différenciation des cellules vers une identité neuronale et active plus efficacement l'expression du récepteur du facteur neurotrophique ciliaire (Cntfr) que les autres isoformes (White & Ziman, 2008; Ziman et al., 2001). Dans les cellules souches isolées du muscle, l'isoforme Pax7 Q+GL- active efficacement le programme myogénique et augmente aussi l'expression de Pax7 endogène (et donc possiblement d'autres isoformes) (Seale, Ishibashi, Scimè, & Rudnicki, 2004).

Chez l'humain, une forme supplémentaire de Pax7 a été trouvée dans différents tissus. Il s'agit du résultat d'une terminaison transcriptionnelle différentielle soit dans l'exon 8 au lieu de l'exon 9 (Proskorovski-Ohayon et al., 2017; Vorobyov & Horst, 2004). Cette troncation de l'extrémité C-terminale mène à la perte du domaine OAR (Pax7 OAR-) et, par conséquent, pourrait présenter des propriétés de transactivation différentes. Cette forme semble être exprimée dans plusieurs tissus (excepté les tissus du cerveau fœtal et adulte), tandis que la forme Pax7 pleine longueur (OAR+) est spécifique aux tissus du cerveau, muscle et testicule (Proskorovski-Ohayon et al., 2017). Ainsi, l'épissage différentiel de l'ARNm Pax7 peut entraîner la production de deux formes protéiques alternatives qui incluent ou excluent le domaine C-terminal conservé de l'évolution et peut potentiellement mener à différentes fonctions de la protéine (Vorobyov & Horst, 2004).

1.5.3 Mutations du gène *PAX7* et maladies

Chez la souris, les mutants Pax7^{-/-} sont viables jusqu'à deux à trois semaines après la naissance avec des retards de développement et une faiblesse musculaire (Buckingham & Relaix, 2015; Mansouri, Stoykova, Torres, & Gruss, 1996; Seale et al., 2000). L'inactivation du gène *Pax7* cause une absence complète de cellules satellites dans le muscle squelettique (Seale et al., 2000) et des défauts de développement de structures cranio-faciales dérivées de la crête neurale (Mansouri et al., 1996). Chez l'humain, des études d'association pangénomique (*genome-wide association studies* - GWAS) ont révélé que des mutations du gène Pax7 constituent un facteur de risque pour des malformations congénitales. Ainsi, plusieurs variants rares et potentiellement pathogènes situés dans ou à proximité du gène *PAX7* ont été détectés chez des patients présentant des fentes orofaciales (soit des fentes labiales ou palatines) (Butali et al., 2013; Gaczkowska et al., 2019; Gowans et al., 2016; Leslie et al., 2015). Les maladies associées à *PAX7* incluent la myopathie congénitale progressive avec scoliose (MYOSCO) où les variants *PAX7* homozygotes associés à ces maladies sont concentrés dans le domaine PD (Feichtinger et al., 2019). Une mutation de *PAX7* a aussi été associée à un syndrome de retard de croissance, d'hypotonie et de retard neurodéveloppemental global (Proskorovski-Ohayon et al., 2017). Cette mutation cible un site d'épissage dans *PAX7* qui empêche l'expression de l'isoforme *PAX7* OAR+. De plus, les FTs *PAX3* et *PAX7* sont impliqués dans la pathogenèse

du rhabdomyosarcome alvéolaire (ARMS), un cancer pédiatrique des tissus mous lié à la lignée musculaire striée (Barr, 2001; Du et al., 2005). Les translocations chromosomiques caractéristiques du ARMS génèrent des fusions PAX3-FKHR et PAX7-FKHR qui contiennent les domaines de liaison à l'ADN de PAX3/PAX7 et le domaine d'activation transcriptionnelle FKHR, ayant pour conséquence la surexpression des cibles de PAX (Bennicelli, Advani, Schäfer, & Barr, 1999).

En définitive, les mécanismes de l'action pionnière montrent leur implication majeure dans la spécification du destin cellulaire, la reprogrammation et les pathologies. Les facteurs de transcription pionniers constituent toutefois un groupe hétérogène, qui, contrairement à d'autres FTs, sont capables de reconnaître et de se lier aux séquences d'ADN cibles dans la chromatine fermée. En conséquence de la liaison du facteur pionnier, la structure locale de la chromatine peut s'ouvrir et faciliter la liaison d'autres protéines établissant la compétence pour l'expression génique. Les mécanismes initiaux de liaison de la chromatine par les facteurs pionniers sont encore peu compris. Les travaux de recherches présentés dans cet ouvrage visent à mieux comprendre les propriétés pionnières de Pax7, notamment par l'étude fonctionnelle de ses deux domaines de liaison à l'ADN, *paired* et *homeo*, ainsi que des isoformes de Pax7.

Chapitre 2 : L'action pionnière de Pax7 nécessite ses deux domaines de liaison à l'ADN : paired et homeo

Ce chapitre est constitué d'un article publié (Pelletier et al., 2021). Il se trouve dans sa langue de publication originale, soit l'anglais.

Contributions

Audrey Pelletier: Apport original, expériences d'EMSA, lignées AtT-20 stables exprimant les mutants Pax7, RT-qPCR transfections, analyses des données, validation et visualisation.

Alexandre Mayran: Expériences de BS-seq, ATAC-seq,

Arthur Gouhier: Analyses bio-informatiques

James G Omichinski: Production des peptides Pax7

Aurelio Balsalobre: Analyses bio-informatiques et supervision

Jacques Drouin: Supervision

Pax7 pioneer factor action requires both paired and homeo DNA binding domains

Audrey Pelletier^{a c}, Alexandre Mayran^{a b}, Arthur Gouhier^a, James G Omichinski^c, Aurelio Balsalobre^a and Jacques Drouin^{a c 1}

^aLaboratory of Molecular Genetics, Institut de recherches cliniques de Montréal (IRCM), Montréal QC H2W 1R7, Canada;

^bEPFL SV ISREC UPDUB, CH-1015 Lausanne, Switzerland;

^cDepartment of Biochemistry, Faculté of Médecine, Université de Montréal, Montréal H3C 3J7, Canada.

Keywords: transcription, chromatin, epigenetics, ChIPseq, ATACseq, pituitary, POMC

¹To whom correspondence should be addressed at: Jacques Drouin, Laboratory of Molecular Genetics, Institut de recherches cliniques de Montréal (IRCM), 110, avenue des Pins Ouest, Montréal QC Canada H2W 1R7; Tel: 514-987-5680, E-mail: jacques.drouin@ircm.qc.ca

Abstract

The pioneer transcription factor Pax7 contains two DNA binding domains (DBD), a paired and a homeo domain. Previous work on Pax7 and the related Pax3 showed that each DBD binds a cognate DNA sequence, thus defining two targets of binding and possibly modalities of action. Genomic targets of Pax7 pioneer action leading to chromatin opening are enriched for composite DNA target sites containing juxtaposed sites for both paired and homeo domains. The present work investigated the implication of the DBDs in pioneer action. We show that the composite sequence is a higher affinity binding site and that efficient binding to this site involves both DBDs of the same Pax7 molecule. This binding is not sensitive to cytosine methylation of the DNA sites consistent with pioneer action within nucleosomal heterochromatin. Introduction of single amino acid mutations in either paired or homeo domain that impair binding to cognate DNA sequences showed that both DBDs must be intact for pioneer action. In contrast, only the paired domain is required for low affinity binding of heterochromatin sites. Thus, Pax7 pioneer action on heterochromatin requires unique protein:DNA interactions that are more complex compared to its simpler DNA binding modalities at accessible enhancer target sites.

Introduction

Pioneer factors are a subset of transcription factors (TFs) that have been dubbed with the unique property of being able to “open chromatin structure” (1-3). While pioneers have the typical TF domain organisation found in most TFs including DNA binding and transactivation domains, the molecular basis of their ability to “open chromatin” remains largely elusive. In fact, recent work clearly separated the chromatin opening function from the pioneer ability and showed that the former can be conferred through recruitment of a nonpioneer factor: this was shown for the pioneer Pax7 that is required to implement the melanocyte enhancer repertoire (4,5) but that is dependent on the cell determination TF Tpit for chromatin opening (6). These findings thus limit the unique pioneer property to recognition of DNA binding sites within heterochromatin. A large body of work also showed that pioneer TFs can bind their target DNA sequence within nucleosomes that would presumably be tightly packed within heterochromatin (7,8). Whereas the ability to recognize target DNA within nucleosomal structures appears to be a prerequisite for pioneer action, this in itself is not sufficient to explain the specificity of pioneer action nor to explain how initial sequence recognition is linked to chromatin remodelling.

Various pioneers activate novel enhancer repertoires that control cell specific gene expression programs. This was shown for the pioneers FoxA (9,10), Pax7 (4,5), C/EBP α (11), EBF1 (12,13), Ascl1 (14), NeuroD1 (15), and GATA3 (16). For all these factors, with the exception of Pax7, the DNA binding sites for the TF at pioneered enhancers (hereafter, we refer to ‘pioneered enhancers/sites’ as those that require chromatin opening for activity) appeared similar to their previously defined transcriptional target sites. In contrast, the initial work on Pax7 pioneer target sites revealed enrichment for a unique composite target sequence that includes motifs recognized by each DNA binding domain (DBD) present in Pax7 (5). Indeed, Pax7 contains two DBDs, a paired domain (PD) and a homeo (HD) domain, and each recognizes different prototypical DNA target sites. Prior work on the DNA binding properties of Pax7 and the closely related Pax3 used cognate binding sites for either the PD or the HD (17), and these appear to show independent binding by the PD or HD domains. However, close interaction of the two DBDs appears important for optimal DNA binding (18,19). Indeed, DNA binding by either the PD or HD causes a structural change in the other DBD (20). In this context, recognition of a unique composite DNA sequence containing binding motifs for both PD and

HD enriched at pioneered targets suggest that Pax7 pioneer action may involve a unique conformation or binding requirement.

It is presently unclear what DNA recognition features may be critical for pioneer target recognition and action. Although higher affinity for DNA sites present at pioneer targets may be a mechanism to trigger structural changes at nucleosomal sites, various studies have shown the ability of pioneer TFs to recognize subsets of their target sequences by recognition of partial and presumably lower affinity DNA motifs. This was proposed to represent a “scanning” ability of pioneers to probe putative DNA targets within nucleosomal heterochromatin (21). This feature results in large numbers of pioneer-bound sites that are unproductive for chromatin remodelling and were thus labelled as “resistant” (4) or as unchanged (21,22). The contribution of DNA binding affinity for pioneer action thus remains unclear. It was recently shown that the FoxA pioneer factor requires a core histone interaction domain for chromatin opening (23). However, the scanning ability and core histone interaction do not explain the selection of site subsets that are productive for chromatin opening.

The present work investigated the DNA binding properties of Pax7 in relation to its various target DNA sequences both *in vitro* using gel mobility shift and *in vivo* using ChIPseq. The highest *in vitro* binding affinity was observed with a composite Pax7 target that includes DNA motifs for recognition by both PD and HD and this composite sequence is enriched at pioneered enhancers. The composite DNA site is bound *in vitro* by a Pax7 monomer and the binding affinity is unchanged when the composite sequence contains a methylated CpG, a feature that would be expected of pioneers targeting regulatory sequences within nucleosomal heterochromatin. Finally using single amino acid substitutions of key residues required for DNA binding of either PD or HD, we show that the DNA binding activity of both domains are required for pioneer action whereas the PD alone is sufficient for binding to low affinity resistant sites. In summary and in contrast with the ability of each DNA binding domain to bind its target sequence in naked DNA, the dependence on both domains in order to trigger chromatin opening at pioneered sites suggests that a unique conformation of the Pax7 DBDs is needed for the initial events leading to chromatin opening.

Materials and methods

Cell culture. AtT-20/D16v-F2 cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and antibiotics (penicillin/streptomycin). To generate stable transgenic AtT-20 Pax7 G43A, R56L and S264E cell populations, retroviruses were packaged using the Platinum-E Retroviral Packaging Cell Line (Cell Biolabs, catalog #RV-101) and infections performed as described (24). Selection of retrovirus-infected cell populations was achieved with 400 $\mu\text{g}/\text{mL}$ Geneticin (Gibco, #11811-031). Resistant colonies were pooled to generate populations of hundreds of independent colonies.

Plasmids and Luciferase assays. The PC2 reporter plasmid (5) and construction of oligonucleotide-driven reporters were described previously (25). Mutagenesis of mouse Pax7 coding sequences was carried out using the Stratagene QuickChange Site-Directed Mutagenesis kit as reported previously (26). AtT-20 cells were plated in 12-well dishes (4 x 10⁵ cells per well) and transfections performed the next day using PEI 25K (Polysciences #23966-1) in triplicates. Transfected DNA (total 2 μg in 150 μL DMEM) was added to the PEI 25K solution (4 μL per 150 μL DMEM) and incubated at room temperature for 30 minutes before adding 100 μL of the mixture to each well. The media was changed 24 hours after transfection and luciferase activity assessed 48 hours after transfection. Growth medium was removed from transfected cells and 200 μL of lysis buffer (Tris 100mM, NP40 0.5% and DTT 0,001M) added to each well. After 10 minutes on a vigorous shaker, 100 μL of cell lysate supernatant was used for analysis using the GloMax Navigator luminometer (Promega) using the kinetics protocol (100 μL injection, speed 100 $\mu\text{L}/\text{s}$, integration time 10 s). Statistical significance was assessed by 2-way ANOVA and Tukey's multiple combinations test.

ChIPseq, ATACseq and MNaseSeq. ChIP experiments were performed in AtT-20 cells as previously described (4,27) using 5 μg of antibody (anti-FLAG-M2, mouse monoclonal, Sigma F3165). ATACseq was performed as previously described (4). The libraries and flow cells were prepared by the IRCM Molecular Biology Core Facility following Illumina recommendations (Illumina, San Diego, CA). The ChIP libraries were sequenced on the Illumina Hi-Seq 2000 sequencer. Supplementary Table 2 provides details of antibodies used for ChIP and sequencing depths. Relative Pax7 ChIP and ATACseq read density profiles were generated by Easeq average signal function (28) (<http://easeq.net>) and data were exported to

Microsoft Excel to create graphs. ChIPseq and ATACseq heatmaps were generated by Easeq. To visualize ChIPseq and ATACseq profiles, we used the Integrative Genome Viewer tool (29).

MNase-seq was performed as described (30) with the following modifications. AtT-20 Neo and Pax7 cells (60 million) were lysed in 1 mL lysis buffer (10mM Tris, pH7.5, 10mM NaCl, 3mM MgCl₂, 0.5% NP-40, 0.15mM spermine, 0.5mM spermidine, 1 mM PMSF, 0.6 μM aprotinin, 0.6 μM leupeptin and 2 μM pepstatin A). Nuclei were resuspended in 1 mL MNase digestion buffer (10 mM Tris-HCl pH 7.4, 15 mM NaCl, 60 mM KCl, 0.15 mM spermine, 0.5 mM spermidine, 1.5 mM CaCl₂, 1 mM PMSF, 0.6 μM aprotinin, 0.6 μM leupeptin and 2 μM pepstatin A). 50 μL of nuclei suspension (□50 μg) were digested in 16U MNase (Micrococcal Nuclease NEB M0247S) at 37 °C for 60 min. The reaction was stopped with 5 μL EDTA 250 mM and to break the nuclei, 5 μL SDS 10 % was added. The volume was adjusted to a total of 200 μL with 140 μL TE and 20 μL NaCl 5 M was added to dissociate histones from DNA. After phenol/chloroform extraction and ethanol precipitation, 2 μL of RNase (10 mg/mL) was added to the mixture and incubated at 37 °C for 30 min. DNA was purified using Qiagen PCR purification kit and eluted with 30 μL EB. Sequencing libraries were performed as previously described (4).

Expression and purification of Pax7 polypeptides from bacteria. For the expression of polypeptide fragments of Pax7 in bacteria, the sequences encoding the PD alone (PD; residues 35-158 of mouse Pax7 Uniprot P47239-1), the HD alone (HD; residues 215-273 of mouse Pax7), the PD plus the linker region (PD-L; residues 35-215 of mouse Pax7) or the PD plus linker plus HD (PD-HD; residues 35-273 of mouse Pax7) were incorporated into the pET15b (Novagen) vector. The vectors were then transformed in *E. coli* host strain BL21DE3 (Novagen) and the cells were grown at 37°C in Luria Broth (LB) media overnight, diluted 1:4 with fresh LB and expression of the His-tagged Pax-7 polypeptides was induced for 4h at 30°C with 0.7 mM isopropyl-β-D-thiogalactopyranoside (IPTG; Inalco). The cells were harvested by centrifugation, suspended in lysis buffer (either 20 mM Tris-HCl pH 7.4, 1M NaCl and 50 mM imidazole for HD, PD-L and PD-HD or 20 mM NaPO₄ pH 7.4, 1M NaCl and 50 mM imidazole for PD), lysed with a French press and centrifuged at 105,000g for 1h at 4°C. The supernatant was collected and incubated with a nickel charged Chelating-Sepharose resin (GE healthcare) for 1h at 4°C. Following incubation, the resin was thoroughly washed with 20 mM Tris-HCl pH

7.4, 1M NaCl and 50 mM imidazole, and the polypeptides were eluted from the resin with elution buffer (20 mM Tris-HCl pH 7.4, 1M NaCl and 500 mM imidazole). The eluted proteins were dialyzed against 20 mM Tris-HCl pH 7.4, 1mM EDTA buffer overnight and further purified on a 50 mL High Performance SP-Sepharose (GE Healthcare) column using a gradient with 20 mM Tris-HCl pH 7.4, 0.75 M NaCl, 1mM EDTA over 5 column volumes. The fractions containing the purified polypeptides were dialyzed against 20 mM Tris-HCl pH 7.4, 1mM EDTA buffer overnight and stored at -80 °C until usage.

Electrophoretic mobility shift assays (EMSA). *Probe labeling.* Double-stranded DNA probes were prepared by annealing oligonucleotides (1 μ M of each strand: see sequences in the Supplementary Figure 1A) in buffer TEN (10 mM Tris pH 8.0, 1 mM EDTA and 50 mM NaCl). Labeling was done at room temperature for 20 min in the following buffer: 50 mM Tris pH 7.5, 10 mM MgCl₂, 1 mM DTT and 0.5 mg/mL BSA supplemented with 0.25 μ M dGTP, dTTP, 0.2 μ M α -³²P-dCTP (3,000 Ci/mmol, 10 mCi/mL, Perkin Elmer) and 1 unit of DNA polymerase I Large (Klenow) fragment from *E. coli* (NEB, catalog # M0210M). Probes were then purified on Micro Bio-Spin P-30 Tris chromatography column (Bio-rad, catalog # 732-6223).

Nuclear extracts. To prepare nuclear extracts, AtT-20 cells from a 100mm petri dish were washed and collected in PBS. Pelleted cells were resuspended in 800 μ L of cold low-salt buffer (10 mM HEPES pH 7.9, 10 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 1 mM DTT, 0.5 mM PMSF, 0.6 μ M aprotinin, 0.6 μ M leupeptin and 2 μ M pepstatin A). After 15 min incubation on ice, 100 μ L NP40 10 % was added. Cytosolic fractions were discarded, and nuclear proteins extracted in 100 μ L of cold high-salt buffer containing 20 mM HEPES pH 7.9, 400 mM KCl, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 1 mM PMSF, 0.6 μ M aprotinin, 0.6 μ M leupeptin and 2 μ M pepstatin A. Protein concentration was quantitated by Bradford assay.

EMSA. Gel shift assays were performed using 1 μ g of nuclear protein extracts from different AtT-20 cell lines pre-incubated 10 min on ice with 1 μ g of nonspecific carrier DNA poly(dI-dC) and poly(dA-dT) in binding buffer (25 mM HEPES pH 7.9, 84 mM KCl, 10% glycerol, 5 mM DTT). For purified polypeptides, 300 fmoles were used per binding reactions. If applicable, cold competitor probe was added to the pre-incubation mix. Equal amounts of radiolabeled DNA probes (50,000 cpm) were then added for a total volume of 20 μ L and

incubated 60 min on ice. If applicable, 1 µg of Flag M2 antibody (F3165, SIGMA) was added to the mix during the last 10 min of incubation.

Electrophoresis. Nondenaturing polyacrylamide gels (40 mM Tris, 195 mM glycine, 0.08% APS, 0.5 µL/mL TEMED, 5% Acrylamide:Bisacrylamide in 19:1 ratio), were prerun in Tris-glycine buffer (40mM Tris and 195 mM glycine) for 60 min at 300 V at 4°C. The gel shift reaction mixture was then loaded, and the gel run for 3 h at 300 V at 4°C. Gels were dried on 3MM Whatman paper in a gel dryer at 80°C during 60 min under vacuum. Autoradiography films (HyBlot CL, catalog # E3018) were then exposed with the dried gels for 10 to 48 h using an intensifier screen.

RT-qPCR. RNA was extracted from AtT-20 cells using the RNeasy Mini kit (Qiagen #74104), and cDNA was synthesized using 5 µg RNA and SuperScript III reverse transcriptase (Invitrogen, #18080-044) following manufacturer's recommendations. The resulting cDNAs (5 µl of 1/50 dilution) were analyzed by qPCR using SYBR Green reagent (ThermoFisher Scientific #A25741) supplemented with 500 nM of each gene-specific primer pair (provided in Supplementary Table 1) in a total volume of 10 µl. At least 3 biological replicates were analyzed in duplicates for each condition using a ViiA™ 7 Real-Time PCR device (ThermoFisher scientific), and results were analyzed using the accompanying software. Gene expression values were normalized to those of GAPDH and statistical significance assessed by 2-way ANOVA and Tukey's multiple combinations test.

NGS data analysis. ChIP-seq and ATAC-seq paired-end sequenced reads were trimmed using Trimmomatic/0.36 (31) and aligned to the mouse mm10 reference genome using Bowtie/2.3.5 (--very-sensitive --no-mixed --no-unal) (32). Bam files were created using view from SAMtools/1.9 (33) and duplicated reads were removed using MarkDuplicates from Picard/2.17.3 (Broad Institute). Coverage tracks were created using bamCoverage (--normalizeUsing RPKM -bs 10 -e) from deepTools/3.3.0 (34). Peaks were called against sequenced input DNA or control Flag ChIP using callpeaks (-f BAMPE -p (1e-3) -g mm) from MACS/2.1.1 (35). Supplementary Table 2 lists the parameters of ChIPseq and ATACseq

experiments. Categories of Pax7 binding sites are based on already published lists (4) but refined as described in Supplementary Table 3 to increase their homogeneity. Pax7 ChIP-seq peaks were categorized using the absence or presence, before or after Pax7 expression, of overlapping peaks from: H3K4me1 ChIP-seq to assess enhancer marking; p300 ChIP-seq to assess enhancer activation; and ATAC-seq to assess DNA accessibility. SMC1 ChIP-seq was used to avoid insulator contamination in the resistant category. In order to minimize threshold effects, the absence of a peak requires no peak to be detected with a low detection threshold whereas its presence requires its detection with a higher threshold.

Motifs analysis. *De novo* motif analysis of Pax7 ChIP-seq data was performed on 60 bp surrounding peak summits. Each subset of Pax7 peak were processed using HOMER *de novo* motif search (36) and MEME (<http://meme-suite.org/tools/meme-chip>), in order to find the motifs with the most significant P-value. Graphical representations of the position weight matrices obtained from these analyses were generated using a custom Python script for motif logos. Motifs frequencies were analyzed using a custom Python based program. Using uncovered matrices, motif scores for each Pax7 binding site were computed with a custom Python script in a 100 bp window around Pax7 peak summits.

Whole-genome bisulfite sequencing (WGBS). The CpG methylation levels were represented using the published data from (4).

Results

Strongest *in vitro* Pax7-binding motif is enriched at pioneered target sites

Prior studies (reviewed in (17)) on the DNA binding properties of the related Pax TFs Pax3 and Pax7 (Fig.1A) revealed cognate binding sites for either of their two DBDs, the PD or the HD domain (Fig.1B). Genome-wide assessment of these TF binding sites obtained by ChIPseq showed enrichment of the PD target sequence core motif TGAC and of different associations of the TAAT motif recognized by the HD (5,37). Prior work identified the H.I2 palindrome as a binding site for Pax3 and Pax7 dimers (38,39). The H.I-2 overlapping TAATTA motif was identified in ChIPseq data for Pax7 (5) but its significance remains poorly explored. We also reported that target enhancers pioneered by Pax7 show enrichment of the composite motif that contains juxtaposed PD and HD motifs (5). Extensive characterization of the Pax7 ChIPseq peaks defined different genomic recruitment sites (4). These included binding at putative enhancer sites that are already accessible before Pax7 expression (Constitutive sites), at sites where Pax7 binding results in transcriptional activation revealed by p300 recruitment (Activated sites / before Pax7: low H3K4me1 and ATAC, no p300 / after Pax7: high H3K4me1 and ATAC, p300 present), at Resistant sites (before and after Pax7: no H3K4me1, ATAC or p300), and finally at Pioneered sites (before Pax7: no H3K4me1, ATAC or p300 / after Pax7: high H3K4me1 and ATAC, p300 present) or Primed sites (before Pax7: no H3K4me1, ATAC or p300 / after Pax7: single peak low H3K4me1 and weak ATAC, no p300). Pax7 motif frequencies for these different subcategories show that the fully activated Pioneer sites have the highest frequency for all motifs (Fig.1C, D) and this correlates with this category of peaks exhibiting the strongest Pax7 ChIPseq signals of all categories (4) as supported by the highest median number of reads or p values for this subset (Fig. 1C). The Pioneer subset also has the highest combined frequencies of all different Pax7 motifs (Fig.1E) consistent with the frequent occurrence of more than one motif under each peak (Fig.1F).

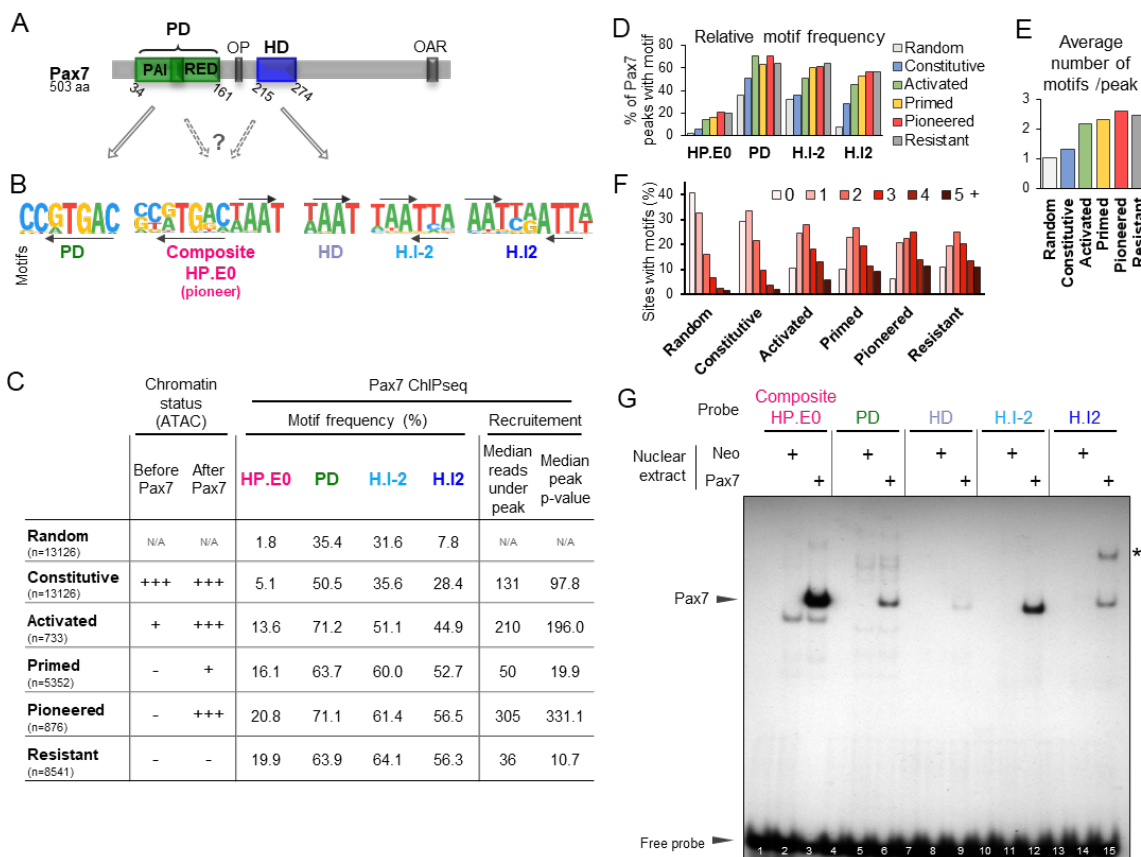


Figure 1. Contribution of paired (PD) and homeodomain (HD) DNA binding domains to Pax7 binding.

(A) Schematic representation of mouse Pax7 structure (Uniprot P47239-1) including the PD and HD DNA binding domains. The positions of two other conserved domains is also indicated: OP, a conserved octapeptide present in the linker region between the two DNA binding domains and OAR, a conserved tetradecapeptide initially identified in the transcription factors otp, aristaless and rax.

(B) Consensus recognition motifs identified in Pax7 ChIPseq (5) and related to recognition by either PD or HD as suggested by arrows linking panels A and B. The composite HP.E0 motif contains everted sequences of the PD and HD recognition motifs and this composite was identified as enriched under Pax7 peaks that trigger pioneer chromatin opening (5).

(C) Distribution of Pax7 interaction motifs in subsets of Pax7 ChIPseq peaks defined by their changes (or not) in pituitary AtT-20 cells before/after Pax7 expression. These categories were defined and discussed extensively in (4), and the subsets presented here (n indicated between parentheses) were trimmed to only include peaks with stringently defined properties compared to other subsets. For reference, the chromatin accessibility (ATAC) status of these subsets is indicated. The average Pax7 recruitment strength for each subset is indicated by the median number of reads and P values of peaks present in each subset.

