Interplay between chromatin and transcription factors drives pituitary cell fate specification

By
Alexandre Mayran

Department of Medicine, Division of Biochemistry
McGill University, Montreal

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

November 2018

Table of contents

List of Figures

Supplementary figures:

Appendix:

Appendix 3:

Main Figure:

Supplementary Figure:

List of Abbreviations

AL: Anterior Lobe

AR: Androgen Receptor

ATACseq: Assay for Transposase of Accessible Chromatin followed by high throughput sequencing

ChIPseq: Chromatin Immunoprecipitation followed by high throughput sequencing

DAR: Differentially Accessible Region

ER: Estrogen Receptor

FSH: Follicle-Stimulating Hormone

GH: Growth Hormone

GR: Glucocorticoid Receptor

IL: Intermediate Lobe

LH: Luteinizing Hormone

NIL: Neuro-Intermediate Lobe

Pax: Paired box

PF: Pioneer factor

Pitx: Pituitary Homeobox

POMC: Pro-Opiomelanocortin

Prl: Prolactin

TAD: Topologically associated Domain

TF: Transcription Factor

TSH: Thyroid-Stimulating Hormone

TSS: Transcription Start Site

À mes grands-parents

Abstract

Complex organisms begin as a single undifferentiated cell, the zygote, which derives a multitude of highly specialized cell types through sequential rounds of differentiation. Cell fate specification is the fundamental process by which cells engage towards specialization. Cell fate transitions during normal development are poorly understood. My work focused on the role of pioneer factors during differentiation. These transcription factors find their target sites even concealed in closed chromatin leading to chromatin opening and binding of nonpioneer transcription factors.

We first studied pioneer driven cell differentiation of two pituitary lineages, melanotropes and corticotropes, that both express the pro-opiomelanocortin (POMC) gene. We showed that Pax7 pioneer action drives melanotrope specification *in vivo* by opening a new enhancer repertoire. We uncovered pioneer action dynamics using an engineered inducible Pax7. Pax7 can locate its target rapidly (less than 30 minutes) but initially binds weakly. Then, Pax7 binds strongly in less than 24h while chromatin opening is slow and progressive over more than three days. Following Pax7 withdrawal, long-term memory of pioneer action is associated with loss of DNA methylation.

We then, investigated the role of nonpioneers during pioneer driven chromatin opening. The typical model of pioneer factor action assumes that nonpioneers are passive in this process and that they opportunistically bind newly accessible chromatin, yet this was never tested. The pioneer factor Pax7 acts as the specifying factor for melanotropes fate while the nonpioneer Tpit acts as the determining factor of both melanotropes and corticotropes. Chromatin accessibility is affected in Pax7 or Tpit deficient mice; they are required for the opening of their cognate program. Strikingly, in the absence of Tpit, Pax7 fails to drive melanotrope chromatin opening. Ectopically expressed Pax7 confirmed that Pax7 bind closed chromatin regardless of Tpit expression while Tpit is unable to bind closed chromatin in absence of Pax7. However, in the absence of Tpit, Pax7 does not open chromatin. In summary, we propose that cooperation between a pioneer factor and a nonpioneer factor drives lineage specific chromatin opening.

Résumé

Tous les organismes multicellulaires débutent par le zygote, une cellule non différenciée qui dans le développement embryonnaire se différencie en multiples types cellulaires spécialisés. La spécification du destin cellulaire est un processus fondamental par lequel la cellule s'engage en différenciation. Les transitions de destin cellulaire durant le développement sont mal comprises. Mon travail s'est concentré sur le rôle des facteurs pionniers durant la différenciation. Ces facteurs sont capables de trouver leurs cibles même lorsque l'ADN est inaccessible dans de la chromatine fermée et permettent l'ouverture de la chromatine et la liaison de facteurs non-pionniers.

D'abord, nous avons étudié l'implication de l'action pionnière lors de la différenciation de deux lignées hypophysaires, les mélanotropes et les corticotropes, qui expriment toute deux le gène de la pro-opiomelanocortine. Nous avons montré que l'action pionnière de Pax7 dirige la spécification des mélanotropes *in vivo* en ouvrant un nouveau répertoire d'enhancers. Nous avons découvert la dynamique de l'action pionnière à l'aide d'une version inductible de Pax7. Pax7 trouve ses cibles rapidement (moins de 30 minutes) mais initialement les lie faiblement. Pax7 est ensuite stabilisé sur ses sites en moins de 24h alors que l'ouverture de la chromatine est lente et progressive sur plus de trois jours. Après retrait de Pax7, l'action pionnière de Pax7 est stable et associées à une perte de la méthylation de l'ADN.

Nous avons ensuite étudié le rôle des facteurs non-pionniers dans l'ouverture de la chromatine dépendante des facteurs pionniers. Le modèle classique de l'action pionnière présume que les non-pionniers sont passifs dans le processus et lient la chromatine nouvellement accessible de manière opportuniste, mais cela n'a jamais été testé. Le facteur pionnier Pax7 agi comme spécificateur du destin mélanotrope alors que le facteur non pionnier Tpit agi comme facteur de détermination pour les mélanotropes et les corticotropes. L'accessibilité de la chromatine est affectée dans les souris déficientes pour Pax7 ou Tpit, chacun est requis pour l'ouverture de son programme respectif. De manière remarquable, Pax7 ne déclenche pas l'ouverture de ses régions cibles en l'absence de Tpit. L'expression ectopique de Pax7 en présence/absence de Tpit confirme que Pax7 lie de manière stable la chromatine fermée en présence/absence de Tpit alors que Tpit ne lie pas la chromatine fermée en l'absence de Pax7. En revanche, en l'absence de Tpit, la liaison de Pax7 ne permet pas l'ouverture de la chromatine. Nous proposons qu'une coopération entre pionniers et non-pionniers permet l'ouverture de la chromatine.

Acknowledgments

Je veux tout d'abord remercier le Dr Jacques Drouin pour m'avoir soutenu, et avoir cru en moi toutes ces années. Merci d'avoir été le mentor dont je n'aurais pas osé rêver. Merci de tes incontestables qualités scientifiques d'une part, mais avant tout pour ton côté humain, ton écoute, tes conseils tes critiques toujours utiles, et tes histoires de voyages dans les quatre coins du monde ou de la France. Merci de m'avoir laissé une grande liberté dans la recherche tout en étant toujours présent et disponible. Merci de m'avoir permis de m'épanouir en m'envoyant présenter dans des congrès internationaux. Grâce à toi, je me suis vu grandir et devenir, je l'espère, un scientifique capable de réflexion critiques vis-à-vis de mon projet ou de moi-même, je ne l'oublierai jamais.

Je veux également remercier le Dr Aurélio Balsalobre qui m'a grandement aidé toute ces années. Merci de m'avoir formé à l'analyse bio-informatique, merci de tes réflexions toujours pertinentes qui m'ont sans cesse permis d'avancer. Tes qualités de scientifique n'ont d'égale que ta bonne humeur.

Je remercie aussi tout particulièrement Yves, Konstantin et Maeva pour m'avoir grandement aidé tout le long de mon doctorat, que ce soit pour les dissections, la réalisation d'expériences ou m'avoir formé à différentes expériences que vous maitrisez. Merci à tous les membres actuels et passé du labo qui m'ont aidé durant mon doctorat que ce soit pour l'analyse, des expériences ou simplement les discussions qui m'ont permis d'avancer. Merci à Audrey, Kevin, Juliette, Arthur, Ryhem, Amandine, Virginie ainsi qu'Evelyne. Merci également aux membres passés dont j'ai croisé la route, Steve N., Deepak, Panojot, Shinobu, David, Lionel, Catherine, Taby et bien sûr Lise. Merci d'avoir rendu ces années agréables et stimulantes.

Je remercie également Dr Marie Kmita et les membres de son laboratoire Claudia Gentile, Dr Yacine Kherdjemil et Dr Ines Desanlis, pour nos collaborations fructueuses et pour m'avoir permis de découvrir des projets passionnants et de développer chez moi un intérêt scientifique élargi.

I also thank Dr Rushikesh Sheth for our scientific discussions and debates during conferences or at the lab thereby developing my critical thinking skills and interest in epigenetics.

I also wish to thank my co-supervisor, Dr Vincent Giguère and my research advisory committee members, Dr Xiang-Jiao Yang and Dre Marlene Oeffinger. Throughout the years, your advices have always been useful and welcomed. I also thank Dr Dillworth and Dr Witcher for taking the time, as internal and external examiners respectively, to read and comment on my thesis.

I thank the IRCM foundation and the department of biochemistry at McGill for providing me with financial support during my PhD for my stipend or for travel awards.

Je voudrais aussi remercier le Dr Nicolas Matt, la Dr Eva Berros, le Dr Daniel Metzger ainsi que la Dr Delphine Duteil, c'est vous qui m'avez fait découvrir le goût de la recherche.

Merci Dr Joern Pütz, de m'avoir offert l'opportunité de partir au Canada durant ma License et par là permis de continuer mon aventure ici.

Maëva, Rushi, Konstantin, Marine, Audrey, Yacine, Claudia, Laura et Ines, un énorme merci à vous, pour avoir été là dans les moments difficiles, les moments joyeux (ou très joyeux). Sans vous, ces années n'auraient pas été les mêmes. Les chalets, les discussions et les débats qui ont animés nos pauses-café et nos sorties font partie de ces souvenirs qu'on ne peut pas estimer. Les inséparables ne resteront pas séparés! Cheers

Merci également à Aude, François, Alex B., Sylvain, Elodie, Siamak, Thomas, Amélie, Amanda, Isabelle, Sandra, Clément, Marine L., Ludivine, Steve M., Jeanne, Mai, merci d'avoir apporté du bonheur dans mon quotidien, à la maison, dans l'institut ou dans nos sorties. J'espère vous revoir d'un côté ou de l'autre de l'Atlantique.

Enfin merci à ma famille, merci Célia, merci Antonin, merci Charles, et merci Raphaël, je me réjouis de vous retrouver sur le vieux continent. Ce fut trop long loin de vous! Merci Nono, c'est en partie grâce à toi que j'ai pu faire mon doctorat à Montréal! Merci Papa, merci Maman, c'est grâce à vous que j'en suis là aujourd'hui. Vous m'avez élevé avec des valeurs, une joie de vivre et m'avez toujours soutenu dans mes choix comme celui de partir loin et d'y rester si longtemps. Merci, je vous aime!

Contribution to original knowledge

In accordance with the “Guidelines for Thesis Preparation”, I present the results of my research in a manuscript format. This thesis was written with input from my supervisor, Dr. Jacques Drouin. I have presented a general introduction in Chapter I, which includes two published review articles. The results described in Chapter II are published and those in chapter III are submitted; both includes their respective methods. Chapter IV contains a general discussion and concluding remarks. Review articles of the introduction and results described in Chapters II and III appear in the following published or soon to be submitted manuscripts:

The published or submitted reviews and manuscripts are as followed:

1. Mayran A. and Drouin J. (2018), *Pioneer factor shape the epigenetic landscape*, JBC, jbc.R117.001232. **Review**
2. Mayran A.*, Pelletier A* and Drouin J. (2015), *Pax factors in transcription and epigenetic remodeling*, Semin. in cell and dev. biol., 2015;44:135–44 *equal contribution. **Review**
3. Mayran A., Khetchoumian K., Hariri F., Pastinen T., Gauthier Y., Balsalobre B. and Drouin J. (2018), *Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate*, Nature genetics, 50(2), 259–269. **Research article**
4. Mayran A., Sochodolsky K., Khetchoumian K., Harris H., Gauthier Y., Bemmo A., Balsalobre A. and Drouin J. *Pioneer and nonpioneers cooperation drives lineage specific chromatin opening,* (submitted), Biorxiv, **Research article**

In an appendix, a submitted manuscript with the laboratory of Dre Kmita is added:

Desenlis I.*, Kherdjemil Y.*, Mayran A.*, Gentile C., Sheth R., Zeller R., Drouin J. and Kmita M., *The pentadactyl state relies on a pioneer activity of the HOXA13 and HOXD13 transcription factors,* (submitted), *equal contribution. **Research article**

Objectives of the thesis

Patients afflicted with cellular defects or organ failure would benefit from cell-based therapy or organ transplantation. This requires finding a compatible donor and forces the patient to take immuno-suppressive drugs for the rest of their lives to avoid organ rejection of the transplanted organs. Regenerative medicine aims to bypass this by using the patient's own cells to replace defective cells or function. The ability to control cell identity is thus critical to achieve appropriate organ function.

Previous work in the laboratory revealed that Pax7 is a selector gene for melanotrope identity and that it is required to prevent the activation of the pituitary corticotrope program in melanotropes in mice. Further, Pax7 act as pioneer factor in AtT-20 cells, a corticotrope model cell line, as its binding to closed chromatin leads to the binding of another factor, Tpit, to newly accessible sites. As such, my PhD project used pituitary cell differentiation as a model system to study the molecular mechanisms of cell specification.

In Chapter II, we sought to expand on this work. We first defined the chromatin and transcriptional landscape using ATACseq and RNAseq in normal corticotropes and melanotropes. Then, we assessed from transcriptional and chromatin accessibility the consequences of Pax7 loss of functions in melanotropes. We also defined the nature of the enhancer repertoire deployed by Pax7 in Pax7 gain-of-function experiments and identify the limits to Pax7 reprogramming. An engineered inducible Pax7 allowed the dissection of Pax7 pioneer action at the level of Pax7 binding, chromatin opening and gene activation. Finally, Pax7-driven remodeling is stability at the level of chromatin accessibility and transcription factor binding.

Chapter III expands on chapter II and investigates the involvement of the nonpioneer factor Tpit during Pax7-driven chromatin remodeling within the context of determination and specification mechanisms. We first developed our transcriptional and chromatin analyses of pituitary lineages using single cell RNAseq and ATACseq analyses of adult murine pituitary cells and of purified anterior lobe cells. We investigated the role of Pax7 and Tpit for the establishment of their cognate lineages using ATACseq analysis of intermediate lobes from mice deficient for Pax7, Tpit or both. We finally defined that Tpit and Pax7 are both essential but with distinct roles during lineage-specific chromatin opening uncovering that nonpioneers are also important for pioneering. By showing that

Tpit provides the chromatin opening function, our work further restricts the unique features of pioneer action to recognition of heterochromatin sites and initial local remodeling.

Chapter I: Introduction

1 A tale of differentiation

From animals, to fungi or green algae, multicellularity has evolved several times during evolution [1]. This suggests that multicellularity provided an evolutionary advantage compared to unicellular species. Acquisition of multicellularity allowed organisms to dedicate certain tasks such as digestion, reproduction or mobility to specialized cells. Like many evolutionary transitions, this progressive specialization is recapitulated during the development of multicellular organisms. Indeed, human development starts from a unique cell, the zygote that comes from the fusion of the paternal and maternal germ cells. The zygote and its initial derivatives are totipotent stem cells, they will divide and replicate themselves many times but this mass of cells remains unspecialized, yet it has the potential to give rise to the whole set of cell types that are required for embryogenesis. The formation of the blastula constitutes the first event of specialization as it is composed of two distinct structures: the inner cell mass and trophoblasts. The inner cell mass will give rise to the embryo while the trophoblast lineage is the source of the extra embryonic structures making this the first event of cell fate specification. Cells of the inner cell mass are pluripotent; they will give rise to the numerous cell types found in the embryo. However, hundreds of cell types compose the numerous tissues and organs. As such, many more cell fate decisions need to accumulate. This creates a lineage tree of cell fate differentiation where each branch is a new decision. Every new decision will be more restrictive than the previous one and will further limit cell fate potential. This ensures a tight control of cell identity preventing cell types to switch between identities. Recent advances in single cell transcriptomics have revealed the underestimated complexity of cell types. Indeed, in the developing mouse brain alone, 100 different cell types were identified [2]. This enormous complexity requires a precise control of cell fate transitions to establish the appropriate program. In 1940, Waddington portrayed this control of cell identity as the epigenetic landscape (Figure 1).

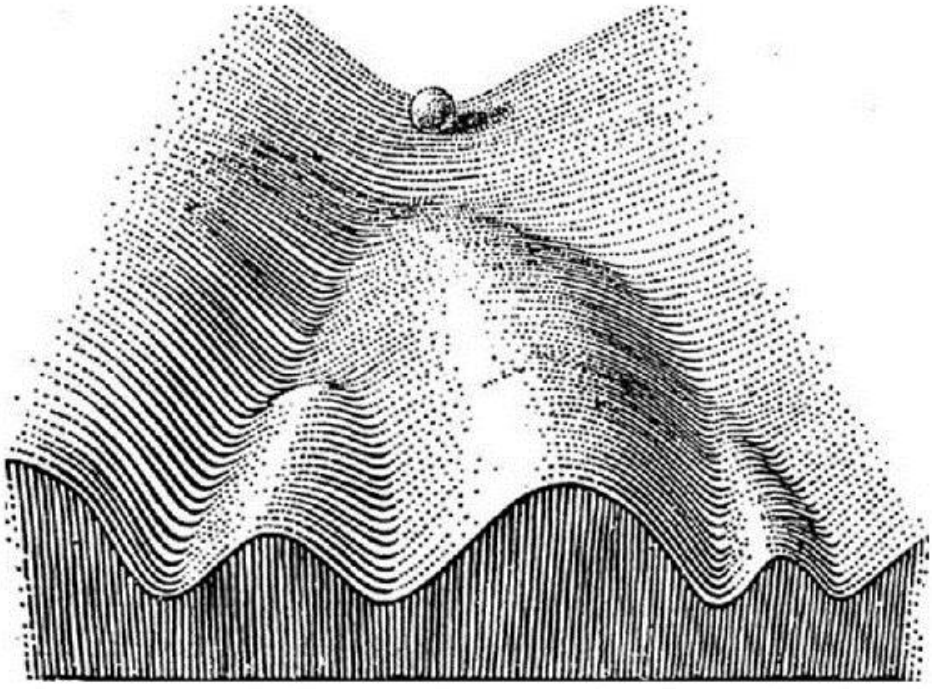


Figure 1: Waddington's view of epigenetic control of cell fate.
(From C.H. Waddington, The Strategy of the Genes; a Discussion of Some Aspects of Theoretical Biology, Allen & Unwin, London 1957)

He considered differentiation as trajectories that originate from the same starting state. According to Waddington, every decision corresponds to cells shifting towards one specific identity rather than another. Hills and valleys ensure that once engaged, each cell must follow its trajectory and that some forces must keep cells in their appropriate lineages. He envisioned that minor disturbance to trajectories would not impair the final state of the cell because of compensatory mechanisms ensuring normal differentiation.

Although the mechanistic basis of cell fate transition was not known at his time, what Waddington described as valleys correspond to transcription factors that drive cell differentiation by activating gene regulatory networks; and hills may correspond to the epigenetic landscape that can limit the reach of transcription factors.

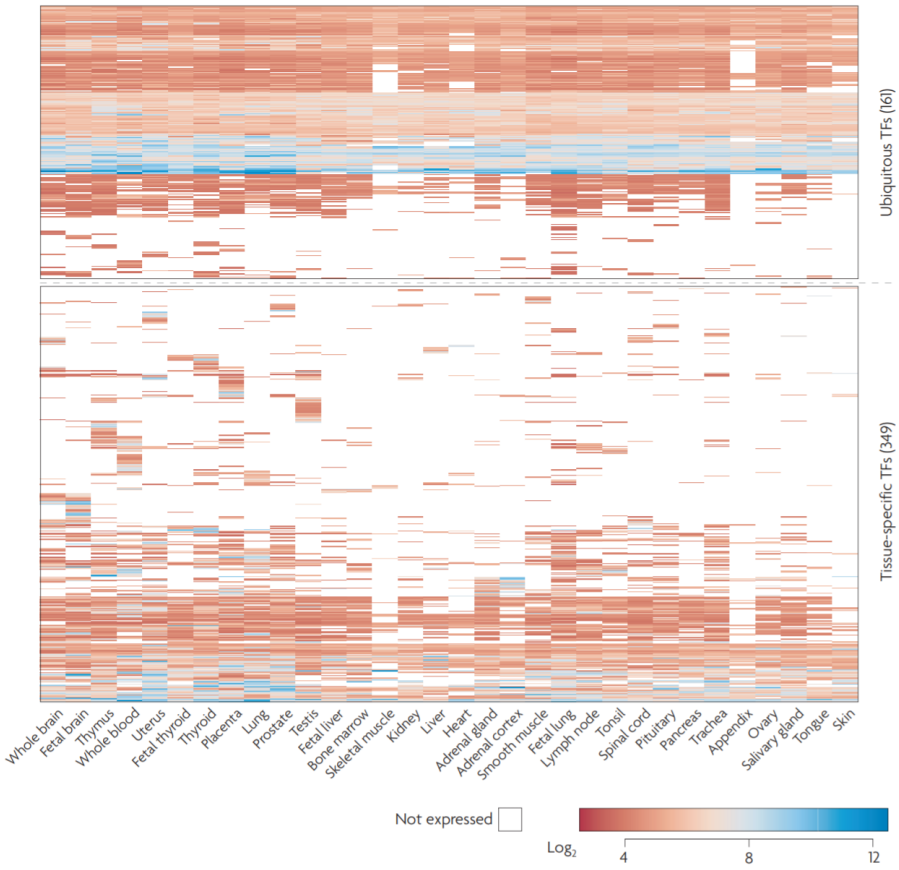


Figure 2 : Expression of human transcription factors across 32 organs and tissue from [3]

Transcription factors are an important class of proteins that binds specific DNA sequences and in turn activate or repress transcription. They are typically modular and contain a specific well-structured DNA binding domain (e.g. Zinc Finger, Helix Loop Helix, T-box, Homeodomains, Paired domain…) that is reflected by the names of transcription factor families. In addition, most transcription factors contain a transactivation domain or a repressor domain that is able to interact with co-activators or co-repressors to activate or repress gene transcription. These factors act on transcription by binding directly to the promoter of their target genes and / or, more frequently, to distal regulatory elements. Typically, enhancer-dependent trans-activation by transcription factors is thought to act by release of paused RNA polymerase that is already loaded onto the targeted promoter. Multiple families of transcription factors exist, and these exert multiple roles ranging from cell cycle

progression, cell differentiation to repression of transposable elements [4]. For the most part, they are highly conserved both at the level of their sequence and their functions. Transcription factors can respond to ligands in the case of nuclear receptor such as the glucocorticoid or the estrogen receptors responding to glucocorticoids and estrogens respectively [5, 6]. In this case, these factors regulate cell responses to external signals. Every cell type is barcoded by a unique combination of transcription factors (Figure 2, [3]).

Although many transcription factors are expressed in a given cell type, usually expression of a limited set of critical transcription factors is sufficient to drive cell fate. These master regulators can drive lineage transition, this is exemplified in the ectopic expression of Myod1 in various cell types which are then reprogramed toward the myogenic lineage [7]. Another case of induced trans-differentiation concerns transformation of fibroblast into hepatocyte-like cells by ectopic expression of HNF4a together with Foxa1/2/3 [8]. Interestingly, Myod1 and HNF4 also play a major role in normal myogenesis and hepatocyte differentiation respectively [9, 10]. Every cell type probably has its own set of critical factors that are required and/or sufficient for differentiation. However, for the sake of conciseness, we will focus on factors that are involved in normal development and differentiation of the pituitary, which was the focus of my doctoral studies.

2 Epigenetic control of transcription

The genome is the fundamental unit of all known organisms. In multicellular organisms, every cell contains the same genome, which in human is 2 meters long and compacted into chromatin fibers within a 10µm nucleus. In complex organisms, from neurons and their unique physiology to endocrine cells and their secretory capacity or the contractile ability of muscle cells, it seems impossible that the same genetic material can give rise to such dramatically different phenotypical features. However, each cell can only access a unique portion of the genome. Chromatin compaction restricts the accessible portion of the genome and occurs at different levels (reviewed in [11]). The smallest unit of chromatin compaction is the nucleosome, a model proposed by Roger Kornberg in 1974. Each nucleosome

is made of pairs of 4 histone proteins Histones H2A, H2B, H3 and H4. The nucleosome is surrounded by 147bp of DNA [12] with linker DNA between two nucleosomes, this is known as the beads-on-a-string-model. Two nucleosomes are connected by the linker histone H1 which allows the formation of higher order structures in chromatin [13] such as the 30 nm chromatin fiber [14]. Nucleosome binding to DNA is strong and stable; as such, it constrains the genome and may prevent access to the transcriptional machinery. Thus, packaging of the genome segregates the DNA into accessible and inaccessible portions.

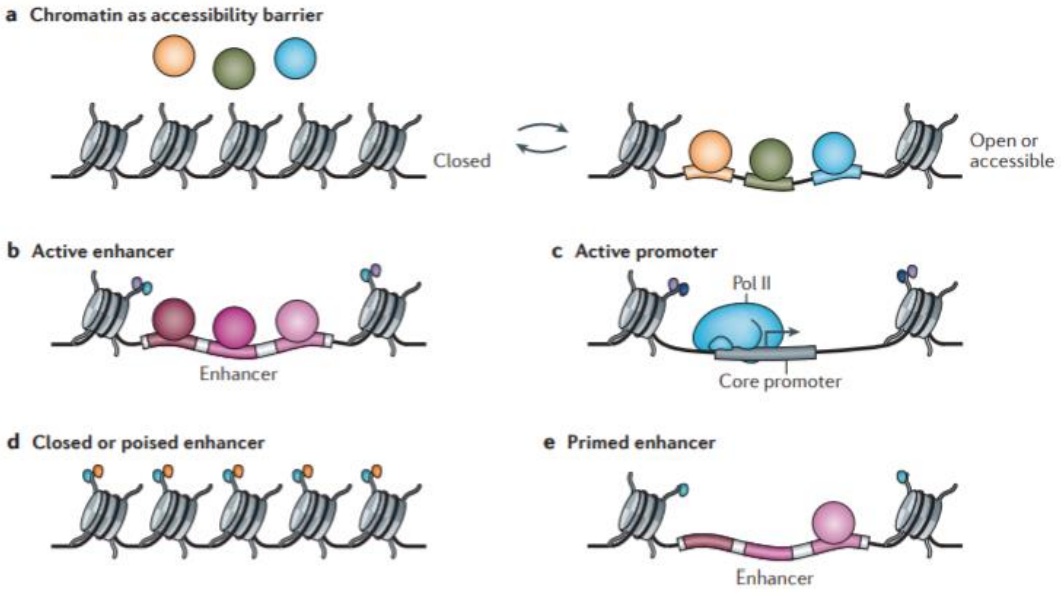


Figure 3: Chromatin accessibility and histone marks at regulatory elements from [15]

Although nucleosomes strongly associate with DNA, this binding can be modulated as nucleosomes can be evicted resulting in naked DNA that becomes accessible to the transcriptional machinery (Figure 3a). Histone variants or post-translational modifications of histone affect how nucleosomes can interact with DNA. For example, H2AZ, a variant of histone H2A and H3.3 a variant of histone H3 are often found at active promoter regions [16] and this was shown to alter nucleosome stability [17]. Post-translational modifications of histones were also associated with

different types of DNA regions establishing a histone code as characterized by David Allis [18] Promoters and enhancers are uniquely marked by a combination of histones post-translational modifications (reviewed in [15, 19], Table 1) . Active enhancers are marked by the combination of histone H3 lysine 4 mono-methyl (H3K4me1) together with H3K27ac while active promoters for example are marked with the H3K4me3 and H3K27ac marks (Figure 3b, c). Incomplete activation of enhancers and promoter corresponds to a primed state where the H3K27ac is missing from promoters or enhancers. These primed enhancers/promoters may allow a fast transcriptional response. In undifferentiated cells such as embryonic stem cells, bivalent promoters present both the active mark H3K4me3 together with the repressive mark H3K27me3 [20]. This is usually resolved during differentiation and ensures gene activation or gene repression once cell fate is decided. In addition, constitutive and facultative heterochromatin are thought to be marked with H3K9me3 and H3K9me2, respectively. Every histone modification is deposited, read or removed by different enzymes as reviewed in [19]. The histone acetyl transferase p300/CBP for example can deposit the H3K27ac mark while Polycomb group proteins deposit the repressive H3K27me3. Multiple enzymes of the MLL/Set1 family can methylate histone H3 and H4. A whole range of post-translational modifications of histone exists, and new modifications are constantly being discovered (for a detailed review, see [21]). These modifications are often highly dynamic during differentiation and understanding the mechanism for their deposition and removal is paramount to understand gene regulation in the context of differentiation and disease.

Histone modification	Genomic location	Activation/repression	Writer
H3K4me1	enhancer	Activation	MLL/Trithorax complex
H3K4m3	promoter	Activation	MLL/Trithorax complex
H3K9me2	Large domains	Repression	G9a/G9al
H3K9me3	Large domains (e.g.:centromeres)	Repression	Suv-39h
H3K27ac	Enhancer and promoter	Activation	P300/CBP
H3K27me3	(Mostly) Promoter	Repression	Polycomb (EZH2)
H3K36me3	Gene body	Activation	SETD2

Table 1. Overview of well-studied histone modifications

At every cell division, chromatin undergoes major changes. During mitosis, most transcription factors are though to be evicted from chromatin and the transcriptional machinery is inactivated by a cell cycle dependent phosphorylation cascade [22]. During mitosis, highly condensed metaphasic chromosomes replace the interphase chromosome. Yet after each mitosis, cell identity must be retained. Similarly, after replication, each newly synthesised DNA must recapitulate the original pattern of histone modifications within reassembled chromatin. It was recently proposed that repressive modifications but not activating ones are kept throughout the different phases of the cell cycle [23].

Another level of epigenetic regulation is on the DNA itself. For example, methyl cytosine (^{5me}C) is found in a wide range of organisms from yeasts to plants and animals. Although, the Drosophila genome is devoid of ^{5me}C, in mammals, ^{5me}C is essential for proper development [24], it is mostly restricted to CpG dinucleotide and found within inaccessible regions of the genome [25]. It is thought to play an active role in gene silencing [26]. DNA methylation is deposited by two mechanisms, DNMT3A/B are responsible for the *de novo* deposition of ^{5me}C [27] while DNMT1 is required for the maintenance of DNA methylation [28]. After each round of replication, DNMT1 methylates the new strand of DNA at hemi-methylated CpG dinucleotide.

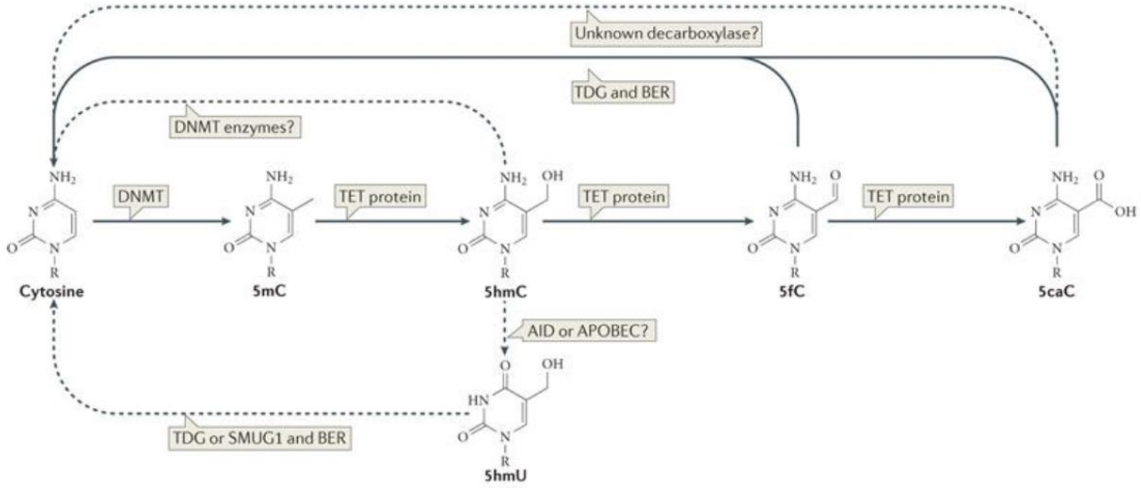


Figure 4: Tet-mediated pathway of DNA demethylation from [29]

Removal of DNA methylation can also occur in different ways. The TET family of dioxygenases can actively remove DNA methylation by altering ^{5me}C into ^{5hm}C, ^{5f}C, $5^{ca}C$ which is then thought to be replaced by a cytosine via a base-excision-repair mechanism (Fig I.2.reviewed in [29]). Another mechanism of DNA methylation removal may be through a passive mechanism that prevents DNMT1 from maintaining the original pattern of DNA methylation. DNA methylation patterns are though to be very robust as they are even transmitted between generations for example during imprinting [30]. Thus, DNA methylation plays a major role as a driver of epigenetic memory in mammalian.

3 The pituitary, a model to study the establishment of cell identity

The pituitary is a key organ responsible for the regulation of the endocrine system. Despite its small size (less than 0.5g in human), this gland impacts many biological processes such as organismal growth, reproduction, stress responses, homeostasis. It was originally described by Gallen who thought its function was to drain the phlegm secreted by the brain to the nose. It is only in the 20th century that its role as a regulator of endocrine function became apparent. The isolation of the multiple pituitary hormones starting with Prolactin (PRL) by Riddle et al. in 1933, as well as Adrenocorticotropin (ACTH) and growth hormones (GH) by Li et al. in 1943, gonadotropin (LH and FSH) by Segaloff and Steelman in 1959 and thyrotropin (TSH) in 1970 by Liao and Pierce, finally let to elucidation of its physiological function. Harrris in 1959 revealed the direct link between the pituitary and the hypothalamus, thus establishing the field of neuroendocrinology that studies the mechanisms of hypothalamic (brain) control of pituitary hormone release. Recent advances in molecular biology led to the identification of cell-specific mechanisms that regulate hormone gene expression at the transcriptional level. Each endocrine lineage is responsible for a distinct physiological function, as such each cell type has a distinct identity to express the proper hormone and

the appropriate gene regulatory network to control production and release of pituitary hormones.

Many transcription factors were found to drive pituitary cell identity (reviewed in [31]; however, cell identity also depends on chromatin environment, which defines the limits within which transcription factors can act. The small size of the pituitary made it challenging to define cell-specific chromatin landscapes through genome wide approaches. Thus, very little is known about chromatin environments in the pituitary. Despite this technical limitation, the pituitary's simple structure makes it a useful model to study cell differentiation. This section summarizes the structure and organogenesis of the pituitary and provides an overview of the different hormone-expressing lineages and the regulators of lineage identity.

The pituitary is located in the *sella turcica*, a cavity of the sphenoid bone, below the hypothalamus. Its role in the regulation of the endocrine system is mediated by the expression of hormones secreted into the bloodstream. Its main functions are to regulate organismal growth, lactation, reproduction, pigmentation, adrenal and thyroid functions: as such it is often characterized as the "master" gland. The discovery that pituitary function is regulated by neuronal signal opened a new field in physiology, neuroendocrinology [32]. Indeed, the pituitary stalk harbors a capillary network that provides direct communication between the hypothalamus and the pituitary [33]. Hypothalamic hypophysiotropic hormones (aka releasing factors/hormones) synthesized in specialized hypothalamic neurons are thus released into these capillaries and directly reach their pituitary target cells.

3.1 Pituitary organogenesis

The pituitary is separated into three lobes, the anterior and intermediate lobes which contain the hormone-expressing cells and the posterior lobe made up by the axonal projections of the hypothalamic neurons (Figure 5). However, despite its close proximity to the central nervous system, the glandular portion of the pituitary does not derive from the neural ectoderm. Rather, it originates from an invagination of the oral ectoderm [34] which become apparent at embryonic day 9 (e9) in mice.

This invagination forms Rathke's pouch named after Martin Rathke who described this structure in 1838. The invagination results from a tight contact between oral and neural ectoderms and the regions of contact becomes the intermediate lobe. This is the reason why the posterior and intermediate lobes are sometimes referred to as the neuro-intermediate pituitary. This contact between the developing pituitary and the diencephalon is essential for proper pituitary morphogenesis as, for example, human mutations in Sox3, a transcription factor that is only expressed in the diencephalon have been associated with hypopituitarism [35]. The essential role of the interaction between forming pituitary and neural tissue was also exemplified by the *in vitro* self-formation of pituitary pouches by culture association with neural cells [36]. At embryonic days e10.5, the forming pituitary quickly severs its connection from the oral ectoderm through apoptosis of the intervening epithelium. Pituitary cells rapidly proliferate in the developing fetal anterior lobe. By e12.5, the three pituitary lobes are well defined. The anterior lobe undergoes an important expansion that will continue after birth. Pituitary cell differentiation is progressive and sequential between e12.5 and e16.5 when all hormone-expressing cells are expressing their respective hormones [31, 37], and have thus acquired their unique cell identity.

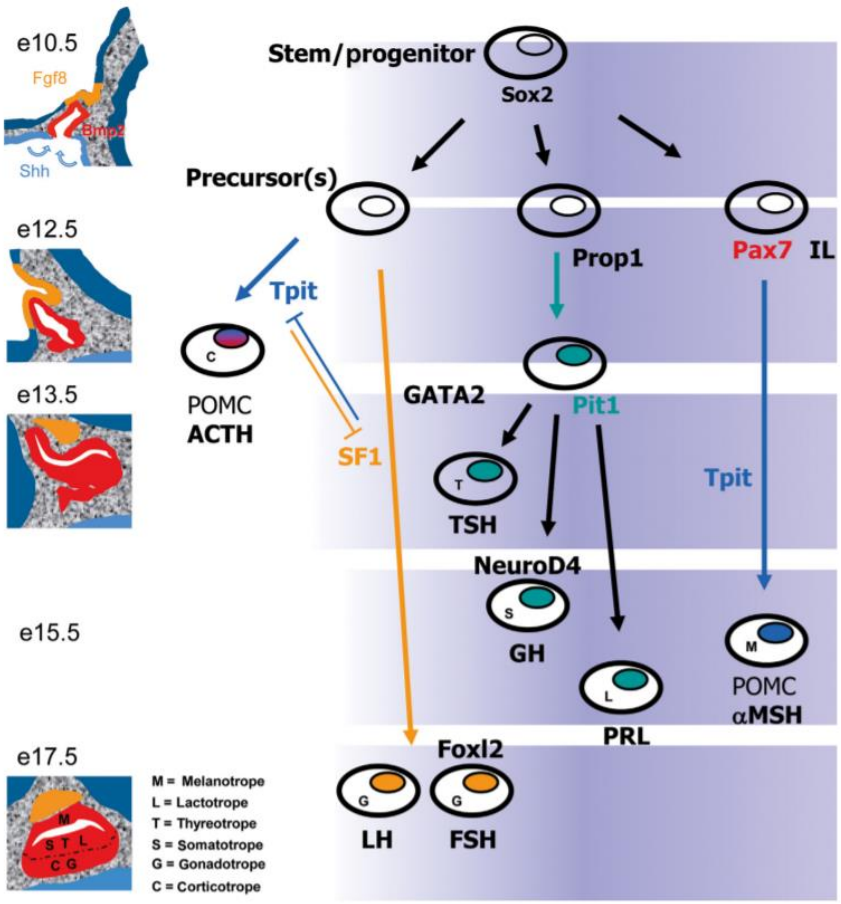


Figure 5: Development and differentiation of the pituitary from [31]

3.2 Composition and differentiation of pituitary cells

3.2.1 The posterior lobe

The posterior lobe is mostly composed of axonal projection of hypothalamic neurons. However, support cells (pituicytes) are also present. These neurones release oxytocin and vasopressin into the pituitary capillary bed [38]. These hormone can act on the pituitary as well as on peripheral organs; for example, when oxytocin is released from the posterior lobe, it is involved in cervical dilatation for parturition in women while in male it stimulates erection and participate in ejaculation (reviewed in [39]). Vasopressin acts on corticotrope cells (that are described below) of the pituitary increasing ACTH secretion [40].

3.2.2 The intermediate lobe

The intermediate lobe or *pars intermedia*, as indicated by its name is located between the posterior and anterior lobes of the pituitary. It is relatively homogeneous, being mostly composed of melanotrope cells as well as some Sox2-expressing stem cells at the lumen close to the anterior lobe. Melanotropes are one of the two pituitary lineage that express the hormone precursor pro-opiomelanocortin (POMC), the other being the corticotrope that is in the anterior pituitary. In melanotropes, the protein convertase 1 (PC1) and 2 (PC2 encoded by the *Pcsk2* gene) cleaves POMC into α-MSH, β-endorphin and γ-lipotropin. αMSH stimulates melanocyte production and release of melanin that regulates pigmentation. In some species such as *Xenopus*, this can serve a function of adaptation to the environment by changing the skin color in response to the color of their surrounding. In a dark background, the *Xenopus* intermediate lobe expands to increase the expression and release of melanin to produce darker skin, thus allowing frogs to mimic their environment [41, 42].

In humans, the pituitary develops during embryogenesis similarly as in mice forming an intermediate lobe. The presence of an intermediate lobe in adult was a subject of controversy. In the 30's, it was described by Rasmussen that colloid-filled cysts in the posterior lobe of the human pituitary are the remains of the human pars intermedia but some have argued that an intermediate zone exists with αMSH-producing cells [43], but this is no longer accepted. In human, pigmentation is regulated locally in response to UV exposure and local production of αMSH. In mice, melanotropes start expressing POMC at e15.5 about 12 hours after the Tbox transcription factor Tpit is expressed [44, 45]. This transcription factor is highly cell-specific. Indeed, to this date, melanotropes and corticotropes (that are discussed in the next section) are the only two lineages where it is expressed. Tpit regulates POMC transcription in cooperation with the pituitary homeobox 1 (Pitx1) transcription factor [44]. Tpit deficient mice fail to express POMC and to differentiate melanotropes and corticotropes. Instead, a large part of the intermediate lobe switches fate and become gonadotropes as well as a few

clusters of thyrotropes [46]. Another part of the intermediate lobe remains in a progenitor stem cell-like state as seen from the invasion Sox2-expressing cells that are normally restrained to the luminal portion of the gland. In addition, melanotropes cells are marked by the expression of the paired-box transcription factor Pax7; Pax7 is expressed just before Tpit is turned on. Pax7-deficient intermediate pituitary cells switch identity towards the corticotrope fate. Finally, over-expression of Pax7 in AtT-20, a corticotrope derived cell line leads to a cell fate change that is melanotrope-like. This happens through chromatin opening. This allowing Tpit binding to newly accessible regulatory elements and activation of melanotrope hallmark genes such as *Pcsk2* [45].Thus, Pax7 acts as a pioneer factor and is a master regulator of the melanotrope fate.

3.2.3 The anterior lobe

Like the intermediate lobe, the anterior lobe is a hormone-expressing structure that derives from the oral ectoderm. It contains five hormone-expressing lineages, the somatotropes, lactotropes, gonadotropes, thyrotropes and corticotropes. They each express a different hormone that regulates a different endocrine system. Hormone production and release are tightly regulated, by the central nervous system through the hypothalamus and in feedback loops from the periphery. Defects in the control or production of these hormones can arise from mutation in the hormone-coding genes but also, and more frequently in genes encoding regulators of the function and differentiation of these cells. Here, we will discuss the functions of the different hormone expressing lineages of the anterior lobe as well as the molecular mechanisms of their differentiation.

3.2.3.1 Corticotropes

The corticotropes represent about ten percent of anterior pituitary cells. With the melanotropes, they are the other pituitary lineage that express the POMC gene as well as the protein convertase PC1 and the transcription factor Tpit. However, unlike melanotropes, they do not express the transcription factor Pax7 or the protein convertase PC2. Consequently, they process the POMC into ACTH (not

αMSH), pro-γ-MSH and β-LPH. ACTH acts on the adrenals and mediates the stress response by triggering the production of glucocorticoids. The hypothalamus can stimulate POMC expression and ACTH secretion from corticotropes by the release of corticotropin-releasing-hormone (CRH) and/or vasopressin [47]. These are recognized by the corticotropin-releasing-hormone receptor 1 (Crhr1) and the vasopressin receptor 1B (Avpr1b), respectively. The transcription factor Nurr77 mediates Crhr1-induced signal and transcriptional action on POMC expression [48]. Conversely, the glucocorticoid receptor (GR) binds glucocorticoids produced by the adrenals and negatively regulates the expression POMC through direct binding to the POMC promoter antagonising Nurr77 function [49]. In addition, glucocorticoids also act on hypothalamic neurons to inhibit the secretion of CRH. Defects in the expression or secretion of ACTH can lead to hypo- or hypercortisolism, ie. insufficient or excess levels of circulating glucocorticoids.

Excess of ACTH can arise from pituitary adenomas such as those found in Cushing disease patients. Those adenomas develop from corticotropes and become resistant to glucocorticoid feedback leading to an increase of ACTH secretion and circulating glucocorticoids. Conversely, two-thirds of isolated ACTH deficiency case are caused by a mutation in the TPIT gene[50]. Indeed, Tpit being required for expression of POMC, mutations that disrupt Tpit DNA binding or function prevents Tpit-dependent expression of POMC. This genetic disease is especially critical in newborns where it typically causes severe hypoglycemia and death. When diagnosed, hormone replacement therapy is an effective and life long treatment.

During development, corticotropes are the first pituitary cells to differentiate. At e12.5, they express Tpit and quickly after, POMC is detected. Expression of POMC also requires E-box transcription factors [51]; indeed, Neurod1, an E-box transcription factor, is expressed in the anterior pituitary during embryonic development and activates POMC transcription early on. At e14.5, in mice deficient for Neurod1, corticotropes fail to activate POMC expression although Tpit is not affected. However by e17.5, an unknown E-box factor provides redundancy and

POMC is then expressed [52]. It was shown that the E-box factor Ascl1/Mash1 also participates in POMC transcription and Ascl1 is expressed during pituitary development [53].It is thus possible that Ascl1 is the redundant E-box factor allowing the late POMC transcription in Neurod1-deficient mice.

3.2.3.2 Somatotropes

The somatotropes are the most abundant cells in adult male pituitary (50% of the anterior lobe) and they are responsible for production of growth hormone (GH). This hormone plays a role in overall organismal growth. It is recognized by the growth hormone receptor in hepatocytes, where this triggers the secretion of insulin-like growth factors (IGF-1.) It is positively regulated by Ghrelin and by the growth hormone releasing hormone (GHRH), and negatively by somatostatin and IGF-1 [54]. Over-production of GH in somatotrope cells can lead to gigantism (acromegaly) while GH deficiency is associated with short stature (dwarfism). This hormone deficiency is often associated with prolactin and TSH deficiency in a syndrome of Combined Pituitary Hormone Deficiency (CPHD). Like lactotropes and a population of thyrotropes (discussed below), somatotropes express the transcription factor Pit1. During pituitary differentiation, Pit1 is turned on at e13.5 in the medial region of the developing anterior lobe. Pit1 activates the transcription of the growth hormone and prolactin genes [55]. Consequently, mutations in PIT1 in mice, human and even in zebrafish lead to CPHD or combined hormone deficiencies of GH, PRL and TSH [56-58]. As its names indicate, the transcription factor Prophet-of-Pit1 (Prop1) is required for the expression of Pit1 [59] and thus, human mutation in PROP1 also leads to CPHD [60, 61]. Neurod4 is also expressed and required for maturation and proliferation of somatotropes and its expression depends on Prop1 [62].

3.2.3.3 Lactotropes

Lactotropes constitute 15-20 percent of the male anterior lobe. However in lactating females, these cells proliferate intensively increasing their proportion in the anterior lobe. They express the hormone prolactin, that is recognized by the

prolactin receptor leading to the expansion of the mammary gland as well as milk secretion. Prolactin expression and release is directed by the prolactin releasing hormone recognized in lactotropes by the prolactin release hormone receptor [63] and negatively regulated by dopamine released from the hypothalamus and recognized by the dopamine D2 receptor (Drd2) [64]. Like somatotropes, lactotrope differentiation and prolactin expression require Pit1 [58]. Pit1 and the estrogen receptor (ERα) synergize on Prolactin enhancer sequences to stimulate PRL gene transcription [65]. ERα gene inactivation leads to a decrease of Prl expression [66].

3.2.3.4 Thyrotropes

Thyrotropes constitute the least abundant hormone-expressing pituitary cells making up 1-5% of the anterior lobe. They express the thyroid-stimulating hormone (TSH) that is composed of two sub-unit, the first one, TSHβ is uniquely expressed in this lineage while the second subunit, αGSU, is also part of the two gonadotrope hormones. As its name indicates, TSH regulates thyroid function leading to the secretion of thyroxine and triiodothyronin. These two hormones negatively regulate TSH expression while it is positively regulated by the tyrotropin-releasing hormone (TRH) secreted from hypothalamic neurons. Thyroid hormones are involved in a wide range of physiological processes such as metabolism and thermogenesis.

While thyrotrope differentiation requires Pit1 as it is one of the lineages affected by the loss of Pit1 [58], a pool of thyrotrope cells seem to be independent of Pit1 [67]. However, thyrotrope differentiation also depends on the transcription factor GATA-2 which is also expressed in gonadotropes. Indeed, conditional deletion of GATA-2 with an αGSU-cre transgenic mouse leads to reduction in the number of thyrotrope and gonadotropes [68]. Interestingly, the thyrotropes are one of the two lineages (with the gonadotropes) that are ectopically found in the intermediate lobe of Tpit-deficient mice [44]. This suggests that the Pit1-independent TSH lineage may be related to the gonadotropes and corticotropes differentiation path.

3.2.3.5 Gonadotropes

Gonadotropes are the fifth endocrine lineage of anterior lobe cells; they represent about 10% of the anterior lobe. They play a major role in the regulation of reproductive function through the combined action of the two gonadotropins: LH and FSH. Like TSH, these two hormones are also made from the combination of two subunits: αGSU (which is shared with TSH) and LHβ or FSHβ as the second subunit. Their synthesis is under control of gonadotropin releasing hormone (GnRH) from the hypothalamus recognized by the gonadotropin releasing hormone receptor (GnRHR). Each hormone plays a distinct role for control of the reproductive system. LH stimulates steroid hormone synthesis in Leydig cells in males and theca cells in females while FSH stimulates spermatogenesis in males and follicle maturation in females. In response to circulating sex hormones, a negative feedback loop decreases expression and secretion of LH and FSH in the pituitary as well as GnRH in the hypothalamus. Insufficient expression of LH and FSH will lead to reproductive defects such as hypogonadism and absence of secondary sex characteristics [69].