(D) Relative frequencies (% of peaks containing indicated motif) for each subset of peaks.

(E) Average total number of all motifs per peak.

(F) Distribution of number of motifs per peak for each subset of Pax7 ChIPseq peaks.

(G) Binding of Pax7 to its different DNA targets shown in A above. Electrophoretic Mobility-Shift Assay (EMSA) was used to assess the *in vitro* binding of Pax7 stably expressed in AtT-20 cells (4). Similar DNA probes (probe sequences in Fig. S1A) were incubated with either control AtT20-neo or AtT20-Pax7 cell nuclear extracts for analyses by EMSA. A single specific band is observed (arrowhead) for each probe but with widely varying intensities. Only the H.I2 palindromic probe shows an additional band (*) that is likely constituted of Pax7 homodimers.

In order to compare Pax7 binding to each target type motifs, we did electrophoretic mobility shift assays (EMSA) using DNA probes corresponding to each motif (Fig. 1B and S1A) and nuclear extracts from AtT-20 cells expressing a Flag-tagged Pax7 or an empty vector (Neo). Complexes present in EMSA using Pax7 compared to Neo cells revealed a single complex present with all probes (arrowhead in Fig. 1G). The apparent strength/stability of this Pax7 complex is however very different for different probes: the highest apparent affinity site is the composite HP.E0 site followed, in decreasing order, by the inverted TAATTA palindrome H.I-2, the PD motif, the TAAT palindrome H.I2, and the single HD. The only other Pax7-specific complex is observed with the H.I2 probe (asterisk in Fig. 1G). This slower migrating band is a likely Pax7 dimer as this sequence was previously shown to form dimers with both Pax3 and Pax7 (38). We did competitive binding experiments in order to assess the relative affinities of complexes formed with the different probes and this supported the interpretation of the highest affinity for the composite followed by PD and HD sites (Fig.S1B and C). The presence of Pax7 in the purported Pax7-DNA complexes was ascertained by supershift EMSA experiments using the Flag antibody: both monomer and dimer Pax7:DNA complexes were indeed confirmed to contain the Flag-Pax7 chimeric protein (Fig. S1D).

Prevalent role of Pax7 paired (PD) domain for DNA binding

In order to assess the relative roles of PD and HD domains for DNA binding, we produced and purified polypeptides containing different Pax7 domains from over-expressing E. Coli. Whereas the PD domain (residues 35-158 of mouse Pax7) is sufficient for *in vitro* DNA binding to the composite HP.E0 and PD probes, the HD domain (aa 215-273) does not bind significantly at the concentrations shown but did at 1000 times greater concentration (Fig. 2A, lanes 1, 3, 5, 7 and not shown). Addition of the linker sequence present between PD and HD to the PD polypeptide appeared to stabilize PD probe binding and to result in a more compact faster migrating complex (lane 6 compared to 5) but the strongest binding is observed for a polypeptide containing both PD and HD (PD.HD, aa 35-273) together with intervening linker sequences (Fig. 2A, lanes 4 and 8). On its own, the HD polypeptide exhibited binding only to the palindromic H.I2 probe compared to the core HD or to the HD H.I-2 probes (Fig. 2A, lanes 19 compared to 11, 15). The PD:HD polypeptide showed the strongest binding to all probes

although this is marginal for the simple HD motif (lanes 4, 8, 12, 16, 20). Further, this PD:HD polypeptide presents an additional band with the H.I2 palindromic probe (lane 20) consistent with homodimer binding reported for both Pax3 and Pax7 with this sequence (38). Collectively, these data indicate that the PD domain is primarily responsible for high affinity DNA binding and that the HD domain may contribute stabilization and specificity of binding, in particular with the composite HP.E0 probe. On the other hand, the HD may increase the repertoire of Pax7 binding sites to include inverted palindromic target sequences of the H.I2 type.

Pax7 binds the composite site as monomer

Binding of Pax7 to the composite element may involve the two DBDs of the same Pax7 molecule or the PD of one associated as dimer with the HD of another Pax7 molecule. In order to address this question, we performed EMSA using AtT20-Pax7 nuclear extracts with and without addition of the bacterially-produced PD.HD polypeptide. These mixing experiments (Fig. 2B) revealed independent binding of the PD.HD polypeptide and AtT20-expressed Flag-Pax7 to the composite (lanes 2-4), PD (lanes 6-7) and HD (lanes 11-13) probes. In contrast, binding of the full-length Pax7 to the palindromic H.I2 probe revealed two bands deemed to represent monomers and dimers (Fig. 2B, lane 14). The PD.HD polypeptide also formed two complexes with the slower migrating band (labelled **) deemed to represent a dimeric complex that is not observed with the other probes (Fig. 2B, lane 18). When both PD.HD polypeptide and full-length Pax7 are present together, a new band (labelled ***) was observed and interpreted as consisting of one molecule of full-length Pax7 together with one molecule of the PD.HD (Fig. 2B, lane 16). This new complex together with the Pax7-containing complexes were all supershifted by addition of the anti-Flag antibody confirming presence of the AtT20-expressed Pax7 in these complexes (Fig. 2B, lane 17). Taken together, these experiments are consistent with the interpretation that a monomer of Pax7 is sufficient to bind the composite target sequence with the PD and HD of the same Pax7 molecule likely to participate in close interactions in order to bind their juxtaposed target DNA sequences.

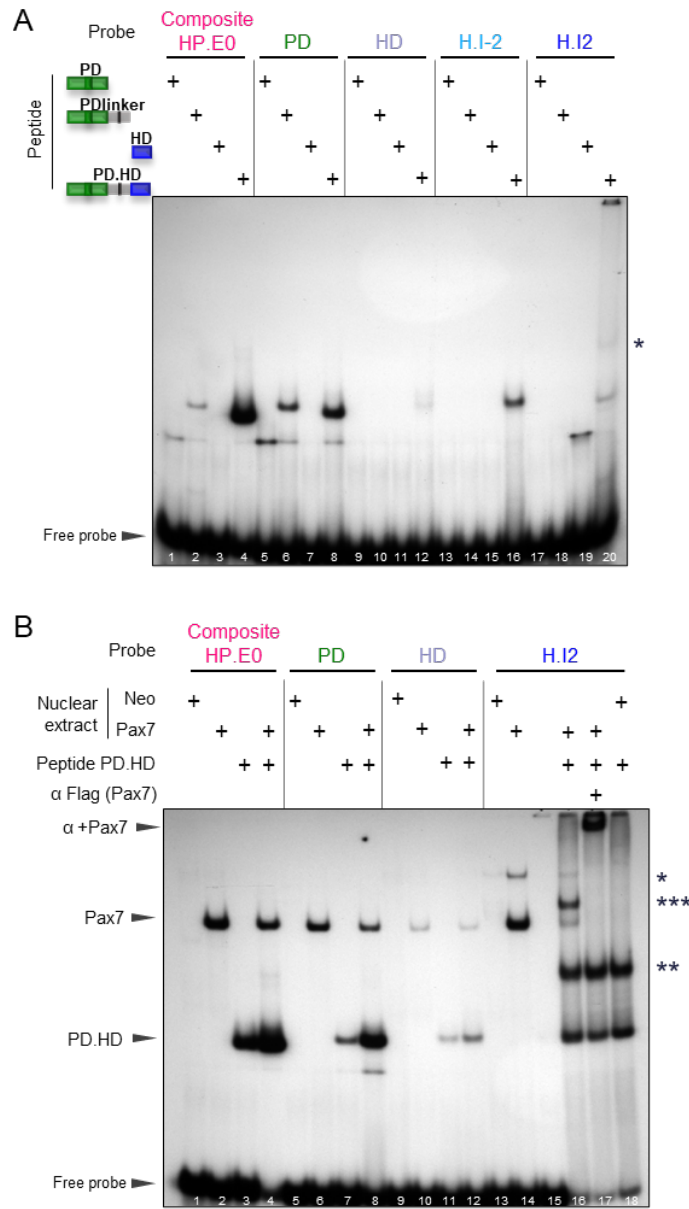


Figure 2. Relative roles of Pax7 PD and HD domains in DNA binding.

(A) *In vitro* binding (EMSA) of various Pax7 polypeptides produced in *E. coli* to contain the represented portions of the Pax7 DNA binding domains. The purified polypeptides were incubated with ³²P-labelled probes before electrophoresis. For each probe, the strongest band likely represents monomer Pax7 polypeptide binding except for the H.I2 probe where a likely dimer band is observed (*).

(B) EMSA mixing experiment showing that Pax7 monomers bind the composite site. EMSA was performed for each DNA probe using either AtT-20 nuclear extracts expressing (or not, neo) Pax7 with / without the PD.HD polypeptide as indicated. Arrowheads indicate the positions of wild-type Pax7 and PD.HD polypeptide monomer complexes. For the H.I2 probe, dimer bands are also observed for Pax7 (*) and PD.HD (**) together with appearance of a new band (***) in the combined reaction (lane 16) that likely represents a heterodimer between intact Pax7 and PD.HD polypeptide. The presence of Flag-Pax7 in complexes is ascertained by supershifts with the Flag M2 antibody.

Cytosine methylation does not interfere with Pax7 DNA binding

Prior to Pax7 binding, the pioneered enhancers are within nucleosomal heterochromatin (4) as determined by MnaseSeq (Fig. S2). Upon Pax7 binding, nucleosomes are displaced from the Pax7 pioneered sites but not from other subsets of Pax7 binding sites (Fig. S2), and this correlates with the loss of histone H3 (4). The Pax7 pioneered sites are also within heterochromatin with high levels of CpG DNA methylation and this includes CpG residues that are within composite sites in some cases (4). These enhancers acquire stable epigenetic memory through demethylation of cytosine residues within targeted enhancers (Fig. 3A, B). Indeed, the average level of CpG methylation drops from ~80% to ~30% after Pax7-dependent chromatin

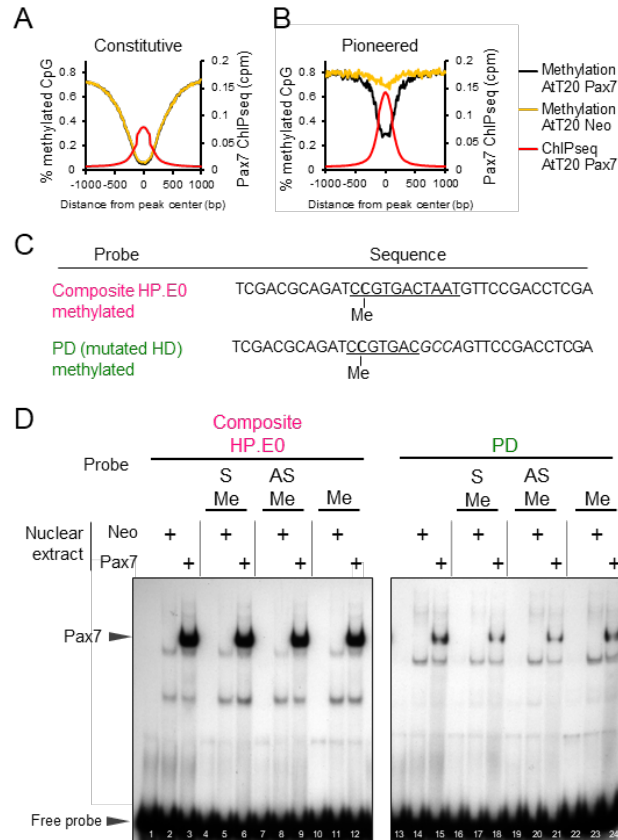


Figure 3. Pax7 binding is independent of DNA methylation.

(A, B) Average profiles of DNA methylation (% methylated CpG) around Pax7 ChIPseq peaks present at Constitutive (A) and Pioneered (B) enhancers. Data are shown for control AtT20-neo compared to AtT20-Pax7 cells and aligned with the average Pax7 ChIPseq profiles.

(C) DNA sequences of probes used in EMSA to test the effect of DNA methylation on Pax7 binding. Methylated cytosine (C) residues are shown on sense (S) strand.

(D) *In vitro* binding (EMSA) of Pax7 to either composite or PD probes methylated or not on sense (S), antisense (AS) or both (Me) strands.

opening at the pioneered enhancers (Fig. 3B). These data suggest that cytosine methylation of DNA target sequences does not interfere with Pax7 interaction, and we directly assessed this by producing DNA probes with cytosine methylation of the CpG dinucleotide present within the composite Pax7 binding site (Fig. 3C). Probes containing methylcytosine on the sense (S, lane 6), the antisense (AS, lane 9) or both strands (lane 12) were compared to an unmethylated probe for Pax7 binding and Pax7 was found to bind equally well to all probes (Fig. 3D). This was also found to be the case for Pax7 binding to the methylated PD probe which showed weaker, but nonetheless methylation-independent, binding (Fig. 3D, lanes 13-25).

The ability of Pax7 to bind its composite and PD sequences independently of cytosine methylation status is entirely consistent with the idea that this pioneer factor can recognize its target sequence within “closed” nucleosomal heterochromatin that is heavily methylated on cytosines.

Both Pax7 DNA binding domains are required for pioneer activity

To ascertain the role of the two Pax7 DBDs in transcription and pioneer action, we introduced single amino acid mutations in either PD or HD (Fig. 4A). For the PD domain, the amino acid substitution R56L in the α 1 helix of the PAI domain was derived from a Waardenburg syndrome mutation identified in the related human PAX3 gene (40,41), and shown to be a loss-of-function mutation (42,43). Another PD mutation G43A was introduced upstream of the α helices of the PAI domain; this mutation was previously shown to lose PD, but not palindromic HD, sequence binding (19). For the HD domain, the critical recognition residue Serine 50 of the HD was changed to glutamic acid (S264E). For comparisons, Supplementary Fig. S4 provides an alignment of sequences for mouse Pax7 (numbering used in the present work), human PAX3 and human PAX6. These mutations were introduced in Flag-Pax7 and stably expressed in AtT-20 cells as previously for Pax7 (5). The three Flag-Pax7 mutant proteins are expressed at similar levels (Fig.S3A). *In vitro* (EMSA), the R56L and G43A PD mutants did not bind the PD probe significantly (Fig. S3B, lanes 8 and 9). Strikingly, all three mutations had much less effect on binding to the composite probe (Fig. S3B, lanes 2-5). For HD probes H.I2 and H.I-2, the HD S264E mutant lost most binding activity but so did the PD mutants (Fig. S3B, lanes 13-15 and 18-20).

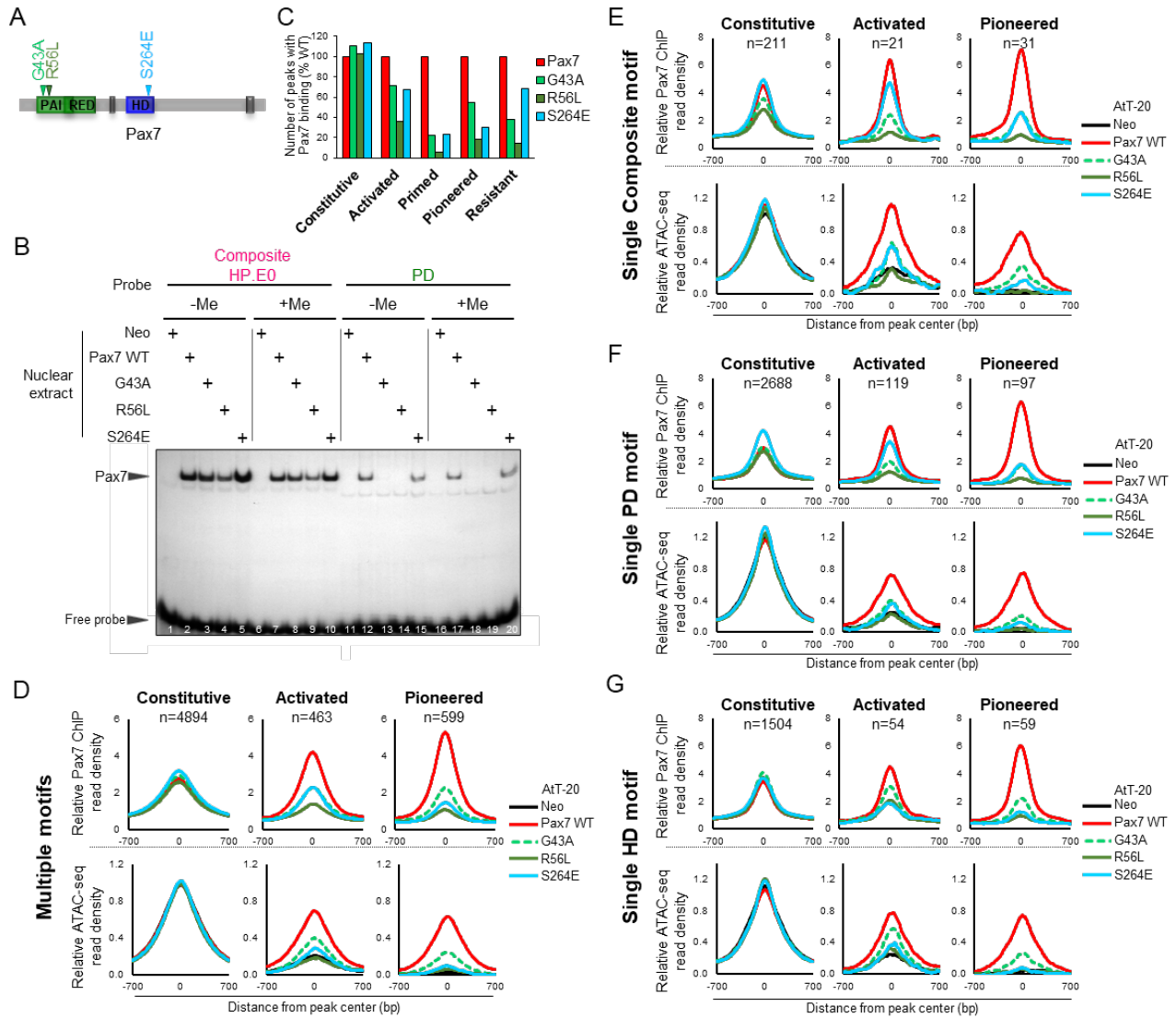


Figure 4. Intact and functional PD and HD domains are both required for recruitment and chromatin opening at Pax7 pioneer sites.

- (A) Schematic representation of Pax7 showing three amino acid substitutions assessed for recruitment by ChIPseq and chromatin opening by ATACseq.
- (B) Effect of three Pax7 mutations on ChIPseq recruitment at different subset of peaks as indicated. The number of peaks (% compared to wild-type Pax7) that retained significant ($p < 10^{-5}$) binding are shown.
- (C) *In vitro* binding (EMSA) of Pax7 and its mutants to either Composite or PD probes methylated or not (+/- Me) on both strands.
- (D) Average profiles of Pax7 ChIPseq and ATACseq signals for the indicated subsets of Pax7 peaks and for the subsets of peaks in each category that have multiple Pax7 recognition motifs.
- (E) Average profiles for the subsets of Pax7 peaks of the indicated categories that only contain one composite motif.
- (F) Average profiles for the subsets of Pax7 peaks of the indicated categories that only contain one PD motif.
- (G) Average profiles for the subsets of Pax7 peaks of the indicated categories that only contain one HD motif.

Since the crystal structure of hPAX6 showed that the Arg residue corresponding to R56L (labeled R23 in Xu et al. (44)) contacts the DNA backbone (but not the methylation site per se) at the level of the CpG dinucleotide present within the PD motif, we assess the hypothesis that the mutations may selectively affect binding of the methylated target sequences. We thus tested all three mutants for binding of the methylated (on both strands) composite and PD sites compared to unmethylated DNA (Fig. 4B); other CpG present in our probes are outside of conserved DNA sequences identified at Pax7 ChIPseq peaks. DNA methylation of the CpG dinucleotide present within the PD sequence of either probe did not affect binding by any mutant.

The transcriptional potential of these mutants was assessed by co-transfection into AtT-20 cells using a Luciferase reporter driven by six copies of the composite element or by the PC2 gene 1.4 Kb enhancer located at -146 Kb upstream of TSS (5). For the composite element reporter, transcriptional activity was affected by the R56L mutation but not by the G43A or S264E mutations. Except for R56L, these results (Fig. S3C) are consistent with the *in vitro* binding results using the composite probe (Fig. S3B). In contrast, the activity of the PC2 enhancer reporter was dependent on both PD and HD as evidenced by significant activity losses for all three mutants (Fig. S3C). The dependence on both the PD and HD for activity at this natural enhancer is thus more complex than the simpler *in vitro* binding patterns observed in EMSA.

The impact of the DBD mutations on *in vivo* genomic recruitment was assessed by ChIPseq (Fig.4C). It is quite striking to observe that key amino acid substitutions in either PD or HD failed to affect the recruitment of Pax7 to the Constitutive enhancers that are already in an open chromatin state: this likely reflects the strong contribution of protein-protein interactions for recruitment of Pax7 in the context of active enhancers. In contrast, recruitment to the Pioneered enhancers is severely curtailed by all three mutations (loss of recruitment at ~80% of sites for R56L, at ~70% sites for S264E and ~50% sites for G43A) with the G43A PD substitution having a partial effect compared to the R56L substitution (Figs. 4C and S3D). Thus, both DBDs must be intact for pioneer site binding whether considering the fully activated Pioneer sites or the Primed sites. Intermediate effects of the different mutations are observed at the group of so-called Activated enhancers, a group of enhancers with open chromatin where

the co-activator p300 is recruited after Pax7 binding. Collectively, these data indicate that the G43A mutant is a partial loss-of-function whereas the other mutants appear complete nulls. Finally, it is noteworthy that binding of the Pax7 mutants to the group of Resistant sites, binding sites that show no chromatin change upon Pax7 binding, is markedly impaired by the PD mutations but relatively unaffected by the HD S264E mutation. This data indicates that the PD domain is primarily involved in scanning heterochromatin sites for binding and that this function is relatively independent of the HD.

The loss of recruitment observed in ChIPseq for the mutants (assessed as the presence of peaks above threshold) is significant for the Pioneered, Primed and Activated sites but it is not complete (Fig. 4C). In order to ascertain the meaning of this residual recruitment, we selected subsets of enhancers (Constitutive, Pioneered or Activated, Fig. 4E-G) based on presence of a single Pax7 recognition motif, either composite, PD or HD motif. For comparisons, average profiles of Pax7 ChIPseq peaks containing multiple motifs are shown in Fig. 4D. The Constitutive Pax7 peaks are relatively insensitive to either PD or HD domain mutations irrespective of the presence of multiple motifs (Fig. 4D) or of only one motif (Fig. 4E-G). In order to correlate Pax7 recruitment with pioneering activity, we performed ATACseq for the Pax7 DBD mutants and compared them to the ATAC profiles of cells expressing wild-type Pax7. These profiles (Fig. 4D-G, bottom panels) illustrate the open chromatin status of Constitutive enhancers independently of Pax7 recruitment. In sharp contrast, the Pioneered peaks show dependence on intact DBDs for both PD and HD, as much for Pax7 recruitment as for chromatin opening (ATAC).

Recruitment to the Activated subset is more dependent on DNA recognition specificity. Indeed, peaks that have either single composite (Fig. 4E) or PD motifs (Fig. 4F) show relatively unaltered Pax7 recruitment for the HD mutant S264E but this is only marginally reflected in ATACseq signals which are similarly decreased by both PD and HD mutations. For sites that have a single HD motif, both Pax7 recruitment and ATAC signals are affected by mutations in either PD or HD. Specific loci examples of these effects are provided in Fig. S5. Collectively, these results indicate that PD and HD domains are required to stably bind Pioneer sites and trigger chromatin opening. For Activated sites that already exhibit some chromatin opening (ATAC signal), there is a similar dependence on both domains for binding targets with a single

HD motif whereas single PD or composite motif-containing targets only show dependence on the intact PD but not intact HD.

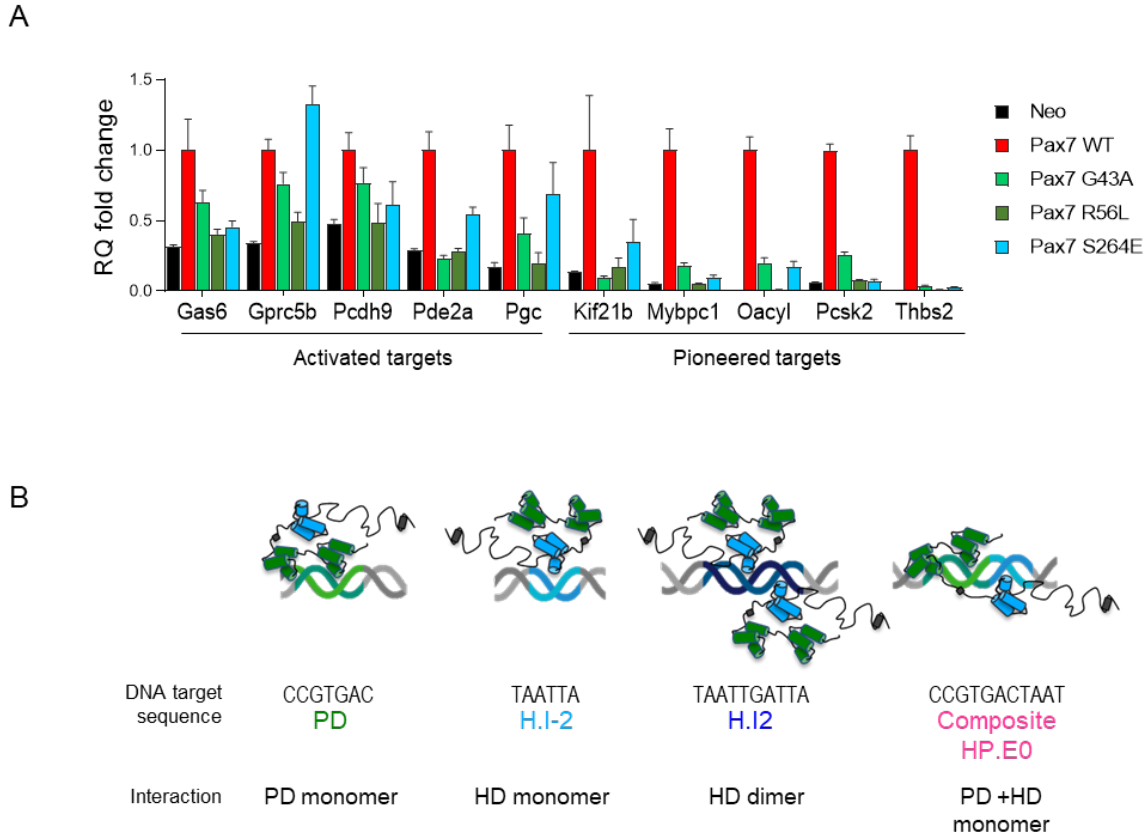


Figure 5. Mutations of either the PD or HD prevent gene activation at targets that require pioneer action whereas they have variable effects on transcriptional target genes.

(A) RT-qPCR measurements of mRNA levels for genes subject to activation through either Pax7 pioneering or regular transcriptional activation as indicated. The gene activation time courses (slow or rapid, respectively) for these genes is shown in Fig. S6A. Data are presented for wild-type Pax7 in comparison with mutations in either PD or HD DNA binding domains.

(B) Models of Pax7 interaction with its different DNA target sequences.

We then asked whether gene activation shows similar dependence on intact PD and HD domains as revealed in ChIP recruitment. This was assessed by RT-qPCR measurements of mRNA levels for gene targets that are either dependent on regular transcriptional activation or on pioneering mechanism (Fig. 5A and S6A). We previously showed rapid transcriptional response to Pax7 of genes that are subject to transcriptional activation compared to the slower (days) response of genes that require pioneer-dependent enhancer opening (4). The mRNA accumulation time courses of genes assessed for the Pax7 mutants are illustrated in Fig. S5A.

Consistent with the Pax7 recruitment and ATAC data, activation/transcription of genes dependent on pioneer action show dependence on both intact PD and intact HD (Fig. 5A). Activation of gene transcription for gene targets of the Activated class of enhancers is more heterogeneous with some showing dependence on both DBDs (Pcdh9, Pde2a, Gas6) and some with lesser dependence on the HD domain (Gprc5, Pgc). Whereas virtually all Pax7 target genes are regulated by Pax7-dependent enhancers, one notable exception is the Oacy1 gene where Pax7 exerts its pioneer action at the promoter (Fig. S6B); this gene shows a quicker response to Pax7 expression compared to most other Pax7 pioneered target genes (Fig. S6A).

Collectively, the mutant analyses reveal a strong dependence on both the PD and HD for pioneer activity independently of underlying DNA motifs present in pioneered enhancers. This may relate to a unique arrangement of the PD and HD consistent with interaction at composite binding sites.

Discussion

This work has queried the requirements for selection of DNA target sites susceptible to pioneer factor chromatin opening. By comparing the DNA sequence motifs present under pioneered sites with those of other subsets of Pax7 binding sites together with mutations within each Pax7 DBD, the unique requirements for pioneer action suggest that structural features required for pioneer activity are far more stringent than for recruitment at enhancers within pre-opened chromatin. Much more than simple DNA sequence recognition is involved for recruitment of TFs and pioneers, whether considering recruitment of the former at different enhancer states or, of the latter, in different chromatin environments.

What may distinguish the Pax7 pioneered sites compared to other binding/recruitment sites? Compared to all other Pax7 binding sites, the pioneered subset is enriched in high affinity binding sites such as the composite motif, but this site is only present in about 20% of all pioneered enhancers. The pioneered sites are also enriched for the PD motif, but this enrichment is not strikingly different when compared to its abundance in other subsets of Pax7 peaks. Nonetheless, Pax7 binding is clearly much stronger at the pioneer subset compared to other subsets (4). Thus, it may be the number of motifs that results in stronger recruitment, for example when comparing the pioneered subset with the constitutive enhancers, but it is not the specific DNA sequence of the motifs present in different subsets that account for their different properties. This argument is strongly supported by the comparison of specific motifs and their abundance in the subset of Resistant compared to Pioneered sites: both subsets have very similar distribution of all PD and HD binding motifs. This observation clearly points towards heterochromatin organization as the defining property that establishes susceptible and resistant heterochromatin for pioneering.

A unique property of pioneer factors appears to be their ability to gain access to their target DNA sequence within nucleosomal and compacted heterochromatin. The molecular basis for this ability remains unexplained although some hypotheses were put forward. For example, it was proposed that FoxA pioneers may have this ability through molecular mimicry where a portion of the winged helix domain may mimic a structure within linker histone H1 (7). However, this model is not likely to apply to many other pioneers. The FoxA pioneers also have the ability to bind low-affinity sites within heterochromatin and the ability to recognize these

low sequence conservation sites was proposed to represent a mechanism to scan putative sites within heterochromatin (21). Further, the presence of a core histone interaction domain within FoxA factors (23) clearly supports the hypothesis of direct chromatin interaction as a unique feature of pioneers providing them access to heterochromatin. Be that as it may, this fails to adequately explain how specific sites are selected for pioneering. It is quite striking that the ability for Pax7 recruitment to the weak binding sites of the Resistant subset is almost entirely dependent on the PD, but not the HD (Fig. 4C). This preference is not reflected in an enrichment for the PD motif within the Resistant subset compared to HD motifs and thus likely reflects a property of the Pax7 PD to recognize heterochromatin.

Whereas the PD appears predominantly responsible for scanning heterochromatin sites, chromatin opening and gene activation by Pax7 pioneering is completely dependent on both intact PD and HD. This dependence is clear for both the fully activated Pioneered sites as well as for the primed sites: hence, this dependence probably reflects requirements for the initial step of the chromatin opening mechanism. The dependence on both intact DBDs for pioneering is not strictly correlated with the presence of cognate DNA sequence motifs for these domains within the pioneered subset of enhancers, although this subset has the highest enrichment for all motifs and particularly the high-affinity composite (HP.E0) sites. Given the higher affinity of the composite site for Pax7 binding *in vitro* and its association with a Pax7 monomer (Fig. 5B), but recognizing that many pioneered enhancers do not contain this high-affinity site, it is reasonable to propose that a unique conformation of the two DBDs relative to each other is associated with pioneering ability, and that this conformation can be achieved through different combinations of DNA sequence motif arrangements. This hypothetical conformation illustrated over the composite motif may differ compared to the Pax7 dimers interacting with the palindromic H.I2 target or even compared to simple PD or HD motifs (Fig. 5B).

The situation at pioneered enhancers is in striking contrast with the absence of dependence on either intact PD or HD for recruitment to the Constitutive subset of enhancers that were already in an open and active chromatin state before Pax7 expression. This independence towards intact DBDs is correlated with the lesser abundance of PD and HD motifs within the Constitutive subset of enhancers and likely reflects the greater importance of protein-protein interactions at these enhancers that are already occupied by an array of different

transcription factors. It is noteworthy that the subset of Pax7 targets that exhibits the greatest dependence on cognate DNA sequence motifs are found within the Activated subgroup of enhancers, those enhancers that are in a primed state at the onset but become fully activated upon Pax7 binding as revealed by their recruitment of the co-activator p300. For this subset, the presence of a cognate DNA motif is most of the time correlated with the loss of this binding for the Pax7 mutant that contains a substitution in the corresponding DBD. It is noteworthy that target DNA sequence recognition is most important for this group of enhancers that are awaiting for a cognate TF for full activation. In this instance, the pioneer properties are likely not required as the chromatin is already slightly accessible, i.e. in the primed state.

In summary, the present work defined unique contributions of the two Pax7 DBDs towards pioneer action. First, pioneer factors demarcate themselves compared to other TFs by the ability to recognize targets within heterochromatin and this property appears to be associated with the PD. Second, pioneered sites exhibit the strongest recruitment in ChIPseq and this correlates with higher frequency of cognate DNA motifs and of sites of high affinity. Finally, a unique conformation that requires both the PD and the HD of Pax7 is required for chromatin opening at pioneer sites and gene activation.

Data Availability

Sequencing data are available under GSE87185 and GSE173743.

Funding

This work was supported by a Foundation grant (FDN154297) to J.D. from the Canadian Institutes of Health Research. Data analyses were possible thanks to the support of Compute Canada.

Acknowledgements

We are grateful to Odile Neyret and her group at the IRCM Sequencing Core Facility for NGS analyses and to Valerie Magoon for her expert secretarial assistance.

Conflict of interest disclosure: None declared.

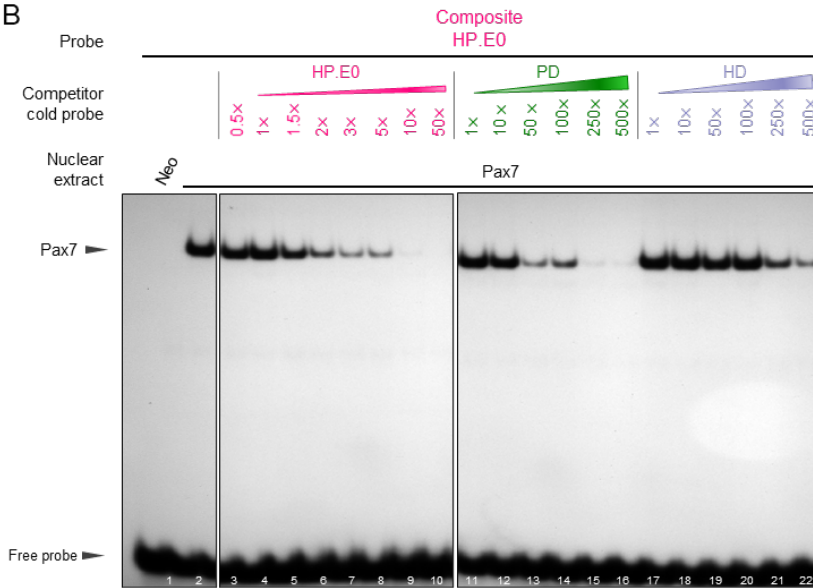
Figures supplémentaires

Supplementary Figure 1

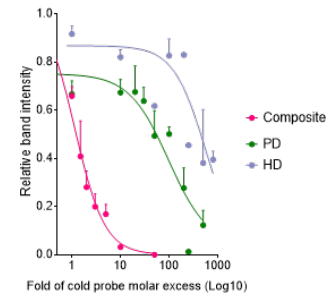
A

Probe sequences	
Composite HP.E0	TCGACGCAGAT <u>CCGTGACTAAT</u> GTTCCGACCTCGA
PD (mutated HD)	TCGACGCAGAT <u>CCGTGAC</u> GCCAGTTCGACCTCGA
HD (mutated PD)	TCGACGCAGATCGTCCGC <u>TAAT</u> TCCAACGACCTCGA
H.I-2	TCGACCAGATCGTCCGC <u>TAATTA</u> CCAACGACCTCGA
H.I2	TCGACTCGTCCGC <u>TAATTGATTA</u> TCCAACGACCTCGA

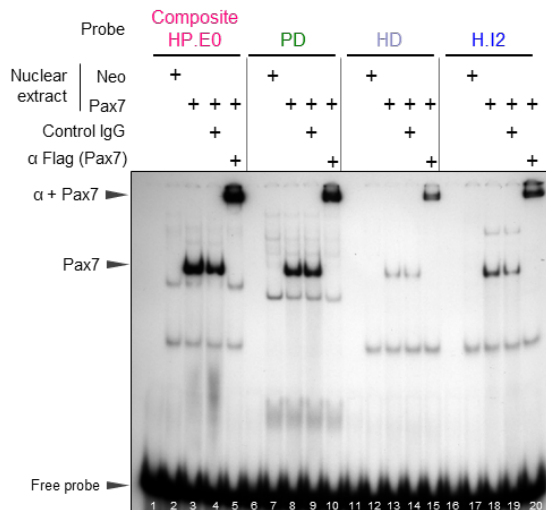
B



C



D



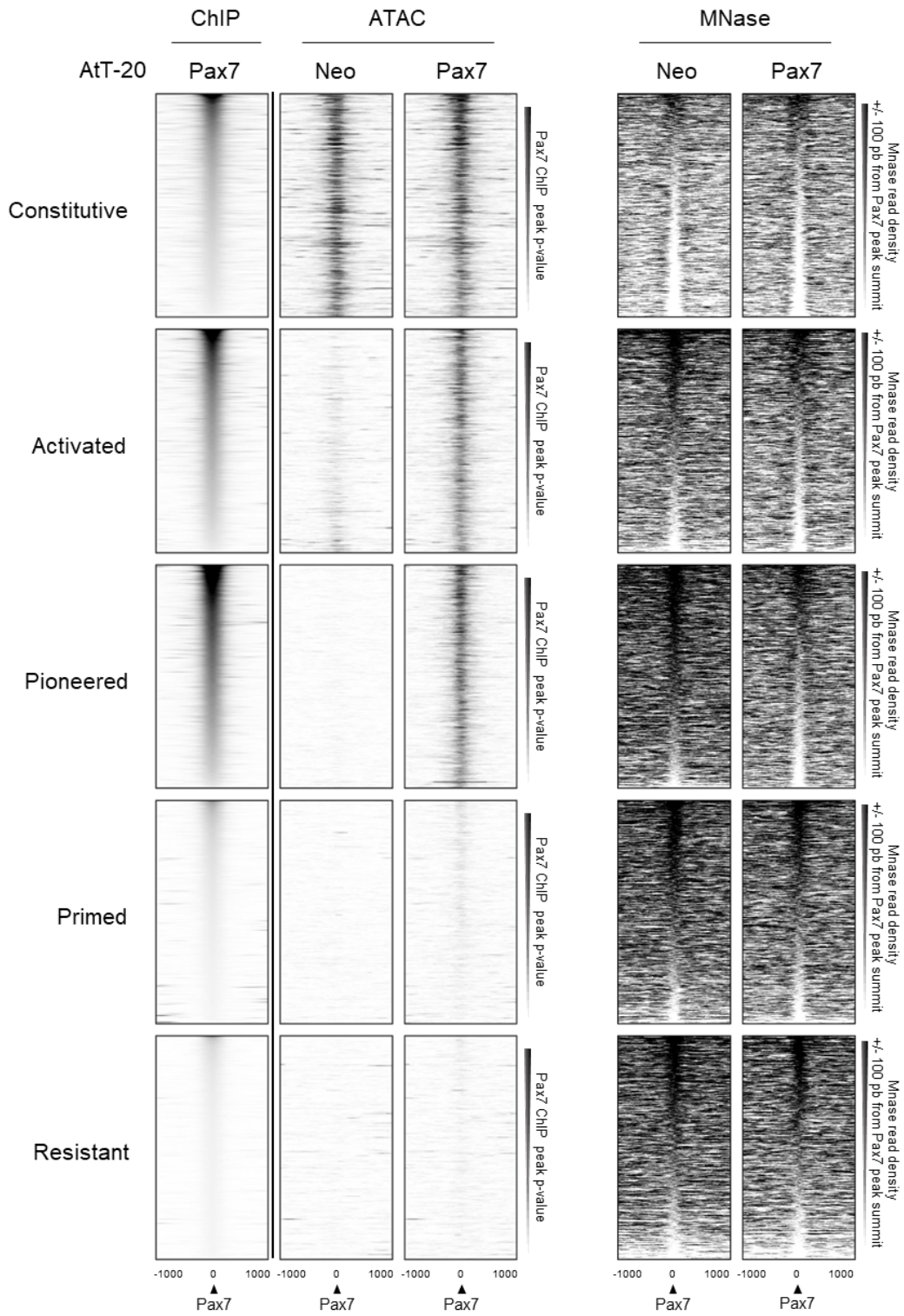
Supplementary Figure 1. Characterization of Pax7 EMSA complexes with various probes.

(A) DNA sequences of probes used in gel retardation assays (EMSA).

(B, C) Competitive EMSA assays to assess relative affinities of composite HP.E0, PD and H.I2 targets. Complexes formed between the HP.E0 probe and Pax7 from AtT20-Pax7 nuclear extracts were competed with the indicated molar excesses of either HP.E0, PD or H.I2 probes. Displacement curves (means \pm SD, n=4) for quantitated band intensities (Image Lab, BioRad) were computed with GraphPad Prism.

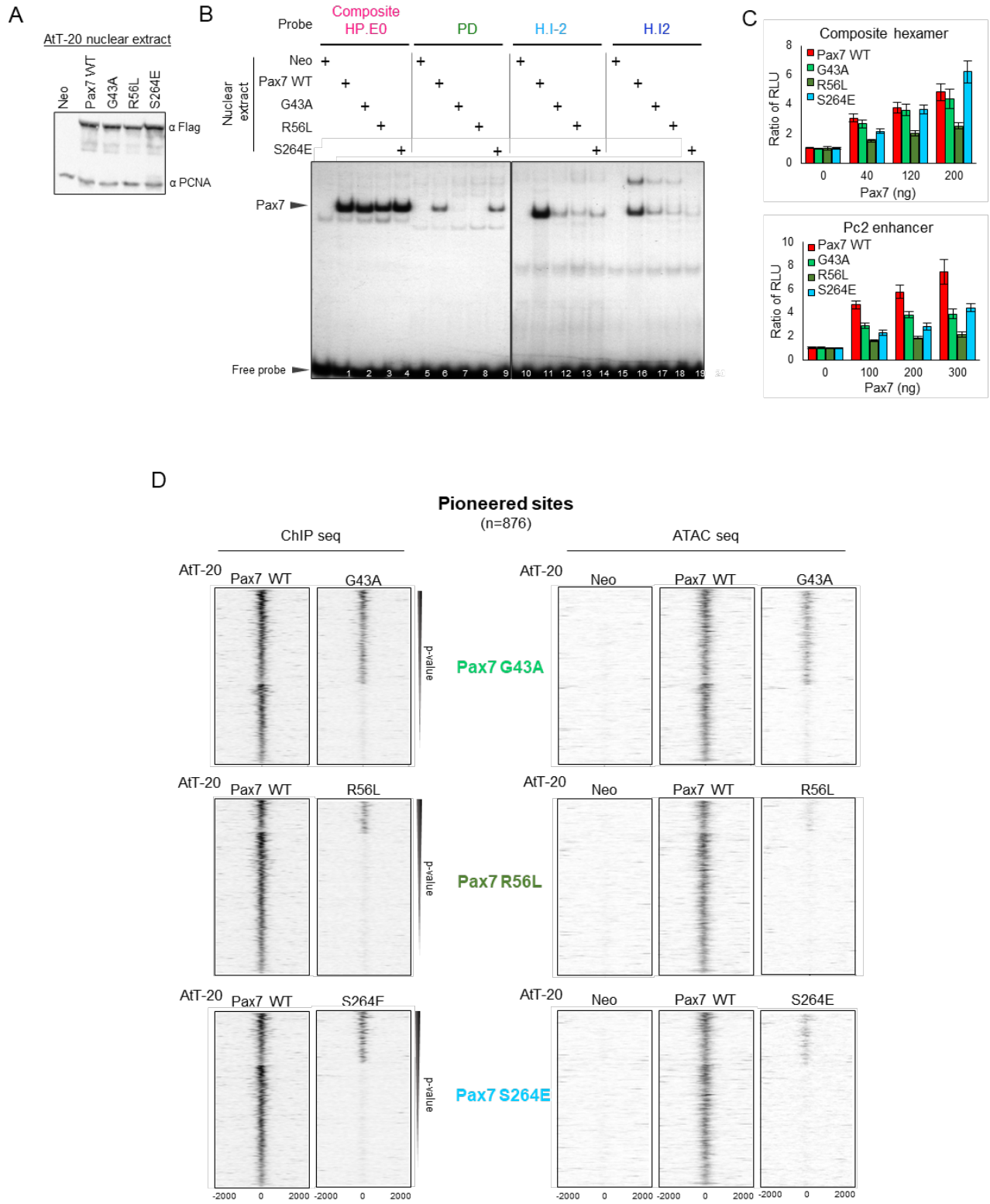
(D) EMSA supershift experiments using the Flag2 antibody to supershift Flag-Pax7-containing DNA complexes.

Supplementary Figure 2



Supplementary Figure 2. Nucleosome positioning relative to Pax7 Pioneer sites was determined by MNaseSeq analyses in AtT-20 Neo and AtT-20 Pax7 cells.

The MNaseSeq heatmaps (right) are sorted according to central read densities (± 100 bp) and centered (± 1000 bp) on the by Pax7 peak summit for each subset of sites. For comparison (left), the Pax7 ChIPseq and ATACseq profiles centered on Pax7 binding are shown (data from ref. 4). Only the Pioneered subset of sites shows a different pattern in AtT-Pax7 cells compared to the Neo control, revealing recruitment to nucleosome-occupied sites and nucleosome depletion after Pax7 binding.



Supplementary Figure 3. Characterization of Pax7 mutants within the PD (G43A and R56L) and HD (S264E) domains.

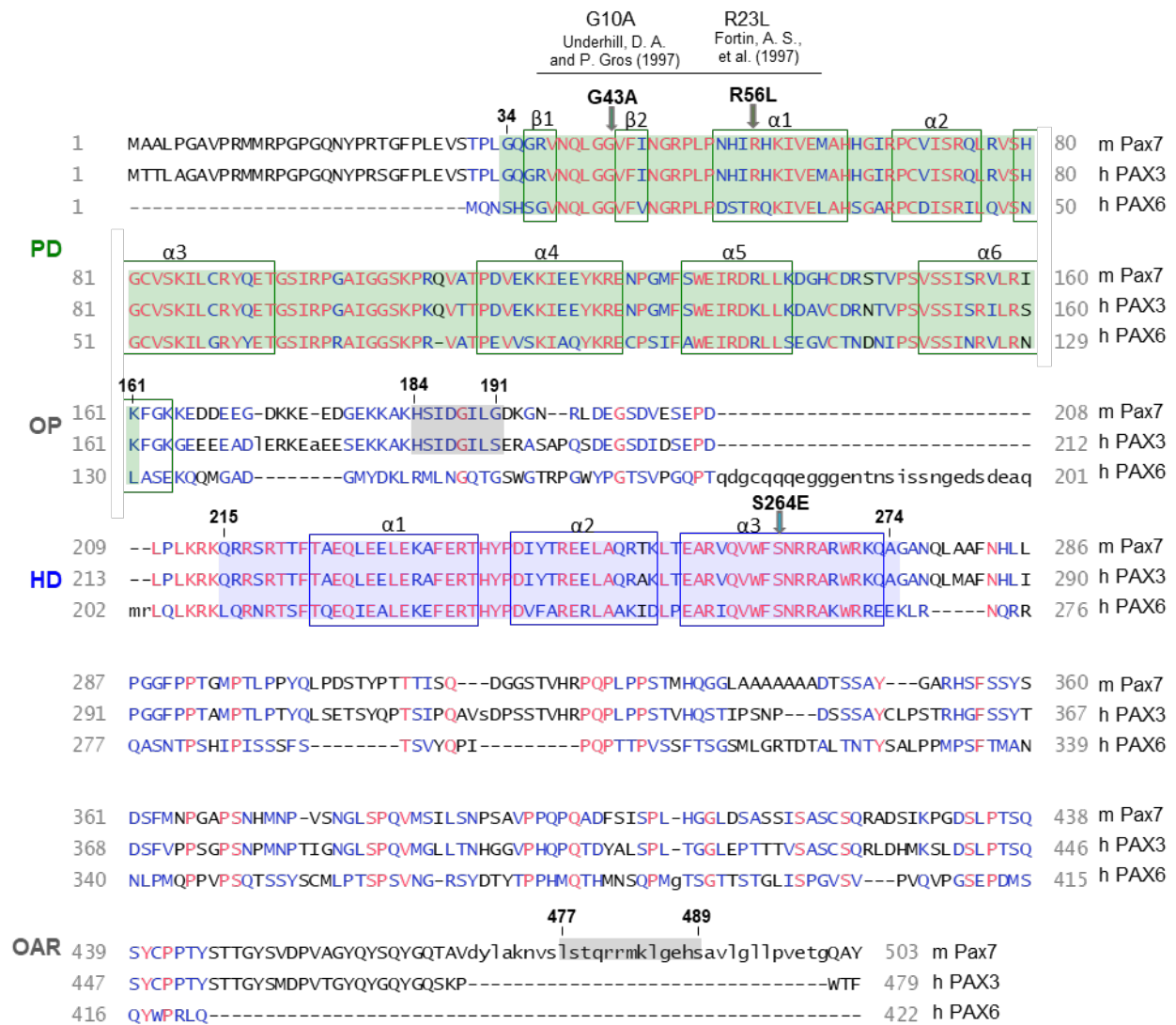
(A) Western blot analysis for expression of Pax7 and three mutants in AtT-20 cells expressing N-terminally flagged Pax7 proteins. PCNA Western blot is used as loading control.

(B) EMSA using the indicated DNA probes incubated with nuclear extracts from AtT-20 cells expressing the indicated Pax7 and mutants.

(C) Transcriptional activity of Pax7 and its DBD mutants assessed by cotransfection of indicated expression plasmids together with a luciferase reporter plasmid driven by six copies of the HP.E0 composite or by the PC2 gene enhancer. Luciferase activity was assessed as previously (5) in extracts of AtT-20 cells prepared 48 hours post-transfection with PEI25K.

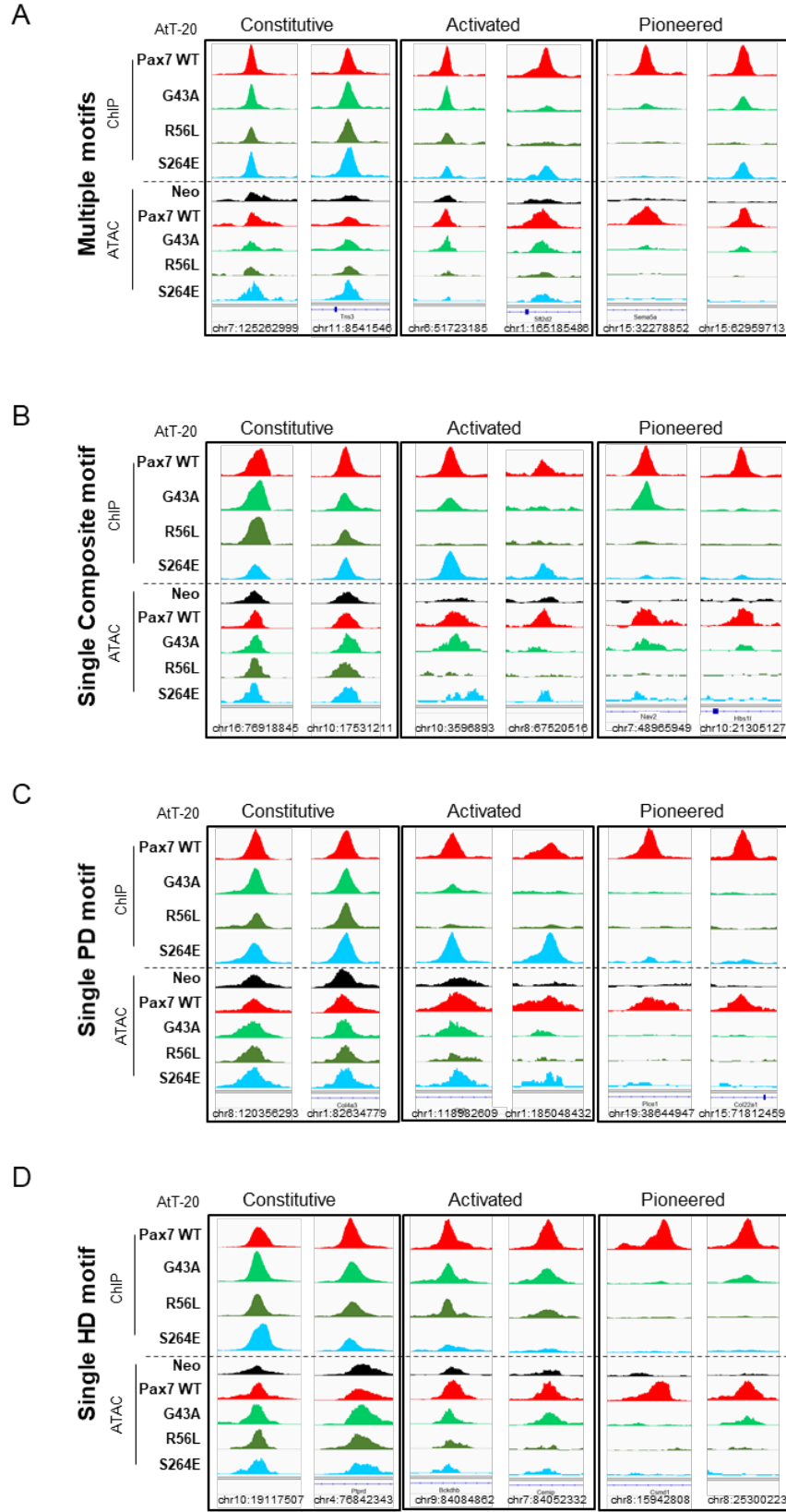
(D) Genomic recruitment (Flag-Pax7 ChIPseq) and ATACseq profiles at the Pax7 Pioneered sites compared to the same sites in AtT-20 cells expressing the three Pax7 DBD mutants

Supplementary Figure 4



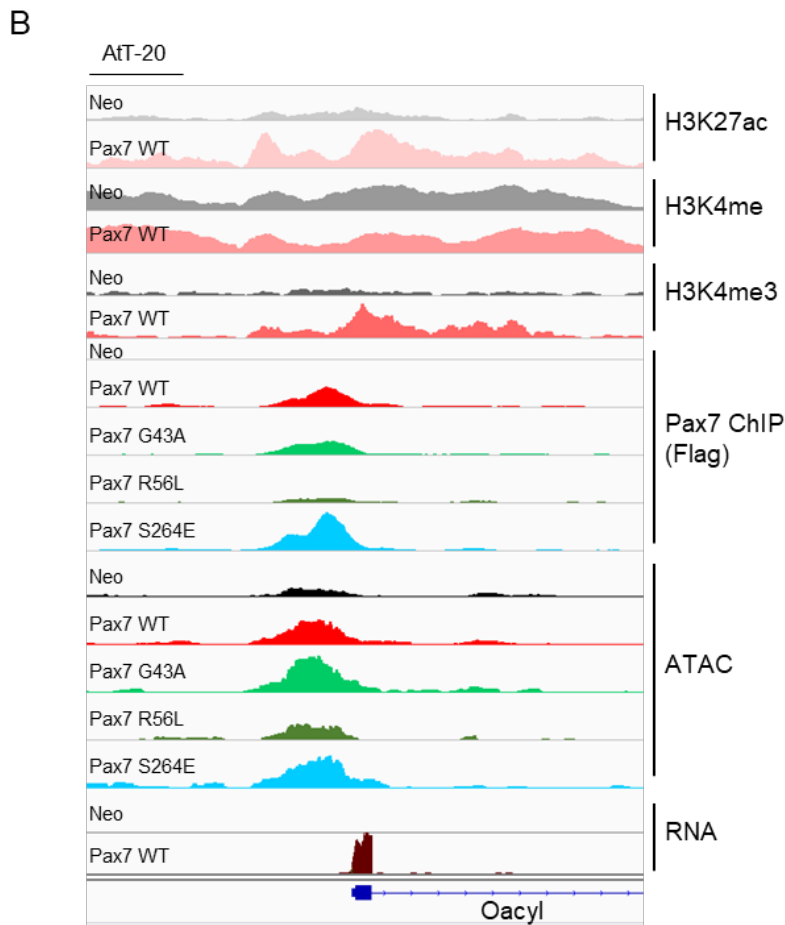
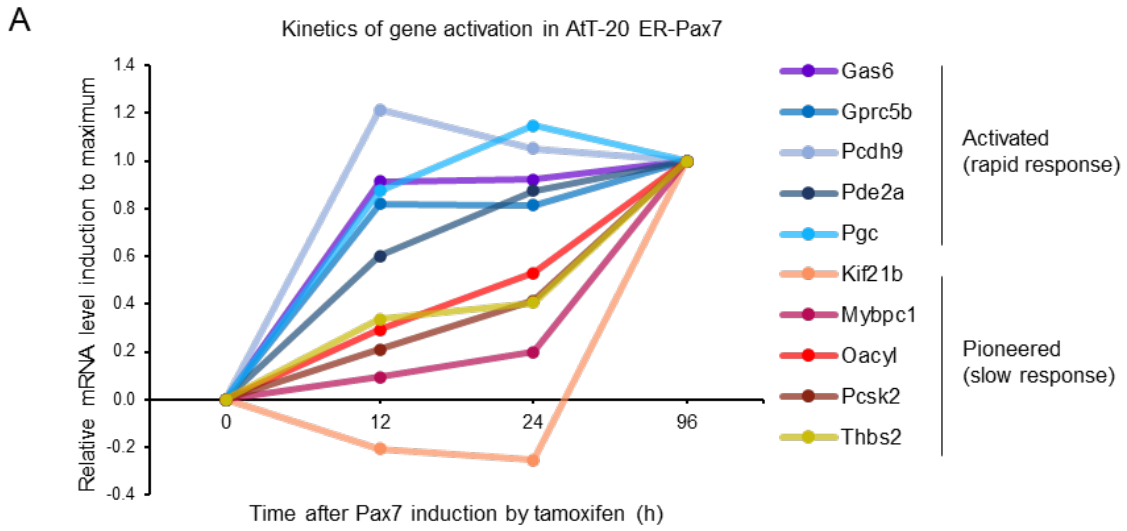
Supplementary Figure 4. Alignment of mouse Pax7, human PAX3 and human PAX6 sequences showing structural features of PD (green) and HD (blue) DNA binding domains.

Sequences (Accession numbers P47239.2, P23760.2, P26367.2) were aligned using NCBI Blast 2. Amino acids legend: Black = all different ; Blue = 2/3 identical ; Red = 3/3 identical, Upper case = corresponding in number ; lower case = specific to one protein. The numbering used in indicated previous publications is shown together with positions in mouse Pax7 used in present work.