At e16.5, gonadotropes are the last lineage of the anterior lobe to express their hallmark hormones [37]. However, they may become specified towards gonadotrope fate earlier by the expression the gonadotrope marker steroidogenic factor 1 (SF-1, Nr5a1) at e13.5 [70]. SF1 is an orphan nuclear receptor that is required for the differentiation of gonadotropes. SF1-deficient gonadotropes fail to express LH and FSH [71, 72]. Interestingly, this phenotype is rescued by administration of GnRH which allow expression of LH and FSH in absence of SF1 [73]. Also, as mentioned above, GATA2 contributes to gonadotrope differentiation as the expression of LH and FSH is reduced in GATA2-deficient gonadotropes [68]. Pitx1 also plays a role in defining the proportion of cells that differentiate towards the gonadotrope fate as the number of gonadotropes is reduced in Pitx1-deficient mice [74].

4 Pioneer factors shape the epigenetic landscape

Alexandre Mayran[1] and Jacques Drouin[1]

J Biol Chem. 2018 Sep 7;293(36):13795-13804. doi: 10.1074/jbc.R117.001232. Epub 2018 Mar 5.

[1]Laboratory of Molecular Genetics, Institut de Recherches Cliniques de Montréal 110, avenue des Pins Ouest, Montréal, QC H2W 1R7, Canada

Running title: Pioneer transcription factors

Correspondence can be addressed to: Jacques Drouin (jacques.drouin@ircm.qc.ca)

Keywords: chromatin, remodeling, nucleosome, transcriptional activation, DNA methylation, development

Pioneer transcription factors have the unique and important role of unmasking chromatin domains during development to allow the implementation of new cellular programs. Compared to those of other transcription factors, this activity implies that pioneer factors can recognize their target DNA sequences in so-called compacted or "closed" heterochromatin and can trigger remodeling of the adjoining chromatin landscape to provide accessibility to non-pioneer transcription factors. Recent studies identified several steps of pioneer action, namely rapid but weak initial binding to heterochromatin, stabilization of binding followed by chromatin opening and loss of CpG methylation that provides epigenetic memory. Whereas CpG demethylation is dependent on replication, chromatin opening is not. In this review, we highlight the unique properties of this transcription factor class and the challenges of understanding their mechanism of action.

In the late nineteen seventies when chromatin structure was beginning to be probed with tools such as DNase (hyper) sensitivity, the concept of pioneer factors was suggested as factors that would have the capacity to bind specific DNA sequences within compacted heterochromatin and initiate the opening of this chromatin, such opening or remodeling being required for implementation of major developmental fate decisions. At the same time, drosophila geneticists developed the notion of selector genes for early developmental regulators that in some way specify the outcome of future cell fates through their action on broad embryonic domains (1). In this context, the notion of pioneer factors offered a possible mechanism to achieve the purpose of selector genes but these ideas remained more in the domain of evening conversations than experimental reality. For clarity, it should be mentioned that in more recent years the term selector has been used by some to identify factors that have the opposite effect in the differentiation scheme compared to the original definition, namely factors that trigger the ultimate step in cell fate decisions (e.g. (2)).

The idea of pioneer action was revived in the late nineties when the transcription factor (TF) FoxA was shown to have the unique ability to bind its target sequence within nucleosomal DNA (3). This unique ability contrasted with many other TFs that will only bind efficiently naked or more readily accessible DNA as observed within active regulatory sequences. However, pioneer factors do not have completely unrestricted access to heterochromatin sites but do exhibit cell-specific actions (4). In parallel, the old binary view of chromatin as either hetero- or euchromatin changed dramatically as the enormous diversity of histone modifications became known, eventually leading to the concept of a histone code (5) that defines a continuum of chromatin flavors associated with regulatory and structural functions. The complexity of this code and the limited tools available to characterize chromatin limit our present ability to define the permissive or restrictive chromatin states that are targeted by pioneer factors. Despite this limitation, the basic features that define pioneer factors (Fig. 1) are the ability 1) to bind specific DNA sequences within “closed” or unmarked chromatin where genomic DNA is not readily accessible, 2) to initiate chromatin remodeling leading

to DNA accessibility and consequently, 3) to allow binding of other transcription factors, and finally 4) to establish stable changes in chromatin structure associated with DNA accessibility and epigenetic stability. Collectively, these features imply that the “act of pioneering” may be a one-shot affair, ie once enacted, its effect on chromatin remains stable. Mechanisms for maintenance of chromatin state at pioneered sites may also exist. This review will discuss the unique aspects of pioneer action and attempt to separate these from the transcriptional actions of the same factors since pioneers do act as transcriptional regulators like other TFs and often at the pioneered as well as other target sites. The list of TFs that share at least some features of pioneers is provided in Table 1.

Since the measure of chromatin features such as DNA accessibility and chromatin marks are not just absent versus present but also present on a continuous scale, the expectation of pioneer function must be more clearly defined. Indeed, DNA accessibility [whether measured by DNase sensitivity(6), formaldehyde-assisted isolation of regulatory elements (FAIRE(7)) or the ATAC procedure(8)] or histone marks such as histone H3K4me1 that marks active enhancer sequences (9) exhibit greater values as the activity of enhancers or the number of bound TFs increases (eg (10)): increments in these marks may reflect quantitative changes in enhancer activity rather that the switch from “closed” naïve chromatin to a state of accessibility. The label pioneer should thus be reserved for factors and actions shown to elicit chromatin opening from a state of complete absence of accessibility marks to the presence of such marks. On the genome scale, it is thus very important to separate the targets of pioneer action from those where the same pioneer factors only exert classical transcriptional activity at already accessible regulatory sequences: this requires an assessment of chromatin status before and after pioneer factor action in an experimental system dependent on cells that have never been exposed (in their developmental history) to the pioneer if epigenetic memory is indeed a pioneer property. Failing that, a pioneer activity may be inferred but formal demonstration requires the before and after comparison.

Pioneers set the stage: assisted-loading and settler factors

Pioneers appear to share the property of interacting with other TFs as do most TFs: although very important from the biological perspective, this property is not a defining feature of pioneers. For example, the pioneer FoxA interacts with nuclear receptors such as glucocorticoid (GR), estrogen (ER) or androgen (AR) receptors and this allows recruitment of these nuclear receptors at subsets of enhancers that (11-14) that establish hormone-responsive gene regulatory networks (15). In this context, FoxA pioneers the opening of subsets of enhancers targeted by the hormone responsive receptors. This subsequent binding of nuclear receptors has been labeled as “assisted loading” (16) and the factors that require the open chromatin state were labeled as “settler factors” (17). The binding of a settler factor may be essential in the biological context, but it does not constitute the core pioneer activity that is restricted to initiation of chromatin opening. However for this specific example, it appears that the interaction between FoxA and nuclear receptor may be reciprocal as nuclear receptors can also recruit FoxA to specific subsets of enhancers (18).

The pluripotency factors

The reprograming of diverse cells such as fibroblasts into induced pluripotent stem cell (iPS) revealed the unique ability of a group of factors to reverse the differentiation process (19) toward a pluripotent state. These pluripotency factors (OSK for Oct4, Sox2 and Klf4) initiate the remodeling (opening) of both enhancers and transcription start sites (TSS). This was revealed by deposition of H3K4me1/2 at targeted enhancers and of H3K4me 2/3 at TSS (20). The initial binding of these factors occurs widely at unmarked (“closed”) chromatin to initiate their remodeling: they thus act as pioneer factors (21). The initial binding of OSK factors is followed by a lengthy period (weeks) of iPS cell selection that leads to the remodeling of large chromatin domains of the epigenome from a somatic to pluripotent state. Interestingly, it was also reported that some of the iPSC reprogrammed enhancers require the expression and binding of more than one of the OSK factors. This suggests that pioneers may also require a cooperative action in order to remodel the epigenome (22). This process opens new sites for OSK binding together with

sites for the accessory factor cMyc (21). There are broad domains where OSK factors cannot bind early in reprograming but only in the iPSC state. These domains have high levels of H3K9me3 and this may thus constitute a barrier that contributes to refractoriness to OSK binding. Indeed, knockdown of the histone methyl transferases SUV39H1/H2 that are responsible for H3K9me3 deposition allows binding at previously inaccessible sites (20,21,23). The pluripotency factors Oct4 and Sox2 have critical roles in normal development to activate the zygote genome (24). In drosophila, the factor Zelda has a similar role for induction of the zygote genome (25) and this is achieved through a pioneer mechanism of chromatin opening (26).

Lineage-specifying pioneer factors

The first indication that FoxA factors have pioneer activity came from showing that liver-specific FoxA binding sites are occupied in endoderm before liver-specification (27). FoxA was then shown to bind nucleosomal DNA (28) and to open compacted chromatin (29). Genome-wide studies then showed its chromatin remodeling activity (4,30,31) as well as the associated nucleosome depletion (32). In *C. Elegans*, the FoxA-related factor PHA-4 is also critical for foregut development and this is achieved through pioneer action (33). Interestingly, the pioneer action of PHA-4 is mostly exerted over promoter regions and this leads to recruitment of RNA polymerase II (34).This recruitment initially leads to a poised state where RNA polymerase is paused on the promoter early on and transcription only occurs later in foregut development.

GATA4 is also present at liver-specific enhancers in early endoderm but its binding appeared supported by FoxA and did not show as strong an ability to bind nucleosomal DNA (29). GATA4 together with GATA6 are required for early liver development (35-37). Hence, GATA factors appear to have pioneer properties although it may not be as effective as FoxA. Nonetheless, GATA4, like FoxA, can induce trans-differentiation into hepatocytes (38,39).

Specification of the lymphoid, in particular macrophage, lineages is dependent on the factors PU.1 and C/EBPα. PU.1 is critical for development of these lineages

(40). It initiates chromatin remodeling and is associated with deposition of active enhancer marks (41). Indeed, PU.1 increases chromatin accessibility and promotes nucleosome depletion (42). C/EBPα can also trigger trans-differentiation into that lineage and its binding to macrophage enhancers during that process is associated with deposition of the active enhancer marks H3K4me1 and H3K27ac. Thus, C/EBPα and PU.1 independently act as pioneer for the other during macrophage differentiation (43).

The pituitary intermediate lobe is specified to a unique developmental fate by the pioneer factor Pax7 (44). This is achieved through binding and chromatin remodeling of a subset of *de novo* active melanotrope specific enhancers. The opening of these enhancers allows for recruitment of the differentiation determination factor Tpit that achieves terminal differentiation of this lineage. Pax7 pioneering results in appearance of DNA accessibility together with deposition of active enhancer histone marks (45).

Establishment of the B cell lineage requires the transcription factor EBF1 (46) and this was associated with chromatin remodeling and increased enhancer H3K4me2 (47). Some EBF1 pioneer actions were shown to depend on an EBF1 C-terminal domain that is required to trigger DNA accessibility and deposition of active chromatin marks at a specific subset of enhancers. Both C-terminus dependent and independent pioneer sites were enriched for the same EBF1 motif suggesting that the EBF1 DNA binding site is not the defining factor between dependence and independence on the C-terminal domain. This supports a model where different pioneer interacting proteins may define functionally distinct subsets of pioneered enhancers (48).

Two neurogenic bHLH transcription factors shown to reprogram fibroblast to the neuronal fate appear to have pioneer activity. Indeed, Ascl1 is the driver of neuronal differentiation in association with Brn2 and Myt11, and its recruitment was associated with increased DNA accessibility (FAIRE) and with increased active chromatin marks H3K4me1 and H3K27ac together with decreases in the repressive mark H3K9me3 (49). The neurogenic factor NeuroD1 was also shown

to induce similar chromatin changes at enhancers and promoters during neuronal reprograming (50).

Pioneer interactions with DNA and chromatin

The pioneer factor activity was inferred from *in vitro* and *in vivo* footprinting experiments that showed FoxA and GATA sites co-occupancy prior to hepatic specification (51). Flanking TF sites were only occupied once cells are specified towards liver identity, suggesting that pioneer factors have the unique ability to bind “closed” or naïve chromatin (21,44). Despite many genome-wide studies, the nature of this naïve or closed chromatin remains vague and the ability of pioneers to bind specific chromatin states is still defined by the negative, ie the absence of recognizable chromatin marks, and in some cases, the presence of methylated cytosines in DNA. Indeed as discussed below, some pioneers can bind methylated target DNA while others appear to be methylation-sensitive.

Pioneer factors tend to have higher residency time or chromatin mobility than other TFs (18,52) suggesting that stable chromatin-pioneer interactions may be critical for pioneer function. These stable pioneer- chromatin interactions may be explained by direct nucleosome binding, as shown for FoxA and the OSK pluripotency factors (29,53). For FoxA, nucleosomal interaction may partly rely on a FoxA domain that resembles a linker histone H1 structure (31,54). For the OSK factors, their ability to target partial consensus motifs may allow their DNA-binding domains to interact directly with nucleosomes (53).

In one instance, target DNA motif preference may play a role in binding stability: indeed, the pioneer Pax7 preferentially recognizes a composite motif composed of binding sites for its two DNA-binding domains, the homeo and paired domains, leading to greater binding stability and possibly allowing for pioneer action (44).

Epigenetic remodeling by pioneer factors

Pioneer factors provide competency for gene expression but their binding to closed chromatin is not in itself sufficient. Indeed, chromatin remodeling is required in order to allow non-pioneer TF binding and transcriptional activation at newly competent regulatory sequences, primarily enhancers. The remodeling or

activation of regulatory sequences from a naïve or “closed” chromatin state appears to be a stepwise process (Fig. 2). None of the pioneers characterized so far have unrestricted access to the genome in heterochromatin: this was shown for FoxA, Pax7, and the pluripotency factors. This aspect is discussed in the next section.

The first step in pioneer action is the initial binding (Fig. 2C) to permissive heterochromatin (Fig. 2B) and it appears to be rapid (eg less than 30 min. for Pax7, (45)). This is followed by a phase of binding stabilization (within 24h for Pax7) that may or may not be paralleled by nucleosomal changes that increase accessibility (31) and to appearance of low levels of the H3K4me1 mark in the center of target enhancers (Fig. 2D). These “Accessible” or “Primed” enhancers can undergo the final step of enhancer activation that involves the binding of other non-pioneer TFs, nucleosome depletion and deposition of the active enhancer mark H3K27ac that is associated with the histone acetylase activity of the general coactivator p300 (Fig. 2E).

As most TFs, pioneers interact with chromatin remodeling proteins that are found within large complexes. These complexes have been associated with the process of transcription itself and/or its activation / initiation; the same complexes or different ones may be critical for the initial act of pioneering as well as for continued transcriptional action of pioneers. The challenge is thus to find experimental systems to separate these two actions. For example, the BRG1 ATPase of the SWI/SNF complex co-occupies many sites together with Oct4 in ES cells (55) and knock-down of BRG1 affects ES cell pluripotency (56). Oct4 is required for maintenance of open chromatin at enhancers in ES cells and its inactivation leads to loss of accessibility at these enhancers (57). Oct4 pioneer function is thus dependent on the chromatin remodeler Brg1. Similarly, the INO80 remodeling complex co-occupies many sites in common with pluripotency factors and its knock-down decreases chromatin accessibility at those sites (58), suggesting that the complex may increase accessibility following recruitment by the pluripotency

factors. Similarly, GATA3 was shown to require BRG1 for cell reprogramming through pioneer action (59).

The Trithorax (drosophila) complex (COMPASS in yeast and MLL in mammals) is involved in activation (opening) of chromatin structure (60). Pax7 was suggested to recruit the MLL1/2 complex through interaction with its component protein WDR5 (61) and FoxA1 directs H3K4me1 deposition through recruitment of MLL3 at enhancers (62). Indeed, this complex has H3K4me1 methylation activity and thus may lead to enhancer activation. It may also be implicated in pioneering as its component protein Ash2l is recruited to sites of Pax7 pioneering (45).

For transcriptional activation, chromatin accessibility is increased at both promoters and enhancers by recruitment of the variants histones H2A.Z and H3.3 that form unstable nucleosomes (63). FoxA factors (32,64,65) and CLOCK:BMAL1 (66) promote recruitment of H2A.Z. This likely contributes to nucleosome instability and loss but it is not clear that this is critical for pioneer action per se. Indeed, FoxA-dependent nucleosome instability is not correlated with H2A.Z deposition and in this particular case, increased nucleosome accessibility may result from displacement of the linker histone H1 (31). FoxA factors have the unique property of containing a H1 mimic region that binds nucleosomes (28).

Barriers to pioneer binding and action

While pioneers have the unique ability to bind their target sequence within nucleosomes in contrast to many TF that cannot, this does not mean that pioneers can bind all their target sequences in the genome. Indeed, pioneers show different binding repertoires in different cell types. For example, Sox2 binds different target subsets in mouse cortex and spinal cord (67), indicating that there are additional constraints on pioneer binding. Further, the pluripotency factors OSK have a large subset of targets that only become accessible in the late phase of reprograming towards iPS (21). The OSK binding sites within these latter binding regions initially have higher levels of the repressive histone mark H3K9me3 and knock-down of the histone modifying enzymes SUV39H1/H2, and SETDB1 to a lesser extent, was sufficient to allow early binding of Oct4 and Sox2 to these sites in fibroblasts. Thus,

the mark H3K9me3 associated with constitutive heterochromatin can constitute a barrier to OSK, and possibly other, pioneer binding. In addition, maintenance of heterochromatin by the histone chaperone CAF-1 is important for stable somatic cell identity as its knockdown accelerates cellular re-programming by pioneer factors (68) Other pioneers such as Pax7 (AM, JD, unpubl.), FoxA and GATA (69) also exhibit lineage-specific binding repertoires. It remains to be seen whether all pioneers are subject to the same barriers or whether some may have unique limitations, and hence different permissive chromatin environments.

Whereas DNA binding by some pioneers like Pax7 is insensitive to CpG methylation within their DNA binding site (45), DNA methylation may be an impediment to binding of TFs that have some properties of pioneers. Indeed, the factor Nrf1, predicted on theoretical bases to have pioneer action (17), will trigger chromatin access (DNase sensitivity) only if its DNA binding site is unmethylated (70). The Nrf1 DNA binding site is very GC-rich and contains two CpG motifs: its DNA interaction may thus be more sensitive to methylation. Another factor with methylation-insensitive DNA binding may thus be required to prime target enhancers through DNA demethylation to allow Nrf1 binding and action. There may thus be a hierarchy of pioneers with differing potencies: “true” pioneers may be considered to be those with methylation-insensitive DNA binding and an ability to induce DNA demethylation but biological context may provide an argument to consider factors such as Nrf1 as pioneer. For example, global DNA de-methylation occurs at two critical stages of mammalian development, in the pre-implantation embryo and during primordial germ cell proliferation and migration (71,72). DNA methylation-sensitive pioneers may thus act as classical TFs in most cells but transiently behave as pioneers during development. Such limitation on pioneer action could explain specific roles played by pioneers in distinct cell types. The detailed assessment of the pioneer mode of action is thus critical to understand their role in lineage specification.

Many pioneers exhibit extensive binding site subsets of low affinity that are resistant to remodeling (45,53). Some of these sites appear to have degenerate

DNA binding site sequences and were proposed to represent a mechanism for scanning targets. Notwithstanding this possibility, this mechanism does not provide an explanation for selection of specific pioneering sites.

Stability of pioneer-induced chromatin remodeling

During development, pioneers stably reprogram the chromatin landscape leading to a stable cell identity. As such they would implement a memory for long term maintenance of cell identity. During mitosis, chromatin is disassembled and reconstituted after replication. There are mechanisms to reconstitute the daughter cell chromatin landscape as in the mother cell (73). It was proposed that pioneers, and possibly other TFs, bookmark the chromatin during mitosis to allow re-establishment of active regulatory networks. Indeed while many TFs were shown to be excluded from mitotic chromosomes (74,75), some pioneers appear to remain bound to mitotic chromosomes; for example, GATA1 binding is maintained in mitotic chromosome at a tissue-specific subset of 5% of its chromatin targets (76). Surprisingly this study also identified mitosis-specific binding of GATA1 at sites that do not contain the consensus GATA motif. Both specific and non-specific binding sites on mitotic chromosomes were also observed for FoxA1 (77) where specific FoxA1 binding occurs at 15% of its interphase targets. Recently, Sox2 and Oct4 were also shown to remain bound during mitosis (78,79). In this last study, the authors also show using live imaging techniques that crosslinking with formaldehyde leads to eviction of most TFs from mitotic chromosomes. They proposed a model where most TFs remain bound during mitosis to maintain the original program despite only showing this for the well-characterized pioneer Sox2.

The most stable epigenetic mark associated with inactive heterochromatin is DNA methylation (80). And indeed, promoters and CpG-rich promoter regions (CpG islands) that are transcriptionally active are largely demethylated and this is required for activity. Similarly, active enhancer sequences are hypomethylated and the patterns of enhancer hypomethylation are associated with cell-specific gene expression programs (81). Following replication, hemi-methylated CpG dinucleotides are recognized and methylated by the Dnmt1-UHRF1 complex

(82,83). Maintenance of DNA methylation patterns by this mechanism thus ensures stability of lineage-specific gene expression programs. As inactive (closed) regions of chromatin that are targeted by pioneers have high DNA methylation, it is expected that pioneers should bind their target sequence independently of DNA methylation and this is indeed the case for FoxA and Pax7 although there may be exceptions as for Nrf1 discussed above (70).

Whereas direct DNA binding by FoxA and Pax7 is not impaired by CpG methylation of their binding site, their action leads to local demethylation of flanking enhancer sequences behond the DNA binding site (45,69). This demethylation is associated with epigenetic memory and maintenance of an open / accessible chromatin environment (45).

A few pioneer factors were investigated for their impact on DNA methylation. FoxA1 can induce DNA demethylation (4) thus demonstrating its impact on the DNA methylation landscape. Active DNA demethylation can be achieved by the Tet enzymes (84) but for FoxA-dependent demethylation activity, it was rather suggested to require recruitment of a FoxA1 DNA repair complex (85). Also, EBF1 and Pax7 pioneer actions lead to loss of DNA methylation (45,48). The mechanism of pioneer-induced DNA demethylation remains uncertain as the known DNA demethylation Tet pathway could not be implicated in either FoxA or EBF1 action (48,85). It is noteworthy that FoxA-dependent chromatin remodeling can occur independently of replication whereas DNA demethylation is impaired by blockade of replication (69). These data clearly separate two steps in pioneer action and these are consistent with the time frames of action defined in an inducible system for Pax7 (45).

Pioneer factors in cancer

In view of their chromatin remodeling activities, pioneer TFs have the potential for significant epigenetic alterations as seen in cancer. And indeed, FOX family genes are involved in several cancers (reviewed in (86). Overexpression of Foxa1 is associated with a poor prognosis in prostate cancer (87) while it is generally a good prognosis of breast cancer (88). Point mutations of FOXA1 were also found in

some prostate cancers and this was associated with decreased androgen signaling and increased tumor growth (89).

In ER^+ breast cancer cells, ER binding requires FOXA1 at many binding sites showing the role of FOXA1 in driving hormone response of these tumors (90). Similarly, AR binding is also influenced by FOXA1; indeed some AR binding sites are lost in cells depleted of FOXA1; however many sites are also gained suggesting a more complex relationship of FOXA1 with AR than with ER (87,91).

Also, FOXM1 is amplified in some breast cancers (92), in non-Hodgkin's lymphomas (93) or in malignant peripheral nerve sheath tumors (94). FOXM1 is activated through post-translational phosphorylation by ERK and FOXM1 activation is associated with a poor prognosis of many human cancers such as lung, medulloblastoma, breast, gastric and pancreatic cancers.

Chromosomal translocations leading to fusion of the N-terminal DNA binding domain of PAX3 or PAX7 with the C-terminal transactivation domain of FOXO1(FKHR) were found in rhabdomyosarcomas. These fusion proteins act as much more potent activators than the native PAX3 or PAX7 (95). PAX3-FOXO1 was shown to lead to activation of genes involved in cancer development and to inappropriate expression of developmental TFs (96). PAX7 and FOXO1 both have pioneer activity (44,97). As such, these fusion proteins may also function as pioneers. Further, FOXO3 or FOXO4 are trans-located to the MLL gene in acute lymphoblastic leukemia leading to increased cell proliferation (98). FOXO proteins function as tumor suppressors (99) and their loss of activity due translocation or deletion may also lead to increased tumorigenesis.

Finally, two studies showed that the tumor suppressor p53 (TP53) can engage inaccessible chromatin. In one study, p53 binding led to deposition of H4K16ac together with H3K27ac at non TSS sites. However, neither gain of chromatin accessibility nor deposition of H3K4me1 accompanied these changes, thus possibly defining a unique chromatin environment specific to p53 (100). A recent study showed that after DNA damage, a subset of p53 binding sites are associated with de novo accessibility assessed by ATACseq possibly highlighting a canonical pioneer action of p53 (101).

Perspective

As exemplified in this review, the critical aspects of pioneer action are still the least understood. First and foremost, the molecular basis for pioneer access to their target DNA sequences in closed chromatin remains obscure. There may be more than one underlying mechanism as the mechanism proposed for FoxA interaction with nucleosomal DNA, namely its putative linker H1 mimicry binding interactions, does not seem to apply to other pioneers. The question thus remains whole for other pioneers and this highlights the fact that different pioneers may not only use different mechanisms but also may differ in their accessibility to various “flavors” of heterochromatin.