Supplementary Figure 5. Representative loci for genomic recruitment (Flag-Pax7 ChIPseq) and ATACseq for wildtype Pax7 and three DBD mutants as indicated.

Two representative loci are shown for each subset of peaks in the Constitutive, Activated and Pioneered categories. Examples are shown for peaks containing multiple motifs (**A**), or only one Composite (**B**), PD (**C**) or HD (**D**) motif.



Supplementary Figure 6. Pax7 target genes.

(A) Time courses of mRNA accumulation determined by RNAseq analyses (5) showing rapid responses for genes regulated through “Pax7 Activated” enhancers compared to the slower responses of “Pioneered” target genes.

(B) A unique example, the Oacyl gene, of Pax7 pioneer action at promoter rather than enhancer sequences. Gene browser views of the Oacyl locus show the indicated ChIPseq and ATACseq analyses performed in AtT-20 Neo, AtT-20 Pax7 WT or mutant cells.

Supplementary Table 1. Sequences of oligonucleotide primers

Gene symbol	Sequence accession number	Amplicon length	Primer sequence	Location
Gapdh	NM_008084.3	110	TGCAGTGGCAAAGTGGAGAT	Exon 2
			ACTGTGCCGTTGAATTTGCC	Exon 2
Gas6	NM_019521.2	107	GCCTCTATCACCTGAATCT	Exon 11
			CCAGTTCCAACCTCCTCAT	Exon 12
Gprc5	NM_022420.2	122	CACCGTTCAGAAGCAATG	Exon 3
			GGAACTTAAGGTCTTTCACC	Exon 4
Pcdh9	NM_001081377.3	150	CAGCCACAGGACGAATTCTAC	Exon 4
			ACCCAAAACCAGACACTCTTG	Exon 5
Pde2a	NM_001143848.2	161	GAGAACCAGGAGGTCATT	Exon 18-19
			CGGTATTGGGCTTCATTC	Exon 20
Pgc	NM_025973.3	168	TGCCCAGTACCTGAATGA	Exon 7
			CCTTCCTCCTGGATGATGTA	Exon 8-9
Kif21b	NM_001039472.2	161	AGAACCAGTCTCGCTATG	Exon 15
			CTTCTTGAGCTGAGCAATC	Exon 16
Mybpc1	NM_175418.5	200	ACGAGGTCCGAATCTTTG	Exon 19
			ATACTCTAGCACATAGCCATC	Exon 20
Oacyl	NM_177028.3	181	GTCCAGATATGTTTGCTTCC	Exon 4
			CTGTTGCTCTCCATGTTTC	Exon 5
Pcsk2	NM_008792.4	123	TGACAAGTGGCCTTTCAT	Exon 12
			ATCAGGGTCCATTCCCTTC	Exon 12
Thbs2	NM_011581.3	183	CACTCAAAGTGGTAAACTCC	Exon 19
			CTCTCATGTAGCCTGTCTTA	Exon 20-21

Supplementary Table 2. Summary of high throughput sequencing datasets and experimental conditions

Experiment	Cell line	Number of cells	Sequencing depth (read number)	Matching control for peak calling	Antibody reference	Antibody quantity (µg)	Replicate number
ATAC-seq	AtT-20 Neo	1,00E+05	95251912	Input AtT Pax7	N/A	N/A	2
ATAC-seq	AtT-20 Pax7	1,00E+05	82583056	Input AtT Pax7	N/A	N/A	2
ATAC-seq	AtT-20 Pax7 G43A	1,00E+05	32502682	Input AtT Pax7	N/A	N/A	1
ATAC-seq	AtT-20 Pax7 R56L	1,00E+05	31497109	Input AtT Pax7	N/A	N/A	1
ATAC-seq	AtT-20 Pax7 S264E	1,00E+05	26377490	Input AtT Pax7	N/A	N/A	1
ChIP-seq	AtT-20 Pax7	1,00E+07	77949054	AtT-20 Neo	Flag M2 Sigma	6	1
ChIP-seq	AtT-20 Pax7 G43A	1,00E+07	94880640	AtT-20 Neo	Flag M2 Sigma	6	1
ChIP-seq	AtT-20 Pax7 R56L	1,00E+07	72825980	AtT-20 Neo	Flag M2 Sigma	6	1
ChIP-seq	AtT-20 Pax7 S264E	1,00E+07	102208451	AtT-20 Neo	Flag M2 Sigma	6	1
MNase-seq	AtT-20 Neo	3,00E+06	220458865	N/A	N/A	N/A	2
MNase-seq	AtT-20 Pax7	3,00E+06	229205088	N/A	N/A	N/A	2

Supplementary Table 3. ChIP and ATAC-seq thresholds to categorize Pax7 peaks

	Pax7	H3K4me1		p300		ATAC		SMC1
	AtT-20 Pax7	AtT-20 Neo	AtT-20 Pax7	AtT-20 Neo	AtT-20 Pax7	AtT-20 Neo	AtT-20 Pax7	AtT-20 Neo
Constitutive n = 13 126	PEAK AT $p \leq 1e-5$	PEAK AT $p \leq 1e-5$	PEAK AT $p \leq 1e-5$	PEAK AT $p \leq 1e-5$	PEAK AT $p \leq 1e-5$	PEAK AT $p \leq 1e-5$	PEAK AT $p \leq 1e-5$	N/A
Activated n = 733	PEAK AT $p \leq 1e-5$	PEAK AT $p \leq 1e-5$	PEAK AT $p \leq 1e-5$	NO PEAK AT $p \leq 1e-3$	PEAK AT $p \leq 1e-5$	N/A	PEAK AT $p \leq 1e-5$	N/A
Pioneered n = 876	PEAK AT $p \leq 1e-5$	NO PEAK AT $p \leq 1e-3$	PEAK AT $p \leq 1e-5$	NO PEAK AT $p \leq 1e-3$	PEAK AT $p \leq 1e-5$	NO PEAK AT $p \leq 1e-3$	PEAK AT $p \leq 1e-5$	N/A
Primed n = 5 352	PEAK AT $p \leq 1e-5$	NO PEAK AT $p \leq 1e-3$	PEAK AT $p \leq 1e-5$	NO PEAK AT $p \leq 1e-3$	NO PEAK AT $p \leq 1e-3$	NO PEAK AT $p \leq 1e-3$	N/A	N/A
Resistant n = 8 541	PEAK AT $p \leq 1e-5$	NO PEAK AT $p \leq 1e-3$	NO PEAK AT $p \leq 1e-3$	NO PEAK AT $p \leq 1e-3$	NO PEAK AT $p \leq 1e-3$	NO PEAK AT $p \leq 1e-3$	NO PEAK AT $p \leq 1e-3$	NO PEAK AT $p \leq 1e-3$

References

1. Mayran, A., Jacques Drouin. (2017) Pioneer transcription factors shape the epigenetic landscape. *Journal of Biological Chemistry*.
2. Klemm, S.L., Shipony, Z. and Greenleaf, W.J. (2019) Chromatin accessibility and the regulatory epigenome. *Nature reviews. Genetics*, 20, 207-220.
3. Zaret, K.S. and Mango, S.E. (2016) Pioneer transcription factors, chromatin dynamics, and cell fate control. *Current opinion in genetics & development*, 37, 76-81.
4. Mayran, A., Khetchoumian, K., Hariri, F., Pastinen, T., Gauthier, Y., Balsalobre, A. and Drouin, J. (2018) Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. *Nat Genet*, 50, 259–269
5. Budry, L., Balsalobre, A., Gauthier, Y., Khetchoumian, K., L'Honore, A., Vallette, S., Brue, T., Figarella-Branger, D., Meij, B. and Drouin, J. (2012) The selector gene Pax7 dictates alternate pituitary cell fates through its pioneer action on chromatin remodeling. *Genes & Development*, 26, 2299-2310.
6. Mayran, A., Sochodolsky, K., Khetchoumian, K., Harris, J., Gauthier, Y., Bemmo, A., Balsalobre, A. and Drouin, J. (2019) Pioneer and nonpioneer factor cooperation drives lineage specific chromatin opening. *Nature Communications*, 10, 3807.
7. Cirillo, L.A., McPherson, C.E., Bossard, P., Stevens, K., Cherian, S., Shim, E.Y., Clark, K.L., Burley, S.K. and Zaret, K.S. (1998) Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *Embo j*, 17, 244-254.
8. Fernandez Garcia, M., Moore, C.D., Schulz, K.N., Alberto, O., Donague, G., Harrison, M.M., Zhu, H. and Zaret, K.S. (2019) Structural Features of Transcription Factors Associating with Nucleosome Binding. *Mol Cell*, 75, 921-932.e926.
9. Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M. and Zaret, K.S. (2002) Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell*, 9, 279-289.
10. Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J.A., Mahony, S., Pugh, B.F., Lee, D., Kaestner, K.H. and Zaret, K.S. (2016) The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. *Mol Cell*, 62, 79-91.
11. van Oevelen, C., Collombet, S., Vicent, G., Hoogenkamp, M., Lepoivre, C., Badaeux, A., Bussmann, L., Sardina, J.L., Thieffry, D., Beato, M. et al. (2015) C/EBPalpha Activates Pre-existing and De Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis. *Stem cell reports*, 5, 232-247.
12. Boller, S., Ramamoorthy, S., Akbas, D., Nechanitzky, R., Burger, L., Murr, R., Schubeler, D. and Grosschedl, R. (2016) Pioneering Activity of the C-Terminal Domain of EBF1 Shapes the Chromatin Landscape for B Cell Programming. *Immunity*, 44, 527-541.
13. Li, R., Cauchy, P., Ramamoorthy, S., Boller, S., Chavez, L. and Grosschedl, R. (2018) Dynamic EBF1 occupancy directs sequential epigenetic and transcriptional events in B-cell programming. *Genes Dev*, 32, 96-111.
14. Wapinski, O.L., Lee, Q.Y., Chen, A.C., Li, R., Corces, M.R., Ang, C.E., Treutlein, B., Xiang, C., Baubet, V., Suchy, F.P. et al. (2017) Rapid Chromatin Switch in the Direct Reprogramming of Fibroblasts to Neurons. *Cell Rep*, 20, 3236-3247.
15. Pataskar, A., Jung, J., Smialowski, P., Noack, F., Calegari, F., Straub, T. and Tiwari, V.K. (2016) NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. *Embo j*, 35, 24-45.

16. Takaku, M., Grimm, S.A., Shimbo, T., Perera, L., Menafra, R., Stunnenberg, H.G., Archer, T.K., Machida, S., Kurumizaka, H. and Wade, P.A. (2016) GATA3-dependent cellular reprogramming requires activation-domain dependent recruitment of a chromatin remodeler. *Genome Biol*, 17, 36.
17. Mayran, A., Pelletier, A. and Drouin, J. (2015) Pax factors in transcription and epigenetic remodelling. *Semin Cell Dev Biol*, 44, 135-144.
18. Jun, S. and Desplan, C. (1996) Cooperative interactions between paired domain and homeodomain. *Development*, 122, 2639-2650.
19. Underhill, D.A. and Gros, P. (1997) The paired-domain regulates DNA binding by the homeodomain within the intact Pax-3 protein. *J Biol Chem*, 272, 14175-14182.
20. Apuzzo, S., Abdelhakim, A., Fortin, A.S. and Gros, P. (2004) Cross-talk between the paired domain and the homeodomain of Pax3: DNA binding by each domain causes a structural change in the other domain, supporting interdependence for DNA Binding. *J Biol Chem*, 279, 33601-33612.
21. Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M. and Zaret, K.S. (2015) Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell*, 161, 555-568.
22. Donaghey, J., Thakurela, S., Charlton, J., Chen, J.S., Smith, Z.D., Gu, H., Pop, R., Clement, K., Stamenova, E.K., Karnik, R. et al. (2018) Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat Genet*.
23. Iwafuchi, M., Cuesta, I., Donahue, G., Takenaka, N., Osipovich, A.B., Magnuson, M.A., Roder, H., Seeholzer, S.H., Santisteban, P. and Zaret, K.S. (2020) Gene network transitions in embryos depend upon interactions between a pioneer transcription factor and core histones. *Nat Genet*, 52, 418-427.
24. Rambaud, J., Desroches, J., Balsalobre, A. and Drouin, J. (2009) TIF1beta/KAP-1 is a coactivator of the orphan nuclear receptor NGFI-B/Nur77. *J. Biol. Chem*, 284, 14147-14156.
25. Khetchoumian, K., Balsalobre, A., Mayran, A., Christian, H., Chénard, V., St-Pierre, J. and Drouin, J. (2019) Pituitary cell translation and secretory capacities are enhanced cell autonomously by the transcription factor Creb3l2. *Nature Communications*, 10, 3960.
26. Vallette-Kasic, S., Couture, C., Balsalobre, A., Gauthier, Y., Metherell, L.A., Dattani, M. and Drouin, J. (2007) The TPIT gene mutation M86R associated with isolated ACTH deficiency interferes in protein:protein interactions. *J. Clin. Endocrinol. Metab*, 92, 3991-3999.
27. Langlais, D., Couture, C. and Drouin, J. (2012) The Stat3/GR interaction code: predictive value of direct/indirect DNA recruitment for transcription outcome. *Molecular Cell*, 47, 38-49.
28. Lerdrup, M., Johansen, J.V., Agrawal-Singh, S. and Hansen, K. (2016) An interactive environment for agile analysis and visualization of ChIP-sequencing data. *Nature structural & molecular biology*, 23, 349-357.
29. Thorvaldsdottir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform*, 14, 178-192.
30. Cui, K. and Zhao, K. (2012) Genome-wide approaches to determining nucleosome occupancy in metazoans using MNase-Seq. *Methods in molecular biology (Clifton, N.J.)*, 833, 413-419.
31. Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114-2120.
32. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 10, R25.
33. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078-2079.

34. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F. and Manke, T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*, 44, W160-165.
35. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol*, 9, R137.
36. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*, 38, 576-589.
37. Soleimani, V.D., Punch, V.G., Kawabe, Y., Jones, A.E., Palidwor, G.A., Porter, C.J., Cross, J.W., Carvajal, J.J., Kockx, C.E., van Ijcken, W.F. et al. (2012) Transcriptional dominance of pax7 in adult myogenesis is due to high-affinity recognition of homeodomain motifs. *Dev Cell*, 22, 1208-1220.
38. Wilson, D., Sheng, G., Lecuit, T., Dostatni, N. and Desplan, C. (1993) Cooperative dimerization of paired class homeo domains on DNA. *Genes Dev*, 7, 2120-2134.
39. Birrane, G., Soni, A. and Ladias, J.A. (2009) Structural basis for DNA recognition by the human PAX3 homeodomain. *Biochemistry*, 48, 1148-1155.
40. Chalepakis, G., Goulding, M., Read, A., Strachan, T. and Gruss, P. (1994) Molecular basis of splotch and Waardenburg Pax-3 mutations. *Proc. Natl Acad Sci. U. S. A*, 91, 3685-3689.
41. Hoth, C.F., Milunsky, A., Lipsky, N., Sheffer, R., Clarren, S.K. and Baldwin, C.T. (1993) Mutations in the paired domain of the human PAX3 gene cause Klein-Waardenburg syndrome (WS-III) as well as Waardenburg syndrome type I (WS-I). *American journal of human genetics*, 52, 455-462.
42. Armoni, M., Quon, M.J., Maor, G., Avigad, S., Shapiro, D.N., Harel, C., Esposito, D., Goshen, Y., Yaniv, I. and Karnieli, E. (2002) PAX3/forkhead homolog in rhabdomyosarcoma oncoprotein activates glucose transporter 4 gene expression in vivo and in vitro. *J Clin Endocrinol Metab*, 87, 5312-5324.
43. Fortin, A.S., Underhill, D.A. and Gros, P. (1997) Reciprocal effect of Waardenburg syndrome mutations on DNA binding by the Pax-3 paired domain and homeodomain. *Hum. Mol. Genet*, 6, 1781-1790.
44. Xu, H.E., Rould, M.A., Xu, W., Epstein, J.A., Maas, R.L. and Pabo, C.O. (1999) Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev*, 13, 1263-1275.

Chapitre 3 : Les propriétés de liaison à l'ADN et transcriptionnelles des isoformes de Pax7

Ce chapitre est constitué d'un manuscrit qui se trouve dans sa langue de rédaction originale, soit l'anglais.

Contributions

Audrey Pelletier: Apport original, expériences d'EMSA, lignées AtT-20 stables exprimant les délétions Pax7, transfections, analyses des données, validation et visualisation.

Virginie Bascunana: Expériences lignées AtT-20 stables exprimant les isoformes Pax7 et Pax3

Jacques Drouin: Supervision

DNA binding and transcriptional properties of Pax7 isoforms

Audrey Pelletier, Virginie Bascunana and Jacques Drouin

Laboratory of Molecular Genetics, Institut de recherches cliniques de Montréal (IRCM)
Montréal QC H2W 1R7, Canada

Correspondence should be addressed at: Jacques Drouin, Laboratory of Molecular Genetics,
Institut de recherches cliniques de Montréal (IRCM), 110, avenue des Pins Ouest, Montréal QC
Canada H2W 1R7; Tel: 514-987-5680, E-mail: jacques.drouin@ircm.qc.ca

Abstract

Pax transcription factors are sequence-specific DNA-binding proteins that share a highly conserved paired domain. In vertebrates, *Pax* genes are involved in development and organogenesis through complex processes of lineage determination in the embryo. Pax7 is essential in neural crest formation, myogenesis and pituitary lineage specification. In addition to the paired domain, Pax7 also has a homeo DNA binding domain and both domains are required for Pax7 action as a pioneer factor. This ability to open target sites in closed chromatin is required to implement a new gene expression program. Previous studies identified several tissue-specific Pax7 alternative splicing events that modify the paired domain with potentially distinct functions. The four Pax7 isoforms differ by the inclusion or exclusion of a single glutamine residue (Q108) and/or glycine and leucine residues (GL151/152). Here, we investigated the DNA binding and transactivation properties of four Pax7 isoforms (Q+GL+, Q+GL-, Q-GL+, Q-GL-). We found that the alternative splicing events of Pax7 have no major effect on the ability of these isoforms to bind Pax7 consensus motifs *in vitro* but that it may adjust the transactivation potential at different target sites.

Introduction

Pax7 is one of nine members of the paired-box (Pax) family of transcription factors (TF) involved in many aspects of mammalian development. Pax genes are master regulators that manifest a dynamic expression patterns throughout embryonic development in a wide variety of tissues and inappropriate expression of these genes was linked to murine and human diseases [1, 2]. As a pioneer TF, Pax7 exemplify the key role of pioneers in development, by its ability to interact with nucleosomal DNA and trigger chromatin remodeling at a repertoire of cell fate specific enhancers [3, 4]. Moreover, alternate transcripts of Pax family members due to alternative splicing increase the level of their regulatory function complexity during evolution [5, 6]. In fact, multiple transcript variants of the Pax genes exhibit differential expression patterns during organogenesis in a spatially and temporally restricted manner and this was recently reviewed by Thompson et al. [7]. Studies of the functional significance of Pax6 and Pax8 isoforms show that variants involving changes in the paired domain (PD) or the C-terminal region exhibit distinct DNA binding properties and transcriptional activity [8, 9]. The PD is a conserved DNA binding domain (DBD) of the Pax family, constituted by a bipartite structure of two helix-turn-helix (HTH) modules, namely the PAI and RED subdomains. Another HTH module, the homeodomain (HD), is present in the Pax4/6 and Pax3/7 family groups, whereas only a partial HD exists in the Pax2/5/8 group [10].

Alternate Pax7 transcripts have been identified in different systems. Pax7 expression is required for development of brain, skeletal muscle and pituitary gland [3, 11-13], and its expression was also observed in adult brain and testis [12, 14]. Although previous knockout and transfection studies testify to the critical role of Pax7 for specification of skeletal muscle satellite cells and a sub-population of dorsal mesencephalic neuron [13, 15-17], the role of the various Pax7 isoforms remains elusive. Pax7 transcripts are subject to two alternate splicing events, producing transcript variants that differ by the inclusion or exclusion of a trinucleotide or a hexanucleotide within the PD sequence. Consequently, the isoforms differ by the presence and/or absence of a single glutamine residue (Q108) or glycine and leucine residues (GL151/152). The Q108 is located in the linker region between PAI and RED subdomains preceding the first alpha helix of the RED subdomain ($\alpha 4$) and GL151/152 are located before the third helix of the RED subdomain ($\alpha 6$) (Fig. 1). These four Pax7 splice variants are detected

at different levels in murine tissues. For instance, in muscle cells, Pax7 Q+GL+ is most abundant followed by Pax7 Q+GL-, yet both Q- isoforms are also detected in this tissue, whereas in brain and pituitary, Pax7 Q+GL- is most highly expressed and Q- isoforms are not detected [18, 19].

In human, PAX7 isoform 3 harboring Q+GL+ is specifically expressed in brain, skeletal muscle and in testis [20-22]. An alternate in-frame splice site results in a short exon 8 and the inclusion of exon 9 with the conserved OAR (Otp, Aristaless, Rax) domain, also called paired tail. PAX7 isoforms 1 and 2, corresponding to Q+GL+ or Q+GL- forms, are transcript variants that exclude the last exon (exon 9 – encoding the OAR domain), resulting in a longer exon 8 with a unique C-terminus [21, 23]. Isoforms without exon 9 exist only in human and were detected at very low level in multiple tissues [20].

The OAR domain is a 13-16 a.a. motif identified in the C-terminal region of numerous proteins with a paired-type homeodomain, yet its precise function is still unclear. Inconsistencies exist in the role associated with the OAR domain. For example, studies on Orthopedia (Otp) showed that the paired tail plays a role in the positive transactivating function [24], whereas for Alx3/Cart1 [25] and Pitx2 [26], the OAR domain intrinsically inhibits DNA binding and transactivation. However, inhibition of Pitx2 by its OAR domain can be overcome by interaction with other TF, such as Pit1, to allow activation during development [26]. The OAR domain also inhibits the transcriptional activation by Prx2 and the extent of inhibition depends on the cell context, supporting a role in protein-protein interactions [27]. When overexpressed in neuroblastoma cells (SH-SY5Y cell line), PAX7 OAR- isoforms tend to accumulate and display high levels of protein [20]. Altogether, it suggests that the OAR domain attenuates or conceals the activity of the DBD and/or activation domain within the same protein.

Although the transcriptional regulatory role of Pax7 isoforms is largely unexplored, study of chemically induced differentiation of neurogenic or myogenic lineages suggested a role for Pax7 alternate transcripts in lineage determination [28]. Also, a study revealed that mouse Pax7 isoforms transfected in P19 cells (embryonic carcinoma cell line) have differential ability to regulate gene expression of the ciliary neurotrophic factor receptor (Cntfr) gene [29]. Moreover, a splice-site mutation causing expression of PAX7 isoforms 1 or 2 (OAR-) at the

expense of PAX7 isoform 3 (OAR+), correlates with a human disorder characterized by failure-to-thrive, hypotonia, and global neurodevelopmental delay [20].

Pax3 and Pax7 are homologous proteins that share 76% identity (372/488 a.a.), but only Pax7 has an OAR domain in its C terminus (Fig. 1). Both have a conserved octapeptide in the linker between PD and HD and their DBDs are very similar. Their PDs share 94% identity (120/128 a.a.), with divergent residues located in the extended linker region between PAI and RED and in the RED subdomain, and their HDs have 96% identity (58/60 a.a.). Pax3 and Pax7 have overlapping but largely non-redundant functions in muscle cell development [30]. In primary myoblasts cells, they show discrete preference in their target motifs: HD type motifs are enriched at Pax7 targets whereas PD motifs are enriched at Pax3 targets [31]. There is a cell context discrepancy in target motif enrichment that was revealed by the comparison of motif frequencies under Pax7 ChIP-seq peaks in myoblasts compared to pituitary cells where PD sites are more represented [10]. The composite HP.E0 motif (juxtaposed PD and HD motifs) is also enriched at Pax7 pioneer sites in AtT-20, a pituitary corticotrope cell model. Unlike Pax7, there is no evidence of pioneer activity for Pax3 reported so far.

We previously showed that Pax7 PD is crucial for high affinity interactions with composite DNA motif and both DBDs are essential for Pax7 pioneer action [32]. Thus, this raises the question of the functional significance of the Pax7 isoforms with variations in PD sequence. In this work, we characterized the DNA binding properties and transcriptional activity of Pax7 isoforms. We show that binding to various Pax7 target motifs involves primarily the PAI subdomain of the PD, and isoforms modifications in RED subdomain residues scarcely change DNA binding affinity *in vitro*. Pax7 isoforms exhibit different transcriptional activities that are not correlated to their DNA binding activity.

In order to understand the role of each Pax7 DBDs and subdomains, we generated AtT-20 cell lines expressing Flag-tagged Pax7 DBD deletions (Fig. 2A and 2B). We then assessed the DNA binding affinity of Pax7 DBD deletions with the different Pax7 target motifs by Electrophoretic Mobility-Shift Assay (EMSA) (Fig. 1C). Not surprisingly, PD is critical for DNA binding *in vitro* as previously demonstrated for other Pax proteins [33, 34]. Pax7 Δ PD loses completely the interaction not only with the PD probe (line 9), but also with the composite (line 3) and the HD type probes (lines 15, 21, 27). As expected, Pax7 Δ HD shows no interaction with any of the HD type probes (lines 18, 24 and 30), but retains a very weak binding to composite and PD probes (lines 6 and 12). These observations are consistent with previous studies demonstrating the interdependence of the two DBDs of Pax factors [35-37].

The Pax7 Δ PD mutant is unique as it forms dimers with the palindromic H.I2 probe (line 27), suggesting that the interaction between the PD and HD is critical for monomer binding to DNA, whereas the HD alone is involved in the dimerization interactions and DNA binding to H.I2. To be more precise, the interaction of HD with one or the other HTH module of the PD appears essential for binding of a Pax7 monomer to the HD probe H.I2 as both deletions of PD subdomains, Δ PAI and Δ RED, retain this ability (lines 28 and 29).

Regarding the PAI subdomain, the DNA binding profile is consistent with the structural studies indicating a primary role of the PAI third alpha helix for specific recognition of DNA bases within the motif TGAC [38, 39]. Indeed, the PAI subdomain is essential for the binding to the PD motif as well as for the binding to the composite motif (line 4 and 10). The interaction of Pax7 Δ PAI with HD probe is disrupted (line 16), while H.I-2 probes is substantially diminished (line 22), in agreement with cooperative interactions between the two DBDs through the second helix of the PAI subdomain regulating DNA binding by the HD [35, 40]. However, Pax7 Δ RED binds all probes with a similar affinity compared to full length Pax7 (lines 5, 11, 17, 23 and 29), indicating that the RED subdomain has virtually no contribution to DNA binding *in vitro*.

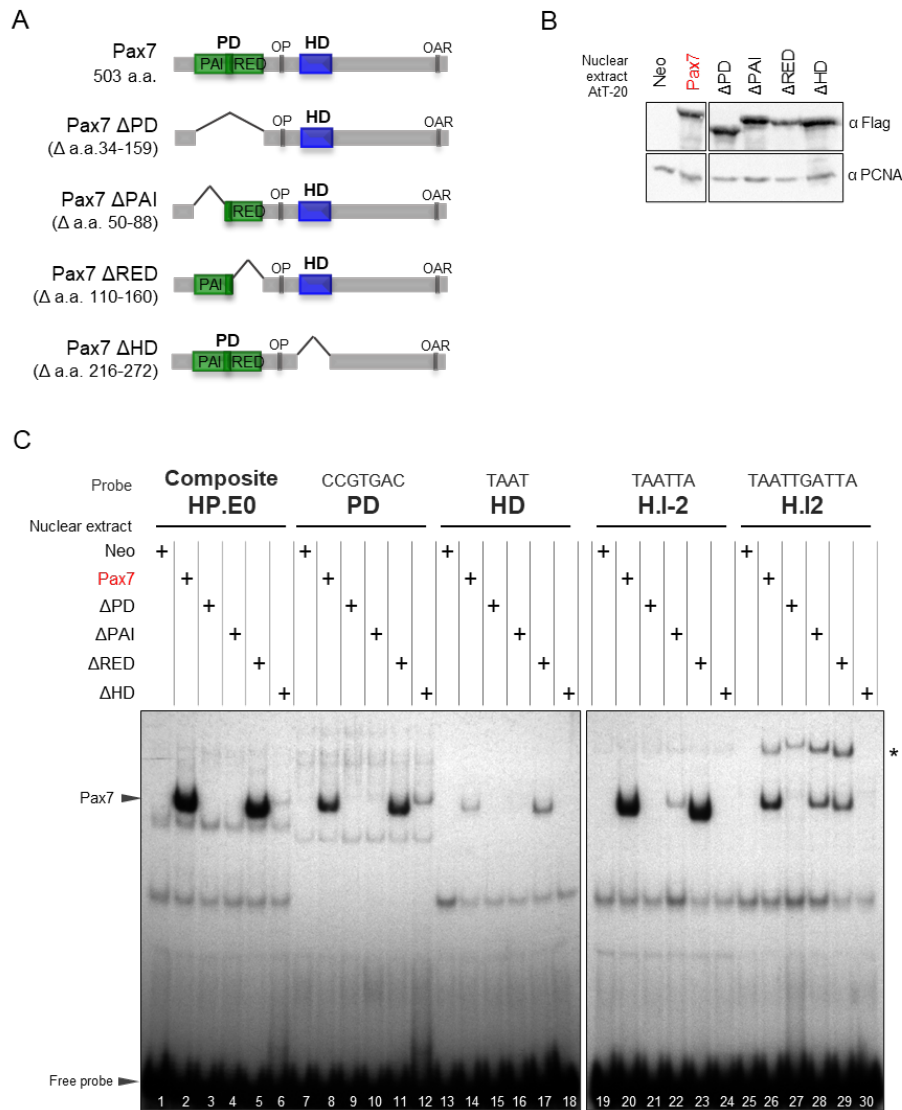


Figure 2. Contribution of HTH modules of Pax7 DBDs to DNA binding.