The initial binding and action of pioneers to closed chromatin regions and the initiation of chromatin remodeling is the critical feature that distinguishes pioneers from other TFs. Is there something unique about pioneer action on chromatin at this initiating event? Or is the recruitment of chromatin remodeling complexes at that initiating event the same as those that occur during activation of enhancer function in transcription? This latter possibility would imply that the only unique aspect of pioneer action is the ability to recognize target sites in “closed” chromatin. Alternatively, this ability may be operating in conjunction with recruitment of a unique set of chromatin remodelers involved in initiating chromatin opening but not necessarily involved in maintenance of this accessible state. To answer these difficult questions requires the availability of experimental systems where the specific steps of pioneer action can be followed and investigated. Is there something unique about the maintenance of chromatin accessibility at pioneer sites, or does this simply result from recruitment of enhancer machinery (combination of TFs, chromatin remodelers and chromatin modifiers) leading ultimately to changes in the most stable epigenetic mark, demethylation of DNA cytosines?

Addressing those questions is paramount to understand pioneer action and to use this knowledge in the context of cell fate reprogramming that may occur during

tumorigenic processes or for therapeutic development in the context of cell therapies.

Conflict of interests

The authors declare that they have no conflicts of interest with the contents of this article.

References

1. Garcia-Bellido, A. (1975) Genetic control of wing disc development in Drosophila. *Ciba Found. Symp* **0**, 161-182
2. Hobert, O. (2008) Regulatory logic of neuronal diversity: terminal selector genes and selector motifs. *Proc. Natl. Acad. Sci. USA* **105**, 20067-20071
3. Cirillo, L. A., and 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**, 961-969
4. Serandour, A. A., Avner, S., Percevault, F., Demay, F., Bizot, M., Lucchetti-Miganeh, C., Barloy-Hubler, F., Brown, M., Lupien, M., Metivier, R., Salbert, G., and Eeckhoute, J. (2011) Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res* **21**, 555-565
5. Gardner, K. E., Allis, C. D., and Strahl, B. D. (2011) Operating on chromatin, a colorful language where context matters. *Journal of molecular biology* **409**, 36-46
6. Song, L., Zhang, Z., Grasfeder, L. L., Boyle, A. P., Giresi, P. G., Lee, B. K., Sheffield, N. C., Graf, S., Huss, M., Keefe, D., Liu, Z., London, D., McDaniell, R. M., Shibata, Y., Showers, K. A., Simon, J. M., Vales, T., Wang, T., Winter, D., Zhang, Z., Clarke, N. D., Birney, E., Iyer, V. R., Crawford, G. E., Lieb, J. D., and Furey, T. S. (2011) Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res* **21**, 1757-1767
7. Giresi, P. G., and Lieb, J. D. (2009) Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). *Methods* **48**, 233-239
8. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., and Greenleaf, W. J. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods* **10**, 1213-1218

9. Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., and Ren, B. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet* **39**, 311-318
10. 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
11. Laganiere, J., Deblois, G., Lefebvre, C., Bataille, A. R., Robert, F., and Giguere, V. (2005) From the Cover: Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response. *Proc. Natl. Acad. Sci. USA* **102**, 11651-11656
12. Lupien, M., Eeckhoute, J., Meyer, C. A., Wang, Q., Zhang, Y., Li, W., Carroll, J. S., Liu, X. S., and Brown, M. (2008) FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**, 958-970
13. Carroll, J. S., Liu, X. S., Brodsky, A. S., Li, W., Meyer, C. A., Szary, A. J., Eeckhoute, J., Shao, W., Hestermann, E. V., Geistlinger, T. R., Fox, E. A., Silver, P. A., and Brown, M. (2005) Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122**, 33-43
14. Wang, Q., Li, W., Zhang, Y., Yuan, X., Xu, K., Yu, J., Chen, Z., Beroukhim, R., Wang, H., Lupien, M., Wu, T., Regan, M. M., Meyer, C. A., Carroll, J. S., Manrai, A. K., Janne, O. A., Balk, S. P., Mehra, R., Han, B., Chinnaiyan, A. M., Rubin, M. A., True, L., Fiorentino, M., Fiore, C., Loda, M., Kantoff, P. W., Liu, X. S., and Brown, M. (2009) Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell* **138**, 245-256
15. Hurtado, A., Holmes, K. A., Ross-Innes, C. S., Schmidt, D., and Carroll, J. S. (2011) FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat. Genet* **43**, 27-33

16. Voss, T. C., Schiltz, R. L., Sung, M. H., Yen, P. M., Stamatoyannopoulos, J. A., Biddie, S. C., Johnson, T. A., Miranda, T. B., John, S., and Hager, G. L. (2011) Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. *Cell* **146**, 544-554
17. Sherwood, R. I., Hashimoto, T., O'Donnell, C. W., Lewis, S., Barkal, A. A., van Hoff, J. P., Karun, V., Jaakkola, T., and Gifford, D. K. (2014) Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol* **32**, 171-178
18. Swinstead, E. E., Miranda, T. B., Paakinaho, V., Baek, S., Goldstein, I., Hawkins, M., Karpova, T. S., Ball, D., Mazza, D., Lavis, L. D., Grimm, J. B., Morisaki, T., Grontved, L., Presman, D. M., and Hager, G. L. (2016) Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. *Cell* **165**, 593-605
19. Okita, K., Ichisaka, T., and Yamanaka, S. (2007) Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313-317
20. Koche, R. P., Smith, Z. D., Adli, M., Gu, H., Ku, M., Gnirke, A., Bernstein, B. E., and Meissner, A. (2011) Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell* **8**, 96-105
21. Soufi, A., Donahue, G., and Zaret, K. S. (2012) Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994-1004
22. Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017) Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell* **168**, 442-459.e420
23. Sridharan, R., Tchieu, J., Mason, M. J., Yachechko, R., Kuoy, E., Horvath, S., Zhou, Q., and Plath, K. (2009) Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136**, 364-377
24. Lee, M. T., Bonneau, A. R., Takacs, C. M., Bazzini, A. A., DiVito, K. R., Fleming, E. S., and Giraldez, A. J. (2013) Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* **503**, 360-364

25. Liang, H. L., Nien, C. Y., Liu, H. Y., Metzstein, M. M., Kirov, N., and Rushlow, C. (2008) The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature* **456**, 400-403
26. Schulz, K. N., Bondra, E. R., Moshe, A., Villalta, J. E., Lieb, J. D., Kaplan, T., McKay, D. J., and Harrison, M. M. (2015) Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. *Genome Res* **25**, 1715-1726
27. Gualdi, R., Bossard, P., Zheng, M., Hamada, Y., Coleman, J. R., and Zaret, K. S. (1996) Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes Dev* **10**, 1670-1682
28. 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
29. 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
30. Li, Z., Schug, J., Tuteja, G., White, P., and Kaestner, K. H. (2011) The nucleosome map of the mammalian liver. *Nat. Struct. Mol. Biol* **18**, 742-746
31. 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
32. Li, Z., Gadue, P., Chen, K., Jiao, Y., Tuteja, G., Schug, J., Li, W., and Kaestner, K. H. (2012) Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation. *Cell* **151**, 1608-1616
33. Fakhouri, T. H., Stevenson, J., Chisholm, A. D., and Mango, S. E. (2010) Dynamic chromatin organization during foregut development mediated by the organ selector gene PHA-4/FoxA. *PLoS. Genet* **6**

34. Hsu, H. T., Chen, H. M., Yang, Z., Wang, J., Lee, N. K., Burger, A., Zaret, K., Liu, T., Levine, E., and Mango, S. E. (2015) TRANSCRIPTION. Recruitment of RNA polymerase II by the pioneer transcription factor PHA-4. *Science* **348**, 1372-1376
35. Holtzinger, A., and Evans, T. (2005) Gata4 regulates the formation of multiple organs. *Development* **132**, 4005-4014
36. Zhao, R., Watt, A. J., Li, J., Luebke-Wheeler, J., Morrisey, E. E., and Duncan, S. A. (2005) GATA6 is essential for embryonic development of the liver but dispensable for early heart formation. *Mol Cell Biol* **25**, 2622-2631
37. Watt, A. J., Zhao, R., Li, J., and Duncan, S. A. (2007) Development of the mammalian liver and ventral pancreas is dependent on GATA4. *BMC developmental biology* **7**, 37
38. Huang, P., He, Z., Ji, S., Sun, H., Xiang, D., Liu, C., Hu, Y., Wang, X., and Hui, L. (2011) Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* **475**, 386-389
39. Sekiya, S., and Suzuki, A. (2011) Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. *Nature* **475**, 390-393
40. Carotta, S., Wu, L., and Nutt, S. L. (2010) Surprising new roles for PU.1 in the adaptive immune response. *Immunological reviews* **238**, 63-75
41. 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
42. Barozzi, I., Simonatto, M., Bonifacio, S., Yang, L., Rohs, R., Ghisletti, S., and Natoli, G. (2014) Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. *Mol Cell* **54**, 844-857
43. van Oevelen, C., Collombet, S., Vicent, G., Hoogenkamp, M., Lepoivre, C., Badeaux, A., Bussmann, L., Sardina, J. L., Thieffry, D., Beato, M., Shi, Y., Bonifer, C., and Graf, T. (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

44. 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 Dev* **26**, 2299-2310
45. 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*, doi10.1038/s41588-41017-40035-41582
46. Gyory, I., Boller, S., Nechanitzky, R., Mandel, E., Pott, S., Liu, E., and Grosschedl, R. (2012) Transcription factor Ebf1 regulates differentiation stage-specific signaling, proliferation, and survival of B cells. *Genes Dev* **26**, 668-682
47. Treiber, T., Mandel, E. M., Pott, S., Gyory, I., Firner, S., Liu, E. T., and Grosschedl, R. (2010) Early B cell factor 1 regulates B cell gene networks by activation, repression, and transcription- independent poising of chromatin. *Immunity* **32**, 714-725
48. 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
49. Wapinski, O. L., Vierbuchen, T., Qu, K., Lee, Q. Y., Chanda, S., Fuentes, D. R., Giresi, P. G., Ng, Y. H., Marro, S., Neff, N. F., Drechsel, D., Martynoga, B., Castro, D. S., Webb, A. E., Sudhof, T. C., Brunet, A., Guillemot, F., Chang, H. Y., and Wernig, M. (2013) Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* **155**, 621-635
50. 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

51. Bossard, P., and Zaret, K. S. (1998) GATA transcription factors as potentiators of gut endoderm differentiation. *Development* **125**, 4909-4917
52. Sekiya, T., Muthurajan, U. M., Luger, K., Tulin, A. V., and Zaret, K. S. (2009) Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. *Genes Dev* **23**, 804-809
53. 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
54. Clark, K. L., Halay, E. D., Lai, E., and Burley, S. K. (1993) Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364**, 412-420
55. de Dieuleveult, M., Yen, K., Hmitou, I., Depaux, A., Boussouar, F., Bou Dargham, D., Jounier, S., Humbertclaude, H., Ribierre, F., Baulard, C., Farrell, N. P., Park, B., Keime, C., Carriere, L., Berlivet, S., Gut, M., Gut, I., Werner, M., Deleuze, J. F., Olaso, R., Aude, J. C., Chantalat, S., Pugh, B. F., and Gerard, M. (2016) Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells. *Nature* **530**, 113-116
56. Kidder, B. L., Palmer, S., and Knott, J. G. (2009) SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells. *Stem Cells* **27**, 317-328
57. King, H. W., and Klose, R. J. (2017) The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. *eLife* **6**
58. Wang, L., Du, Y., Ward, J. M., Shimbo, T., Lackford, B., Zheng, X., Miao, Y. L., Zhou, B., Han, L., Fargo, D. C., Jothi, R., Williams, C. J., Wade, P. A., and Hu, G. (2014) INO80 facilitates pluripotency gene activation in embryonic stem cell self-renewal, reprogramming, and blastocyst development. *Cell Stem Cell* **14**, 575-591
59. 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

60. Schuettengruber, B., Martinez, A. M., Iovino, N., and Cavalli, G. (2011) Trithorax group proteins: switching genes on and keeping them active. *Nat. Rev. Mol Cell Biol* **12**, 799-814
61. McKinnell, I. W., Ishibashi, J., Le Grand, F., Punch, V. G., Addicks, G. C., Greenblatt, J. F., Dilworth, F. J., and Rudnicki, M. A. (2008) Pax7 activates myogenic genes by recruitment of a histone methyltransferase complex. *Nat. Cell Biol* **10**, 77-84
62. Jozwik, K. M., Chernukhin, I., Serandour, A. A., Nagarajan, S., and Carroll, J. S. (2016) FOXA1 Directs H3K4 Monomethylation at Enhancers via Recruitment of the Methyltransferase MLL3. *Cell Rep* **17**, 2715-2723
63. Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009) H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat Genet* **41**, 941-945
64. Updike, D. L., and Mango, S. E. (2006) Temporal regulation of foregut development by HTZ-1/H2A.Z and PHA-4/FoxA. *PLoS. Genet* **2**, e161
65. Gevry, N., Hardy, S., Jacques, P. E., Laflamme, L., Svotelis, A., Robert, F., and Gaudreau, L. (2009) Histone H2A.Z is essential for estrogen receptor signaling. *Genes Dev* **23**, 1522-1533
66. Menet, J. S., Pescatore, S., and Rosbash, M. (2014) CLOCK:BMAL1 is a pioneer-like transcription factor. *Genes Dev* **28**, 8-13
67. Hagey, D. W., Zaouter, C., Combeau, G., Lendahl, M. A., Andersson, O., Huss, M., and Muhr, J. (2016) Distinct transcription factor complexes act on a permissive chromatin landscape to establish regionalized gene expression in CNS stem cells. *Genome Res* **26**, 908-917
68. Cheloufi, S., Elling, U., Hopfgartner, B., Jung, Y. L., Murn, J., Ninova, M., Hubmann, M., Badeaux, A. I., Euong Ang, C., Tenen, D., Wesche, D. J., Abazova, N., Hogue, M., Tasdemir, N., Brumbaugh, J., Rathert, P., Jude, J., Ferrari, F., Blanco, A., Fellner, M., Wenzel, D., Zinner, M., Vidal, S. E.,

Bell, O., Stadtfeld, M., Chang, H. Y., Almouzni, G., Lowe, S. W., Rinn, J., Wernig, M., Aravin, A., Shi, Y., Park, P. J., Penninger, J. M., Zuber, J., and Hochedlinger, K. (2015) The histone chaperone CAF-1 safeguards somatic cell identity. *Nature* **528**, 218-224

69. Donaghey, J., Thakurela, S., Charlton, J., Chen, J. S., Smith, Z. D., Gu, H., Pop, R., Clement, K., Stamenova, E. K., Karnik, R., Kelley, D. R., Gifford, C. A., Cacchiarelli, D., Rinn, J. L., Gnirke, A., Ziller, M. J., and Meissner, A. (2018) Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat Genet*
70. Domcke, S., Bardet, A. F., Adrian Ginno, P., Hartl, D., Burger, L., and Schubeler, D. (2015) Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* **528**, 575-579
71. Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012) The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol Cell* **48**, 849-862
72. Messerschmidt, D. M., Knowles, B. B., and Solter, D. (2014) DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev* **28**, 812-828
73. Bernstein, B. E., Meissner, A., and Lander, E. S. (2007) The mammalian epigenome. *Cell* **128**, 669-681
74. Gottesfeld, J. M., and Forbes, D. J. (1997) Mitotic repression of the transcriptional machinery. *Trends Biochem Sci* **22**, 197-202
75. Martinez-Balbas, M. A., Dey, A., Rabindran, S. K., Ozato, K., and Wu, C. (1995) Displacement of sequence-specific transcription factors from mitotic chromatin. *Cell* **83**, 29-38
76. Kadauke, S., Udugama, M. I., Pawlicki, J. M., Achtman, J. C., Jain, D. P., Cheng, Y., Hardison, R. C., and Blobel, G. A. (2012) Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. *Cell* **150**, 725-737

77. Caravaca, J. M., Donahue, G., Becker, J. S., He, X., Vinson, C., and Zaret, K. S. (2013) Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. *Genes Dev* **27**, 251-260
78. Deluz, C., Friman, E. T., Strebinger, D., Benke, A., Raccaud, M., Callegari, A., Leleu, M., Manley, S., and Suter, D. M. (2016) A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. *Genes Dev* **30**, 2538-2550
79. Teves, S. S., An, L., Hansen, A. S., Xie, L., Darzacq, X., and Tjian, R. (2016) A dynamic mode of mitotic bookmarking by transcription factors. *eLife* **5**
80. Almouzni, G., and Cedar, H. (2016) Maintenance of Epigenetic Information. *Cold Spring Harbor perspectives in biology* **8**
81. Varley, K. E., Gertz, J., Bowling, K. M., Parker, S. L., Reddy, T. E., Pauli-Behn, F., Cross, M. K., Williams, B. A., Stamatoyannopoulos, J. A., Crawford, G. E., Absher, D. M., Wold, B. J., and Myers, R. M. (2013) Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res* **23**, 555-567
82. Hermann, A., Goyal, R., and Jeltsch, A. (2004) The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem* **279**, 48350-48359
83. Bostick, M., Kim, J. K., Esteve, P. O., Clark, A., Pradhan, S., and Jacobsen, S. E. (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317**, 1760-1764
84. Rasmussen, K. D., and Helin, K. (2016) Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev* **30**, 733-750
85. Zhang, Y., Zhang, D., Li, Q., Liang, J., Sun, L., Yi, X., Chen, Z., Yan, R., Xie, G., Li, W., Liu, S., Xu, B., Li, L., Yang, J., He, L., and Shang, Y. (2016) Nucleation of DNA repair factors by FOXA1 links DNA demethylation to transcriptional pioneering. *Nat Genet* **48**, 1003-1013
86. Katoh, M., Igarashi, M., Fukuda, H., Nakagama, H., and Katoh, M. (2013) Cancer genetics and genomics of human FOX family genes. *Cancer Lett* **328**, 198-206

87. Sahu, B., Laakso, M., Ovaska, K., Mirtti, T., Lundin, J., Rannikko, A., Sankila, A., Turunen, J. P., Lundin, M., Konsti, J., Vesterinen, T., Nordling, S., Kallioniemi, O., Hautaniemi, S., and Janne, O. A. (2011) Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. *Embo j* **30**, 3962-3976

88. Thorat, M. A., Marchio, C., Morimiya, A., Savage, K., Nakshatri, H., Reis-Filho, J. S., and Badve, S. (2008) Forkhead box A1 expression in breast cancer is associated with luminal subtype and good prognosis. *Journal of clinical pathology* **61**, 327-332

89. Grasso, C. S., Wu, Y. M., Robinson, D. R., Cao, X., Dhanasekaran, S. M., Khan, A. P., Quist, M. J., Jing, X., Lonigro, R. J., Brenner, J. C., Asangani, I. A., Ateeq, B., Chun, S. Y., Siddiqui, J., Sam, L., Anstett, M., Mehra, R., Prensner, J. R., Palanisamy, N., Ryslik, G. A., Vandin, F., Raphael, B. J., Kunju, L. P., Rhodes, D. R., Pienta, K. J., Chinnaiyan, A. M., and Tomlins, S. A. (2012) The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239-243

90. Carroll, J. S., Meyer, C. A., Song, J., Li, W., Geistlinger, T. R., Eeckhoute, J., Brodsky, A. S., Keeton, E. K., Fertuck, K. C., Hall, G. F., Wang, Q., Bekiranov, S., Sementchenko, V., Fox, E. A., Silver, P. A., Gingeras, T. R., Liu, X. S., and Brown, M. (2006) Genome-wide analysis of estrogen receptor binding sites. *Nat. Genet* **38**, 1289-1297

91. Pomerantz, M. M., Li, F., Takeda, D. Y., Lenci, R., Chonkar, A., Chabot, M., Cejas, P., Vazquez, F., Cook, J., Shivdasani, R. A., Bowden, M., Lis, R., Hahn, W. C., Kantoff, P. W., Brown, M., Loda, M., Long, H. W., and Freedman, M. L. (2015) The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat Genet* **47**, 1346-1351

92. Curtis, C., Shah, S. P., Chin, S. F., Turashvili, G., Rueda, O. M., Dunning, M. J., Speed, D., Lynch, A. G., Samarajiwa, S., Yuan, Y., Graf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., Langerod, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P., Markowetz, F., Murphy, L., Ellis, I., Purushotham, A., Borresen-Dale, A. L., Brenton, J. D.,

Tavare, S., Caldas, C., and Aparicio, S. (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346-352

93. Green, M. R., Aya-Bonilla, C., Gandhi, M. K., Lea, R. A., Wellwood, J., Wood, P., Marlton, P., and Griffiths, L. R. (2011) Integrative genomic profiling reveals conserved genetic mechanisms for tumorigenesis in common entities of non-Hodgkin's lymphoma. *Genes, chromosomes & cancer* **50**, 313-326
94. Yu, J., Deshmukh, H., Payton, J. E., Dunham, C., Scheithauer, B. W., Tihan, T., Prayson, R. A., Guha, A., Bridge, J. A., Ferner, R. E., Lindberg, G. M., Gutmann, R. J., Emnett, R. J., Salavaggione, L., Gutmann, D. H., Nagarajan, R., Watson, M. A., and Perry, A. (2011) Array-based comparative genomic hybridization identifies CDK4 and FOXM1 alterations as independent predictors of survival in malignant peripheral nerve sheath tumor. *Clin Cancer Res* **17**, 1924-1934
95. Bennicelli, J. L., Advani, S., Schafer, B. W., and 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**, 4348-4356
96. Cao, L., Yu, Y., Bilke, S., Walker, R. L., Mayeenuddin, L. H., Azorsa, D. O., Yang, F., Pineda, M., Helman, L. J., and Meltzer, P. S. (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
97. Hatta, M., and Cirillo, L. A. (2007) Chromatin opening and stable perturbation of core histone:DNA contacts by FoxO1. *J Biol Chem* **282**, 35583-35593
98. So, C. W., and Cleary, M. L. (2003) Common mechanism for oncogenic activation of MLL by forkhead family proteins. *Blood* **101**, 633-639
99. Dansen, T. B., and Burgering, B. M. (2008) Unravelling the tumor-suppressive functions of FOXO proteins. *Trends Cell Biol* **18**, 421-429

100. Sammons, M. A., Zhu, J., Drake, A. M., and Berger, S. L. (2015) TP53 engagement with the genome occurs in distinct local chromatin environments via pioneer factor activity. *Genome Res* **25**, 179-188
101. Younger, S. T., and Rinn, J. L. (2017) p53 regulates enhancer accessibility and activity in response to DNA damage. *Nucleic Acids Res* **45**, 9889-9900
102. 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., Sankar, V., Sim, S., Quake, S. R., Dahmane, N., Wernig, M., and Chang, H. Y. (2017) Rapid Chromatin Switch in the Direct Reprogramming of Fibroblasts to Neurons. *Cell Rep* **20**, 3236-3247
103. Raposo, A. A., Vasconcelos, F. F., Drechsel, D., Marie, C., Johnston, C., Dolle, D., Bithell, A., Gillotin, S., van den Berg, D. L., Ettwiller, L., Flicek, P., Crawford, G. E., Parras, C. M., Berninger, B., Buckley, N. J., Guillemot, F., and Castro, D. S. (2015) Ascl1 Coordinately Regulates Gene Expression and the Chromatin Landscape during Neurogenesis. *Cell Rep*
104. Feng, R., Desbordes, S. C., Xie, H., Tillo, E. S., Pixley, F., Stanley, E. R., and Graf, T. (2008) PU.1 and C/EBPalpha/beta convert fibroblasts into macrophage-like cells. *Proc Natl Acad Sci U S A* **105**, 6057-6062
105. Festuccia, N., Dubois, A., Vandormael-Pournin, S., Gallego Tejeda, E., Mouren, A., Bessonnard, S., Mueller, F., Proux, C., Cohen-Tannoudji, M., and Navarro, P. (2016) Mitotic binding of Esrrb marks key regulatory regions of the pluripotency network. *Nat Cell Biol* **18**, 1139-1148
106. Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008) Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat. Biotechnol* **26**, 101-106
107. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-872

Figures and table

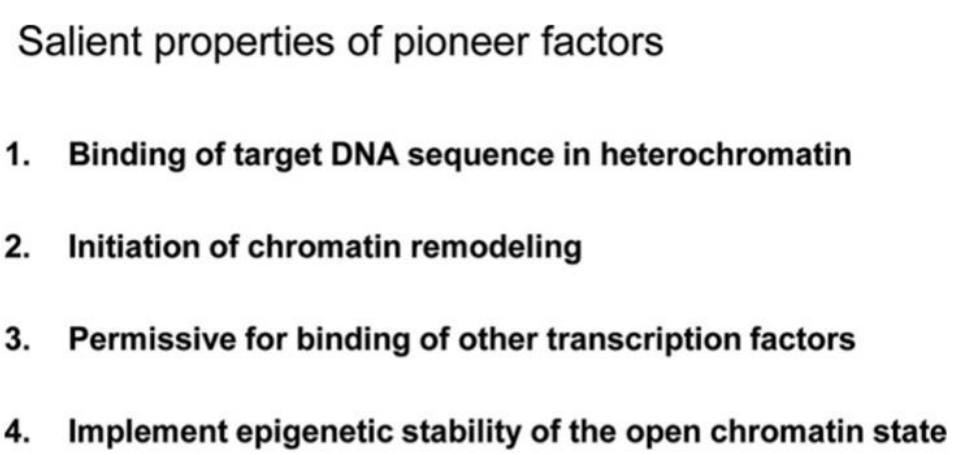


Figure 1. Salient properties of pioneer factors.

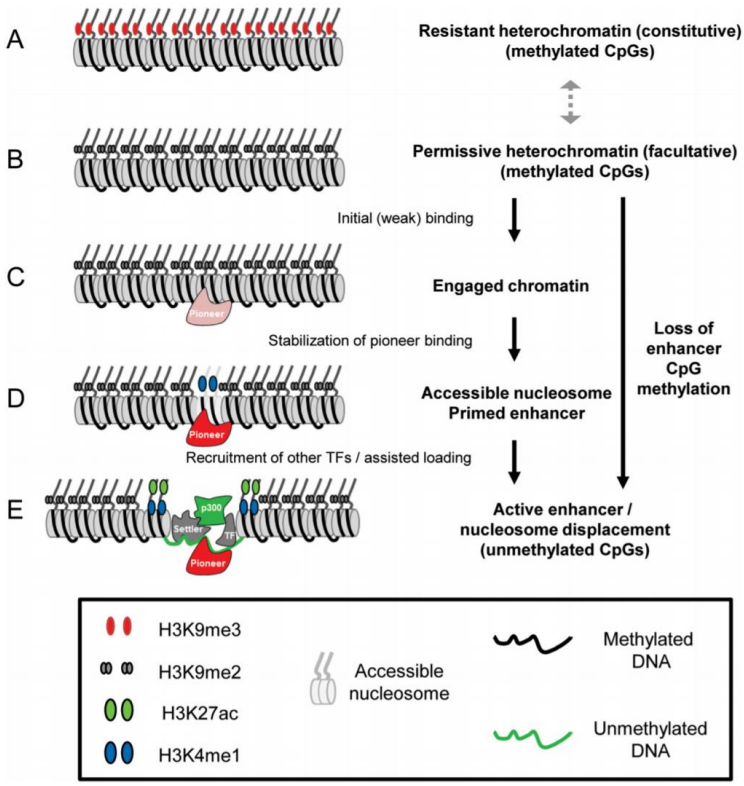


Figure 2. Current scheme of pioneer action. The permissive chromatin state for pioneer action appears to be facultative heterochromatin. Following initial weak binding of the pioneer, target site chromatin (mostly characterized at enhancers) undergoes a first transition where a central nucleosome becomes more accessible and this may (or not) overlap with a state of Primed enhancer characterized by a weak H3K4me1 signal. Complete activation of enhancers is characterized by nucleosome depletion, bimodal distribution of H3K4me1 and H3K27ac, together with recruitment of the general coactivator p300 and other transcription factors. Whereas the ability to bind methylated DNA target sites is not a unique feature of all pioneers, for most pioneers current evidence correlates pioneer-dependent chromatin remodeling with loss of CpG methylation at the newly accessible DNA / enhancers.

List of transcription factors that share at least some features of pioneers

For each feature listed at the top of the table are references within the table that provide supporting evidence.

Factor	Binding to heterochromatin	Chromatin activation	Epigenetic memory: DNA demethylation	Cell fate reprogramming	Nucleosome binding	Mitotic bookmarking
Ascl1/Mash1	102	102		102, 103		
C/EBPα		43		104		
Ebf1	47, 48	47, 48	48			
Esrrb						105
Foxa	3	3, 4, 28, 31, 32	4, 69, 85	38, 39	28	77
Gata	59	59		38, 39	29	76
GR/AR	18	18				
Klf4	21, 22	21, 22		106, 107	53	
Neurod1				50, 70		
Nrf1	70	70	Inhibitory (70)			
Oct4	21, 22	21, 22		106, 107	53	
p53	100, 101	100, 101				
Pax7	44, 45	44, 45, 61	45	44		
PU.I	41, 42	41, 42		104		
Sox2	21, 22	21, 22		106, 107	53	78, 79

Table 1. List of transcription factors that share at least some features of pioneers. For each feature listed at top, the Table indicates references that provide supporting evidence.

5 Pax factors in transcription and epigenetic remodelling

Alexandre Mayran[1], Audrey Pelletier[1], Jacques Drouin[2]

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].

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 ß-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, ie 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].

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).

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.

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 where 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 co-activators recruited by Pax factors have been identified (Fig. 5).

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 Pax5-dependent 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 transaction 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 αA-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 V_H-to-DJ_H recombination that normally occurs in committed B cells. Non-committed B cells show high level of H3K9 methylation at the V_H locus consistent with inactive chromatin, and Pax5 is necessary and sufficient for removal of H3K9me3 at the V_H 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 actions but it may also draw on the activity of similar coregulatory protein partners.

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 unmasks 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, Whitington 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 A 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, Irmler 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, Georlette 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 .

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 .

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 .

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, McDaniell 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, Conners 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, Birshtein BK (2002) The interaction of Pax5 (BSAP) with Daxx can result in transcriptional activation in B cells. J Biol Chem 277: 11156-11164.

Figures

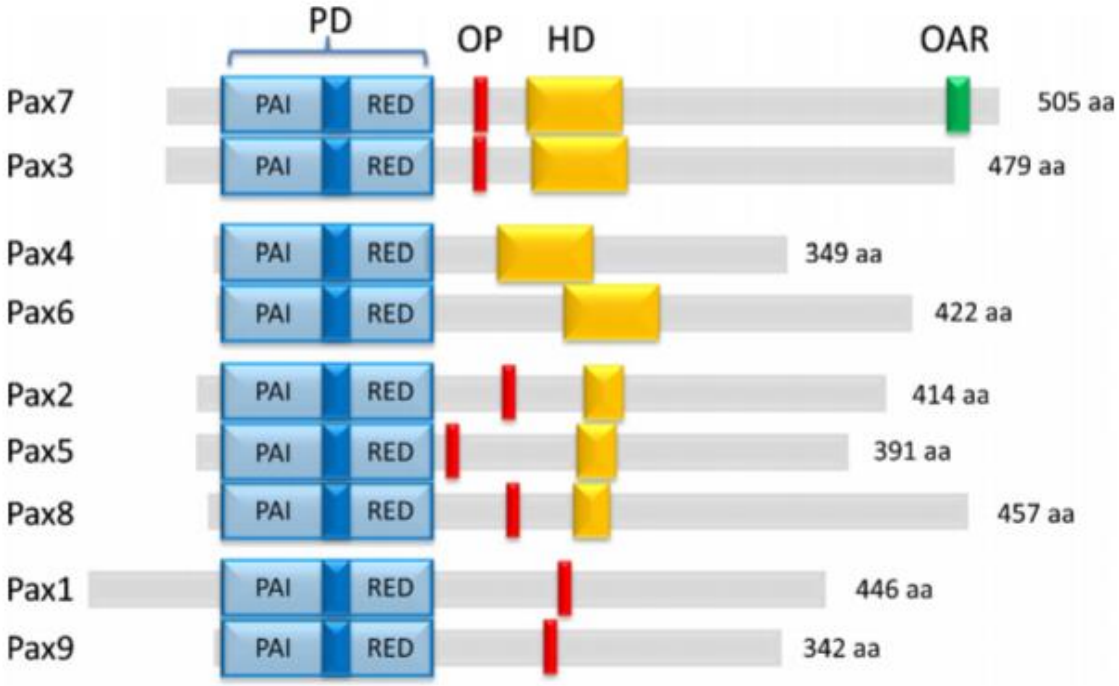


Figure 1. Structure of Pax transcription factors. Schematic drawings represent the conserved features of the mouse Pax factors. These include the paired (PD) DNA binding domain (blue) present in all Pax factors and subdivided into two helix-turn-helix (HTH) motif (PAI and RED). The Pax family is classified 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.

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**$\frac{T}{C}$**GC**$\frac{A}{G}$**T**	[102]
5	FL	P	**GT**$\frac{C}{T}\frac{A}{C}$**CGC**	[12,49]
8	FL	P	**GTCAC**$\frac{G}{C}$**C**	[7]
4	Isolated HD	H	**TAATT**	[103]
4	Isolated PD	P	$\frac{G}{T}$**T**$\frac{C}{G}$**A**$\frac{T}{A}$**GC** GTG	[47]
6	FL	P	**TTCA**CGC	[102,104]
6	Isolated PD	P.I2	**TTCACGC**AT**GAGTGCC**	[102,104]
6-5a	Isolated PD	P.I6	**CTCAGTG**AATGTT**CATTGAC**	[19]
6	Isolated HD	H.I3 H.I2	**TAAT**$\frac{G}{T}$**CGATTA** **TAATCGATTA**	[49]
3	Isolated HD	H.I2	**TAAT**$\frac{T}{C}$**GATTA**	
3	FL	P	**gTCA**CG$\frac{G}{A}$**C**	[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.

Pax	Motif	Natural sites	Reference
1 / 9	P	*Bapx1* enhancer TCCAATCCAATCCA**ATCAAAC**GTAAAAT	[105]
2	P	*Pax5* enhancer GCTACC*TAAT*TTCATGC**ATCAGAG**	[106]
2	P P.D0	*Engrailed-2* enhancer BS-I **GGCACGG** BS-II **GTCAAGGCTCAGTG**	[107]
5	P.D0	*CD19* promoter GCGGTG**GTCACGCCTCAGGC**CCCA	[108]
8	P	*Tg* promoter CACTGCCCA**GTCAAGT**GTTCTTGA	[109]
4	P.I-2	Glucagon promoter **TTCACGCCTGAC**	[110]
6	P.I10	Rat insulin I, C2 element **AAATGAG**GTGGAAAATG**CTCAGCC**	[111]
6	P.I2	Glucagon, G3 element **TTCACGC**CT**GACTGAG**	[111]
3	PH.E0 (PH.I4, PH.D6)	*Mitf* promoter **TTTAGTTCCAG**TAGTAT**TAAT**	[8]
	P	*Trp-1* promoter **GTCACAC**TT	[8]
	PH.E0	*Myf5* enhancer **GCATGACTAAT**	[31]
	PH.E0	*Dmrt2* enhancer, site 2 **TGTGTGACTAAT**	[32]
	PH.E5	*Evenskipped* **ATTA**GCACC**GTTCCGC**	[42]
7	PH.E0	*Pcsk2* (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.

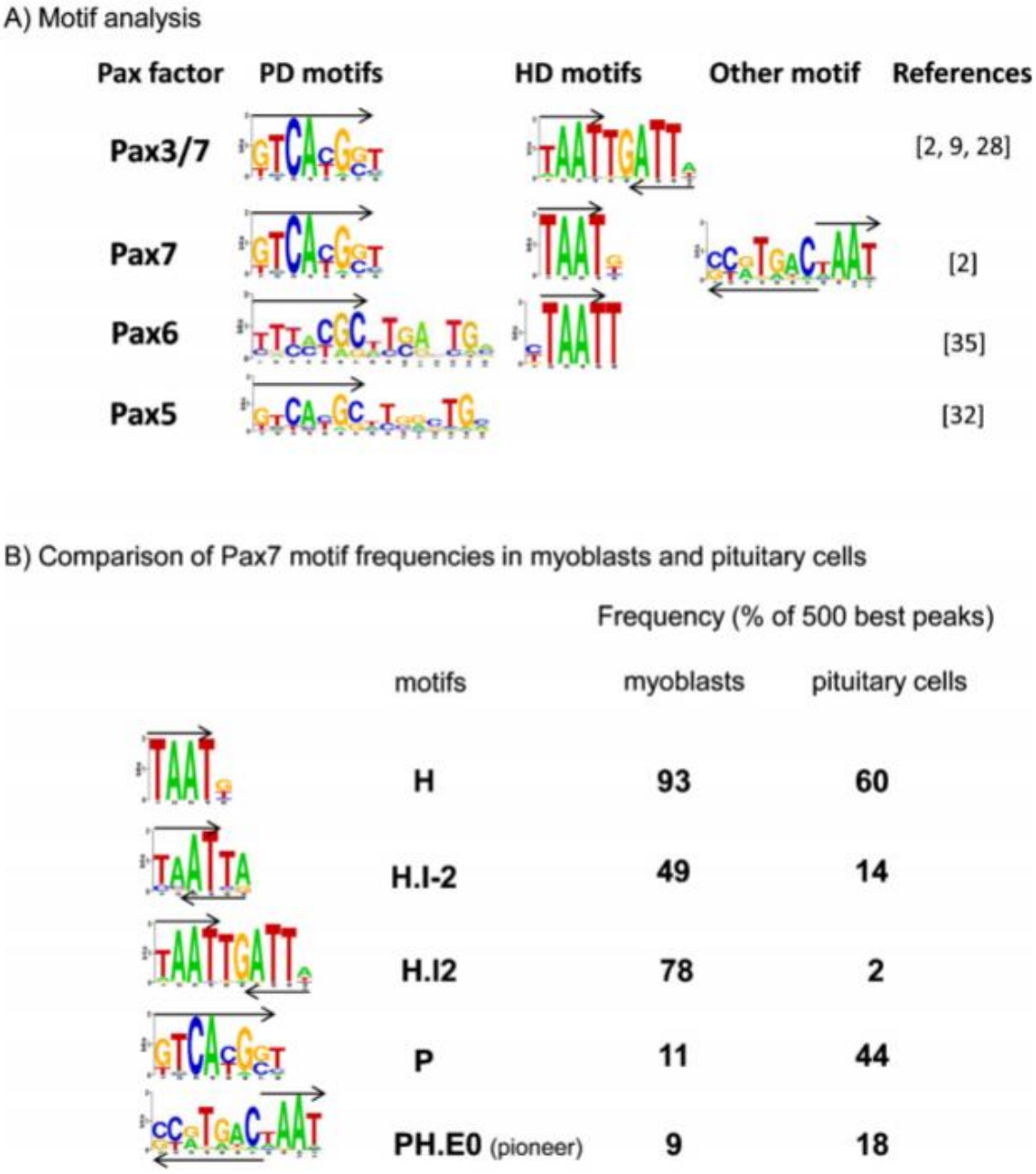


Figure 4. ChIPseq analyses of Pax genomic binding sites. (A) The results of *de novo* motif analyses performed on ChIPseq datasets are shown. (B) Comparison of motif frequencies present under Pax7 peaks in ChIPseq data from myoblasts [9] and pituitary [2] cells. Motif occurrence was determined for the 500 stronger peaks of each dataset using the HOMER program.

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.

Chapter II: Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate

Alexandre Mayran[1,2], Konstantin Khetchoumian[1], Fadi Hariri[3], Tomi Pastinen[4], Yves Gauthier[1], Aurelio Balsalobre[1] and Jacques Drouin[1,2,5]

[1] Laboratoire de génétique moléculaire, Institut de Recherches Cliniques de Montréal (IRCM), Montreal, Québec, Canada,

[2] Department of Biochemistry, McGill University, Montreal, Québec, Canada

[3] McGill Genome Innovation Centre, McGill University, Montreal, Québec, Canada

[4] Canada Research Chair in Human Genomics, McGill University, Montreal, Québec, Canada

[5]Correspondence: jacques.drouin@ircm.qc.ca

AUTHOR CONTRIBUTIONS

A.M., K.K., F.H., and Y.G. performed experiments. A.M., T.P. and J.D. conceived and designed the experiments. A.M., A.B. and J.D. analyzed the data. A.M. and J.D. wrote the manuscript.

ABSTRACT

Pioneer transcription factors establish new cell fate competences by triggering chromatin remodeling. However, many features of pioneer action such as their kinetics and stability remain poorly defined. Here, we show that Pax7 is necessary and sufficient for specification of one pituitary lineage by opening a unique enhancer repertoire. Pax7 binds its targeted enhancers rapidly but chromatin remodeling and gene activation are slower. Enhancers opened by Pax7 show loss of DNA methylation and acquire stable epigenetic memory revealed by binding of non-pioneer factors after Pax7 withdrawal. The present work shows that transient Pax7 expression is sufficient for stable specification of cell identity.

Keywords: transcription, chromatin, DNA methylation, differentiation, pituitary, POMC

INTRODUCTION

Cell specification and differentiation occur throughout development of multicellular organisms leading, in complex life forms, to a huge diversity of cells. The differentiation of cells into different lineages is implemented by a unique combination of transcription factors (TF) that collectively control the program of cell-specific gene expression. During development, progenitors are specified by the action of selector genes that establish the differentiation competence of downstream lineages. Pioneer factors are a class of TFs that fulfil this expectation as they have the unique ability to bind condensed, otherwise inaccessible, chromatin and to open this chromatin for access by other TFs. The three pluripotency factors, Oct4, Klf4 and Sox2, have pioneer activity on a global scale by widely affecting the epigenome beyond their initial binding sites[1]. In addition, a handful of TFs were shown to have more localized pioneer activity. The ability of FoxA to specify liver fate relies on its property to alter chromatin organization upon binding nucleosomal target DNA sequences[2]. We showed that the factor Pax7 specifies intermediate pituitary melanotrope cell identity through a pioneer activity[3]. The pioneer TFs C/EBPα was implicated in macrophage differentiation[4] and EBF1 in B cell development[5]. Recently, two neurogenic bHLH TFs, NeuroD1 and MASH1 (Ascl1), were also suggested to have pioneer activity during neural development[6,7].

Here, we used the Pax7 pituitary model to define the mechanism and stability of cell specification through pioneer action. The pituitary has two lineages that use the same hormone precursor, pro-opiomelanocortin (POMC), for entirely different biological functions. Indeed, POMC-expressing corticotropes produce ACTH that controls adrenal glucocorticoid synthesis whereas melanotropes process POMC into αMSH that regulates pigmentation (Fig. 1a). Transcriptional regulation of the POMC gene is unique to each lineage but the same highly cell-restricted factor Tpit drives terminal differentiation of both lineages[8,9]. Pax7 acts before Tpit in intermediate lobe progenitors to implement a unique tissue identity upon which Tpit-driven differentiation establishes the melanotrope identity[3]. We characterized the actions of Pax7 within the framework of the critical properties of pioneer TFs. These include the ability to: 1) bind DNA target sites in heterochromatin; 2) initiate remodeling of the surrounding chromatin; 3) facilitate binding of other TFs; 4) trigger stable changes in chromatin accessibility, thus ensuring epigenetic stability, ie long-term access by non-pioneer factors.

We now show that the bulk of differentially accessible chromatin regions between the two alternate POMC cell fates are at putative enhancers. Pax7 binding is rapid at uniquely

marked heterochromatin pioneer sites and initiates a slower process of chromatin opening that ultimately provides access to both developmental and signal-dependent TFs. Finally, Pax7 pioneer action is stable and ensures epigenetic memory even after Pax7 withdrawal. This long-term stability is supported by loss of DNA hypermethylation at pioneered enhancers. In summary, the present work defines the critical features of Pax7 pioneer action on tissue-specific progenitors for establishment and maintenance of cell identity.

RESULTS

Lineage-specific chromatin accessibility at distal, but not promoter, elements

In order to address the relative importance of cell-specific chromatin accessibility in establishment of differentiation-specific programs of gene expression, we subjected FACS-purified pituitary corticotropes and melanotropes from *POMC-EGFP* transgenic mice[10,11] (Fig. 1a) to RNAseq and ATACseq analyses (Supplementary Fig. 1a-d). Differential expression analysis of the transcriptomes showed 504 differentially expressed genes (FDR<0.05) with similar transcript proportions enriched in each lineage, 243 cortico-specific and 261 melanotrope-specific genes. The most significantly melanotrope gene (Fig. 1b) is *Pcsk2* (19th most abundant, Supplementary Fig. 2a, b) that encodes the protein convertase PC2 responsible for cleavage of ACTH into αMSH, the hallmark hormone produced by melanotropes. ATACseq data were analysed to identify differentially accessible regions (DARs) in each lineage using a high stringency filter to exclude cell contamination issues (*P* value<1e^{-20}). In contrast to the general distribution of ATACseq peaks present at both promoters and enhancers (Fig. 1c), DARs are found at distal elements (Fig. 1d). Interestingly, the DAR repertoire is much richer in melanotropes compared to corticotropes with 2891 and 558 DARs, respectively (Fig. 1d). Further, melanotrope and corticotrope DARs tend to be spatially associated with melanotrope and corticotrope genes, respectively (Fig. 1e and Supplementary Fig. 2c). Both DAR repertoires are evolutionarily conserved and melanotrope DARs are the most conserved suggesting that they may be subjected to more evolutionary pressure (Supplementary Fig. 2d). The vast majority of lineage-specific genes show accessibility at the transcription start site (+/-200bp) in both lineages in contrast to DARs (Fig. 1f). Only 8 promoters of corticotrope-specific genes are only ATAC-sensitive (p<0.05) in corticotropes whereas 13 melanotrope-specific promoters are only accessible in melanotropes. The

presence/absence of accessibility at promoters is thus a poor predictor of cell identity. The *Pcsk2* locus illustrates well how melanotrope DARs may control expression of cell specific transcripts. Despite the high melanotrope specificity of the *Pcsk2* gene, its promoter also shows accessibility in corticotrope cells (Fig. 1g). Many melanotrope-specific DARs are present at this locus, both intronic and intergenic, including the known -146Kb *Pcsk2* enhancer[3]. Other examples of melano- and cortico-specific loci support these conclusions (Supplementary Fig. 2e-h). Together, these findings indicate that the main differences in chromatin accessibility between these two lineages are at known and putative enhancers.

Pax7 is required for deployment of the melanotrope enhancer repertoire

To identify factors responsible for establishment of cell-specific DARs, we searched for enriched DNA motifs in each DAR repertoire (Supplementary Fig. 3a, b). The corticotrope DARs are enriched for motifs of the glucocorticoid receptor (GR), Pitx and Tpit. GR is not expressed in melanotropes, hence the high frequency of its motif in corticotrope DARs is consistent. The melanotrope repertoire is enriched for the Pax7 motifs: paired box, homeobox, as well as a composite motif containing both (Fig. 2a and Supplementary Fig. 3a, b) in agreement with the restricted expression of Pax7 in melanotropes[3]. To confirm the involvement of cognate factors in targeting DARs, we showed enrichment of GR binding to corticotrope DARs by ChIPseq (Fig. 2b) and similarly, Pax7 binding is enriched at the melanotrope DARs in Pax7-expressing AtT-20 cells (Fig. 2c). Using microarray data[3], we assessed the extent of the switch in identity of melanotrope cells in *Pax7*$^{-/-}$ mice (Fig. 2d). Remarkably, 69% of melano-specific transcripts (181 of 261) are down-regulated in absence of Pax7 while 64% of cortico-specific transcripts are up-regulated, 160 out of 249 (Supplementary Fig. 3c). To better define the relative roles of Pax7 and Tpit in melanotrope identity, we assessed gene expression in *Tpit*$^{-/-}$ mice. Most *Pax7*-dependent melanotrope genes are also *Tpit*-dependent (87%, or 157 of 181 genes) whereas only 37% of corticotrope genes up-regulated in *Pax7*$^{-/-}$ pituitaries are also up-regulated in *Tpit*$^{-/-}$ mice (Fig. 2e). Thus, Pax7 and Tpit targets overlap extensively for activation of the melanotrope program.

It is noteworthy that corticotrope-specific gene loci exhibit very few DARs (Supplementary Fig. 2g showing *GR* and *CRHR* loci) in contrast to melano-specific genes that have multiple melano-specific DARs (Supplementary Fig. 2e for *Drd2* and *Gpr6* loci). This would be consistent with the corticotrope fate representing the default pathway of pituitary differentiation as previously suggested by the *Pitx1/2* and *Lhx3/4* double knockout

mice[12,13], and the Pax7-driven melanotrope fate constituting an epigenetic re-programming of the default state.

In order to assess the dependence on Pax7 for opening melanotrope-specific enhancers, we performed ATACseq on *Pax7*$^{-/-}$ intermediate lobes. Thus, ATACseq peaks present at enhancers of the *Pcsk2*, *Drd2* and *Kif21b* loci in melanotropes, but not in corticotropes, are lost in *Pax7*$^{-/-}$ mice (Fig. 2f), indicating that chromatin opening at these enhancers requires Pax7 *in vivo*.