(A) Schematic representation of Pax7 DNA binding domains (DBD) deletions. DBDs illustrated are the paired domain (PD), including PAI and RED subdomain, and homeodomain (HD). Two other conserved domains are illustrated: OP (octapeptide) and OAR (named after otp, aristaless and rax).

(B) Western blot showing Pax7 protein expression in AtT-20 stable cell lines expressing Flag-tagged Pax7 deletions.

(C) EMSA was used to assess binding of Pax7 deletions to Pax7 target motifs *in vitro*. Nuclear extract from either control AtT-20 neo or AtT-20 stably expressing Pax7 full length and Pax7 deletions were incubated with each radioactive probe as indicated. Electrophoretic mobility of the Pax7-nucleic acids complex is indicated (arrowhead) and the Pax7 homodimer with the palindromic H.I2 probe shows an additional band (*).

DNA binding affinity of Pax7 isoforms

Recognition of DNA sites in heterochromatin and nucleosomes is the hallmark of pioneer factors. In a previous study, we showed that mutations of a single residue in PD (more precisely in the PAI subdomain) or HD affecting DNA binding of respective target motifs, has a deleterious impact on the recruitment to pioneer sites [32]. We therefore sought to investigate whether Pax7 isoforms have different DNA binding properties. To do so, we first generated AtT-20 cell lines expressing each of the Pax7 isoforms, in addition to its protein homologue Pax3 (Q+GL- OAR-) and a truncated Pax7 at the residue 465 that is deleted of its OAR domain (Q+GL- OAR-) (Fig. 3A and 3B). The expression of all Pax forms is equivalent except for Pax7 465 that seems rather stabilized (Fig. 3B). This phenomenon was also noted when various truncations and C-terminal isoforms of the human PAX7 were overexpressed in SH-SY5Y cells [20]. Proskorovski-Ohayon et al. observed by confocal microscopy and by western blot a significant accumulation of both PAX7 mutants lacking the OAR domain compared to PAX7 WT, suggesting a role of this domain for Pax7 protein stability, possibly by controlling the protein degradation rate.

We next investigated the impact of variations in Pax7 isoforms (Q and GL) in comparison with Pax3 and the truncated Pax7 465 on DNA binding. We thus performed EMSA experiments (Fig. 3C) to determine the effect of the various Pax7 isoforms on DNA binding to the different Pax7 consensus target motifs (probe sequences are shown in Fig. S1A). Considering the previous results (Fig. 2C), it is not surprising that all 4 Pax7 isoforms, which divergent residues located within the RED subdomain, display similar affinities for the various Pax7 motif probes. Nevertheless, subtle differences in affinity can be distinguished, notably between the Pax7 GL+ and GL- isoforms. Indeed, the Pax7 GL- forms bind all probes with slightly more affinity than GL+ forms including the composite motif (compare lanes 2 and 4 to 3 and 5), the PD motif (compare lanes 9 and 11 to 10 and 12), the HD motif (compare lanes 16 and 18 to 17 and 19), the H.I-2 motif (compare lanes 23 and 25 to 24 and 26) and the H.I2 motif (compare lanes 30 and 32 to 31 and 33). The ability to dimerize on the palindromic H.I2 probe is not affected by the various Pax7 isoforms (* in lanes 30-33), in agreement with the dimer binding of the different PD deletions Δ PD, Δ PAI and Δ RED (Fig. 1C, * in lanes 28-29).

PAI subdomain (probe e5) [41, 42]. However, Pax3 Q- isoforms shows increased affinity for PD sites with strong consensus elements interacting with both PAI and RED subdomains (probes P6CON and CD19-2/A). DNA binding properties of PAX7 isoforms with this P6CON site showed the lowest affinity associated with PAX7 Q+GL+ and Q-GL+, in agreement with our results, although PAX7Q-GL- has the strongest binding to P6CON site [42].

Even though Pax3 is a GL- form, its binding strength is similar to Pax7 GL+ isoforms on composite and PD probes (lanes 7 and 14) and shows decreased affinity on the three HD type probes (lanes 21, 28 and 35). The similarity of DNA binding affinity is not surprising since Pax3 and Pax7 DBDs share high sequence identity (Fig. 1). Moreover, the PAI subdomain that is primarily involved in DNA binding is identical in both Pax3 and Pax7 protein.

On the other hand, Pax7 465 lacking the OAR domain exhibits slightly stronger binding to all probes (lanes 6, 13, 20, 27 and 34), yet this could be due to the moderately higher accumulation of Pax7 465 protein in AtT-20 (Fig. 3B). This indicates that the OAR domain does not significantly affect Pax7 DNA binding properties. To summarize, the *in vitro* DNA binding properties of all the Pax forms is similar, but removal of glycine-leucine residues in the RED subdomain slightly enhances affinity for binding sites.

Transcriptional activity of Pax7 isoforms

In order to analyze transcriptional activity of Pax7 isoforms, we performed luciferase assays by transient transfections in AtT-20 cells (Fig. 4). In these experiments, increasing amounts of Pax effectors were cotransfected with a fixed amount of reporter plasmids containing a hexamer of the indicated Pax7 motif (sequence in Fig. S1A) or the natural 1.4 kb PC2 enhancer located at -146 kb of the PC2 gene TSS containing multiple Pax7 binding sites (Fig. S1B). For all reporters, Pax3 is a more potent transactivator compared to Pax7 isoforms (Fig. 4A-F). The difference in transcriptional activity contrasts with the DNA binding properties where Pax3 affinity to the various probes is like Pax7. This is possibly due to their C-terminal region or transactivation domain, that differ for most in their sequence (Fig. 1).

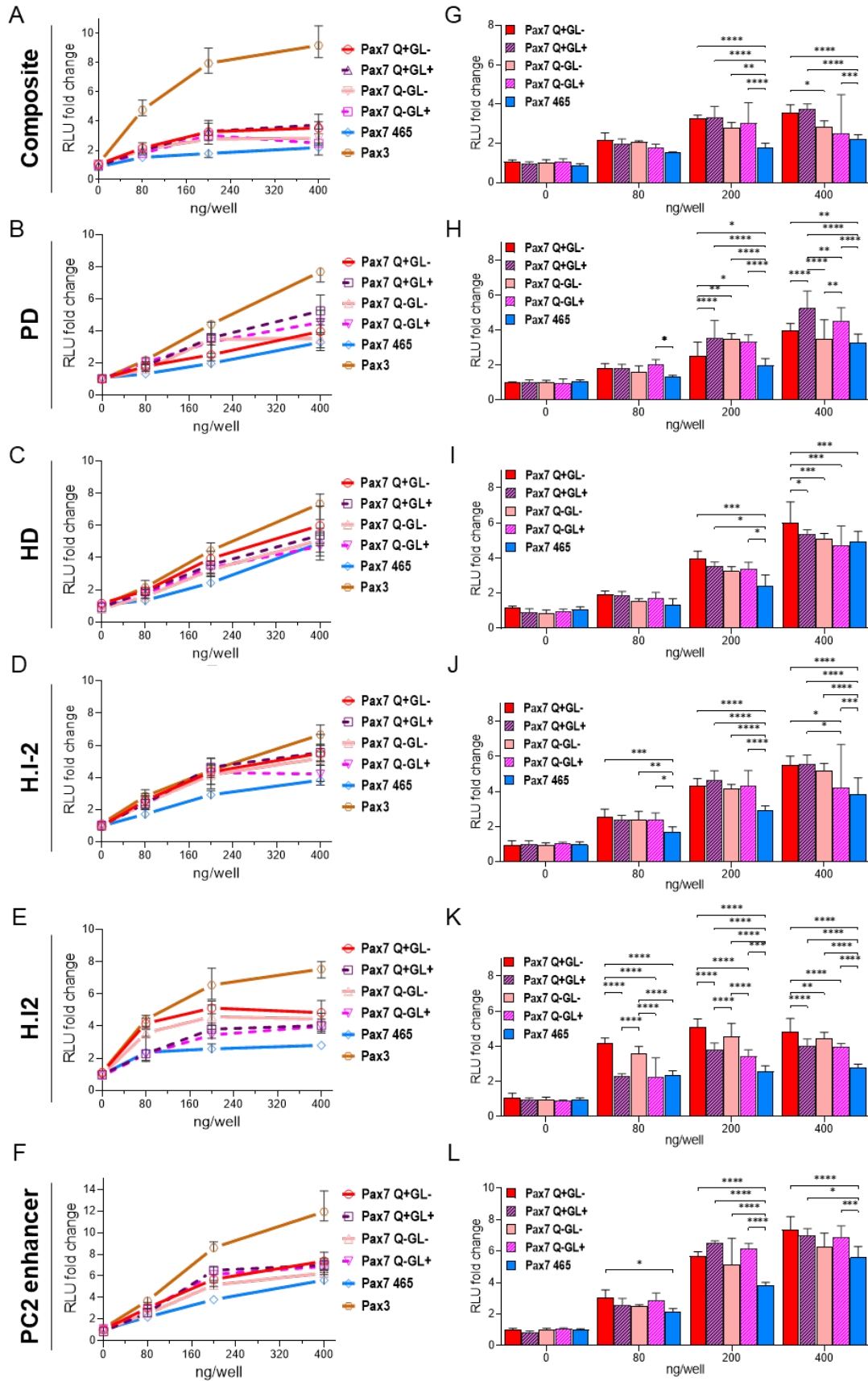


Figure 4. Transcriptional activity of Pax7 isoforms.

Transcriptional activation by Pax7 isoforms was assessed by cotransfection of expression plasmids together with luciferase reporter plasmids driven by six copies of the indicated motif or by the PC2 gene enhancer as described in [3]. Different amounts of effector plasmids DNA were tested (0, 80, 200, 400 ng/well). For each reporter plasmid, all the data are relative to basal activity (without effector plasmid). Assays were done in triplicate for the following number of experiments: Composite n=4, PD n=5, HD n=5, H.I-2 n=6, H.I2 n=5, PC2 enhancer n=3.

(A-F) Dose-response curve where dots correspond to median value of RLU fold change whereas error bars indicate interquartile range.

(G-L) Bar graph of median fold change with interquartile range (error bars) where statistical significance was calculated by two-way ANOVA (Analysis of Variance) and Tukey's multiple comparisons test.

Yet, it is not due to the absence of OAR domain in Pax3. In contrast to studies involving OAR in inhibition of transcription for Prx2, Pitx2 and Alx3/Cart1 proteins [25-27], the Pax7 465 (OAR-) shows the weakest transcriptional activity for all target motifs. As it was suggested by Underhill, the OAR motif could affect the adjacent proline-serine-threonine-rich region in the C-terminus of Pax7 [43]: the truncation of the last 38 residues could thus disrupt the transactivation domain.

At first sight, the dose-response curves of the four Pax7 isoforms (Q+/-, GL+/-) are very similar, especially with the composite, HD and H.I-2, but also the PC2 enhancer (Fig. 4A, 4C, 4D and 4F). Otherwise, subtle but significant differences are observed at high dose (400ng) with the PD showing a greater activation by the GL+ isoforms (Fig. 4B and 4H), whereas GL- isoforms appear transcriptionally dominant at low dose (80ng) with H.I2 (Fig. 4E and 4K). Pax7 Q+GL- shows stronger activation at HD and H.I2 sites (Fig. 4C and 4I; 4E and 4K). For Pax7, alternate isoforms differentially regulate expression of the ciliary neurotrophic factor receptor (Cntfr) in P19 cells [29]. Pax7 Q+GL+ is a more potent transactivator and Pax7 Q-GL- activate the least Cntfr gene expression where multiple PD motifs are found at Pax7 CHIP site. Our data differ from study on Pax3 Q+/- isoforms showing an increased transactivation potential for Pax3 Q- on a PD binding site (CD19-2/A), supported by their finding of enhanced DNA binding by the Q- isoform *in vitro* [41].

The study of the functional significance of Pax7 isoforms was conducted with consensus motifs for Pax7 Q+GL- DNA binding and perhaps other isoforms have different motif

preference or sensibility for suboptimal binding sites. For example, DNA binding affinity of Pax3 Q^{+/-} isoforms was dependent on the length and optimal bases of PD-type targets [41]. This idea was taken further suggesting that the Q insertion influence the relative binding affinity of the PAI and RED, whereas GL insertion was predicted to affect the binding specificity [18]. Enhanced recognition of degenerated motifs may be an important feature *in vivo* to generate a greater diversity of DNA binding specificity of Pax7 and to recognize a broader range of target sites. The capacity of Pax7 isoforms to stimulate transcription would allow for fine tuning regulation of gene expression. Alternate Pax7 isoforms may also bind with different affinity to discrete downstream target genes on the basis of unique protein:protein interactions, thus possess differential functions.

Materials and methods

Cell culture. AtT-20/D16v-F2 cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and antibiotics (penicillin/streptomycin). To generate stable transgenic AtT-20 Pax7 deletions (Δ PD, Δ PAI, Δ RED and Δ HD) and Pax isoforms (Q+GL+, Q+GL-, Q-GL+, Q-GL-, Pax7 465 and Pax3) cell populations, retroviruses were packaged using the Platinum-E Retroviral Packaging Cell Line (Cell Biolabs, catalog #RV-101) and infections performed as described [44]. Selection of retrovirus-infected cell populations was achieved with 400 μ g/mL Geneticin (Gibco, #11811-031). Resistant colonies were pooled to generate populations of thousands of independent colonies.

Electrophoretic mobility shift assays (EMSA). Probe labeling. Double-stranded DNA probes were prepared by annealing oligonucleotides (1 μ M of each strand: see sequences in the Supplementary Figure 1A) as previously described [32]. Nuclear extracts and Electrophoresis were performed as previously described [32].

Luciferase assays. AtT20 cells were plated and transfected as described in [32].

References

1. Dahl, E., H. Koseki, and R. Balling, Pax genes and organogenesis. *Bioessays*, 1997. 19(9): p. 755-65.
2. Tremblay, P. and P. Gruss, Pax: genes for mice and men. *Pharmacol Ther*, 1994. 61(1-2): p. 205-26.
3. Budry, L., et al., The selector gene Pax7 dictates alternate pituitary cell fates through its pioneer action on chromatin remodeling. *Genes Dev*, 2012. 26(20): p. 2299-310.
4. Mayran, A., et al., Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. *Nat Genet*, 2018. 50(2): p. 259-269.
5. Fabian, P., et al., Pax2/5/8 and Pax6 alternative splicing events in basal chordates and vertebrates: a focus on paired box domain. *Front Genet*, 2015. 6: p. 228.
6. Short, S. and L.Z. Holland, The evolution of alternative splicing in the Pax family: the view from the Basal chordate amphioxus. *J Mol Evol*, 2008. 66(6): p. 605-20.
7. Thompson, B., et al., Overview of PAX gene family: analysis of human tissue-specific variant expression and involvement in human disease. *Hum Genet*, 2021. 140(3): p. 381-400.
8. Poleev, A., et al., Distinct functional properties of three human paired-box-protein, PAX8, isoforms generated by alternative splicing in thyroid, kidney and Wilms' tumors. *Eur J Biochem*, 1995. 228(3): p. 899-911.
9. Kozmik, Z., T. Czerny, and M. Busslinger, Alternatively spliced insertions in the paired domain restrict the DNA sequence specificity of Pax6 and Pax8. *Embo j*, 1997. 16(22): p. 6793-803.
10. Mayran, A., A. Pelletier, and J. Drouin, Pax factors in transcription and epigenetic remodelling. *Semin Cell Dev Biol*, 2015. 44: p. 135-44.
11. Relaix, F., et al., A Pax3/Pax7-dependent population of skeletal muscle progenitor cells. *Nature*, 2005. 435(7044): p. 948-53.
12. Stoykova, A. and P. Gruss, Roles of Pax-genes in developing and adult brain as suggested by expression patterns. *J Neurosci*, 1994. 14(3 Pt 2): p. 1395-412.
13. Jostes, B., C. Walther, and P. Gruss, The murine paired box gene, Pax7, is expressed specifically during the development of the nervous and muscular system. *Mech Dev*, 1990. 33(1): p. 27-37.
14. Aloisio, G.M., et al., PAX7 expression defines germline stem cells in the adult testis. *J Clin Invest*, 2014. 124(9): p. 3929-44.
15. Seale, P., et al., Pax7 is required for the specification of myogenic satellite cells. *Cell*, 2000. 102(6): p. 777-86.
16. Matsunaga, E., I. Araki, and H. Nakamura, Role of Pax3/7 in the tectum regionalization. *Development*, 2001. 128(20): p. 4069-77.
17. Thompson, J.A., et al., Pax7 is requisite for maintenance of a subpopulation of superior collicular neurons and shows a diverging expression pattern to Pax3 during superior collicular development. *BMC Dev Biol*, 2008. 8: p. 62.

18. Ziman, M.R. and P.H. Kay, Differential expression of four alternate Pax7 paired box transcripts is influenced by organ- and strain-specific factors in adult mice. *Gene*, 1998. 217(1-2): p. 77-81.
19. Ziman, M.R., S. Fletcher, and P.H. Kay, Alternate Pax7 transcripts are expressed specifically in skeletal muscle, brain and other organs of adult mice. *Int J Biochem Cell Biol*, 1997. 29(7): p. 1029-36.
20. Proskorovski-Ohayon, R., et al., PAX7 mutation in a syndrome of failure to thrive, hypotonia, and global neurodevelopmental delay. *Hum Mutat*, 2017. 38(12): p. 1671-1683.
21. Brown, G.R., et al., Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res*, 2015. 43(Database issue): p. D36-42.
22. Banerji, C.R.S., et al., PAX7 target genes are globally repressed in facioscapulohumeral muscular dystrophy skeletal muscle. *Nature communications*, 2017. 8(1): p. 2152-2152.
23. Schäfer, B.W., et al., Molecular cloning and characterization of a human PAX-7 cDNA expressed in normal and neoplastic myocytes. *Nucleic Acids Res*, 1994. 22(22): p. 4574-82.
24. Simeone, A., et al., Orthopedia, a novel homeobox-containing gene expressed in the developing CNS of both mouse and Drosophila. *Neuron*, 1994. 13(1): p. 83-101.
25. Brouwer, A., et al., The OAR/aristaless domain of the homeodomain protein Cart1 has an attenuating role *in vivo*. *Mechanisms of Development*, 2003. 120(2): p. 241-252.
26. Amendt, B.A., L.B. Sutherland, and A.F. Russo, Multifunctional role of the Pitx2 homeodomain protein C-terminal tail. *Mol Cell Biol*, 1999. 19(10): p. 7001-10.
27. Norris, R.A. and M.J. Kern, Identification of domains mediating transcription activation, repression, and inhibition in the paired-related homeobox protein, Prx2 (S8). *DNA and cell biology*, 2001. 20(2): p. 89-99.
28. Ziman, M., et al., A Key Role for Pax7 Transcripts in Determination of Muscle and Nerve Cells. *Experimental cell research*, 2001. 268: p. 220-9.
29. White, R.B. and M.R. Ziman, Genome-wide discovery of Pax7 target genes during development. *Physiol Genomics*, 2008. 33(1): p. 41-9.
30. Relaix, F.d.r., et al., Pax3 and Pax7 have distinct and overlapping functions in adult muscle progenitor cells. *Journal of Cell Biology*, 2005. 172(1): p. 91-102.
31. Soleimani, V.D., et al., Transcriptional dominance of Pax7 in adult myogenesis is due to high-affinity recognition of homeodomain motifs. *Dev Cell*, 2012. 22(6): p. 1208-20.
32. Pelletier, A., et al., Pax7 pioneer factor action requires both paired and homeo DNA binding domains. *Nucleic Acids Res*, 2021. 49(13): p. 7424-7436.
33. Czerny, T., G. Schaffner, and M. Busslinger, DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site. *Genes Dev*, 1993. 7(10): p. 2048-61.
34. Treisman, J., E. Harris, and C. Desplan, The paired box encodes a second DNA-binding domain in the paired homeo domain protein. *Genes Dev*, 1991. 5(4): p. 594-604.
35. Apuzzo, S. and P. Gros, Cooperative interactions between the two DNA binding domains of Pax3: helix 2 of the paired domain is in the proximity of the amino terminus of the homeodomain. *Biochemistry*, 2007. 46(11): p. 2984-93.

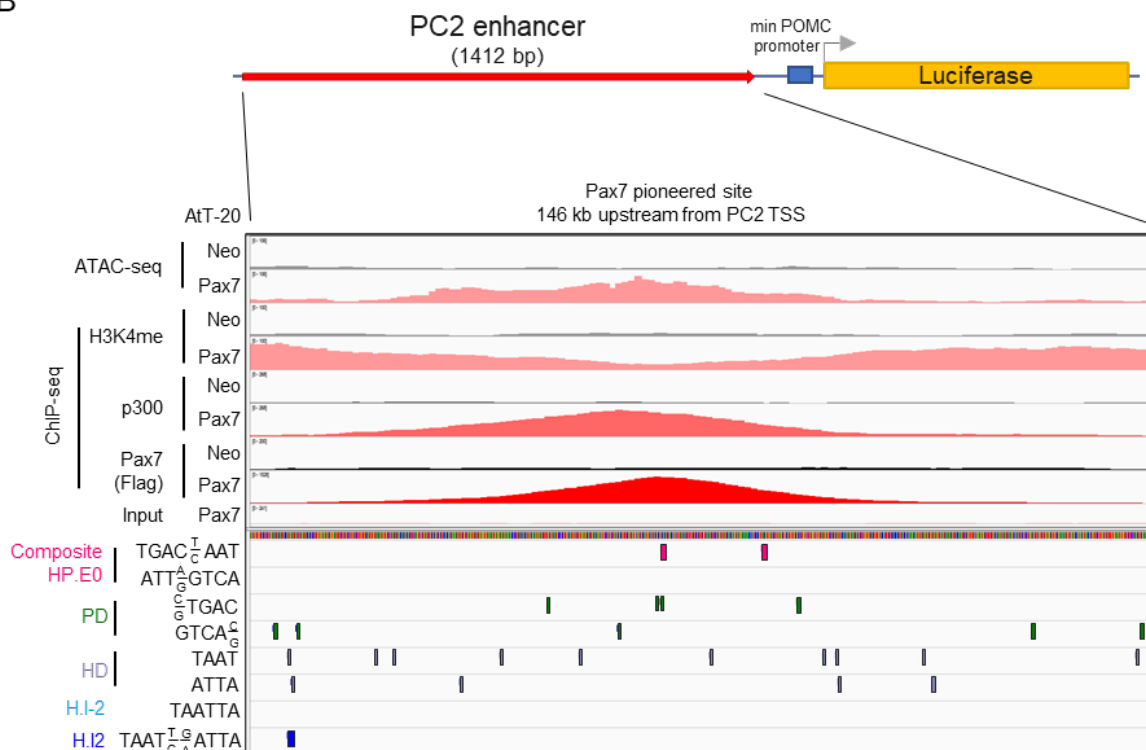
36. Underhill, D.A. and P. Gros, The paired-domain regulates DNA binding by the homeodomain within the intact Pax-3 protein. *J Biol Chem*, 1997. 272(22): p. 14175-82.
37. Jun, S. and C. Desplan, Cooperative interactions between paired domain and homeodomain. *Development*, 1996. 122(9): p. 2639-50.
38. Xu, H.E., et al., Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev*, 1999. 13(10): p. 1263-75.
39. Wintjens, R. and M. Rooman, Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. *J Mol Biol*, 1996. 262(2): p. 294-313.
40. Fortin, A.S., D.A. Underhill, and P. Gros, Helix 2 of the paired domain plays a key role in the regulation of DNA-binding by the Pax-3 homeodomain. *Nucleic Acids Res*, 1998. 26(20): p. 4574-81.
41. Vogan, K.J., D.A. Underhill, and P. Gros, An alternative splicing event in the Pax-3 paired domain identifies the linker region as a key determinant of paired domain DNA-binding activity. *Mol Cell Biol*, 1996. 16(12): p. 6677-86.
42. Du, S., et al., Co-expression of alternatively spliced forms of PAX3, PAX7, PAX3-FKHR and PAX7-FKHR with distinct DNA binding and transactivation properties in rhabdomyosarcoma. *Int J Cancer*, 2005. 115(1): p. 85-92.
43. Underhill, D.A., PAX proteins and fables of their reconstruction. *Crit Rev Eukaryot Gene Expr*, 2012. 22(2): p. 161-77.
44. Rambaud, J., et al., TIF1beta/KAP-1 is a coactivator of the orphan nuclear receptor NGFI-B/Nur77. *J. Biol. Chem*, 2009. 284(21): p. 14147-14156.
45. Birrane, G., A. Soni, and J.A. Ladias, Structural basis for DNA recognition by the human PAX3 homeodomain. *Biochemistry*, 2009. 48(6): p. 1148-55.

Figure supplémentaire

A

Probe sequences	
Composite HP.E0	TCGACGCAGAT <u>CCGTGACTAAT</u> GTTCCGACCTCGA
PD (mutated HD)	TCGACGCAGAT <u>CCGTGAC</u> GCCAGTTCGACCTCGA
HD (mutated PD)	TCGACGCAGATCGTCCGC <u>TAAT</u> TCCAACGACCTCGA
H.I-2	TCGACCAGATCGTCCGC <u>TAATTA</u> CCAACGACCTCGA
H.I2	TCGACTCGTCCGC <u>TAATTGATTA</u> TCCAACGACCTCGA

B



Supplementary Figure 1. Pax7 target probes and natural PC2 enhancer.

A) DNA sequences of oligonucleotides used for probes in EMSA and hexamer reporter in luciferase assay.
 B) Schematic representation of the natural PC2 enhancer reporter plasmid and overview of the pioneered PC2 enhancer locus in AtT-20 cells. AtT-20 epigenetic status before and after Pax7 pioneer action revealed by chromatin opening (ATAC-seq) and genomic recruitment of Pax7 (Flag-Pax7 ChIP-seq) and active marks (H3K4me and p300 ChIP-seq). Pax7 motifs occurrence within the enhancer sequence used in luciferase reporter including composite (TGACYAAT), PD (STGAC), HD (TAAT), H.I-2 (TAATTA) and H.I2 (TAATYRATTA).

Chapitre 4 : Discussion

4.1 Les modalités de liaison à l'ADN du facteur pionnier Pax7

4.1.1 L'affinité différentielle de Pax7 à ses séquences cibles

Pax7 possède un éventail de possibilité d'interactions avec l'ADN à travers divers motifs cibles d'affinité différentes. Tandis que les motifs canoniques PD (TGAC) et HD (TAAT) ainsi que le palindrome H.I2 (TAAT(N)₂ATTA) (Czerny et al., 1993; D. Wilson et al., 1993) avaient déjà été étudiés, le motif composite (TGACTAAT) enrichi aux sites pionniers de Pax7 a été identifié plus récemment (Budry et al., 2012). Mes travaux ont permis de mieux comprendre la nature des interactions de Pax7 avec la diversité de motifs de séquence et le rôle des deux domaines de liaison à l'ADN de Pax7 pour l'action pionnière.

Les expériences de liaison à l'ADN *in vitro* ont montré que l'interaction de Pax7 avec le site composite est d'affinité supérieure aux autres séquences cibles et bien que ce motif soit formé de deux sites canoniques, soit PD et HD juxtaposés, c'est sous forme de monomère que Pax7 s'y lie (Chapitre 2 Fig.1 et 2). Comme Pax7 possède la capacité de dimériser sur le palindrome HD, il est possible que différents espacements entre les deux sites (PD et HD) permettent une interaction dimérique, mais de tels sites n'ont pas été trouvés. En fait, il n'y a pas d'espacement spécifique entre les différents motifs qui sont ressortis de nos analyses. En somme, cela suggère qu'une conformation particulière de la protéine est impliquée pour permettre aux deux domaines de liaison à l'ADN d'interagir avec leur site sur l'ADN (Chapitre 2 Fig.5). La structure tertiaire de Pax7 n'est pas déterminée et les seules références qui existent actuellement sont le PD de Pax5 et PAX6 ainsi que le HD de Pax3 (Birrane et al., 2009; Garvie et al., 2001; H. E. Xu et al., 1999). Les difficultés de produire une structure cristallographique de Pax7 est probablement parce qu'il s'agit d'une protéine qui possède une forte plasticité. Il serait toutefois pertinent de pouvoir comparer les différentes configurations putatives de Pax7 associée à l'ADN de chaque motif cible. Ces modes de liaison à l'ADN distincts devraient déterminer quelles surfaces sont disponibles pour les interactions protéine-protéine. Différentes conformations de Pax7 pourraient permettre des interactions distinctes avec des co-régulateurs

et, par conséquent, sont susceptibles d'influencer l'expression de gènes cibles de différentes manières.

L'utilisation de peptides Pax7 correspondant aux domaines de liaison à l'ADN individuels ainsi que les délétions de chacun des modules HTH de Pax7 a permis de mettre en lumière l'interaction dominante du PD avec l'ADN en comparaison au HD (Chapitre 2 Fig.2 et Chapitre 3 Fig.1). C'est plus précisément le sous-domaine PAI du PD qui lie fortement les différentes cibles étudiées. L'interaction plus efficace avec l'ADN par le PD avait été observé aussi pour Pax6 (Czerny & Busslinger, 1995) et Pax3 qui lie aussi les séquences PD principalement avec le PAI (Vogan et al., 1996). Des études analysant plusieurs mutations du PD de Pax3, notamment dans le sous-domaine PAI et les régions des brins beta, ont indiqué qu'une série de résidus qui font des contacts avec le squelette phosphate d'ADN sont nécessaires pour permettre la liaison à l'ADN par le HD (Fortin, Underhill, & Gros, 1997; Underhill & Gros, 1997). Cependant, les études sur Pax3 ont montré que les sites PD allongés de type P6CON et CD19-2/A contiennent des séquences consensus pour les deux sous-domaines PAI et RED (Vogan et al., 1996). De telles séquences n'ont pas été retrouvées dans les recherches de motifs aux cibles de Pax7, cela suggère que ce n'est pas un mode de liaison à l'ADN privilégié par Pax7. Malgré tout, il serait intéressant d'évaluer si la délétion du sous-domaine RED, qui n'a aucun impact sur l'affinité de liaison *in vitro*, influence la capacité de Pax7 à être recruté aux différentes cibles et se stabiliser sur des sites nucléosomaux. Les protéines Pax utilisent plusieurs mécanismes pour la reconnaissance de l'ADN comme en témoigne la diversité de cibles pouvant adopter un certain nombre d'orientations différentes les uns par rapport aux autres.