Pax7 is sufficient for deployment of the melanotrope enhancer repertoire

We previously showed that Pax7 opens enhancers[3] and ATACseq confirmed that Pax7 is sufficient to open the *in vivo* Pax7-dependent melanotrope enhancers (Fig. 2f). To broadly define the action of Pax7, we assessed the chromatin status at Pax7 binding sites in AtT-20 cells before and after Pax7 expression. We used the enhancer mark H3K4me1 to identify putative enhancers and we scored for presence of the coactivator p300 to further separate the subset of transcriptionally active enhancers[14]. The detailed analysis of chromatin changes at all Pax7 peaks is described in Supplementary Figure 4a and Online Methods. We took the Pax7-dependent gain of H3K4me1 at previously unmarked sites to define the subset of newly pioneered enhancers (Fig. 3a, b). These targets were subdivided into two subsets. A first subset of 2313 Pioneer Activated targets acquire the chromatin marks H3K4me1, p300, H3K27ac and ATACseq sensitivity after strong Pax7 binding (Supplementary Fig. 4b). This is associated with *de novo* nucleosomal depletion as revealed by ChIPseq for histone H3 (Fig. 3a) at sites that were DNase-resistant before Pax7 action (Supplementary Fig. 4c). A second subset of 8016 Pioneer Primed enhancers acquire H3K4me1 but not p300 or H3K27ac after Pax7 binding (Fig. 3b): these have a single weaker peak of H3K4me1 compared to the bimodal distribution of H3K4me1 at Pioneer Activated enhancers (Fig. 3a, b and Supplementary Fig. 4d). The subset of 2171 enhancers activated by Pax7 to acquire p300 show a switch from a single weaker peak to a bimodal distribution of H3K4me1; this is associated with appearance of H3K27ac and nucleosomal depletion (Fig. 3c and Supplementary Fig. 4d) consistent with the status of 16113 Constitutively Active enhancers (Fig. 3d).

To test whether the pioneering action of Pax7 increases accessibility to other TFs, we assessed recruitment of the critical developmental regulator of corticotrope and melanotrope differentiation, the Tbox factor Tpit, together with recruitment of the signal-dependent factor STAT3 that is activated in response to LIF treatment[15]. A large number

of Pioneer Activated sites (Fig. 3a) recruit Tpit and STAT3 after Pax7 opening but not before (Fig. 3e and Supplementary Fig. 4e). This occurs on a much smaller proportion for the Pioneer Primed subset (Fig. 3f). In contrast, the Pax7-activated (Fig. 3c) and Constitutively Active enhancers (Fig. 3d) are bound by STAT3 and/or Tpit before/after Pax7 binding (Fig. 3g, h). Similar data were obtained for GR (Supplementary Fig. 4f).

Interestingly, the Pioneer Activated enhancers (Fig. 3a) largely correspond to melanotrope-specific DARs defined by ATACseq in normal pituitary cells (Fig. 3i). The Pax7 Primed (Fig. 3b) and Pax7 Activated (Fig. 3c) enhancers show slightly greater ATACseq signals in melanotropes compared to corticotropes (Fig. 3j, k). Thus, the Pax7-pioneered enhancers defined in AtT-20 cells match quite well the putative enhancers (DARs) defined by ATACseq in normal pituitary cells. To validate the status of Pax7-pioneered enhancers in normal pituitary cells, we showed H3K4me1 at the *Pcsk2* and *Kif21b* enhancers (depicted in Fig. 2f) in intermediate but not anterior pituitary cells (Supplementary Fig. 4g, h).

Histone H3K4 methylation is associated with recruitment of the MLL complexes that include the Ash2l histone methyl transferase[16] and Pax7 was shown to interact with this complex[17]. To assess involvement of the MLL complex in Pax7-dependent chromatin remodeling, we performed ChIPseq for Ash2l. Ash2l is not present at pioneer target sites in AtT-20 cells and is recruited following Pax7 binding to both fully activated and primed (weaker) enhancers (Fig. 3q). Thus, Pax7 may remodel chromatin through recruitment of the MLL complex.

We then performed RNAseq before/after Pax7 to relate chromatin changes with gene expression. We identified 304 Pax7-repressed and 298 Pax7-induced genes (*P* value<0.05) (Fig. 3m). Promoters of both Pax7-induced and repressed genes are in similarly active state before/after Pax7 action (Fig. 3n). Pioneer Activated targets are more spatially associated with Pax7-induced than repressed genes (Fig. 3o) and they show higher sequence conservation than other targets (Fig. 3p). Collectively, the data suggest at least two steps of pioneer action with sets of enhancers that are either primed or become fully active after Pax7 pioneering; this latter subset matches the melanotrope-specific DAR candidate enhancers (Fig. 1d-f) and they are associated with activation of expression.

A unique chromatin environment for pioneering

In order to define the chromatin state that is permissive (or not) for Pax7 pioneer action, we identified a subset of Pax7 binding sites that are Resistant to Pax7 remodeling (Supplementary Fig. 4a). This Resistant subset exhibits weaker Pax7 binding compared to Pioneered and Constitutive enhancers (Fig. 4a). To address the contribution of binding site sequence to Pax7 binding and action, we assessed enriched motifs at these subsets. The Pax7 motifs are enriched at these three subsets (Supplementary Fig. 5). In addition, we found a strong enrichment of CTCF sequences at Resistant sites (Fig. 4b, c and Supplementary Fig. 5). We performed CTCF ChIPseq in AtT-20 cells and found that CTCF binds ~30% of the Resistant loci (Fig. 4d). These could be border elements (Insulators) and cohesin (SMC1) ChIPseq confirmed this[18]. We assessed the repressive histone marks H3K9me3 present in definitive heterochromatin together with H3K9me2 and K3K27me3 of facultative heterochromatin (Fig. 4d,e) in order to characterize the remaining Resistant sites. These marks were validated by comparison of TSS chromatin at expressed and non-expressed genes and by cross-correlation analyses (Supplementary Fig. 6). While the Constitutive enhancers and Insulator subsets have the lowest H3K9me3 levels, the CTCF-devoid Resistant subset exhibits the highest levels of H3K9me3 while the pioneered enhancers exhibit intermediate H3K9me3 levels (Fig. 4e). The overlap in H3K9me3 levels across these subsets suggests that, on its own, this mark may not be sufficient to explain pioneering ability or resistance. However, H3K9me2 levels are high and the same at pioneered and Resistant sites, and they show a strong depletion after Pax7 action at pioneered sites (Fig. 4e). In summary, the subset permissive for pioneering (intermediate H3K9me3 and high H3K9me2) appears to be facultative heterochromatin in contrast to both Insulator elements and definitive heterochromatin (high H3K9me2 and me3) that are resistant to Pax7 action.

Pax7 binds quickly but acts slowly at pioneering targets

In order to determine the relative timing of Pax7 binding to its pioneering sites relative to chromatin remodeling and gene activation, we engineered a Tamoxifen(Tam)-inducible Pax7 chimera (ER-Pax7, Fig. 5a and Supplementary Fig. 7a). The inducible ER-Pax7 targets a subset of pioneer (n=213) and constitutive (n=8399) Pax7 sites defined in stable Pax7-expressing cells and these subsets were used for analyses. ER-Pax7 cells do not show Pax7 DNA binding in absence of Tam (Fig. 5b, c). After Tam treatment for 30 or 60 minutes, ER-Pax7 binding is detected at both active enhancers (Fig. 5b) and pioneer

targets (Fig. 5c). At active enhancers, the binding is similarly high at 30 minutes and throughout the next 3 days (Fig. 5b). However at pioneer targets, Pax7 binding is weak at 30/60 minutes and is stable after 24h (Fig. 5c, d). Interestingly, the Resistant sites show similar binding compared to Pioneered sites at 30 min. but no change at 3 days (Fig. 5d). Chromatin accessibility was assessed at pioneered and constitutive enhancers by ATACseq and found to be similar after long term (>20 days) with Tam (Fig. 5e). Pioneered sites do not show significant ATACseq signal at 30 min. compared to untreated cells but they show increasing accessibility over the next 3 days (Fig. 5e). Thus, Pax7 pioneer targets are bound quickly but the action on chromatin accessibility is delayed. Accordingly, we identified genes with rapid and slow kinetics of transcriptional activation (Fig. 5f). For example, the melanotrope hallmark *Pcsk2* gene associated with pioneering (Fig. 1g and 2f) exhibits slow activation kinetics with highest expression after 5 days of Tam. Similar profiles were obtained for larger groups of Pax7-induced genes identified by RNAseq after 12, 24 or 96 h of Tam treatment (Supplementary Fig. 7b). In summary, despite the rapid binding of Pax7 to its pioneer targets, chromatin opening and gene activation are slower, in stark contrast to the rapid process observed at direct transcriptional targets of Pax7.

Pax7 pioneer action leads to loss of DNA methylation and long-term chromatin accessibility

DNA methylation being the hallmark of stable gene repression, we assessed whether Pax7 pioneer action may regulate this process. Thus, we measured its impact on DNA CpG methylation by whole-genome bisulfite sequencing (WGBS) before and after Pax7 action. Before Pax7, most CpGs at pioneer targets have >80% methylation while active enhancers have <20% CpG methylation at +/-400bp from Pax7 peak summits (Fig. 6a). Thus, Pax7 pioneer sites are within methylated regions indicating that DNA methylation does not prevent Pax7 binding at these sites (Fig. 6b). After Pax7 pioneering, DNA methylation is greatly reduced at pioneer sites although in a more locally restricted manner than at active enhancers (<40% methylation at +/- 160bp). Further, the canonical Pax7 binding site contains a CpG that is itself highly methylated before Pax7 action and that becomes demethylated after (Fig. 6c). The loss of DNA methylation at pioneered enhancers would be expected to provide epigenetic memory. To investigate this, we used the ER-Pax7 system in order to effectively remove Pax7 from its targets as shown by ChIP-qPCR (Fig. 6d) and assess if Pax7 remodeling is reversible. We found that ER-Pax7 pioneered regions show decreased accessibility three days after removing Tam but remain ATACseq accessible for at least 18 days (Fig. 6e); this represents at least twelve

cell divisions for these cells. Consistent with this long-term accessibility, the signal-dependent factor STAT3 can access sites that were pioneered by Pax7 18 days before its acute (20 min. in presence of LIF) activation (Fig. 6f). In sum, Pax7 can initiate chromatin remodeling at CpG-methylated enhancers leading to loss of DNA methylation and memory of this reprogramming event.

DISCUSSION

In eukaryotes, developmental processes are tightly controlled by chromatin organization and epigenetic regulation of genome accessibility is critical for normal development. Pioneer factors have a critical role for epigenome remodeling and implementation of new developmental fates. In the present work, we used a simple system to define the hallmark properties of Pax7, a pioneer factor that confers pituitary intermediate lobe identity and thus presets the epigenome prior to terminal differentiation. Despite also being driven by Tpit, the anterior pituitary corticotrope fate has a very different developmental history compared to the intermediate lobe where Pax7 expression is the end point of its unique developmental history as the only site of continued contact between neuro and surface ectoderms[20,21]. Through its pioneer action, Pax7 is necessary *in vivo* and sufficient in AtT-20 cells for specification of the melanotrope cell fate.

The mechanism of pioneer recognition of target DNA sequences has created much speculation in recent years. As factors that trigger developmental switches through opening of naïve chromatin, pioneers have been assumed, and shown, to bind their targets within nucleosomal DNA[2,22]. The present work is consistent with a rapid binding of Pax7 to nucleosomal DNA targets, and indeed, nucleosome displacement is observed after Pax7 action (Fig. 3 and 7). One unique feature of pioneer chromatin recruitment may be their ability for widespread low affinity interactions[23,24]: these were interpreted as a possible scanning mechanism. The lower affinity sites may correspond to the rapid on-off subgroup of binding sites that were identified for the pioneer FoxA and also for the nuclear receptors GR and ER[25]. Pax7 also has ~30% low affinity binding peaks that are not marked by either H3K4me1 or p300 (Supplementary Fig. 4a) and that are resistant to Pax7-dependent chromatin remodeling (Fig. 4). These Resistant sites bind Pax7 with a low apparent affinity that resembles the initial (30/60 min) Pax7 binding to pioneered sites (Fig. 5d); binding to the latter group is stabilized within 24h whereas this does not occur at Resistant sites. Clearly, there must be permissive conditions that characterize the facultative heterochromatin pioneered by Pax7 (Fig. 7). The repressive marks profiles are

not strikingly different at pioneered compared to Resistant sites: the higher H3K9me3 levels at Resistant sites may qualify these as definitive heterochromatin while the similarly high H3K9me2 levels are suggestive of the pioneer-permissive environment as facultative heterochromatin (Fig. 7). At this time, it is not clear what chromatin reader may distinguish this difference in H3K9 di- versus tri-methylation but it is noteworthy that the H3K4 methyl transferase Ash2l that is recruited to Pax7-pioneered enhancers is inhibited by H3K9 trimethylation of flanking residues[26].

Pax7 action on the melanotrope enhancer repertoire is a slow process (Fig. 5 and 7) since it takes a few days to activate genes that depend on pioneering for expression; in contrast, less than 24h are required for transcriptional activation (Fig. 5f). This slow time course suggests that implementation of the long term effects of Pax7 pioneering may require passage through cell division which occurs every 30-36 hours in AtT-20 cells. This slow process is also reflected at the level of DNA accessibility as the ATACseq signal slowly increases over at least 3 days at Pioneered sites but not within the 30 minutes that are sufficient for Pax7 binding. These slow kinetics are however comparable to the transcriptional response of pioneered genes and they are consistent with a model where a permissive chromatin environment is established at replication.

In addition to the one-shot event that represents pioneering, Pax7 acts as a usual transcriptional regulator with a maintenance function for a large transcriptome. This maintenance role includes preferential recruitment and interaction with other TFs such as Tpit. This type of interaction was described in details between FoxA and nuclear receptors such as the estrogen, glucocorticoïd and androgen receptors[25,27-29]. These interactions, sometime labelled as assisted-loading[23] or as « settler » factors[30], take effect within short time frames (30 minutes treatments for ligand activation of nuclear receptors) and thus represent facilitation of enhancer activation. The enhancer targets of these interactions may be in the primed state since weak DHS signals are present at these sites before binding of either FoxA or ER (eg the recent work of [25]). This is quite different when compared to Pax7 pioneered enhancers that have no ATACseq (Fig. 3a, b), FAIREseq[3] or DHS signal before Pax7 action (Supplementary Fig. 4c). Assisted-loading may thus contribute to establishment and maintenance of developmental programs. In contrast, chromatin pioneering is a slower process that may require DNA replication, that provides long term stable memory as reflected by changes in DNA methylation and that is quite different from assisted-loading or factor cooperativity.

As gatekeeper of epigenetic stability, DNA methylation provides an ultimate memory of cell identity. In the present work, we showed loci-specific CpG demethylation associated with long term enhancer accessibility. Whether implementation of this epigenetic memory depends solely on DNA hypomethylation or whether unique chromatin components are also involved remains to be investigated. DNA methylation-dependent memory requires passage through replication to be altered[31]. Thus, this process is coherent with the slow chromatin opening of Pax7-pioneered enhancers. The 3-5 days required for pioneer activation of gene expression (Fig. 5f and Supplementary Fig. 7b) represent more than a single cell division. This time frame also argues against DNA accessibility to pioneered sites being provided by the simple passage through replication as Pax7 binding already reached its maximum at 24h. The kinetics are however more consistent with stepwise changes introduced at the time of cell division such as impaired maintenance of DNA methylation or active demethylation. In this respect, it is noteworthy that Pax7 appears capable of binding the composite DNA binding site (enriched at pioneer sites) even when methylated. Indeed, this site often includes a CpG dinucleotide and it was found to be methylated before Pax7 action but demethylated after (Fig. 6c). Thus, the slow kinetics of pioneer site opening and gene activation may reflect the time required to stably establish new methylation patterns and associated chromatin components. Be that as it may, the accessibility of Pax7-pioneered enhancers is stably maintained once established, consistent with DNA methylation playing a critical role in this maintenance.

In summary, Pax7 fulfills its role of selector gene for pituitary intermediate lobe identity through a pioneer activity that has all the mechanistic hallmarks expected of this class of factors. Specifically, Pax7 binds its pioneer target sequences in naïve chromatin irrespective of nucleosomes or CpG methylation; it does so rapidly (less than 30 minutes) but requires longer than a cell division to implement its effect on chromatin organization. The melanotrope-specific enhancer repertoire acquires stable epigenetic memory providing long-term access for other non-pioneer transcription factors.

ACKNOWLEDGMENTS

We are grateful to our colleagues Nicole Francis for critical comments on the manuscript, Lionel Budry and Steve Nemec for some profiling data and Pitx1 ChIPseq, respectively, Odile Neyret for NGS analyses, Eric Massicotte for FACS sorting, Alexis Blanchet-Cohen for WGBS data analysis, and to Evelyne Joyal for her expert secretarial assistance.

Alexandre Mayran was supported by an IRCM Challenge fellowship. This work was supported by grants to JD from the Canadian Institutes of Health Research.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

ONLINE METHODS

Mice, cells and tissue culture. After dissection, EGFP-positive cells of pituitary intermediate and anterior lobes were FACS-sorted from POMC-EGFP mice by the IRCM FACS core facility[10]. For RNAseq, two biological replicates of pools of eleven C57Bl/6 pituitaries were used whereas similar pools of seven pituitaries were used for ATACseq. Two biological replicates, each a pool of three 129/sv *Pax7*$^{-/-}$ and *Pax7*$^{+/+}$ intermediate lobes, were used for ATACseq analyses. AtT-20 cells (obtained from the late E. Herbert in 1981 and maintained in this lab since with yearly negative mycoplasma tests) were grown and selected as described[3]. Tamoxifen treatment used a final concentration of 400 nM and 0.1% ethanol for the specified duration and control cells were treated with 0.1% ethanol. For cells treated more than 24h, cell medium containing 400nM Tamoxifen was renewed every day. Dexamethasone and LIF treatment was performed for 20 minutes at 10^{-7}M and 10 ng/ml respectively, on cells used for assessment of GR and STAT3 binding. *Pax7*$^{-/-}$ and *Tpit*$^{-/-}$ mice were described previously[3,9].

All animal experimentation was approved by the IRCM Animal Ethics Committee in accordance with Canadian regulations.

Genome-wide analyses. Table S1 provides all experimental conditions and reagents for ATACseq, ChIPseq, RNAseq and WGBS samples. All data have been deposited on GEO as GSE87185.

ATACseq. ATACseq was performed according to[32] with small alterations to the original protocol to diminish mitochondrial contamination and increase signal to noise ratio. Briefly, we isolated 100, 000 nuclei using serial 30 minutes incubation at 4 °C, first for 30 minutes in a hypotonic cell lysis buffer (0.1% $^{w}/_{v}$ Sodium Citrate Tribasic Dihydrate and 0.1% $^{v}/_{v}$ Triton X-100) followed by 30 minutes in normal cell lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% $^{v}/_{v}$ IGEPAL CA-630). Transposition was performed directly on nuclei using 25 µl of a transposase Master Mix (2.5 µl 10x TD Buffer, 10 µl H_2O, + 12.5 µl Enzyme from Illumina Nextera Kit; FC-121-1031). DNA was then purified and enriched by PCR, and the library was recovered using GeneRead Purification column (Qiagen) and sequenced on Illumina Hi-seq 2000.

ChIPseq. ChIP-qPCR and ChIP-seq were performed and analyzed as described[15]. At least 3 independent ChIPs were pooled before library preparation. For Tam induction experiments, we used ChIPexo conditions in order to optimize signal-to-noise ratios using

the Active Motif ChIP exo kit[33]. 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 S1 provides details of antibodies used for ChIP and sequencing depth and Table S2 lists the PCR primers.

RNAseq. For each biological replicate, RNA was extracted from 1,000,000 cultured cells or from 250,000 FACS-purified corticotropes or melanotropes using Qiagen RNeasy plus mini kit for cultured cells and the Zymo Research quick-RNA MiniPrep kit for normal pituitary cells. Ribosomal depletion, library preparation and flow cells preparation for sequencing were performed by the IRCM Molecular Biology Core Facility following Illumina recommendations (Illumina, San Diego, CA).

Whole Genome Bisulfite Sequencing (WGBS). Genomic DNA was extracted from 1,000,000 AtT-20Neo and AtT-20Pax7cells using Qiagen DNeasy Blood & Tissue Kit. Bisulfite conversion was done with Zymo Research EZ DNA Methylation-Lightning Kit; library and flow cell preparations were performed by the IRCM Biology Core Facility following Illumina recommendations (Illumina, San Diego, CA).

ChIPseq and ATACseq peak analyses. All ChIPseq and ATACseq used 50bp paired-end sequencing reads, except the Tpit ChIPseq that used 50bp single-end reads, and the ATACseq in AtT-20Neo and AtT-20Pax7 cells that used 100bp paired-end reads. We mapped ChIPseq and ATACseq reads on the mouse genome assembly mm10 using bowtie v1.1.2 using the following setting: bowtie -t -p 4 --trim5 1 --best mm10 –S[34]. For the Pax7-based analyses in Figures 3, 4 and 5, to identify significant binding/presence of TF, co-activator, histone modification and accessibility by ATACseq, we processed the mapped sequence reads with MACS version 2.1.1 against their matching control samples (see details in Table S1) using the parameters: --bw 250 -g mm --mfold X X -p 1e-5 [35]). The MACS option –mfold was determined independently for each experiment. We kept peaks with *P* values $<10^{-5}$ for further analyses. Quantification of input samples at Pax7 peaks using HOMER[36] allowed removal of repeated or duplicated regions to extract single copy loci. Single copy Pax7 peaks were matched with H3K4me1 peaks within a 2kb window; for TF and p300 peaks, we used a 1kb window to define overlaps. For DARs analyses in purified melanotropes and corticotropes, we used each sample as control for the other and kept differential peaks with $P <10^{-20}$.

Motif analyses. Both *de novo* motif searches and known motifs were identified within 200bp windows around DAR summits using the HOMER findMotifsGenome command[36]. Motif densities were assessed using the HOMER annotatePeaks command with the matrices of the motifs that resulted from the *de novo* motif analyses.

RNAseq analyses. Strand specific RNAseq paired-end reads were trimmed using trimmomatic and mapped on the ensembl GRCm38.77 genome with tophat2[37] with the parameters: --rg-library "L" --rg-platform "ILLUMINA" --rg-platform-unit "X" --rg-id "run#many" --no-novel-juncs --library-type fr-firststrand -p 12. Gene expression was quantified using HOMER analyzeRepeats command[36] and differential expression assessed using EdgeR 3.12.1[38].

WGBS analysis. We generated ~750,000,000 100bp paired-end reads per sample and aligned them to the bisulfite converted reference genome GRCm38 with Bismarck 0.14.3[39]. After alignment, the deduplicate script (Bismark) was used to remove duplicate reads. The methylation counts were then extracted with methylKit 0.9.4[41]. MethylKit was also used to calculate differential methylation between the AtTNeo and AtTpax7 samples. Methylation was computed in all three contexts, CpG, CHG, and CHH (where H is A, C or T). Methylation was also computed at single base level.

Heatmaps, box-plots, average profiles, dispersion plot generation and statistical analyses. We generated the figures using a combination of easeq[40], (http://easeq.net), and HOMER commands.

Statistical analyses were performed using unpaired two-sided T-test or unpaired two-sided Wilcoxon rank-sum test as indicated.

Data Availability. All genomic data have been deposited on Gene Expression Omnibus (GEO) under accession number GSE87185.

REFERENCES

1. Soufi, A., Donahue, G. & Zaret, K.S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994-1004 (2012).
2. Cirillo, L.A. *et al.* Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279-289 (2002).

3. Budry, L. *et al.* The selector gene Pax7 dictates alternate pituitary cell fates through its pioneer action on chromatin remodeling. *Genes Dev* **26**, 2299-2310 (2012).
4. van Oevelen, C. *et al.* C/EBPalpha Activates Pre-existing and De Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis. *Stem Cell Reports* **5**, 232-47 (2015).
5. Boller, S. *et al.* Pioneering Activity of the C-Terminal Domain of EBF1 Shapes the Chromatin Landscape for B Cell Programming. *Immunity* **44**, 527-41 (2016).
6. Wapinski, O.L. *et al.* Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* **155**, 621-35 (2013).
7. Pataskar, A. *et al.* NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. *Embo j* **35**, 24-45 (2016).
8. Lamolet, B. *et al.* A pituitary cell-restricted T-box factor, Tpit, activates POMC transcription in cooperation with Pitx homeoproteins. *Cell* **104**, 849-859 (2001).
9. Pulichino, A.M. *et al.* Tpit determines alternate fates during pituitary cell differentiation. *Genes Dev* **17**, 738-747 (2003).
10. Lavoie, P.L., Budry, L., Balsalobre, A. & Drouin, J. Developmental Dependence on NurRE and EboxNeuro for Expression of Pituitary POMC. *Mol. Endocrinol* **22**, 1647-1657 (2008).
11. Langlais, D., Couture, C., Kmita, M. & Drouin, J. Adult pituitary cell maintenance: lineage-specific contribution of self-duplication. *Mol. Endocrinol* **27**, 1103-1112 (2013).
12. Sheng, H.Z. & Westphal, H. Early steps in pituitary organogenesis. *Trends Genet* **15**, 236-240 (1999).
13. Charles, M.A., Suh, H., Drouin, J., Camper, S.A. & Gage, P.J. PITX genes are required for cell survival and *Lhx3* activation. *Mol. Endocrinol* **19**, 1893-1903 (2005).
14. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol. Cell* **49**, 825-837 (2013).