4.1.2 Les deux domaines de liaison à l'ADN intact sont essentiels à l'action pionnière de Pax7

Les expériences de liaison *in vitro* par les peptides, les mutations ponctuelles et les délétions des domaines de liaison à l'ADN de Pax7 supportent la notion d'interdépendance des deux domaines de liaison à l'ADN de Pax7 (Chapitre 2 Fig. 2 et S3; Chapitre 3 Fig.1). Pour les peptides, les deux domaines de liaison à l'ADN doivent se retrouver dans la même molécule pour avoir une liaison efficace à l'ADN. La mutation d'un des domaines de liaison à l'ADN de

Pax7 influence non seulement la liaison du domaine de liaison à l'ADN ciblé mais aussi l'interaction de l'autre domaine de liaison à l'ADN avec l'ADN. Cela est aussi le phénomène observé avec les délétions, à l'exception de la délétion du sous-domaine RED qui n'a pas d'effet sur la liaison *in vitro*. Les mutations au PD (G43A et R56L) affectent spécifiquement la liaison aux séquences PD et diminuent l'affinité aux séquences HD, tandis que la mutation au HD (S264E) perturbe la liaison aux séquences HD seulement (Chapitre 2 Fig. S3). Dans les expériences d'essais luciférase, bien que les trois mutations aux domaines de liaison à l'ADN de Pax7 ne changent pas la liaison au site composite *in vitro*, Pax7 R56L semble perdre son potentiel de transactivation au site composite et à l'enhancer PC2. Cela suggère que la mutation R56L a un impact autre que simplement sur la liaison à l'ADN, que ce soit sur la structure de Pax7 ou les interactions protéine-protéine avec des co-facteurs. De plus, Pax7 G43A et Pax7 S264E voient une diminution de l'activité transcriptionnelle à l'enhancer PC2, mais conserve l'habileté d'activer la transcription.

L'impact le plus marquant des mutations est observable par leur recrutement dans les cellules AtT-20 (Chapitre 2 Fig.4). D'abord, ce sont les cibles dans la chromatine fermée qui sont le plus affectées par les mutations aux domaines de liaison à l'ADN, c'est-à-dire les sites pionniers, *primed* et résistants. Au contraire, les cibles constitutives dans la chromatine ouverte où des interactions protéine-protéine peuvent être suffisantes au recrutement sont conservées. En général, les motifs enrichis pour chacun des mutants correspondent à ce que l'on observe *in vitro* et qui a été décrit précédemment, à la différence que les mutants PD perdent le recrutement au séquence composite. Les préférences de motifs sont éloquentes aux sites activés où l'accessibilité à l'ADN est réduite dans les AtT-20. L'étude du recrutement et de l'ouverture de la chromatine dans les lignées exprimant les mutants Pax7 montrent que les deux domaines de liaison à l'ADN intacts sont essentiels à l'action pionnière de Pax7. Mis ensemble, ces résultats sont compatibles avec un modèle d'interactions coopératives entre les deux domaines de liaison à l'ADN, mais dont les déterminants structuraux sont encore peu compris.

Certains facteurs pionniers tel que FoxA et Sox2 ont la capacité de favoriser une flexion de l'axe de l'hélice d'ADN lors de la liaison (Cirillo & Zaret, 2007; Reményi et al., 2003). Cette caractéristique participerait à l'ouverture partielle de la chromatine afin de faciliter le recrutement d'autres facteurs de transcription. Des études de permutation circulaire permettant

d'évaluer les changements conformationnels de l'ADN résultant de la liaison *in vitro*, ont révélé que le PD des protéines Pax a la capacité d'induire une courbure à l'ADN (Chalepakis, Wijnholds, & Gruss, 1994). Pax7 ne faisant pas partie des membres de la famille Pax testés dans cette étude, il serait pertinent d'investiguer en ce sens, afin de mieux comprendre les mécanismes de liaison et de stabilisation à l'ADN nucléosomal.

4.2 Les conditions et obstacles à l'action pionnière de Pax7

4.2.1 Les sites pionniers sont enrichis de motifs de forte affinité

Ce travail apporte des éléments de réponse aux questions entourant la propriété unique des facteurs pionniers de reconnaître leur cible dans la chromatine fermée. La sélection de sites d'ADN cibles dans un environnement réfractaire à la plupart des facteurs de transcription est multifactoriel. Au début de ce projet, Budry et al. avaient découvert la capacité de Pax7 à reprogrammer les cellules AtT-20 corticotropes en mélanotropes par son action pionnière, et le nouveau motif composite, enrichi aux sites pionniers de Pax7 (Budry et al., 2012). L'analyse combinée des données génomiques dans ce modèle a permis de mettre en évidence deux modes d'action de Pax7: pour certaines cibles, il agit comme facteur de transcription classique et pour d'autres, comme facteur de transcription pionnier.

Tout d'abord, en comparant les différentes catégories de sites Pax7 (constitutifs, activés, pionniers), les sites pionniers montrent un plus grand nombre de motifs Pax7, sont enrichis de tous les motifs, mais surtout des motifs de haute affinité (composite) et le recrutement de Pax7 y est plus robuste (Chapitre 2 Fig.1). De plus, les expériences avec les peptides et les mutations ponctuelles des domaines de liaison à l'ADN de Pax7 montrent qu'une forte interaction avec l'ADN est essentielle à l'action pionnière de Pax7, ce qui nécessite l'implication de ses deux domaines de liaison à l'ADN, *paired* et *homeo* (Chapitre 2 Fig.2 et Fig.4). L'aspect du nombre de motifs aux sites pionniers est intéressant car cela augmente la probabilité que Pax7 reconnaisse une séquence cible sur l'ADN enroulé sur le nucléosome et puisse s'y lier de façon coopérative. En plus de cela, des simulations de dynamiques moléculaires des nucléosomes sur différentes séquences d'ADN ont suggéré que les nucléosomes avec plusieurs sites de liaison de facteur de transcription sont plus mobiles et moins compact, ce qui faciliterait la liaison de facteurs pionniers (Huertas et al., 2020).

L'enrichissement de motifs composites de haute affinité aux sites pionniers de Pax7, contraste avec les facteurs pionniers FoxA, Oct4, Sox2 et Klf4 qui reconnaissent des sites partiels ou dégénérés sur l'ADN nucléosomal (Meers et al., 2019). Ces observations ne sont pas incompatibles car lorsqu'on observe le recrutement de Pax7 dans un système inductible, la liaison initiale aux sites pionniers est d'abord faible (30 minutes) puis se stabilise après 24h

(Mayran et al., 2018). Les étapes initiales de reconnaissance des séquences sur l'ADN nucléosomal pourraient impliquer l'interaction partielle des domaines de liaison à l'ADN de Pax7, mais une stabilisation est nécessaire pour l'ouverture de la chromatine. Cela est en accord avec l'idée que les facteurs pionniers tolèrent des interactions protéine-ADN réduites lors de la reconnaissance initiale sur le nucléosome (Soufi et al., 2015). Ainsi, les motifs de séquence exposés à la surface du nucléosome pourraient être reconnus par les domaines de liaison à l'ADN de Pax7 possédant des hélices de reconnaissance courte, mais pas par les facteurs non-pionniers (Fernandez Garcia et al., 2019). Comme Pax7 possède deux domaines de liaison à l'ADN dont trois modules HTH, et se lie à une diversité de motifs cible, cela suggère qu'il peut avoir une adaptabilité inhérente des domaines de liaison à l'ADN à la reconnaissance du nucléosome. Ceci est non négligeable puisque l'orientation et la conformation de l'ADN sur le nucléosome peut être favorable ou non à la reconnaissance d'un motif cible par le facteur pionnier.

4.2.2 L'environnement chromatinien aux sites résistants

À l'aide du système Pax7 inductible, nous avons déterminé que la vitesse à laquelle se produit l'ouverture de la chromatine est proportionnelle à la force de liaison de Pax7 aux cibles pionnières (Mayran et al., 2018). La gradation de l'action pionnière vers les sites *primed* puis résistants (où la liaison de Pax7 n'induit pas de changement épigénétique) est aussi proportionnelle à la force de liaison de Pax7. Il est important de noter que le recrutement initial de Pax7 aux sites dans la chromatine fermée, autant aux sites pionniers, *primed* et résistants, se fait dans les 30 minutes suivant l'activation. Cette liaison faible aux cibles contenues dans le nucléosome représente la capacité de Pax7 de sonder les régions de chromatine condensée. Contrairement aux sites pionniers et *primed*, où les deux domaines de liaison à l'ADN sont essentiels au recrutement de Pax7, la liaison du mutant HD aux cibles résistantes a montré que le PD est plus important pour cette interaction (Chapitre 2 Fig.4). En revanche, le nombre et la nature des motifs n'expliquent pas le fait que les sites résistants ne sont pas liés plus fortement et ne subissent pas d'ouverture de la chromatine ou de dépôt de marques actives (Chapitre 2 Fig.1). Cela suggère que les mécanismes de stabilisation de la liaison sont distincts, soit que la liaison par le PD mène à une conformation de Pax7 qui permet l'interaction avec d'autres protéines à ces sites.

Les sites résistants sont globalement enrichis en marques d'histones associées à l'hétérochromatine tel que H3K9me3, mais présentent des niveaux similaires de H3K9me2 comparativement aux sites pionniers (Mayran et al., 2018). Les régions d'hétérochromatine marquées de H3K9me3 représenteraient donc un obstacle au recrutement ou à la stabilisation de Pax7. En revanche, la méthylation des cytosines de l'ADN qui est présente aux sites fermés avant l'action pionnière de Pax7 n'interfère pas avec la liaison à l'ADN de Pax7 *in vitro* (Chapitre 2 Fig. 3). On trouve quelques exemples de blocage de l'action pionnière par la marque H3K9me3, par exemple lors de l'induction de l'expression ectopique de FOXA2 dans les fibroblastes ou les sites FOXA2 enrichis sont déplétés de la marque H3K9me3 (Donaghey et al., 2018). La marque H3K9me3 a également été observée lors de la reprogrammation des fibroblastes humains en cellule IPS, où de grandes régions de la taille de mégabases de chromatine bloquent la liaison des facteurs de pluripotence OSKM à l'ADN (Soufi et al., 2012). Or, ce sont des régions que les facteurs OSKM occupent normalement dans les cellules ES et abritent des gènes nécessaires aux stades avancés de la reprogrammation cellulaire vers la pluripotence. Dans ce cas, H3K9me3 empêche la liaison initiale des facteurs OSKM, mais ceux-ci finissent par surmonter cet obstacle dans le processus de reprogrammation.

Cependant, dans le modèle hypophysaire, Pax7 est recruté aux sites résistants présentant la marque H3K9me3, mais faiblement et il n'y a pas de changement épigénétique subséquent. Comme les facteurs Oct4 et Sox2 lient typiquement leurs sites près de l'entrée/sortie du nucléosome et de la dyade du nucléosome respectivement, donc près des histones H3 et H4 (Dodonova et al., 2020; Echigoya et al., 2020), il est ainsi possible que la modification de l'histone H3 et l'architecture des nucléosomes condensés perturbent physiquement leur recrutement. Il serait intéressant d'explorer le recrutement de Pax7 avec différentes séquences sur le nucléosome pour évaluer le positionnement et les mécanismes moléculaires par lesquels Pax7 interagit avec son ADN cible dans la chromatine.

Les sites résistants ne sont pas associés avec quelques sous-groupes de gènes distinctifs. Par exemple, les sites myogéniques-spécifiques de Pax7 sont complètement absents des cibles dans les cellules hypophysaires; ce qui suggère que l'environnement chromatinien aux sites spécifiques de Pax7 dans les cellules musculaires présente des particularités qui empêchent même le balayage (*scanning*) des séquences par Pax7. Cela pourrait correspondre, par exemple,

à des régions distinctives d'hétérochromatine associées à la protéine HP1 α qui conduit à la séparation de phase (Larson et al., 2017). Ce modèle de liaison cellule-spécifique suggère que les facteurs pionniers peuvent être activement exclus de certains domaines. Malgré la capacité des facteurs pionniers à reconnaître leurs cibles sur le nucléosome, il est clair qu'ils n'occupent pas tous leurs sites ou motifs cibles apparentés dans une cellule. Autrement dit, Pax7 ne régule pas le même ensemble de gènes dans les différents types cellulaires où il est exprimé (crête neurale, cerveau, muscle squelettique, hypophyse); c'est-à-dire qu'ils présentent une liaison à l'ADN spécifique aux cellules. Ainsi, cela peut être expliqué par des caractéristiques répressives de l'environnement chromatinien en place, y compris le compactage d'ordre supérieur imposé par l'hétérochromatine, et par la présence ou l'absence de cofacteurs spécifiques à une cellule.

4.2.3 Coopération entre facteurs pionniers et non-pionniers

Les exemples de facteurs qui possèdent une action pionnière dans différents modèles ont soulignés l'importance de la coopération et souvent la synergie de plusieurs facteurs de transcription, pionniers et non-pionniers. Tandis que seul le facteur pionnier peut reconnaître et lier ses cibles dans la chromatine fermée, l'ouverture de la chromatine nécessite le recrutement de plusieurs facteurs. Bien qu'une liaison coopérative soit typique de tous les facteurs de transcription, il demeure que la capacité de ciblage des nucléosomes soit unique aux facteurs pionniers. Il est probable que lorsque les facteurs pionniers sondent la chromatine et analysent les sites, ils restent liés plus stablement aux sites où il y a une liaison coopérative avec d'autres facteurs de transcription. Ce concept a été démontré avec les facteurs de pluripotence OSKM durant la reprogrammation (Chronis et al., 2017), FOXA et GATA4 pour la spécification de l'endoderme (Cirillo et al., 2002), FOXA coopère avec d'autres facteurs de transcription tel que HNF1 α ou HNF4 α dans les cellules hépatiques (Huang et al., 2011; S. Sekiya & Suzuki, 2011) et GATA6 dans la différenciation des cellules pancréatiques (K. Lee et al., 2019).

Dans l'hypophyse, la spécification du destin mélanotrope nécessite la coopération du facteur de transcription non-pionnier Tpit avec le facteur pionnier Pax7 pour l'ouverture de la chromatine et la différenciation cellulaire (Mayran et al., 2019). L'expression ectopique de Pax7 dans les cellules α T3 (un modèle de cellule gonadotrope hypophysaire qui n'exprime pas Tpit), montre que Pax7 peut lier fortement la chromatine fermée, mais cette liaison n'est pas

productive et ces sites restent fermés (Mayran et al., 2019). Ceci indique que la coopération entre Tpit et Pax7 n'est pas au niveau de la liaison au site nucléosomal, contrairement aux facteurs OSKM qui sont des partenaires de liaison coopératifs par exemple. Tpit est requis pour les étapes subséquentes de l'action pionnière, c'est-à-dire l'augmentation de l'accessibilité à l'ADN. La liaison initiale de Pax7 favoriserait ainsi des interactions coopératives avec d'autres facteurs qui sont associés à l'ouverture de la chromatine permettant secondairement à d'autres facteurs de se lier. Ce type de complexe régulateur ne pourrait pas être établi par des facteurs qui ne possèdent pas la capacité de cibler les nucléosomes. En somme, l'action pionnière de Pax7 dépend du contexte cellulaire où le paysage de la chromatine locale et l'interaction avec des cofacteurs spécifiques est essentielle pour ouvrir la chromatine aux *enhancers* signatures de l'identité.

Il sera intéressant d'évaluer si la fonction de Pax7 est régulée par des modifications post-traductionnelles. Bien que le domaine de transactivation de Pax7 soit riche en résidus phosphorylables, les modifications post-traductionnelles de Pax7 sont peu documentées. Les modifications connues sont surtout situées dans la région des domaines de liaison à l'ADN; la méthylation d'arginines dans la partie N-terminale de Pax7 dans les cellules satellites musculaires (Kawabe, Wang, McKinnell, Bedford, & Rudnicki, 2012), la sumoylation de la lysine 85 essentielle au développement de la crête neurale et des muscles (Luan, Liu, Stuhlmiller, Marquez, & García-Castro, 2013), l'acétylation des lysines 105 et 193 dans les cellules satellites musculaire (Sincennes et al., 2021), la phosphorylation des sérines 201 et 205 requise pour le maintien des niveaux de protéines Pax7 dans les cellules progénitrices musculaires (Dick et al., 2015; González et al., 2016).

Il est possible que la fonction pionnière de Pax7 nécessite des modifications post-traductionnelles spécifiques pour l'interaction avec des partenaires de liaison ou pour le recrutement des complexes de remodelage de la chromatine et des enzymes qui contribuent à l'activation transcriptionnelle. Par exemple, il pourrait y avoir des modifications post-traductionnelles spécifiques à l'hypophyse, au muscle ou au cerveau. Ce mécanisme par lequel des enzymes tissu-spécifiques réguleraient l'activité de Pax7 pourrait fournir une explication supplémentaire à l'aspect contexte-dépendant de l'action de Pax7.

4.3 Les isoformes Pax7 ont des propriétés intrinsèques de liaison à l'ADN et de transactivation similaires

L'expression différentielle de quatre transcrits alternatifs de Pax7 dans différents tissus au cours du développement suggère qu'il y a une signification fonctionnelle aux différents isoformes (Vogan et al., 1996; Ziman & Kay, 1998). Nos travaux montrent que les propriétés de liaison aux différents motifs cibles de Pax7 sont globalement similaires pour tous les isoformes de Pax7 (Q+/-, GL+/-) (Chapitre 3 Fig.2). Toutefois, les formes qui excluent le GL (GL-) montrent une affinité renforcée à tous les motifs. Selon la structure connue du PD de PAX6, la troisième hélice du sous-domaine RED interagit avec le sillon majeur de l'ADN, devenant une hélice de reconnaissance additionnelle au PD. De plus, des prédictions de la structure secondaire suggèrent que l'inclusion de GL allongeraient la troisième hélice du sous-domaine RED, ce qui aurait pour effet de perturber les interactions avec l'ADN (Ziman & Kay, 1998). Cependant, les expériences de délétions des domaines de liaison à l'ADN et sous-domaines PAI et RED de Pax7 ont montrés que le RED n'est pas nécessaire aux interactions spécifiques avec ces séquences d'ADN *in vitro* (Chapitre 3 Fig.1). Cela suggère qu'il fonctionne plus indépendamment que les deux autres modules HTH de Pax7 (PAI et HD). Cela est supporté par les études de *cross-linking* avec Pax3 qui montrent que les interactions entre les domaines PD et HD implique le sous-domaine PAI seulement (Apuzzo & Gros, 2007), ainsi que les études de liaison de Pax3 qui montrent que le RED est impliqué seulement dans la reconnaissance de séquences motifs PD allongées (Vogan & Gros, 1997).

Le potentiel de transactivation des isoformes Pax7 aux différents motifs cibles a des différences subtiles (Chapitre 3 Fig. 3). Principalement, les formes Pax7 GL+ active mieux la cible PD et les formes Pax7 Q+GL-, et dans une moindre mesure Pax7 Q-GL-, sont dominants aux cibles HD (HD et H.I2). Ces résultats démontrent que la liaison à l'ADN *in vitro* ne corrèle pas nécessairement avec l'activité transcriptionnelle. Par exemple, les sondes HD sont d'affinité faible, mais l'activation transcriptionnelle est comparable aux autres motifs, et les formes GL+ lient plus faiblement la sonde PD mais y active mieux la transcription. Cela peut être dû à la configuration hexamérique des motifs, espacés d'environ 20 pb (2 tours d'hélice d'ADN), qui favorise la liaison de plusieurs facteurs. Ainsi, les facteurs peuvent possiblement se lier par coopération car ils se retrouveraient du même côté de la double hélice.

Il sera intéressant d'analyser le recrutement des différents isoformes Pax7 et leur potentielle action pionnière ainsi que la régulation génique *in vivo*. Cela permettrait d'évaluer les préférences au niveau des motifs cibles, à savoir si certains isoformes peuvent reconnaître des motifs sous-optimaux, ou si les formes Q- peuvent reconnaître des sites PD allongés comme il a été suggéré par Vogan et al. (Vogan et al., 1996). Aussi, les changements Q+/- pourraient réguler des ensembles distincts de gènes cibles, tandis que les changements GL+/- pourraient influencer le niveau d'expression des gènes cibles. Comme nous avons montré que les motifs de haute affinité sont enrichis aux sites pionniers où le recrutement de Pax7 est le plus robuste, et le fait que les isoformes GL+ se lient moins fortement aux sites consensus de Pax7, on pourrait s'attendre que l'interaction avec l'ADN nucléosomal soit affecté, et la capacité pionnière diminuée.

Bien que ces résultats suggèrent que les isoformes Pax7 offrent un potentiel de régulation fine de l'expression génique, la comparaison des isoformes montre des propriétés intrinsèques de liaison à l'ADN et de transactivation somme toute similaires. Si tant est que l'action différentielle des isoformes Pax7 doive s'expliquer à d'autres niveaux, cela pourrait inclure des interactions protéine-protéine particulières. Les acides aminés variants pourraient être impliqués directement dans de telles interactions ou ils pourraient également affecter la conformation inhérente de Pax7 et de manière allostérique, modifier la présentation d'autres résidus essentiels aux interactions.

Une autre possibilité est que le résidu variant favorise ou interfère avec des modifications post-traductionnelles de Pax7. Comme mentionné dans le sous-chapitre précédent, certaines modifications post-traductionnelles pourraient jouer un rôle clé dans les fonctions transcriptionnelle et/ou pionnière de Pax7. Par exemple, il a été démontré dans les cellules souches musculaire que l'acétylation de PAX7 aux résidus K105 et K193 est essentielle à son activité dans ce type cellulaire, associant spécifiquement la lysine acétyltransférase MYST1 et la désacétylase SIRT2 à la régulation de l'activité de PAX7 (Sincennes et al., 2021). La glutamine 108 de Pax7 étant à proximité de la lysine 105, le motif protéique que reconnaissent les enzymes de modifications post-traductionnelles pourrait être affecté ayant pour conséquence d'empêcher la régulation de l'activité de Pax7 par cette modification post-traductionnelle.

À l'heure actuelle, les rôles des différentes isoformes de Pax7 dans la régulation transcriptionnelle chez l'embryon et chez l'adulte restent à élucider. À savoir, quelles sont les variations d'expression durant le développement et chez l'adulte et si les différents isoformes peuvent avoir des rôles distincts dans différents tissus. Cela soulève ainsi la possibilité qu'il y ait des facteurs d'épissage tissu-spécifiques qui influencent la spécificité de la machinerie d'épissage et qui régulent l'expression différentielle des variants de Pax7.

Chapitre 5 : Conclusion et perspectives

Le facteur de transcription pionnier Pax7 est un sélecteur du lobe intermédiaire. Il met en place le programme génétique mélanotrope grâce à sa capacité de reconnaître ses sites cibles dans la chromatine fermée et en augmente l'accessibilité permettant le recrutement d'autres facteurs de transcription. La capacité de Pax7 à reprogrammer les cellules corticotropes vers une identité mélanotrope le démontre : les propriétés uniques du facteur pionnier Pax7 sont la clé du destin cellulaire mélanotrope. Mes travaux ont permis de mettre en lumière les propriétés de liaison à l'ADN distinctives de Pax7 lui permettant son action pionnière tel que l'affinité forte pour les motifs aux sites pionniers et le rôle essentiel des deux domaines de liaison à l'ADN intacts pour l'action pionnière. Cet ouvrage relève aussi les différences qui existent entre les facteurs pionniers connus dans les mécanismes de reconnaissance des sites dans la chromatine fermée. La compréhension des mécanismes par lesquels les facteurs pionniers peuvent lier la chromatine fermée permettrait d'avancer notre capacité à reprogrammer le destin des cellules.

En effet, la réversibilité de la différenciation cellulaire, sous les conditions des facteurs de pluripotence, et la reprogrammation de cellules souches permet d'envisager la production de cellules différenciées matures et fonctionnelles qui auraient un potentiel thérapeutique en médecine régénérative. Des applications cliniques sont actuellement envisageables grâce à des cellules souche dont on manipule l'identité pour restaurer une fonction déficiente ou régénérer la lésion d'un tissu ou, éventuellement, d'organes du corps humain. Des études de médecine régénérative personnalisée sont déjà en cours pour développer des cellules souches à partir des cellules de la peau des patients afin de soigner différentes pathologies tel que la maladie de Parkinson, la fibrose pulmonaire, des myopathies et la neuropathie périphérique héréditaire. Pour le développement de telles thérapies cellulaires, il est nécessaire d'approfondir les connaissances fondamentales sur les mécanismes de reprogrammation cellulaire par les facteurs pionniers.

Bibliographie

- Adams, C. C., & Workman, J. L. (1995). Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol Cell Biol*, *15*(3), 1405-1421. doi:10.1128/mcb.15.3.1405
- Allshire, R. C., & Madhani, H. D. (2018). Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol*, *19*(4), 229-244. doi:10.1038/nrm.2017.119
- Aloia, L., Di Stefano, B., & Di Croce, L. (2013). Polycomb complexes in stem cells and embryonic development. *Development*, *140*(12), 2525-2534. doi:10.1242/dev.091553
- Anderson, J. D., & Widom, J. (2001). Poly(dA-dT) Promoter Elements Increase the Equilibrium Accessibility of Nucleosomal DNA Target Sites. *21*(11), 3830-3839. doi:10.1128/MCB.21.11.3830-3839.2001 %J Molecular and Cellular Biology
- Andrabi, M., Mizuguchi, K., & Ahmad, S. (2014). Conformational changes in DNA-binding proteins: relationships with precomplex features and contributions to specificity and stability. *Proteins*, *82*(5), 841-857. doi:10.1002/prot.24462
- Angrand, P.-O. (1993). Les domaines de liaison à l'ADN des facteurs de transcription eucaryotes. *Med Sci (Paris)*, *9*(6-7), 725-736. Retrieved from <http://hdl.handle.net/10608/2983>
- Apostolou, E., & Stadtfeld, M. (2018). Cellular trajectories and molecular mechanisms of iPSC reprogramming. *Current Opinion in Genetics & Development*, *52*, 77-85. doi:10.1016/j.gde.2018.06.002
- Apuzzo, S., Abdelhakim, A., Fortin, A. S., & Gros, P. (2004). Cross-talk between the paired domain and the homeodomain of Pax3: DNA binding by each domain causes a structural change in the other domain, supporting interdependence for DNA Binding. *J Biol Chem*, *279*(32), 33601-33612. doi:10.1074/jbc.M402949200
- Apuzzo, S., & Gros, P. (2007). Cooperative interactions between the two DNA binding domains of Pax3: helix 2 of the paired domain is in the proximity of the amino terminus of the homeodomain. *Biochemistry*, *46*(11), 2984-2993. doi:10.1021/bi062107q
- Baralle, F. E., & Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nature reviews. Molecular cell biology*, *18*(7), 437-451. doi:10.1038/nrm.2017.27
- Barnhart, K. M., & Mellon, P. L. (1994). The orphan nuclear receptor, steroidogenic factor-1, regulates the glycoprotein hormone alpha-subunit gene in pituitary gonadotropes. *Mol Endocrinol*, *8*(7), 878-885. doi:10.1210/mend.8.7.7527122
- Barr, F. G. (2001). Gene fusions involving PAX and FOX family members in alveolar rhabdomyosarcoma. *Oncogene*, *20*(40), 5736-5746. doi:10.1038/sj.onc.1204599
- Baumann, H., Paulsen, K., Kovács, H., Berglund, H., Wright, A. P., Gustafsson, J. A., & Härd, T. (1993). Refined solution structure of the glucocorticoid receptor DNA-binding domain. *Biochemistry*, *32*(49), 13463-13471. doi:10.1021/bi00212a011
- Bednar, J., Horowitz, R. A., Grigoryev, S. A., Carruthers, L. M., Hansen, J. C., Koster, A. J., & Woodcock, C. L. (1998). Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc Natl Acad Sci U S A*, *95*(24), 14173-14178. doi:10.1073/pnas.95.24.14173
- Bennicelli, J. L., Advani, S., Schäfer, B. W., & Barr, F. G. (1999). PAX3 and PAX7 exhibit conserved cis-acting transcription repression domains and utilize a common gain of function mechanism in alveolar rhabdomyosarcoma. *Oncogene*, *18*(30), 4348-4356. doi:10.1038/sj.onc.1202812
- Bernstein, B. E., Meissner, A., & Lander, E. S. (2007). The mammalian epigenome. *Cell*, *128*(4), 669-681. doi:10.1016/j.cell.2007.01.033

- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., . . . Lander, E. S. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, *125*(2), 315-326. doi:10.1016/j.cell.2006.02.041
- Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., . . . Canaider, S. (2013). An estimation of the number of cells in the human body. *Ann Hum Biol*, *40*(6), 463-471. doi:10.3109/03014460.2013.807878
- Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J. M., . . . Shen, R. (2011). High density DNA methylation array with single CpG site resolution. *Genomics*, *98*(4), 288-295. doi:10.1016/j.ygeno.2011.07.007
- Bilokapic, S., Strauss, M., & Halic, M. (2018). Structural rearrangements of the histone octamer translocate DNA. *Nature Communications*, *9*(1), 1330. doi:10.1038/s41467-018-03677-z
- Birrane, G., Soni, A., & Ladias, J. A. (2009). Structural basis for DNA recognition by the human PAX3 homeodomain. *Biochemistry*, *48*(6), 1148-1155. doi:10.1021/bi802052y
- Blomquist, P., Li, Q., & Wrangé, O. (1996). The affinity of nuclear factor 1 for its DNA site is drastically reduced by nucleosome organization irrespective of its rotational or translational position. *J Biol Chem*, *271*(1), 153-159. doi:10.1074/jbc.271.1.153
- Bolouri, H., & Davidson, E. H. (2003). Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(16), 9371-9376. doi:10.1073/pnas.1533293100
- Bommarito, P. A., & Fry, R. C. (2019). Chapter 2-1 - The Role of DNA Methylation in Gene Regulation. In S. D. McCullough & D. C. Dolinoy (Eds.), *Toxicoepigenetics* (pp. 127-151): Academic Press.
- Bonev, B., Cohen, N. M., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., . . . Tanay, A. (2017). Multiscale 3D genome rewiring during mouse neural development. *Cell*, *171*(3), 557-572. e524.
- Bowman, G. D., & Poirier, M. G. (2015). Post-translational modifications of histones that influence nucleosome dynamics. *Chemical reviews*, *115*(6), 2274-2295. doi:10.1021/cr500350x
- Buckingham, M., & Relaix, F. (2015). PAX3 and PAX7 as upstream regulators of myogenesis. *Semin Cell Dev Biol*, *44*, 115-125. doi:10.1016/j.semcdb.2015.09.017
- Budry, L., Balsalobre, A., Gauthier, Y., Khetchoumian, K., L'Honore, A., Vallette, S., . . . Drouin, J. (2012). The selector gene Pax7 dictates alternate pituitary cell fates through its pioneer action on chromatin remodeling. *Genes Dev*, *26*(20), 2299-2310. doi:10.1101/gad.200436.112
- Budry, L., Lafont, C., El Yandouzi, T., Chauvet, N., Conéjero, G., Drouin, J., & Mollard, P. (2011). Related pituitary cell lineages develop into interdigitated 3D cell networks. *Proc Natl Acad Sci U S A*, *108*(30), 12515-12520. doi:10.1073/pnas.1105929108
- Burow, D. A., Umeh-Garcia, M. C., True, M. B., Bakhaj, C. D., Ardell, D. H., & Cleary, M. D. (2015). Dynamic regulation of mRNA decay during neural development. *Neural development*, *10*, 11-11. doi:10.1186/s13064-015-0038-6
- Butali, A., Suzuki, S., Cooper, M. E., Mansilla, A. M., Cuenco, K., Leslie, E. J., . . . Murray, J. C. (2013). Replication of genome wide association identified candidate genes confirm the role of common and rare variants in PAX7 and VAX1 in the etiology of nonsyndromic CL(P). *Am J Med Genet A*, *161a*(5), 965-972. doi:10.1002/ajmg.a.35749
- Canzio, D., Chang, E. Y., Shankar, S., Kuchenbecker, K. M., Simon, M. D., Madhani, H. D., . . . Al-Sady, B. (2011). Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. *Mol Cell*, *41*(1), 67-81. doi:10.1016/j.molcel.2010.12.016

- Caravaca, J. M., Donahue, G., Becker, J. S., He, X., Vinson, C., & Zaret, K. S. (2013). Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. *Genes Dev*, 27(3), 251-260. doi:10.1101/gad.206458.112
- Carotta, S., Wu, L., & Nutt, S. L. (2010). Surprising new roles for PU.1 in the adaptive immune response. *Immunol Rev*, 238(1), 63-75. doi:10.1111/j.1600-065X.2010.00955.x
- Carroll, J. S., Liu, X. S., Brodsky, A. S., Li, W., Meyer, C. A., Szary, A. J., . . . Brown, M. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell*, 122(1), 33-43. doi:10.1016/j.cell.2005.05.008
- Cawley, N. X., Li, Z., & Loh, Y. P. (2016). 60 YEARS OF POMC: Biosynthesis, trafficking, and secretion of pro-opiomelanocortin-derived peptides %J Journal of Molecular Endocrinology. 56(4), T77-T97. doi:10.1530/jme-15-0323
- Cazaly, E., Charlesworth, J., Dickinson, J. L., & Holloway, A. F. (2015). Genetic Determinants of Epigenetic Patterns: Providing Insight into Disease. *Molecular Medicine*, 21(1), 400-409. doi:10.2119/molmed.2015.00001
- Chalepakis, G., Stoykova, A., Wijnholds, J., Tremblay, P., & Gruss, P. (1993). Pax: gene regulators in the developing nervous system. *J Neurobiol*, 24(10), 1367-1384. doi:10.1002/neu.480241009
- Chalepakis, G., Wijnholds, J., & Gruss, P. (1994). Pax-3-DNA interaction: flexibility in the DNA binding and induction of DNA conformational changes by paired domains. *Nucleic Acids Res*, 22(15), 3131-3137. doi:10.1093/nar/22.15.3131
- Charles, M. A., Saunders, T. L., Wood, W. M., Owens, K., Parlow, A. F., Camper, S. A., . . . Gordon, D. F. (2006). Pituitary-specific Gata2 knockout: effects on gonadotrope and thyrotrope function. *Mol Endocrinol*, 20(6), 1366-1377. doi:10.1210/me.2005-0378
- Chávez, S., & Beato, M. (1997). Nucleosome-mediated synergism between transcription factors on the mouse mammary tumor virus promoter. *Proc Natl Acad Sci U S A*, 94(7), 2885-2890. doi:10.1073/pnas.94.7.2885
- Chaya, D., Hayamizu, T., Bustin, M., & Zaret, K. S. (2001). Transcription factor FoxA (HNF3) on a nucleosome at an enhancer complex in liver chromatin. *J Biol Chem*, 276(48), 44385-44389. doi:10.1074/jbc.M108214200
- Chen, J., Chen, X., Li, M., Liu, X., Gao, Y., Kou, X., . . . Gao, S. (2016). Hierarchical Oct4 Binding in Concert with Primed Epigenetic Rearrangements during Somatic Cell Reprogramming. *Cell Rep*, 14(6), 1540-1554. doi:10.1016/j.celrep.2016.01.013
- Chen, J., Hersmus, N., Van Duppen, V., Caesens, P., Deneff, C., & Vankelecom, H. (2005). The adult pituitary contains a cell population displaying stem/progenitor cell and early embryonic characteristics. *Endocrinology*, 146(9), 3985-3998. doi:10.1210/en.2005-0185
- Chen, K., Long, Q., Xing, G., Wang, T., Wu, Y., Li, L., . . . Liu, X. (2020). Heterochromatin loosening by the Oct4 linker region facilitates Klf4 binding and iPSC reprogramming. *The EMBO journal*, 39(1), e99165-e99165. doi:10.15252/embj.201899165
- Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., . . . Ng, H. H. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, 133(6), 1106-1117. doi:10.1016/j.cell.2008.04.043
- Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., . . . Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell*, 168(3), 442-459.e420. doi:10.1016/j.cell.2016.12.016
- Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M., & Zaret, K. S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell*, 9(2), 279-289. doi:10.1016/s1097-2765(02)00459-8

- Cirillo, L. A., McPherson, C. E., Bossard, P., Stevens, K., Cherian, S., Shim, E. Y., . . . Zaret, K. S. (1998). Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *Embo j*, *17*(1), 244-254. doi:10.1093/emboj/17.1.244
- Cirillo, L. A., & Zaret, K. S. (1999). An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol Cell*, *4*(6), 961-969. doi:10.1016/s1097-2765(00)80225-7
- Cirillo, L. A., & Zaret, K. S. (2007). Specific interactions of the wing domains of FOXA1 transcription factor with DNA. *J Mol Biol*, *366*(3), 720-724. doi:10.1016/j.jmb.2006.11.087
- Clapier, C. R., Iwasa, J., Cairns, B. R., & Peterson, C. L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol*, *18*(7), 407-422. doi:10.1038/nrm.2017.26
- Clark, K. L., Halay, E. D., Lai, E., & Burley, S. K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature*, *364*(6436), 412-420. doi:10.1038/364412a0
- Cohen, L. E., & Radovick, S. (2002). Molecular Basis of Combined Pituitary Hormone Deficiencies. *Endocrine Reviews*, *23*(4), 431-442. doi:10.1210/er.2001-0030
- Cote, T. E., Felder, R., Keabian, J. W., Sekura, R. D., Reisine, T., & Affolter, H. U. (1986). D-2 dopamine receptor-mediated inhibition of pro-opiomelanocortin synthesis in rat intermediate lobe. Abolition by pertussis toxin or activators of adenylate cyclase. *J Biol Chem*, *261*(10), 4555-4561.
- Cuesta, I., Zaret, K. S., & Santisteban, P. (2007). The forkhead factor FoxE1 binds to the thyroperoxidase promoter during thyroid cell differentiation and modifies compacted chromatin structure. *Mol Cell Biol*, *27*(20), 7302-7314. doi:10.1128/mcb.00758-07
- Cui, J. Y., Fu, Z. D., & Dempsey, J. (2019). Chapter 1-2 - The Role of Histone Methylation and Methyltransferases in Gene Regulation. In S. D. McCullough & D. C. Dolinoy (Eds.), *Toxicoepigenetics* (pp. 31-84): Academic Press.
- Czerny, T., & Busslinger, M. (1995). DNA-binding and transactivation properties of Pax-6: three amino acids in the paired domain are responsible for the different sequence recognition of Pax-6 and BSAP (Pax-5). *Mol Cell Biol*, *15*(5), 2858-2871. doi:10.1128/mcb.15.5.2858
- Czerny, T., Schaffner, G., & Busslinger, M. (1993). DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site. *Genes Dev*, *7*(10), 2048-2061. doi:10.1101/gad.7.10.2048
- Davis, S. W., Mortensen, A. H., & Camper, S. A. (2011). Birthdating studies reshape models for pituitary gland cell specification. *Developmental biology*, *352*(2), 215-227. doi:10.1016/j.ydbio.2011.01.010
- Day, R. N., Koike, S., Sakai, M., Muramatsu, M., & Maurer, R. A. (1990). Both Pit-1 and the estrogen receptor are required for estrogen responsiveness of the rat prolactin gene. *Mol Endocrinol*, *4*(12), 1964-1971. doi:10.1210/mend-4-12-1964
- Demeneix, B. A., Desaulles, E., Feltz, P., & Loeffler, J. P. (1984). Dual population of GABAA and GABAB receptors in rat pars intermedia demonstrated by release of alpha MSH caused by barium ions. *Br J Pharmacol*, *82*(1), 183-190. doi:10.1111/j.1476-5381.1984.tb16457.x
- Di Croce, L., & Helin, K. (2013). Transcriptional regulation by Polycomb group proteins. *Nat Struct Mol Biol*, *20*(10), 1147-1155. doi:10.1038/nsmb.2669
- Dick, S. A., Chang, N. C., Dumont, N. A., Bell, R. A., Putinski, C., Kawabe, Y., . . . Megeney, L. A. (2015). Caspase 3 cleavage of Pax7 inhibits self-renewal of satellite cells. *Proc Natl Acad Sci U S A*, *112*(38), E5246-5252. doi:10.1073/pnas.1512869112

- Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J., Lee, A. Y., . . . Xie, W. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature*, *518*(7539), 331-336.
- Dodonova, S. O., Zhu, F., Dienemann, C., Taipale, J., & Cramer, P. (2020). Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. *Nature*, *580*(7805), 669-672. doi:10.1038/s41586-020-2195-y
- Donaghey, J., Thakurela, S., Charlton, J., Chen, J. S., Smith, Z. D., Gu, H., . . . Meissner, A. (2018). Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat Genet*, *50*(2), 250-258. doi:10.1038/s41588-017-0034-3
- Drouin, J. (2011). Chapter 1 - Pituitary Development. In S. Melmed (Ed.), *The Pituitary (Third Edition)* (pp. 3-19). San Diego: Academic Press.
- Drouin, J. (2014). Minireview: pioneer transcription factors in cell fate specification. *Mol Endocrinol*, *28*(7), 989-998. doi:10.1210/me.2014-1084
- Du, S., Lawrence, E. J., Strzelecki, D., Rajput, P., Xia, S. J., Gottesman, D. M., & Barr, F. G. (2005). Co-expression of alternatively spliced forms of PAX3, PAX7, PAX3-FKHR and PAX7-FKHR with distinct DNA binding and transactivation properties in rhabdomyosarcoma. *Int J Cancer*, *115*(1), 85-92. doi:10.1002/ijc.20844
- Dubois, P. M., el Amraoui, A., & Héritier, A. G. (1997). Development and differentiation of pituitary cells. *Microsc Res Tech*, *39*(2), 98-113. doi:10.1002/(sici)1097-0029(19971015)39:2<98::Aid-jemt2>3.0.Co;2-s
- Eberharter, A., & Becker, P. B. (2002). Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. *EMBO Rep*, *3*(3), 224-229. doi:10.1093/embo-reports/kvf053
- Echigoya, K., Koyama, M., Negishi, L., Takizawa, Y., Mizukami, Y., Shimabayashi, H., . . . Kurumizaka, H. (2020). Nucleosome binding by the pioneer transcription factor OCT4. *Sci Rep*, *10*(1), 11832. doi:10.1038/s41598-020-68850-1
- Elgin, S. C., Elgin, S. C. R., & Workman, J. L. (2000). *Chromatin Structure and Gene Expression*: Oxford University Press.
- Ellsworth, B. S., Egashira, N., Haller, J. L., Butts, D. L., Cocquet, J., Clay, C. M., . . . Camper, S. A. (2006). FOXL2 in the Pituitary: Molecular, Genetic, and Developmental Analysis. *Molecular Endocrinology*, *20*(11), 2796-2805. doi:10.1210/me.2005-0303
- Epstein, J., Cai, J., Glaser, T., Jepeal, L., & Maas, R. (1994). Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes. *J Biol Chem*, *269*(11), 8355-8361.
- Epstein, J. A., Glaser, T., Cai, J., Jepeal, L., Walton, D. S., & Maas, R. L. (1994). Two independent and interactive DNA-binding subdomains of the Pax6 paired domain are regulated by alternative splicing. *Genes Dev*, *8*(17), 2022-2034. doi:10.1101/gad.8.17.2022
- Ericson, J., Rashbass, P., Schedl, A., Brenner-Morton, S., Kawakami, A., van Heyningen, V., . . . Briscoe, J. (1997). Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling. *Cell*, *90*(1), 169-180. doi:10.1016/s0092-8674(00)80323-2
- Esch, D., Vahokoski, J., Groves, M. R., Pogenberg, V., Cojocaru, V., Vom Bruch, H., . . . Schöler, H. R. (2013). A unique Oct4 interface is crucial for reprogramming to pluripotency. *Nat Cell Biol*, *15*(3), 295-301. doi:10.1038/ncb2680
- Escobar, T. M., Oksuz, O., Saldaña-Meyer, R., Descostes, N., Bonasio, R., & Reinberg, D. (2019). Active and Repressed Chromatin Domains Exhibit Distinct Nucleosome Segregation during DNA Replication. *Cell*, *179*(4), 953-963.e911. doi:10.1016/j.cell.2019.10.009

- Eslami-Mossallam, B., Schiessel, H., & van Noort, J. (2016). Nucleosome dynamics: Sequence matters. *Advances in Colloid and Interface Science*, 232, 101-113. doi:10.1016/j.cis.2016.01.007
- Fakhouri, T. H., Stevenson, J., Chisholm, A. D., & Mango, S. E. (2010). Dynamic chromatin organization during foregut development mediated by the organ selector gene PHA-4/FoxA. *PLoS Genet*, 6(8). doi:10.1371/journal.pgen.1001060
- Fauquier, T., Rizzoti, K., Dattani, M., Lovell-Badge, R., & Robinson, I. C. (2008). SOX2-expressing progenitor cells generate all of the major cell types in the adult mouse pituitary gland. *Proc Natl Acad Sci U S A*, 105(8), 2907-2912. doi:10.1073/pnas.0707886105
- Feichtinger, R. G., Mucha, B. E., Hengel, H., Orfi, Z., Makowski, C., Dort, J., . . . Dumont, N. A. (2019). Biallelic variants in the transcription factor PAX7 are a new genetic cause of myopathy. *Genet Med*, 21(11), 2521-2531. doi:10.1038/s41436-019-0532-z
- Fernandez Garcia, M., Moore, C. D., Schulz, K. N., Alberto, O., Donague, G., Harrison, M. M., . . . Zaret, K. S. (2019). Structural Features of Transcription Factors Associating with Nucleosome Binding. *Mol Cell*, 75(5), 921-932.e926. doi:10.1016/j.molcel.2019.06.009
- Ferri, A. L., Lin, W., Mavromatakis, Y. E., Wang, J. C., Sasaki, H., Whitsett, J. A., & Ang, S. L. (2007). Foxa1 and Foxa2 regulate multiple phases of midbrain dopaminergic neuron development in a dosage-dependent manner. *Development*, 134(15), 2761-2769. doi:10.1242/dev.000141
- Fillingham, J., Keogh, M. C., & Krogan, N. J. (2006). GammaH2AX and its role in DNA double-strand break repair. *Biochem Cell Biol*, 84(4), 568-577. doi:10.1139/o06-072
- Fortin, A. S., Underhill, D. A., & Gros, P. (1997). Reciprocal effect of Waardenburg syndrome mutations on DNA binding by the Pax-3 paired domain and homeodomain. *Hum Mol Genet*, 6(11), 1781-1790. doi:10.1093/hmg/6.11.1781
- Fuchs, Y., & Steller, H. (2011). Programmed cell death in animal development and disease. *Cell*, 147(4), 742-758. doi:10.1016/j.cell.2011.10.033
- Gaczkowska, A., Biedziak, B., Budner, M., Zadurska, M., Lasota, A., Hozyasz, K. K., . . . Mostowska, A. (2019). PAX7 nucleotide variants and the risk of non-syndromic orofacial clefts in the Polish population. *Oral Dis*, 25(6), 1608-1618. doi:10.1111/odi.13139
- Gao, N., LeLay, J., Vatamaniuk, M. Z., Rieck, S., Friedman, J. R., & Kaestner, K. H. (2008). Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. *Genes Dev*, 22(24), 3435-3448. doi:10.1101/gad.1752608
- Gao, R., Liang, X., Cheedipudi, S., Cordero, J., Jiang, X., Zhang, Q., . . . Dobрева, G. (2019). Pioneering function of Isl1 in the epigenetic control of cardiomyocyte cell fate. *Cell Res*, 29(6), 486-501. doi:10.1038/s41422-019-0168-1
- García-Bellido, A. (1975). Genetic control of wing disc development in *Drosophila*. *Ciba Found Symp*, 0(29), 161-182. doi:10.1002/9780470720110.ch8
- Garvie, C. W., Hagman, J., & Wolberger, C. (2001). Structural studies of Ets-1/Pax5 complex formation on DNA. *Mol Cell*, 8(6), 1267-1276. doi:10.1016/s1097-2765(01)00410-5
- Garvie, C. W., & Wolberger, C. (2001). Recognition of specific DNA sequences. *Mol Cell*, 8(5), 937-946. doi:10.1016/s1097-2765(01)00392-6
- Gates, L. A., Foulds, C. E., & O'Malley, B. W. (2017). Histone Marks in the 'Driver's Seat': Functional Roles in Steering the Transcription Cycle. *Trends in biochemical sciences*, 42(12), 977-989. doi:10.1016/j.tibs.2017.10.004
- González, N., Moresco, J. J., Cabezas, F., de la Vega, E., Bustos, F., Yates, J. R., 3rd, & Olguín, H. C. (2016). Ck2-Dependent Phosphorylation Is Required to Maintain Pax7 Protein Levels in Proliferating Muscle Progenitors. *PLoS One*, 11(5), e0154919. doi:10.1371/journal.pone.0154919

- Gordon, D. F., Lewis, S. R., Haugen, B. R., James, R. A., McDermott, M. T., Wood, W. M., & Ridgway, E. C. (1997). Pit-1 and GATA-2 interact and functionally cooperate to activate the thyrotropin beta-subunit promoter. *J Biol Chem*, 272(39), 24339-24347. doi:10.1074/jbc.272.39.24339
- Gottesfeld, J. M., & Forbes, D. J. (1997). Mitotic repression of the transcriptional machinery. *Trends Biochem Sci*, 22(6), 197-202. doi:10.1016/s0968-0004(97)01045-1
- Gowans, L. J. J., Adeyemo, W. L., Eshete, M., Mossey, P. A., Busch, T., Aregbesola, B., . . . Butali, A. (2016). Association Studies and Direct DNA Sequencing Implicate Genetic Susceptibility Loci in the Etiology of Nonsyndromic Orofacial Clefts in Sub-Saharan African Populations. *Journal of dental research*, 95(11), 1245-1256. doi:10.1177/0022034516657003
- Goytisolo, F. A., Packman, L. C., & Thomas, J. O. (1996). Photoaffinity labelling of a DNA-binding site on the globular domain of histone H5. *Eur J Biochem*, 242(3), 619-626. doi:10.1111/j.1432-1033.1996.0619r.x
- Gremeaux, L., Fu, Q., Chen, J., & Vankelecom, H. (2012). Activated phenotype of the pituitary stem/progenitor cell compartment during the early-postnatal maturation phase of the gland. *Stem Cells Dev*, 21(5), 801-813. doi:10.1089/scd.2011.0496
- Hagolani, P. F., Zimm, R., Marin-Riera, M., & Salazar-Ciudad, I. (2019). Cell signaling stabilizes morphogenesis against noise. *Development*, 146(20). doi:10.1242/dev.179309
- Härd, T., Kellenbach, E., Boelens, R., Maler, B. A., Dahlman, K., Freedman, L. P., . . . Kaptein, R. (1990). Solution structure of the glucocorticoid receptor DNA-binding domain. *Science*, 249(4965), 157-160. doi:10.1126/science.2115209
- Hashimoto, K., Yamada, M., Monden, T., Satoh, T., Wondisford, F. E., & Mori, M. (2005). Thyrotropin-releasing hormone (TRH) specific interaction between amino terminus of P-Lim and CREB binding protein (CBP). *Mol Cell Endocrinol*, 229(1-2), 11-20. doi:10.1016/j.mce.2004.10.005
- Hatta, M., & Cirillo, L. A. (2007). Chromatin opening and stable perturbation of core histone:DNA contacts by FoxO1. *J Biol Chem*, 282(49), 35583-35593. doi:10.1074/jbc.M704735200
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., . . . Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*, 38(4), 576-589. doi:10.1016/j.molcel.2010.05.004
- Hermann, A., Gowher, H., & Jeltsch, A. (2004). Biochemistry and biology of mammalian DNA methyltransferases. *Cell Mol Life Sci*, 61(19-20), 2571-2587. doi:10.1007/s00018-004-4201-1
- Herz, H. M., Mohan, M., Garruss, A. S., Liang, K., Takahashi, Y. H., Mickey, K., . . . Shilatifard, A. (2012). Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *Genes Dev*, 26(23), 2604-2620. doi:10.1101/gad.201327.112
- Hoffmann, A., Zimmermann, C. A., & Spengler, D. (2015). Molecular epigenetic switches in neurodevelopment in health and disease. *Frontiers in behavioral neuroscience*, 9, 120-120. doi:10.3389/fnbeh.2015.00120
- Huang, P., He, Z., Ji, S., Sun, H., Xiang, D., Liu, C., . . . Hui, L. (2011). Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature*, 475(7356), 386-389. doi:10.1038/nature10116
- Huertas, J., MacCarthy, C. M., Schöler, H. R., & Cojocaru, V. (2020). Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding. *Biophys J*, 118(9), 2280-2296. doi:10.1016/j.bpj.2019.12.038
- Ikeda, H., & Yoshimoto, T. (1991). Developmental changes in proliferative activity of cells of the murine Rathke's pouch. *Cell Tissue Res*, 263(1), 41-47. doi:10.1007/bf00318398

- Ioshikhes, I. P., Albert, I., Zanton, S. J., & Pugh, B. F. (2006). Nucleosome positions predicted through comparative genomics. *Nat Genet*, *38*(10), 1210-1215. doi:10.1038/ng1878
- Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J. A., Mahony, S., Pugh, B. F., . . . Zaret, K. S. (2016). The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. *Molecular Cell*, *62*(1), 79-91. doi:10.1016/j.molcel.2016.03.001
- Japón, M. A., Rubinstein, M., & Low, M. J. (1994). In situ hybridization analysis of anterior pituitary hormone gene expression during fetal mouse development. *J Histochem Cytochem*, *42*(8), 1117-1125. doi:10.1177/42.8.8027530
- Johnson, T. A., Chereji, R. V., Stavreva, D. A., Morris, S. A., Hager, G. L., & Clark, D. J. (2018). Conventional and pioneer modes of glucocorticoid receptor interaction with enhancer chromatin in vivo. *Nucleic Acids Res*, *46*(1), 203-214. doi:10.1093/nar/gkx1044
- Jun, S., & Desplan, C. (1996). Cooperative interactions between paired domain and homeodomain. *Development*, *122*(9), 2639-2650.
- Kapali, J., Kabat, B. E., Schmidt, K. L., Stallings, C. E., Tippy, M., Jung, D. O., . . . Ellsworth, B. S. (2016). Foxo1 Is Required for Normal Somatotrope Differentiation. *Endocrinology*, *157*(11), 4351-4363. doi:10.1210/en.2016-1372
- Kastin, A. J., Plotnikoff, N. P., Viosca, S., Anderson, M. S., & Schally, A. V. (1973). MSH-release inhibiting factors: recent studies. *The Yale journal of biology and medicine*, *46*(5), 617-622.
- Kawabe, Y., Wang, Y. X., McKinnell, I. W., Bedford, M. T., & Rudnicki, M. A. (2012). Carm1 regulates Pax7 transcriptional activity through MLL1/2 recruitment during asymmetric satellite stem cell divisions. *Cell Stem Cell*, *11*(3), 333-345. doi:10.1016/j.stem.2012.07.001
- Knaupp, A. S., Buckberry, S., Pflueger, J., Lim, S. M., Ford, E., Larcombe, M. R., . . . Polo, J. M. (2017). Transient and Permanent Reconfiguration of Chromatin and Transcription Factor Occupancy Drive Reprogramming. *Cell Stem Cell*, *21*(6), 834-845.e836. doi:10.1016/j.stem.2017.11.007
- Koivunen, P., & Laukka, T. (2018). The TET enzymes. *Cell Mol Life Sci*, *75*(8), 1339-1348. doi:10.1007/s00018-017-2721-8
- Kornberg, R. D., & Lorch, Y. (2020). Primary Role of the Nucleosome. *Molecular Cell*, *79*(3), 371-375. doi:10.1016/j.molcel.2020.07.020
- Kozmik, Z., Czerny, T., & Busslinger, M. (1997). Alternatively spliced insertions in the paired domain restrict the DNA sequence specificity of Pax6 and Pax8. *Embo j*, *16*(22), 6793-6803. doi:10.1093/emboj/16.22.6793
- Laganière, J., Deblois, G., Lefebvre, C., Bataille, A. R., Robert, F., & Giguère, V. (2005). From the Cover: Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(33), 11651-11656. doi:10.1073/pnas.0505575102
- Lamolet, B., Pulichino, A., Lamonerie, T., Gauthier, Y., Brue, T., Enjalbert, A., & Drouin, J. (2001). A pituitary cell-restricted T-box factor, Tpit, activates POMC transcription in cooperation with Pitx homeoproteins. *Cell*, *104*, 849-859.
- Larson, A. G., Elnatan, D., Keenen, M. M., Trnka, M. J., Johnston, J. B., Burlingame, A. L., . . . Narlikar, G. J. (2017). Liquid droplet formation by HP1 α suggests a role for phase separation in heterochromatin. *Nature*, *547*(7662), 236-240. doi:10.1038/nature22822
- Larson, A. G., & Narlikar, G. J. (2018). The Role of Phase Separation in Heterochromatin Formation, Function, and Regulation. *Biochemistry*, *57*(17), 2540-2548. doi:10.1021/acs.biochem.8b00401

- Lee, C. S., Friedman, J. R., Fulmer, J. T., & Kaestner, K. H. (2005). The initiation of liver development is dependent on Foxa transcription factors. *Nature*, *435*(7044), 944-947. doi:10.1038/nature03649
- Lee, K., Cho, H., Rickert, R. W., Li, Q. V., Pulecio, J., Leslie, C. S., & Huangfu, D. (2019). FOXA2 Is Required for Enhancer Priming during Pancreatic Differentiation. *Cell Rep*, *28*(2), 382-393.e387. doi:10.1016/j.celrep.2019.06.034
- Leslie, E. J., Taub, M. A., Liu, H., Steinberg, K. M., Koboldt, D. C., Zhang, Q., . . . Murray, J. C. (2015). Identification of functional variants for cleft lip with or without cleft palate in or near PAX7, FGFR2, and NOG by targeted sequencing of GWAS loci. *Am J Hum Genet*, *96*(3), 397-411. doi:10.1016/j.ajhg.2015.01.004
- Li, D., Liu, J., Yang, X., Zhou, C., Guo, J., Wu, C., . . . Pei, D. (2017). Chromatin Accessibility Dynamics during iPSC Reprogramming. *Cell Stem Cell*, *21*(6), 819-833.e816. doi:10.1016/j.stem.2017.10.012
- Li, S., Crenshaw, E. B., 3rd, Rawson, E. J., Simmons, D. M., Swanson, L. W., & Rosenfeld, M. G. (1990). Dwarf locus mutants lacking three pituitary cell types result from mutations in the POU-domain gene pit-1. *Nature*, *347*(6293), 528-533. doi:10.1038/347528a0
- Luan, Z., Liu, Y., Stuhlmiller, T. J., Marquez, J., & García-Castro, M. I. (2013). SUMOylation of Pax7 is essential for neural crest and muscle development. *Cell Mol Life Sci*, *70*(10), 1793-1806. doi:10.1007/s00018-012-1220-1
- Lugo, D. I., & Pintar, J. E. (1996). Ontogeny of basal and regulated secretion from POMC cells of the developing anterior lobe of the rat pituitary gland. *Developmental biology*, *173*(1), 95-109. doi:10.1006/dbio.1996.0009
- Luscombe, N. M., Laskowski, R. A., & Thornton, J. M. (2001). Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res*, *29*(13), 2860-2874. doi:10.1093/nar/29.13.2860
- Mallo, M. (2018). Reassessing the Role of Hox Genes during Vertebrate Development and Evolution. *Trends Genet*, *34*(3), 209-217. doi:10.1016/j.tig.2017.11.007
- Mansouri, A., & Gruss, P. (1998). Pax3 and Pax7 are expressed in commissural neurons and restrict ventral neuronal identity in the spinal cord. *Mech Dev*, *78*(1-2), 171-178. doi:10.1016/s0925-4773(98)00168-3
- Mansouri, A., Stoykova, A., & Gruss, P. (1994). *Pax* genes in development. *Journal of Cell Science*, *1994*(Supplement 18), 35-42. doi:10.1242/jcs.1994.Supplement_18.5
- Mansouri, A., Stoykova, A., Torres, M., & Gruss, P. (1996). Dysgenesis of cephalic neural crest derivatives in Pax7^{-/-} mutant mice. *Development*, *122*(3), 831-838.
- Margueron, R., Justin, N., Ohno, K., Sharpe, M. L., Son, J., Drury, W. J., 3rd, . . . Gambin, S. J. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature*, *461*(7265), 762-767. doi:10.1038/nature08398
- Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C. L., . . . Reinberg, D. (2008). Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol Cell*, *32*(4), 503-518. doi:10.1016/j.molcel.2008.11.004
- Masse, J. E., Wong, B., Yen, Y. M., Allain, F. H., Johnson, R. C., & Feigon, J. (2002). The *S. cerevisiae* architectural HMGB protein NHP6A complexed with DNA: DNA and protein conformational changes upon binding. *J Mol Biol*, *323*(2), 263-284. doi:10.1016/s0022-2836(02)00938-5
- Mayran, A., & Drouin, J. (2018). Pioneer transcription factors shape the epigenetic landscape. *J Biol Chem*, *293*(36), 13795-13804. doi:10.1074/jbc.R117.001232