15. Langlais, D., Couture, C. & Drouin, J. The Stat3/GR interaction code: predictive value of direct/indirect DNA recruitment for transcription outcome. *Molecular Cell* **47**, 38-49 (2012).
16. Schuettengruber, B., Bourbon, H.M., Di Croce, L. & Cavalli, G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* **171**, 34-57 (2017).
17. Kawabe, Y., Wang, Y.X., McKinnell, I.W., Bedford, M.T. & Rudnicki, M.A. Carm1 regulates Pax7 transcriptional activity through MLL1/2 recruitment during asymmetric satellite stem cell divisions. *Cell Stem Cell* **11**, 333-45 (2012).
18. Ghirlando, R. & Felsenfeld, G. CTCF: making the right connections. *Genes Dev* **30**, 881-91 (2016).
19. Liu, F., Wang, L., Perna, F. & Nimer, S.D. Beyond transcription factors: how oncogenic signalling reshapes the epigenetic landscape. *Nat Rev Cancer* **16**, 359-72 (2016).
20. Drouin, J. Chap 1 - Pituitary Development. in *The Pituitary (Fourth Edition)* (ed. Melmed, S.) 3-22, Academic Press, 2017.
21. Rizzoti, K. Genetic regulation of murine pituitary development. *J Mol Endocrinol* **54**, R55-73 (2015).
22. Zaret, K.S., Lerner, J. & Iwafuchi-Doi, M. Chromatin Scanning by Dynamic Binding of Pioneer Factors. *Mol Cell* **62**, 665-7 (2016).
23. Voss, T.C. *et al.* Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. *Cell* **146**, 544-554 (2011).
24. Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555-68 (2015).
25. Swinstead, E.E. *et al.* Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. *Cell* **165**, 593-605 (2016).
26. Wysocka, J., Myers, M.P., Laherty, C.D., Eisenman, R.N. & Herr, W. Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4

methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. *Genes Dev* **17**, 896-911 (2003).

27. Laganiere, J. *et al.* From the Cover: Location analysis of estrogen receptor {alpha} target promoters reveals that FOXA1 defines a domain of the estrogen response. *Proc. Natl. Acad. Sci. USA* **102**, 11651-11656 (2005).
28. Wang, Q. *et al.* A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. *Mol. Cell* **27**, 380-392 (2007).
29. Carroll, J.S. *et al.* Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122**, 33-43 (2005).
30. Sherwood, R.I. *et al.* Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol* **32**, 171-178 (2014).
31. Almouzni, G. & Cedar, H. Maintenance of Epigenetic Information. *Cold Spring Harb Perspect Biol* **8**(2016).
32. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-8 (2013).
33. Rhee, H.S. & Pugh, B.F. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408-19 (2011).
34. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
35. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
36. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589 (2010).

37. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36 (2013).
38. McCarthy, D.J., Chen, Y. & Smyth, G.K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288-97 (2012).
39. Krueger, F. & Andrews, S.R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-2 (2011).
40. Lerdrup, M., Johansen, J.V., Agrawal-Singh, S. & Hansen, K. An interactive environment for agile analysis and visualization of ChIP-sequencing data. *Nat Struct Mol Biol* **23**, 349-57 (2016).
41. Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol* **13**, R87 (2012).

FIGURES

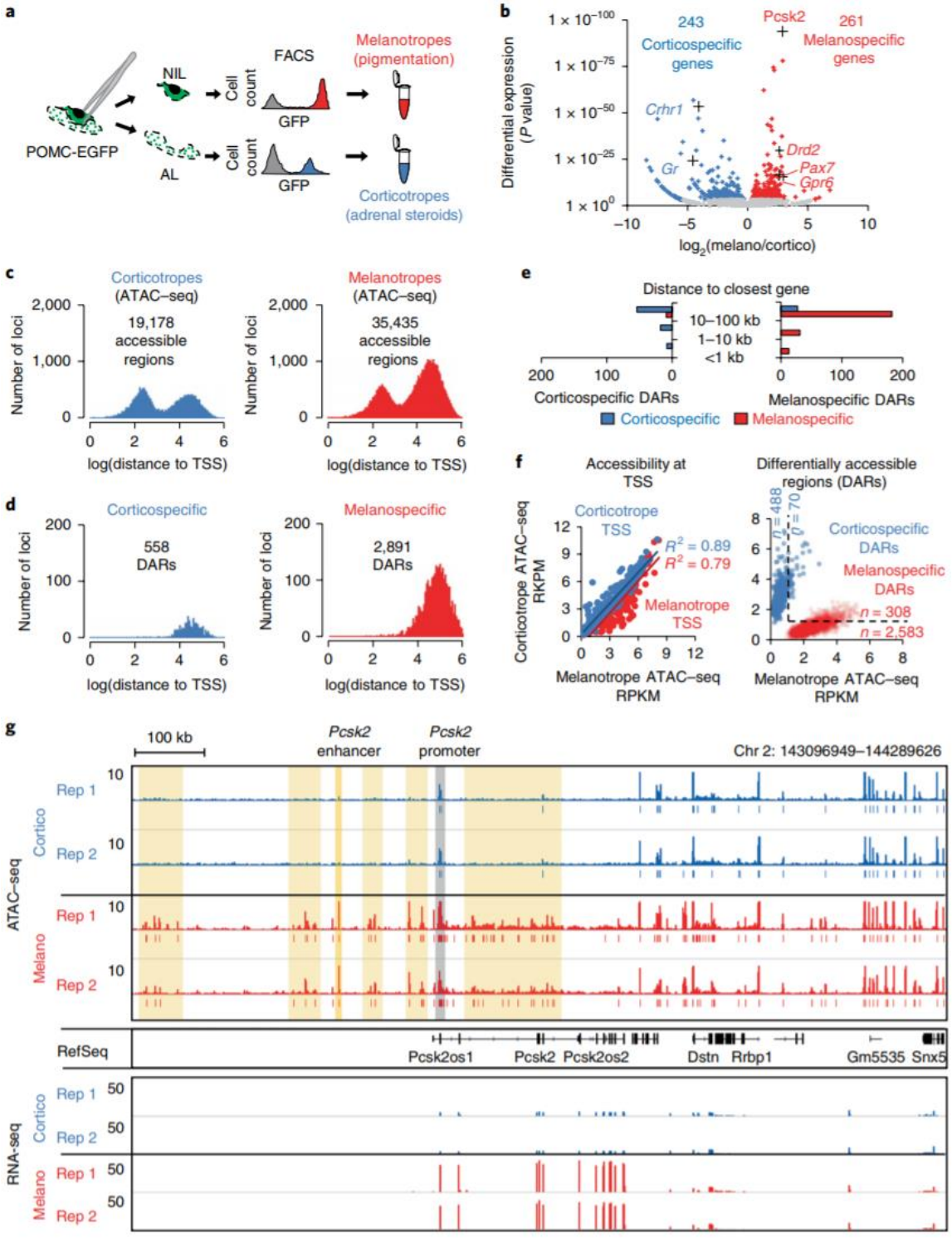


Figure 1. Unique enhancer, but not promoter, repertoires define lineage-specific chromatin landscapes.

Figure 1. Unique enhancer, but not promoter, repertoires define lineage-specific chromatin landscapes.

a) Schematic representation of pituitary tissue and experimental procedure for purification of POMC-expressing melanotropes of the intermediate lobe (NIL) and corticotropes of the anterior lobe (AL).

b) Volcano plot of differential gene expression determined by RNAseq analysis of two independent primary cell replicates of pituitary corticotropes and melanotropes. Blue and Red dots represent genes that are differentially expressed (FDR<0.05) in corticotropes and melanotropes, respectively. Genes that are not significantly differentially expressed are shown in grey. The number of cell-specific genes is indicated; the *P* values for their enrichment derive from analyses using the edgeR tool.

c) Genomic distribution of the distance between ATACseq peaks and the closest TSS. Accessible regions segregate in two groups, proximal (<1kb from TSS) and distal regions (>1kb from TSS).

d) Genomic distribution of the distance between lineage-specific Differentially Accessible Regions (DARs, $P < 10^{-20}$) and the closest TSS.

e) Distance from cell-specific DARs to the nearest cortico- or melano-specific gene, as indicated.

f) Dispersion plots of ATACseq signals from the average of two independent primary cell replicates of purified corticotropes and melanotropes at the TSS (± 200 bp) of differentially expressed genes (left) and at cell specific DARs (right), blue for corticotrope and red for melanotrope genes. The DARs strongly cluster according to lineage. Dotted lines indicate threshold of minimal RPKM signal at statistically significant ($P < 10^{-20}$) DARs.

g) RNAseq and ATACseq profiles of the melanotrope-specific *Pcsk2* locus. Highlighted regions represent the promoter (grey) and DARs (yellow).

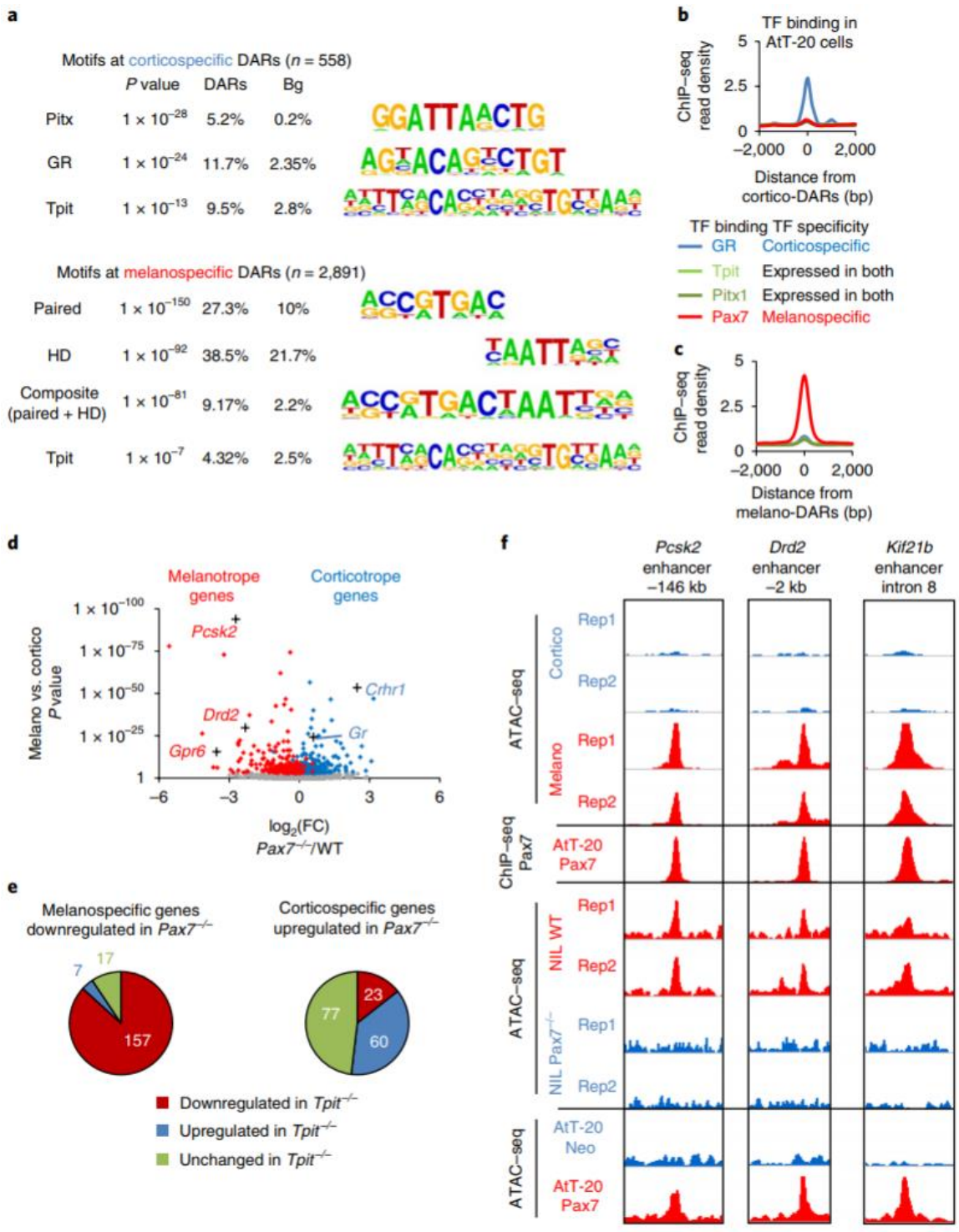


Figure 2. Pax7 binds melanotrope-specific DARs and is required for expression of melanotrope-specific genes.

Figure 2. Pax7 binds melanotrope-specific DARs and is required for expression of melanotrope-specific genes.

a) Motif enrichments at cortico- and melano-specific DARs derived from analyses using HOMER.

b-c) Average profiles of ChIPseq data for GR (from Dex-induced AtT-20 cells), Tpit and Pitx1 (in AtT-20 cells) and Pax7 (in Pax7-expressing AtT-20 cells) at cortico-(b) and melano(c)-specific DARs. Read densities refer to reads per 10^7 reads.

d) Volcano plot of changes in gene expression assessed by microarrays[3] of two tissue replicates of *Pax7*$^{-/-}$ *c*ompared to wild-type intermediate pituitaries (X axis) compared to *P* values of melanotrope versus corticotrope differential expression (Y axis, from Fig. 1B). Hallmark genes of each lineage are identified on the diagram.

e) Pie charts showing proportion of Pax7-regulated[3] (*P* <0.05) melanotrope and corticotrope genes (from Supplementary Fig. 3c) also affected (*P* <0.05) in *Tpit*$^{-/-}$ pituitaries.

f) Genome browser views of ATACseq signals in purified corticotropes and melanotropes, in wild-type and Pax7$^{-/-}$ intermediate lobes and in AtT-20 cells before/after Pax7 expression. Pax7 ChIPseq from AtT20-Pax7 cells is also shown at three putative enhancers of melanotrope genes.

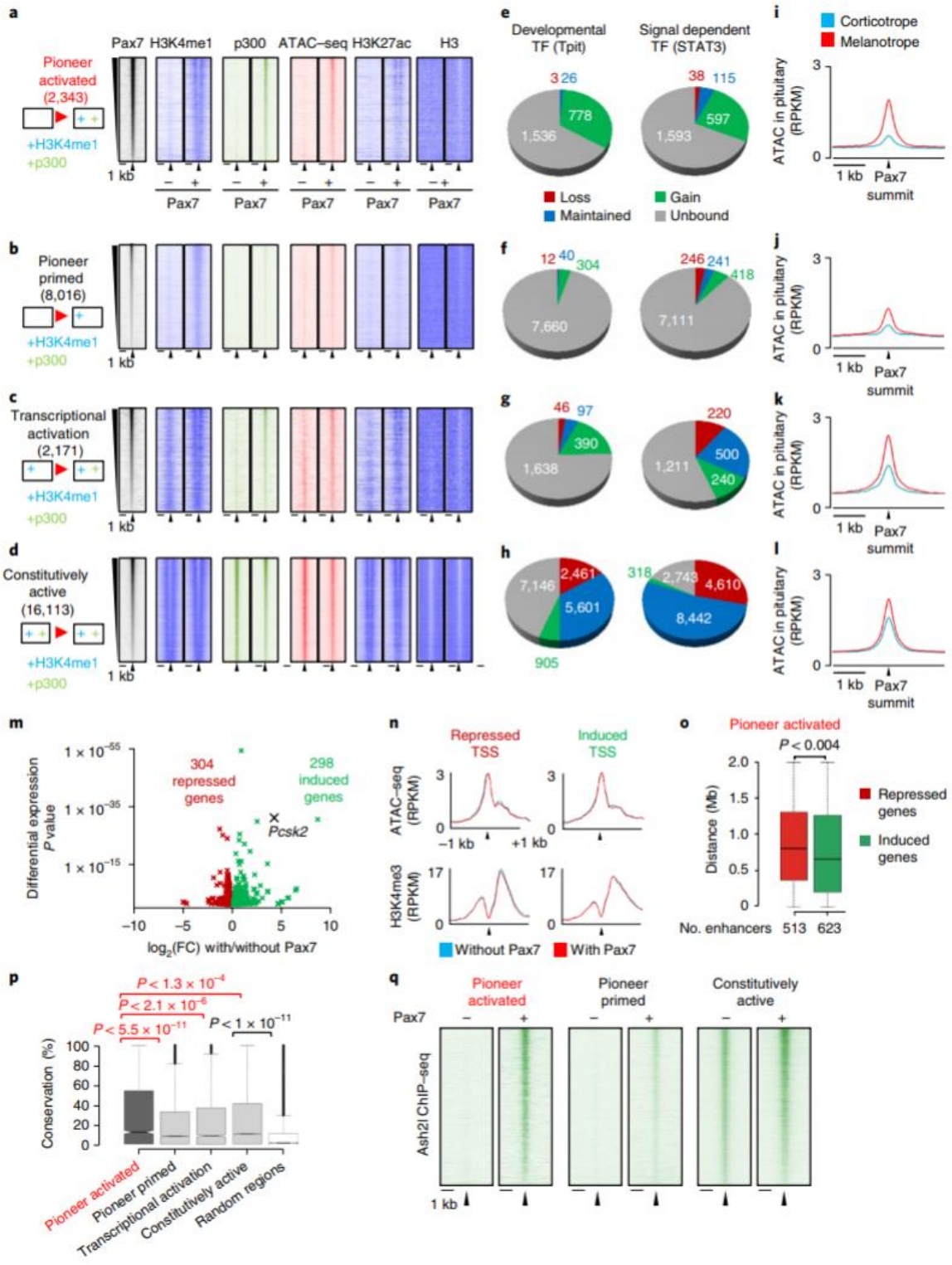


Figure 3. Pax7 pioneers chromatin opening at a subset of sites that had no previous recognizable chromatin mark.

Figure 3. Pax7 pioneers chromatin opening at a subset of sites that had no previous recognizable chromatin mark.

Pax7 binding sites determined by ChIPseq were clustered according to Pax7-dependent changes in the chromatin mark H3K4me1 and with recruitment of p300 measured by ChIPseq; details of the clustering are provided in Supplementary Figure 3a.

a-d) Heatmaps for ChIPseq of Pax7, H3K4me1, p300, H3K27ac, H3 and ATACseq at different subsets of Pax7 targets: Pioneer Activation (a), Pioneer Priming (b), Transcriptional Activation (c) and Constitutively Active enhancers (d). In each case, data are shown for AtT-20 cells before and after Pax7 expression; the number of peaks in each subset is indicated and metaplots are provided in Supplementary Figure 3d.

e-h) Pie charts showing changes in binding (determined by ChIPseq, $P < 10^{-5}$ derived from MACS analyses) for the developmental TF Tpit and signal-dependent TF STAT3 at the different subsets of Pax7 targets, Pioneer Activation (e), Pioneer Priming (f), Transcriptional Activation (g) and at Constitutively Active enhancers (h). Similar results were obtained for the glucocorticoid receptor GR (Supplementary Fig. 3e).

i-l) ATACseq profiles in normal pituitary cells for the different subsets of Pax7 targets, Pioneer Activation (i), Pioneer Priming (j), Transcriptional Activation (k) and at Constitutively Active enhancers (l).

m) Volcano plot of differentially (P <0.05) expressed genes assessed by RNAseq before/after Pax7 expression in two independent cell replicates of AtT-20 cells; the P values for their enrichment derive from analyses using the edgeR tool.

n) Average plots showing ATACseq and H3K4me3 ChIPseq at TSS of Pax7 repressed and induced genes.

o) Box-plot of distance (Mb) between Activated Pioneered enhancers and the closest repressed, induced or switched-on genes. Center lines show medians; box limits indicate the 25^{th} and 75^{th} percentiles; whiskers extend 1.5 times the interquartile range from the 25^{th} to 75^{th} percentiles; outliers are represented by dots. P value was assessed by unpaired two-sided Wilcoxon rank-sum test.

p) Average Phastcons sequence conservation for different groups of Pax7-bound enhancers. P values were assessed by unpaired two-sided Wilcoxon rank-sum test.

q) Heatmaps of Ash2l ChIPseq at the Pioneer Activated, Primed and Constitutively Active targets of Pax7 in AtT-20 cells before/after Pax7 action.

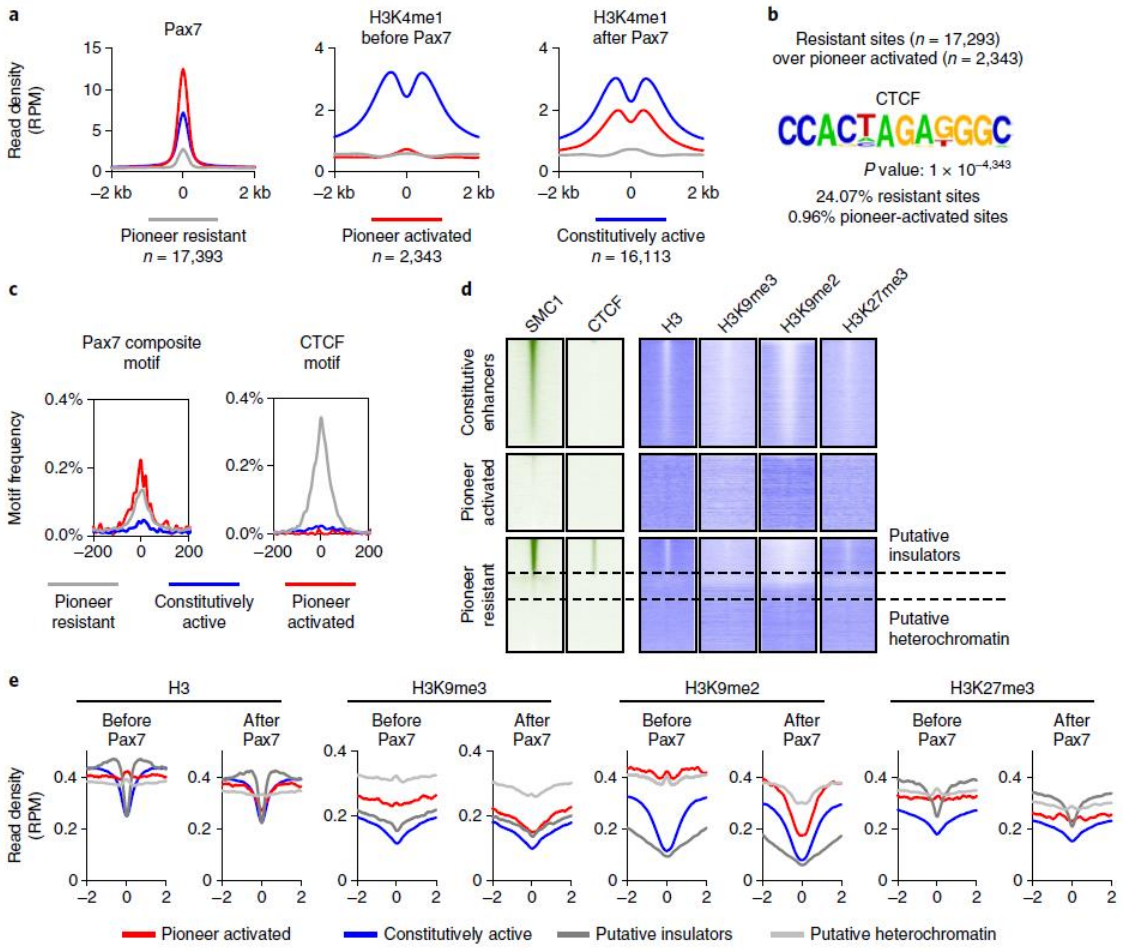


Figure 4. A unique chromatin environment for pioneering by Pax7.

Figure 4. A unique chromatin environment for pioneering by Pax7.

a) Average plots of Pax7 and H3K4me1 ChIPseq in AtT-20 cells before/after Pax7 expression at Pioneer Resistant, Pioneer Activated and Constitutively Active Pax7 subsets. RPM, reads per million.

b) De novo motif search (derived from analyses using HOMER) of the Pioneer Resistant subset identifies the CTCF motif as the most significantly enriched motif compared to Pioneer Activated targets.

c) Average plots of motif frequencies for Pax7 composite and CTCF motifs at the Pioneer Resistant, Pioneer Activated and Constitutively active subsets.

d) Heatmaps of SMC1, CTCF, H3, H3K9me3, H3K9me2, H3K27me3 at the Pioneer Resistant, Pioneer Activated and Constitutively active subsets.

e) Average plots of H3, H3K9me3, H3K9me2 and H3K27me3 ChIPseq in AtT-20 before/after Pax7 expression at Pioneer Resistant, Pioneer Activated and Constitutively Active Pax7 subsets. RPM, reads per million.