- Mayran, A., Khetchoumian, K., Hariri, F., Pastinen, T., Gauthier, Y., Balsalobre, A., & Drouin, J. (2018). Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. *Nat Genet*, *50*(2), 259-269. doi:10.1038/s41588-017-0035-2
- Mayran, A., Pelletier, A., & Drouin, J. (2015). Pax factors in transcription and epigenetic remodelling. *Semin Cell Dev Biol*, *44*, 135-144. doi:10.1016/j.semcdb.2015.07.007
- Mayran, A., Sochodolsky, K., Khetchoumian, K., Harris, J., Gauthier, Y., Bemmo, A., . . . Drouin, J. (2019). Pioneer and nonpioneer factor cooperation drives lineage specific chromatin opening. *Nat Commun*, *10*(1), 3807. doi:10.1038/s41467-019-11791-9
- McPherson, C. E., Shim, E. Y., Friedman, D. S., & Zaret, K. S. (1993). An active tissue-specific enhancer and bound transcription factors existing in a precisely positioned nucleosomal array. *Cell*, *75*(2), 387-398. doi:10.1016/0092-8674(93)80079-t
- Meers, M. P., Janssens, D. H., & Henikoff, S. (2019). Pioneer Factor-Nucleosome Binding Events during Differentiation Are Motif Encoded. *Mol Cell*, *75*(3), 562-575.e565. doi:10.1016/j.molcel.2019.05.025
- Mellor, J. (2006). Dynamic nucleosomes and gene transcription. *Trends Genet*, *22*(6), 320-329. doi:10.1016/j.tig.2006.03.008
- Merino, F., Bouvier, B., & Cojocaru, V. (2015). Cooperative DNA Recognition Modulated by an Interplay between Protein-Protein Interactions and DNA-Mediated Allostery. *PLoS Comput Biol*, *11*(6), e1004287. doi:10.1371/journal.pcbi.1004287
- Michael, A. K., Grand, R. S., Isbel, L., Cavadini, S., Kozicka, Z., Kempf, G., . . . Thomä, N. H. (2020). Mechanisms of OCT4-SOX2 motif readout on nucleosomes. *Science*, *368*(6498), 1460-1465. doi:10.1126/science.abb0074
- Mirny, L. A. (2010). Nucleosome-mediated cooperativity between transcription factors. *Proc Natl Acad Sci U S A*, *107*(52), 22534-22539. doi:10.1073/pnas.0913805107
- Mollard, P., Hodson, D. J., Lafont, C., Rizzoti, K., & Drouin, J. (2012). A tridimensional view of pituitary development and function. *Trends Endocrinol Metab*, *23*(6), 261-269. doi:10.1016/j.tem.2012.02.004
- Monsoro-Burq, A. H. (2015). PAX transcription factors in neural crest development. *Semin Cell Dev Biol*, *44*, 87-96. doi:10.1016/j.semcdb.2015.09.015
- Moyle-Heyrman, G., Tims, H. S., & Widom, J. (2011). Structural constraints in collaborative competition of transcription factors against the nucleosome. *J Mol Biol*, *412*(4), 634-646. doi:10.1016/j.jmb.2011.07.032
- Natoli, G. (2010). Maintaining cell identity through global control of genomic organization. *Immunity*, *33*(1), 12-24. doi:10.1016/j.immuni.2010.07.006
- Ninova, M., Fejes Tóth, K., & Aravin, A. A. (2019). The control of gene expression and cell identity by H3K9 trimethylation. *Development*, *146*(19). doi:10.1242/dev.181180
- Pabo, C. O., & Sauer, R. T. (1984). PROTEIN-DNA RECOGNITION. *Annual Review of Biochemistry*, *53*(1), 293-321. doi:10.1146/annurev.bi.53.070184.001453
- Palmer, D. K., O'Day, K., Wener, M. H., Andrews, B. S., & Margolis, R. L. (1987). A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *The Journal of cell biology*, *104*(4), 805-815. doi:10.1083/jcb.104.4.805
- Papadimitriou, A., & Priftis, K. N. (2009). Regulation of the hypothalamic-pituitary-adrenal axis. *Neuroimmunomodulation*, *16*(5), 265-271. doi:10.1159/000216184
- Parks, J. S., Adess, M. E., & Brown, M. R. (1997). Genes regulating hypothalamic and pituitary development. *Acta Paediatr Suppl*, *423*, 28-32. doi:10.1111/j.1651-2227.1997.tb18365.x

- Paull, T. T., Haykinson, M. J., & Johnson, R. C. (1993). The nonspecific DNA-binding and -bending proteins HMG1 and HMG2 promote the assembly of complex nucleoprotein structures. *Genes Dev*, 7(8), 1521-1534. doi:10.1101/gad.7.8.1521
- Peers, B., Voz, M. L., Monget, P., Mathy-Hartert, M., Berwaer, M., Belayew, A., & Martial, J. A. (1990). Regulatory elements controlling pituitary-specific expression of the human prolactin gene. *Mol Cell Biol*, 10(9), 4690-4700. doi:10.1128/mcb.10.9.4690
- Pelletier, A., Mayran, A., Gouhier, A., Omichinski, J. G., Balsalobre, A., & Drouin, J. (2021). Pax7 pioneer factor action requires both paired and homeo DNA binding domains. *Nucleic Acids Res*, 49(13), 7424-7436. doi:10.1093/nar/gkab561
- Perino, M., & Veenstra, G. J. C. (2016). Chromatin Control of Developmental Dynamics and Plasticity. *Developmental Cell*, 38(6), 610-620. doi:10.1016/j.devcel.2016.08.004
- Phillips-Cremins, J. E., Sauria, M. E. G., Sanyal, A., Gerasimova, T. I., Lajoie, B. R., Bell, J. S. K., . . . Sun, Y. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*, 153(6), 1281-1295.
- Pio, F., Ni, C.-Z., Mitchell, R. S., Knight, J., McKercher, S., Klemsz, M., . . . Ely, K. R. (1995). Co-crystallization of an ETS Domain (PU.1) in Complex with DNA: ENGINEERING THE LENGTH OF BOTH PROTEIN AND OLIGONUCLEOTIDE (∗). *Journal of Biological Chemistry*, 270(41), 24258-24263. doi:10.1074/jbc.270.41.24258
- Potok, M. A., Cha, K. B., Hunt, A., Brinkmeier, M. L., Leitges, M., Kispert, A., & Camper, S. A. (2008). WNT signaling affects gene expression in the ventral diencephalon and pituitary gland growth. *Developmental dynamics : an official publication of the American Association of Anatomists*, 237(4), 1006-1020. doi:10.1002/dvdy.21511
- Poulin, G., Lebel, M., Chamberland, M., Paradis, F. W., & Drouin, J. (2000). Specific protein-protein interaction between basic helix-loop-helix transcription factors and homeoproteins of the Pitx family. *Mol Cell Biol*, 20(13), 4826-4837. doi:10.1128/mcb.20.13.4826-4837.2000
- Poulin, G., Turgeon, B., & Drouin, J. (1997). NeuroD1/beta2 contributes to cell-specific transcription of the proopiomelanocortin gene. *Mol Cell Biol*, 17(11), 6673-6682. doi:10.1128/mcb.17.11.6673
- Proskorovski-Ohayon, R., Kadir, R., Michalowski, A., Flusser, H., Perez, Y., HersHKovitz, E., . . . Birk, O. S. (2017). PAX7 mutation in a syndrome of failure to thrive, hypotonia, and global neurodevelopmental delay. *Hum Mutat*, 38(12), 1671-1683. doi:10.1002/humu.23310
- Proulx-Ferland, L., Meunier, H., Côté, J., Dumont, D., Gagné, B., & Labrie, F. (1983). Multiple factors involved in the control of ACTH and alpha-MSH secretion. *J Steroid Biochem*, 19(1b), 439-445. doi:10.1016/0022-4731(83)90201-7
- Pulichino, A. M., Vallette-Kasic, S., Couture, C., Brue, T., & Drouin, J. (2004). [Tpit mutations reveal a new model of pituitary differentiation and account for isolated ACTH deficiency]. *Med Sci (Paris)*, 20(11), 1009-1013. doi:10.1051/medsci/200420111009
- Pulichino, A. M., Vallette-Kasic, S., Couture, C., Gauthier, Y., Brue, T., David, M., . . . Drouin, J. (2003). Human and mouse TPIT gene mutations cause early onset pituitary ACTH deficiency. *Genes Dev*, 17(6), 711-716. doi:10.1101/gad.1065603
- Pulichino, A. M., Vallette-Kasic, S., Tsai, J. P., Couture, C., Gauthier, Y., & Drouin, J. (2003). Tpit determines alternate fates during pituitary cell differentiation. *Genes Dev*, 17(6), 738-747. doi:10.1101/gad.1065703
- Quinodoz, S. A., Bhat, P., Ollikainen, N., Jachowicz, J. W., Banerjee, A. K., Chovanec, P., . . . Guttman, M. (2020). RNA promotes the formation of spatial compartments in the nucleus. *bioRxiv*, 2020.2008.2025.267435. doi:10.1101/2020.08.25.267435
- Quinonez, S. C., & Innis, J. W. (2014). Human HOX gene disorders. *Mol Genet Metab*, 111(1), 4-15. doi:10.1016/j.ymgme.2013.10.012

- Raetzman, L. T., Cai, J. X., & Camper, S. A. (2007). *Hes1* is required for pituitary growth and melanotrope specification. *Developmental biology*, 304(2), 455-466. doi:10.1016/j.ydbio.2006.11.010
- Raetzman, L. T., Ward, R., & Camper, S. A. (2002). *Lhx4* and *Prop1* are required for cell survival and expansion of the pituitary primordia. *Development*, 129(18), 4229-4239.
- Raisner, R. M., Hartley, P. D., Meneghini, M. D., Bao, M. Z., Liu, C. L., Schreiber, S. L., . . . Madhani, H. D. (2005). Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell*, 123(2), 233-248. doi:10.1016/j.cell.2005.10.002
- Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L., & Sweet, R. M. (1993). Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature*, 362(6417), 219-223. doi:10.1038/362219a0
- Reinberg, D., & Vales, L. D. (2018). Chromatin domains rich in inheritance. *Science*, 361(6397), 33-34. doi:10.1126/science.aat7871
- Relaix, F., Rocancourt, D., Mansouri, A., & Buckingham, M. (2005). A Pax3/Pax7-dependent population of skeletal muscle progenitor cells. *Nature*, 435(7044), 948-953. doi:10.1038/nature03594
- Reményi, A., Lins, K., Nissen, L. J., Reinbold, R., Schöler, H. R., & Wilmanns, M. (2003). Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev*, 17(16), 2048-2059.
- Rizzoti, K. (2015). Genetic regulation of murine pituitary development. *J Mol Endocrinol*, 54(2), R55-73. doi:10.1530/jme-14-0237
- Robert, M. F., Morin, S., Beaulieu, N., Gauthier, F., Chute, I. C., Barsalou, A., & MacLeod, A. R. (2003). DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells. *Nat Genet*, 33(1), 61-65. doi:10.1038/ng1068
- Saland, L. C. (2001). The mammalian pituitary intermediate lobe: an update on innervation and regulation. *Brain Res Bull*, 54(6), 587-593. doi:10.1016/s0361-9230(01)00471-3
- Sauvageau, M., & Sauvageau, G. (2010). Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. *Cell Stem Cell*, 7(3), 299-313. doi:10.1016/j.stem.2010.08.002
- Schaeffner, M., Mrozek-Gorska, P., Buschle, A., Woellmer, A., Tagawa, T., Cernilogar, F. M., . . . Hammerschmidt, W. (2019). BZLF1 interacts with chromatin remodelers promoting escape from latent infections with EBV. *Life Sci Alliance*, 2(2). doi:10.26508/lsa.201800108
- Schäfer, B. W., Czerny, T., Bernasconi, M., Genini, M., & Busslinger, M. (1994). Molecular cloning and characterization of a human PAX-7 cDNA expressed in normal and neoplastic myocytes. *Nucleic Acids Res*, 22(22), 4574-4582. doi:10.1093/nar/22.22.4574
- Schild-Poulter, C., Sassone-Corsi, P., Granger-Schnarr, M., & Schnarr, M. (1996). Nucleosome assembly on the human c-fos promoter interferes with transcription factor binding. *Nucleic Acids Res*, 24(23), 4751-4758. doi:10.1093/nar/24.23.4751
- Schones, D. E., Cui, K., Cuddapah, S., Roh, T. Y., Barski, A., Wang, Z., . . . Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. *Cell*, 132(5), 887-898. doi:10.1016/j.cell.2008.02.022
- Schultz, S. C., Shields, G. C., & Steitz, T. A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science*, 253(5023), 1001-1007. doi:10.1126/science.1653449
- Seale, P., Ishibashi, J., Scimè, A., & Rudnicki, M. A. (2004). Pax7 is necessary and sufficient for the myogenic specification of CD45⁺:Sca1⁺ stem cells from injured muscle. *PLoS Biol*, 2(5), E130. doi:10.1371/journal.pbio.0020130

- Seale, P., Sabourin, L. A., Girgis-Gabardo, A., Mansouri, A., Gruss, P., & Rudnicki, M. A. (2000). Pax7 is required for the specification of myogenic satellite cells. *Cell*, *102*(6), 777-786. doi:10.1016/s0092-8674(00)00066-0
- Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I. K., . . . Widom, J. (2006). A genomic code for nucleosome positioning. *Nature*, *442*(7104), 772-778. doi:10.1038/nature04979
- Seidah, N. G., Benjannet, S., Hamelin, J., Mamarbachi, A. M., Basak, A., Marcinkiewicz, J., . . . Marcinkiewicz, M. (1999). The subtilisin/kexin family of precursor convertases. Emphasis on PC1, PC2/7B2, POMC and the novel enzyme SKI-1. *Ann N Y Acad Sci*, *885*, 57-74. doi:10.1111/j.1749-6632.1999.tb08665.x
- Sekiya, S., & Suzuki, A. (2011). Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. *Nature*, *475*(7356), 390-393. doi:10.1038/nature10263
- Sekiya, T., Muthurajan, U. M., Luger, K., Tulin, A. V., & Zaret, K. S. (2009). Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. *Genes Dev*, *23*(7), 804-809. doi:10.1101/gad.1775509
- Shen, X., Liu, Y., Hsu, Y. J., Fujiwara, Y., Kim, J., Mao, X., . . . Orkin, S. H. (2008). EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol Cell*, *32*(4), 491-502. doi:10.1016/j.molcel.2008.10.016
- Shen, Y., Wei, W., & Zhou, D. X. (2015). Histone Acetylation Enzymes Coordinate Metabolism and Gene Expression. *Trends Plant Sci*, *20*(10), 614-621. doi:10.1016/j.tplants.2015.07.005
- Shipony, Z., Mukamel, Z., Cohen, N. M., Landan, G., Chomsky, E., Zeligler, S. R., . . . Tanay, A. (2014). Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. *Nature*, *513*(7516), 115-119. doi:10.1038/nature13458
- Sincennes, M.-C., Brun, C. E., Lin, A. Y. T., Rosembert, T., Datzkiw, D., Saber, J., . . . Rudnicki, M. A. (2021). Acetylation of PAX7 controls muscle stem cell self-renewal and differentiation potential in mice. *Nature Communications*, *12*(1), 3253. doi:10.1038/s41467-021-23577-z
- Smith, G. W., Aubry, J. M., Dellu, F., Contarino, A., Bilezikjian, L. M., Gold, L. H., . . . Lee, K. F. (1998). Corticotropin releasing factor receptor 1-deficient mice display decreased anxiety, impaired stress response, and aberrant neuroendocrine development. *Neuron*, *20*(6), 1093-1102. doi:10.1016/s0896-6273(00)80491-2
- Smith, M., & Flodman, P. L. (2018). Expanded Insights Into Mechanisms of Gene Expression and Disease Related Disruptions. *Frontiers in molecular biosciences*, *5*, 101-101. doi:10.3389/fmolb.2018.00101
- Solnica-Krezel, L. (2005). Conserved patterns of cell movements during vertebrate gastrulation. *Curr Biol*, *15*(6), R213-228. doi:10.1016/j.cub.2005.03.016
- Soufi, A., Donahue, G., & Zaret, K. S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell*, *151*(5), 994-1004. doi:10.1016/j.cell.2012.09.045
- Soufi, A., Garcia, M. F., Jaroszewicz, A., Osman, N., Pellegrini, M., & Zaret, K. S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell*, *161*(3), 555-568. doi:10.1016/j.cell.2015.03.017
- Steger, D. J., & Workman, J. L. (1996). Remodeling chromatin structures for transcription: what happens to the histones? *Bioessays*, *18*(11), 875-884. doi:10.1002/bies.950181106
- Stoykova, A., & Gruss, P. (1994). Roles of Pax-genes in developing and adult brain as suggested by expression patterns. *J Neurosci*, *14*(3 Pt 2), 1395-1412. doi:10.1523/jneurosci.14-03-01395.1994

- Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, *403*(6765), 41-45. doi:10.1038/47412
- Stricker, S. H., Köferle, A., & Beck, S. (2017). From profiles to function in epigenomics. *Nat Rev Genet*, *18*(1), 51-66. doi:10.1038/nrg.2016.138
- Strom, A. R., Emelyanov, A. V., Mir, M., Fyodorov, D. V., Darzacq, X., & Karpen, G. H. (2017). Phase separation drives heterochromatin domain formation. *Nature*, *547*(7662), 241-245. doi:10.1038/nature22989
- Sugawara, T., Nishino, K., Umezawa, A., & Akutsu, H. (2012). Investigating cellular identity and manipulating cell fate using induced pluripotent stem cells. *Stem Cell Research & Therapy*, *3*(2), 8. doi:10.1186/scrt99
- Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, *126*(4), 663-676. doi:10.1016/j.cell.2006.07.024
- Takahashi, K., & Yamanaka, S. (2016). A decade of transcription factor-mediated reprogramming to pluripotency. *Nat Rev Mol Cell Biol*, *17*(3), 183-193. doi:10.1038/nrm.2016.8
- Tanabe, Y., & Jessell, T. M. (1996). Diversity and Pattern in the Developing Spinal Cord. *274*(5290), 1115-1123. doi:10.1126/science.274.5290.1115 %J Science
- Taylor, I. C., Workman, J. L., Schuetz, T. J., & Kingston, R. E. (1991). Facilitated binding of GAL4 and heat shock factor to nucleosomal templates: differential function of DNA-binding domains. *Genes Dev*, *5*(7), 1285-1298. doi:10.1101/gad.5.7.1285
- Thompson, J. F., & Landy, A. (1988). Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes. *Nucleic Acids Res*, *16*(20), 9687-9705. doi:10.1093/nar/16.20.9687
- Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., . . . Stamatoyannopoulos, J. A. (2012). The accessible chromatin landscape of the human genome. *Nature*, *489*(7414), 75-82. doi:10.1038/nature11232
- Tjian, R., & Maniatis, T. (1994). Transcriptional activation: a complex puzzle with few easy pieces. *Cell*, *77*(1), 5-8. doi:10.1016/0092-8674(94)90227-5
- Tremblay, Y., Tretjakoff, I., Peterson, A., Antakly, T., Zhang, C. X., & Drouin, J. (1988). Pituitary-specific expression and glucocorticoid regulation of a proopiomelanocortin fusion gene in transgenic mice. *Proc Natl Acad Sci U S A*, *85*(23), 8890-8894. doi:10.1073/pnas.85.23.8890
- Tsunaka, Y., Kajimura, N., Tate, S., & Morikawa, K. (2005). Alteration of the nucleosomal DNA path in the crystal structure of a human nucleosome core particle. *Nucleic Acids Res*, *33*(10), 3424-3434. doi:10.1093/nar/gki663
- Underhill, D. A., & Gros, P. (1997). The paired-domain regulates DNA binding by the homeodomain within the intact Pax-3 protein. *J Biol Chem*, *272*(22), 14175-14182. doi:10.1074/jbc.272.22.14175
- Updike, D. L., & Mango, S. E. (2006). Temporal regulation of foregut development by HTZ-1/H2A.Z and PHA-4/FoxA. *PLoS Genet*, *2*(9), e161. doi:10.1371/journal.pgen.0020161
- Vallette-Kasic, S., Brue, T., Pulichino, A. M., Gueydan, M., Barlier, A., David, M., . . . Drouin, J. (2005). Congenital isolated adrenocorticotropin deficiency: an underestimated cause of neonatal death, explained by TPIT gene mutations. *J Clin Endocrinol Metab*, *90*(3), 1323-1331. doi:10.1210/jc.2004-1300
- van der Vliet, P. C., & Verrijzer, C. P. (1993). Bending of DNA by transcription factors. *Bioessays*, *15*(1), 25-32. doi:10.1002/bies.950150105

- van Holde, K., & Zlatanova, J. (1995). Chromatin higher order structure: chasing a mirage? *J Biol Chem*, *270*(15), 8373-8376. doi:10.1074/jbc.270.15.8373
- Vashee, S., Melcher, K., Ding, W. V., Johnston, S. A., & Kodadek, T. (1998). Evidence for two modes of cooperative DNA binding in vivo that do not involve direct protein-protein interactions. *Curr Biol*, *8*(8), 452-458. doi:10.1016/s0960-9822(98)70179-4
- Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., . . . Pennacchio, L. A. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, *457*(7231), 854-858. doi:10.1038/nature07730
- Vogan, K. J., & Gros, P. (1997). The C-terminal subdomain makes an important contribution to the DNA binding activity of the Pax-3 paired domain. *J Biol Chem*, *272*(45), 28289-28295. doi:10.1074/jbc.272.45.28289
- Vogan, K. J., Underhill, D. A., & Gros, P. (1996). An alternative splicing event in the Pax-3 paired domain identifies the linker region as a key determinant of paired domain DNA-binding activity. *Mol Cell Biol*, *16*(12), 6677-6686. doi:10.1128/mcb.16.12.6677
- von Maltzahn, J., Jones, A. E., Parks, R. J., & Rudnicki, M. A. (2013). Pax7 is critical for the normal function of satellite cells in adult skeletal muscle. *Proc Natl Acad Sci U S A*, *110*(41), 16474-16479. doi:10.1073/pnas.1307680110
- Vorobyov, E., & Horst, J. (2004). Expression of two protein isoforms of PAX7 is controlled by competing cleavage-polyadenylation and splicing. *Gene*, *342*(1), 107-112. doi:10.1016/j.gene.2004.07.030
- Voss, T. C., Schiltz, R. L., Sung, M. H., Yen, P. M., Stamatoyannopoulos, J. A., Biddie, S. C., . . . Hager, G. L. (2011). Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. *Cell*, *146*(4), 544-554. doi:10.1016/j.cell.2011.07.006
- Wan, H., Dingle, S., Xu, Y., Besnard, V., Kaestner, K. H., Ang, S. L., . . . Whitsett, J. A. (2005). Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. *J Biol Chem*, *280*(14), 13809-13816. doi:10.1074/jbc.M414122200
- Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T. W., Greven, M. C., . . . Weng, Z. (2012). Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome research*, *22*(9), 1798-1812. doi:10.1101/gr.139105.112
- Ward, R. D., Raetzman, L. T., Suh, H., Stone, B. M., Nasonkin, I. O., & Camper, S. A. (2005). Role of PRO1 in pituitary gland growth. *Mol Endocrinol*, *19*(3), 698-710. doi:10.1210/me.2004-0341
- Wen, S., Ai, W., Alim, Z., & Boehm, U. (2010). Embryonic gonadotropin-releasing hormone signaling is necessary for maturation of the male reproductive axis. *Proc Natl Acad Sci U S A*, *107*(37), 16372-16377. doi:10.1073/pnas.1000423107
- White, R. B., & Ziman, M. R. (2008). Genome-wide discovery of Pax7 target genes during development. *Physiol Genomics*, *33*(1), 41-49. doi:10.1152/physiolgenomics.00256.2007
- Wilson, D., Sheng, G., Lecuit, T., Dostatni, N., & Desplan, C. (1993). Cooperative dimerization of paired class homeo domains on DNA. *Genes Dev*, *7*(11), 2120-2134. doi:10.1101/gad.7.11.2120
- Wilson, D. S., Guenther, B., Desplan, C., & Kuriyan, J. (1995). High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. *Cell*, *82*(5), 709-719. doi:10.1016/0092-8674(95)90468-9
- Xu, H. E., Rould, M. A., Xu, W., Epstein, J. A., Maas, R. L., & Pabo, C. O. (1999). Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev*, *13*(10), 1263-1275. doi:10.1101/gad.13.10.1263

- Xu, W., Rould, M. A., Jun, S., Desplan, C., & Pabo, C. O. (1995). Crystal structure of a paired domain-DNA complex at 2.5 Å resolution reveals structural basis for Pax developmental mutations. *Cell*, *80*(4), 639-650. doi:10.1016/0092-8674(95)90518-9
- Xue, Y., Wong, J., Moreno, G. T., Young, M. K., Côté, J., & Wang, W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol Cell*, *2*(6), 851-861. doi:10.1016/s1097-2765(00)80299-3
- Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., . . . Taipale, J. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*, *356*(6337). doi:10.1126/science.aaj2239
- Yu, X., & Buck, M. J. (2019). Defining TP53 pioneering capabilities with competitive nucleosome binding assays. *Genome research*, *29*(1), 107-115. doi:10.1101/gr.234104.117
- Zaret, K. S. (1999). Developmental competence of the gut endoderm: genetic potentiation by GATA and HNF3/fork head proteins. *Developmental biology*, *209*(1), 1-10. doi:10.1006/dbio.1999.9228
- Zaret, K. S., & Carroll, J. S. (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes Dev*, *25*(21), 2227-2241. doi:10.1101/gad.176826.111
- Zaret, K. S., Lerner, J., & Iwafuchi-Doi, M. (2016). Chromatin Scanning by Dynamic Binding of Pioneer Factors. *Molecular Cell*, *62*(5), 665-667. doi:10.1016/j.molcel.2016.05.024
- Zaret, K. S., & Mango, S. E. (2016). Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr Opin Genet Dev*, *37*, 76-81. doi:10.1016/j.gde.2015.12.003
- Zaret, K. S., Watts, J., Xu, J., Wandzioch, E., Smale, S. T., & Sekiya, T. (2008). Pioneer factors, genetic competence, and inductive signaling: programming liver and pancreas progenitors from the endoderm. *Cold Spring Harbor symposia on quantitative biology*, *73*, 119-126. doi:10.1101/sqb.2008.73.040
- Zeng, Y., & Chen, T. (2019). DNA Methylation Reprogramming during Mammalian Development. *Genes (Basel)*, *10*(4). doi:10.3390/genes10040257
- Zhang, H., Roberts, D. N., & Cairns, B. R. (2005). Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell*, *123*(2), 219-231. doi:10.1016/j.cell.2005.08.036
- Zhang, Y., LeRoy, G., Seelig, H. P., Lane, W. S., & Reinberg, D. (1998). The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell*, *95*(2), 279-289. doi:10.1016/s0092-8674(00)81758-4
- Zhao, L., Bakke, M., Krimkevich, Y., Cushman, L. J., Parlow, A. F., Camper, S. A., & Parker, K. L. (2001). Steroidogenic factor 1 (SF1) is essential for pituitary gonadotrope function. *Development*, *128*(2), 147-154.
- Ziman, M. R., Fletcher, S., & Kay, P. H. (1997). Alternate Pax7 transcripts are expressed specifically in skeletal muscle, brain and other organs of adult mice. *Int J Biochem Cell Biol*, *29*(7), 1029-1036. doi:10.1016/s1357-2725(97)00023-x
- Ziman, M. R., & Kay, P. H. (1998). Differential expression of four alternate Pax7 paired box transcripts is influenced by organ- and strain-specific factors in adult mice. *Gene*, *217*(1-2), 77-81. doi:10.1016/s0378-1119(98)00353-9
- Ziman, M. R., Rodger, J., Chen, P., Papadimitriou, J. M., Dunlop, S. A., & Beazley, L. D. (2001). Pax genes in development and maturation of the vertebrate visual system: implications for optic nerve regeneration. *Histol Histopathol*, *16*(1), 239-249. doi:10.14670/hh-16.239