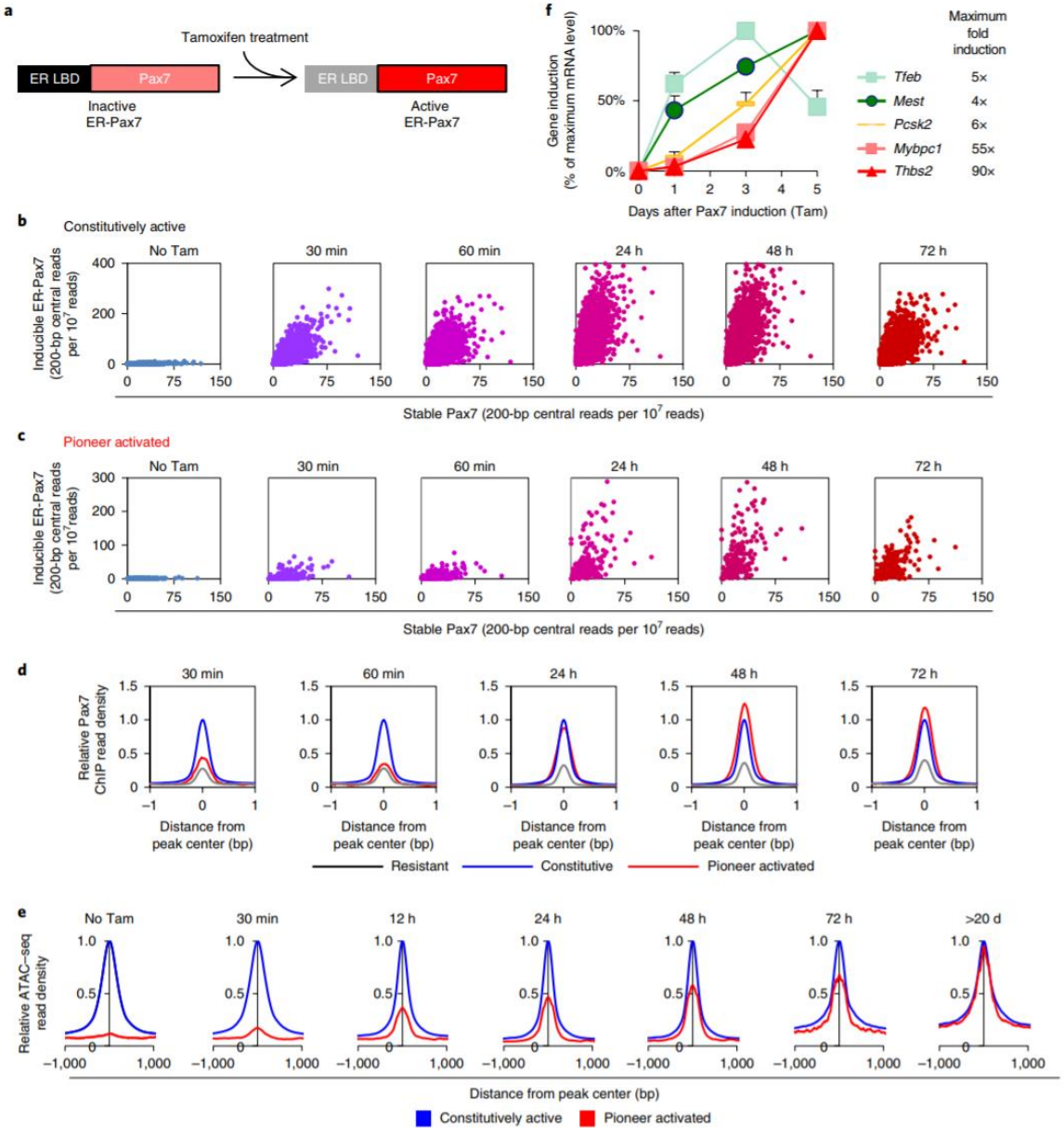


Figure 5. Pax7 binds quickly but acts slowly at pioneer sites.

Figure 5. Pax7 binds quickly but acts slowly at pioneer sites.

a) Schematic of the ER^{Tam}-Pax7 chimera used to assess the kinetics of Pax7 binding and action.

b) Dispersion plots of Pax7 binding at Constitutively Active enhancers (n=8399) in Tamoxifen (Tam)-induced ER-Pax7 cells (for 0, .5, 1, 24, 48, 72h) compared to stable Pax7-expressing cells as determined by ChIPseq.

c) Dispersion plots of Pax7 binding at Pioneer Activated enhancers (n=214) in Tam-induced compared to stably-expressing Pax7 cells.

d) Average profiles of Pax7 ChIPseq signals at Pioneer Activated (red), Pioneer Resistant (grey) and Constitutively Active (blue) subsets normalized to the summit of Constitutive enhancers (blue) without induction and after 0.5, 1, 12, 24, 48, 72h in presence of Tam.

e) Average profiles of ATACseq signals at Pioneer Activated (red) relative to summit of Constitutive enhancers (blue) without induction, after 30 min, 12, 24, 48, 72h and more than 20 days of Tam treatment.

f) Time-course of mRNA induction for two transcriptionally-activated Pax7 target genes (*Tcfeb* and *Mest*) and for Pax7 pioneer-dependent target genes (*Pcsk2*, *Mybpc1*, *Thbs2*). RT-qPCR was used to assess mRNA levels (normalized to *GAPDH* mRNA levels) at different times following addition of Tam and the maximal fold induction for each mRNA is indicated next to the gene list. Data are means ±sem of four separate experiments, each assessed in duplicates.

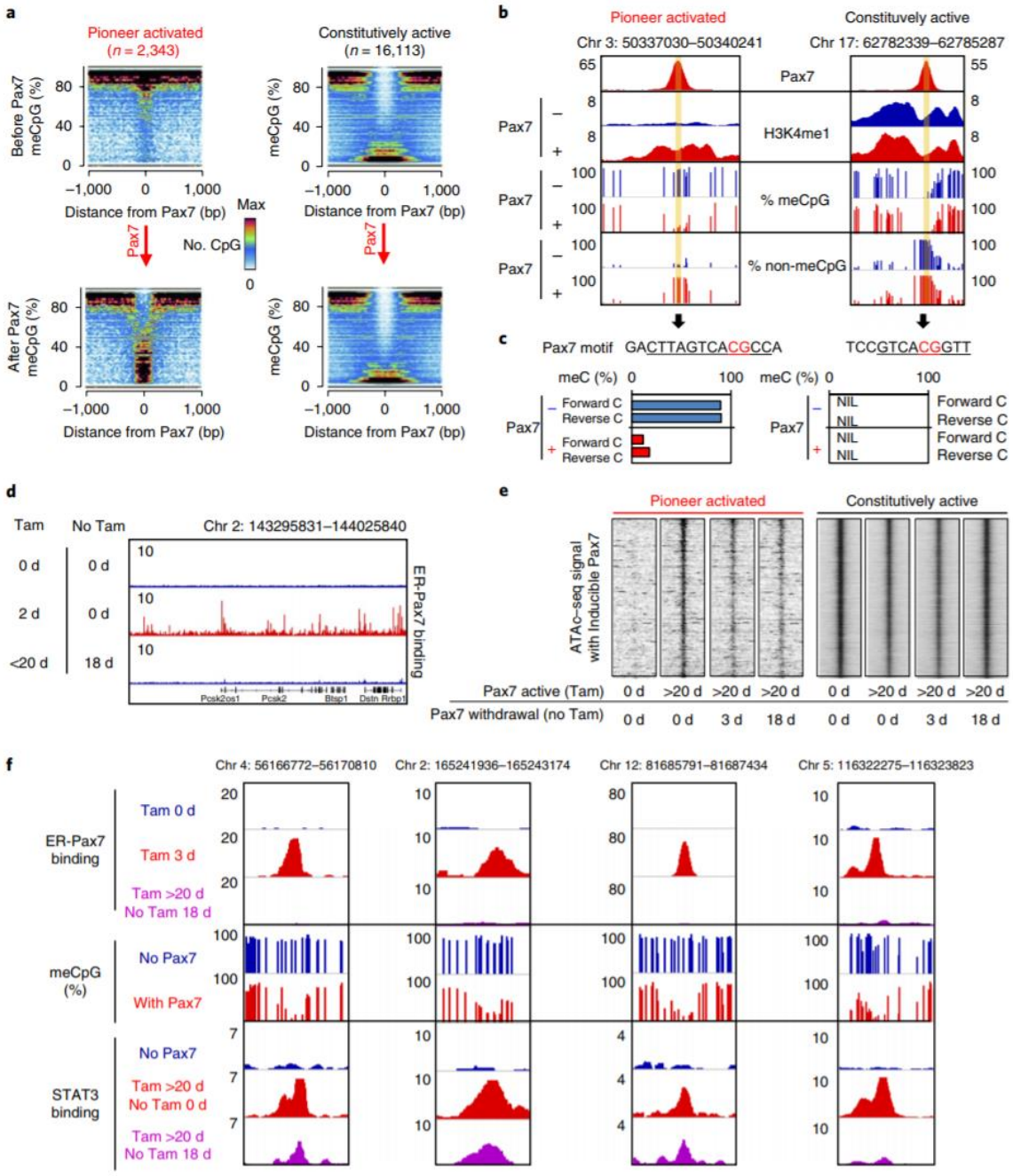


Figure 6. Long term epigenetic memory and DNA demethylation.

Figure 6. Long term epigenetic memory and DNA demethylation.

a) Whole genome bisulfite sequencing (WGBS) was performed on control AtT-20 and Pax7-expressing AtT-20 cells and used to generate dispersion plots (easeq) of CpG methylation levels at enhancer subsets centered on Pax7 summits. Data are shown for the reference constitutively active enhancers that exhibit DNA hypomethylation at their center and for the fully activated Pioneer sites that are highly methylated (> 80%) before Pax7 action but undergo loss of CpG methylation in AtT20-Pax7 cells.

b) Genome Browser representation of Pioneer Activated and Constitutively Active enhancer loci showing the extent of methylation at individual CpGs together with ChIPseq profiles for Pax7 and H3K4me1 before/after Pax7.

c) Specific Pax7 binding sites (underlined) present at loci depicted in b showing CpG dinucleotides in red. Histograms show %CpG methylation for each C residue of the Pax7 binding sites.

d) Pax7 binding assessed by ChIPseq at the *Pcsk2* locus in untreated and AtT-20 cells treated with Tamoxifen compared to cells treated for more than 20 days followed by 18 days of Pax7 withdrawal.

e) Heatmaps of ATACseq signals at pioneered and constitutively active enhancers in Tam-treated and withdrawn ER-Pax7 cells. ER-Pax7 cells were grown in presence of Tam for over 20 days and following removal of Tam for 3 or18 days.

f) Four representative Pax7-pioneered loci showing LIF-induced (20min.) STAT3 binding 18 days after removal of Pax7 (Tam). The fourth locus shows marginal non-significant binding after removal.

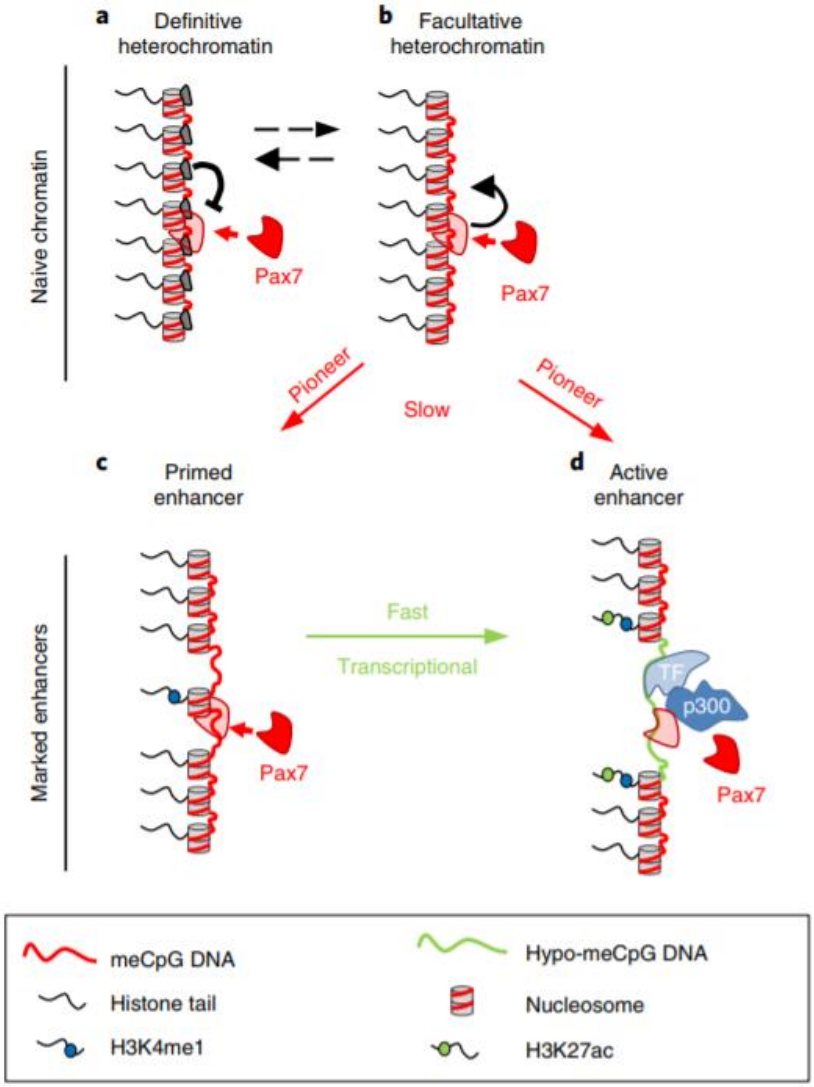


Figure 7. Pioneer action and chromatin states defined by Pax7-dependent transitions.

Figure 7. Pioneer action and chromatin states defined by Pax7-dependent transitions.

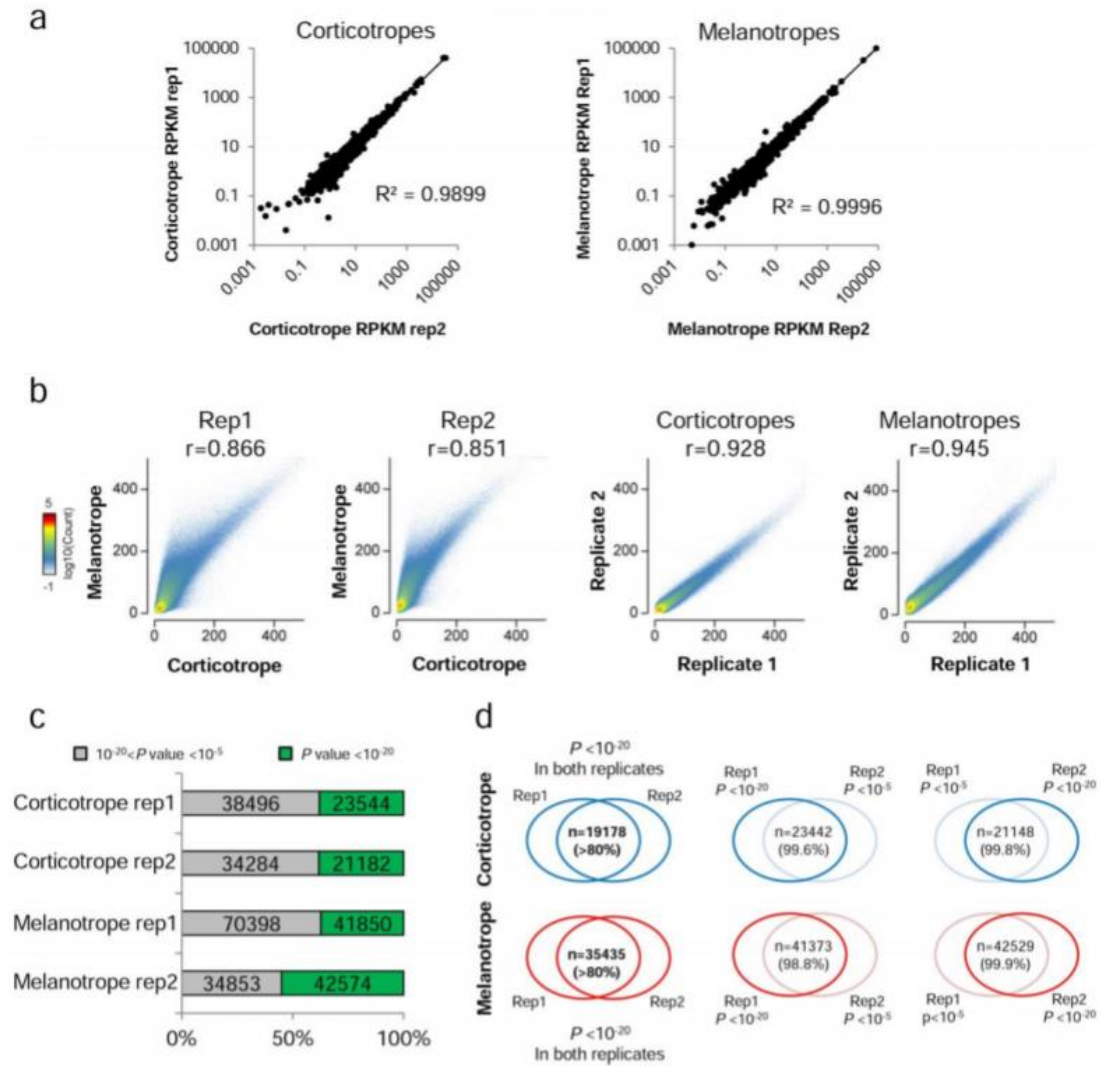
a) Definitive Heterochromatin contains Pax7 binding sites of low affinity that are resistant to pioneering.

b) Pioneer competent sites are found within Facultative Heterochromatin. These sites bind Pax7 with high apparent (ChIPseq) affinity and undergo slow (2-5 days) chromatin opening.

c) Primed enhancers are marked by low levels of centrally-located H3K4me1 and weak DNA accessibility.

d) Active enhancers have accessible DNA (ATAC, DHS or FAIRE), nucleosome depletion with flanking H3K4me1 and H3K27ac. They are occupied by p300 and their DNA is hypomethylated.

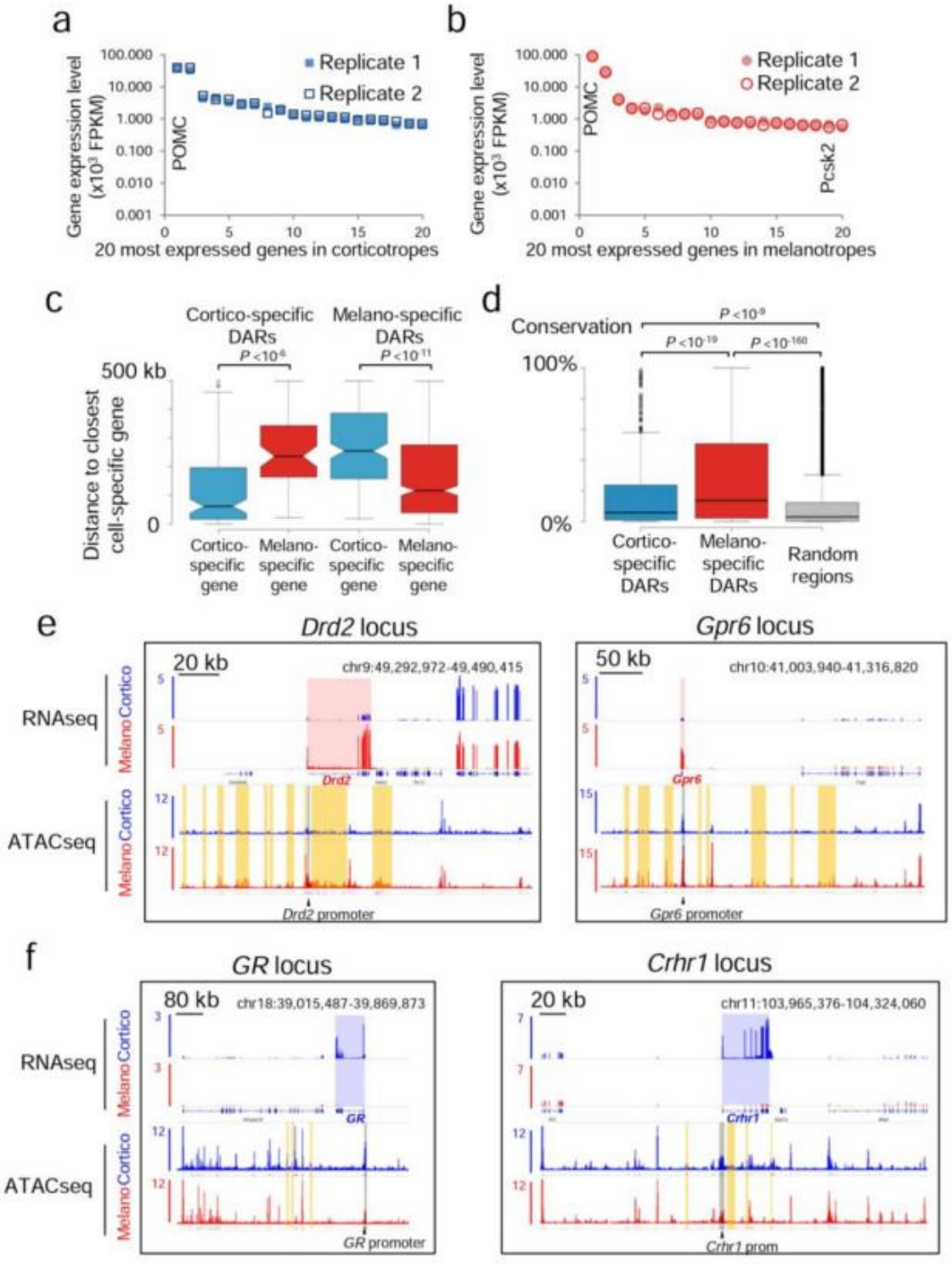
Supplementary figures



Supplementary figure 1. High degree of correlation between RNA-seq and ATAC–seq replicates of FACS-enriched pituitary cells.

Supplementary Figure 1. High degree of correlation between RNA-seq and ATAC–seq replicates of FACS-enriched pituitary cells.

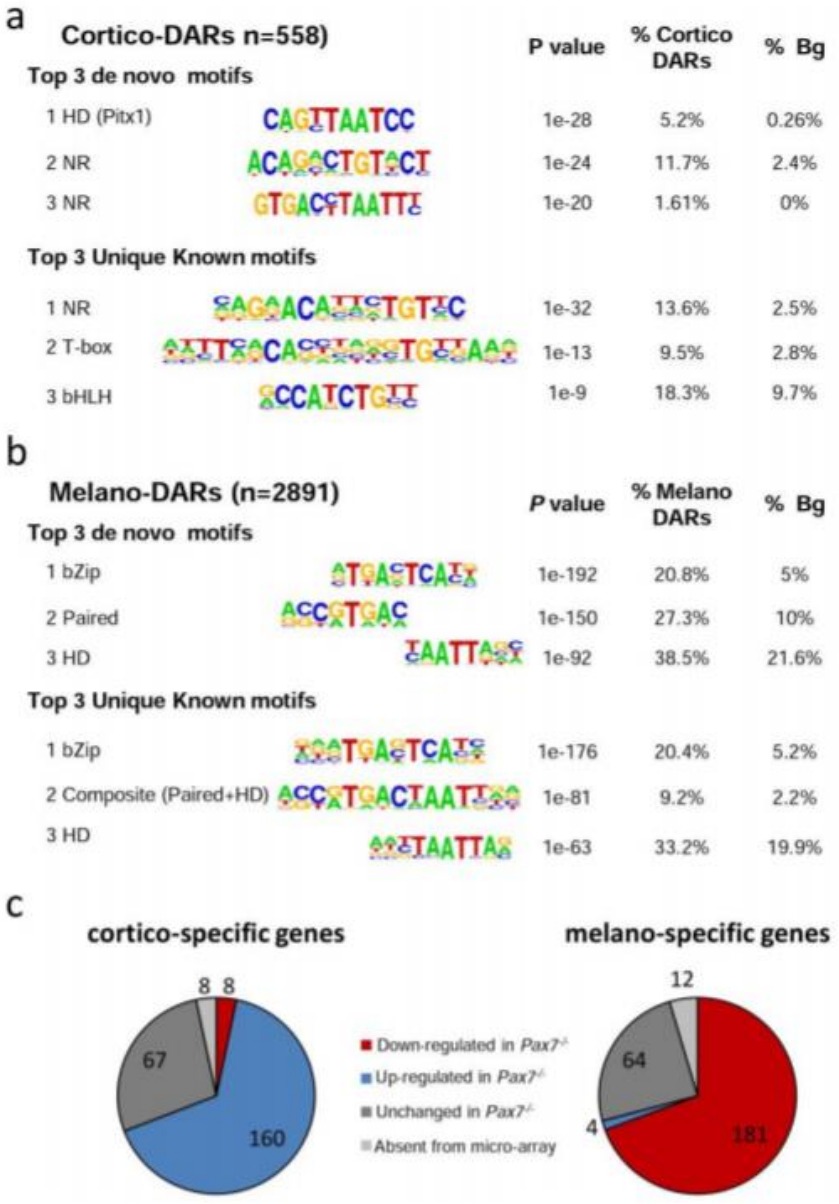
(a) Dispersion plots showing correlation between RNA-seq values for all expressed genes (FPKM >1 in at least one sample) of two independent primary cell replicates of pituitary corticotropes and melanotropes. (b) Density plots showing correlation between two independent primary cell replicates of ATAC-seq's performed in corticotropes and melanotropes. Only 200-bp bins with more than 50 reads in one sample were used for calculating the Pearson coefficient. Bins with more than 200 reads are shown on the plot as 201 reads. (c) Bar graphs showing the proportion of ATAC-seq peaks with P values (derived from MACS) between 10-5 and 10-20 (grey) and less than 10-20 (green). (d) Cross comparisons of ATAC-seq peaks in corticotropes (blue) and melanotropes (red) between the different replicates with the indicated P-value (derived from MACS) thresholds.



Supplementary figure 2. Lineage-specific transcriptional program controlled by cell-specific DARs.

Supplementary Figure 2. Lineage-specific transcriptional program controlled by cell-specific DARs.

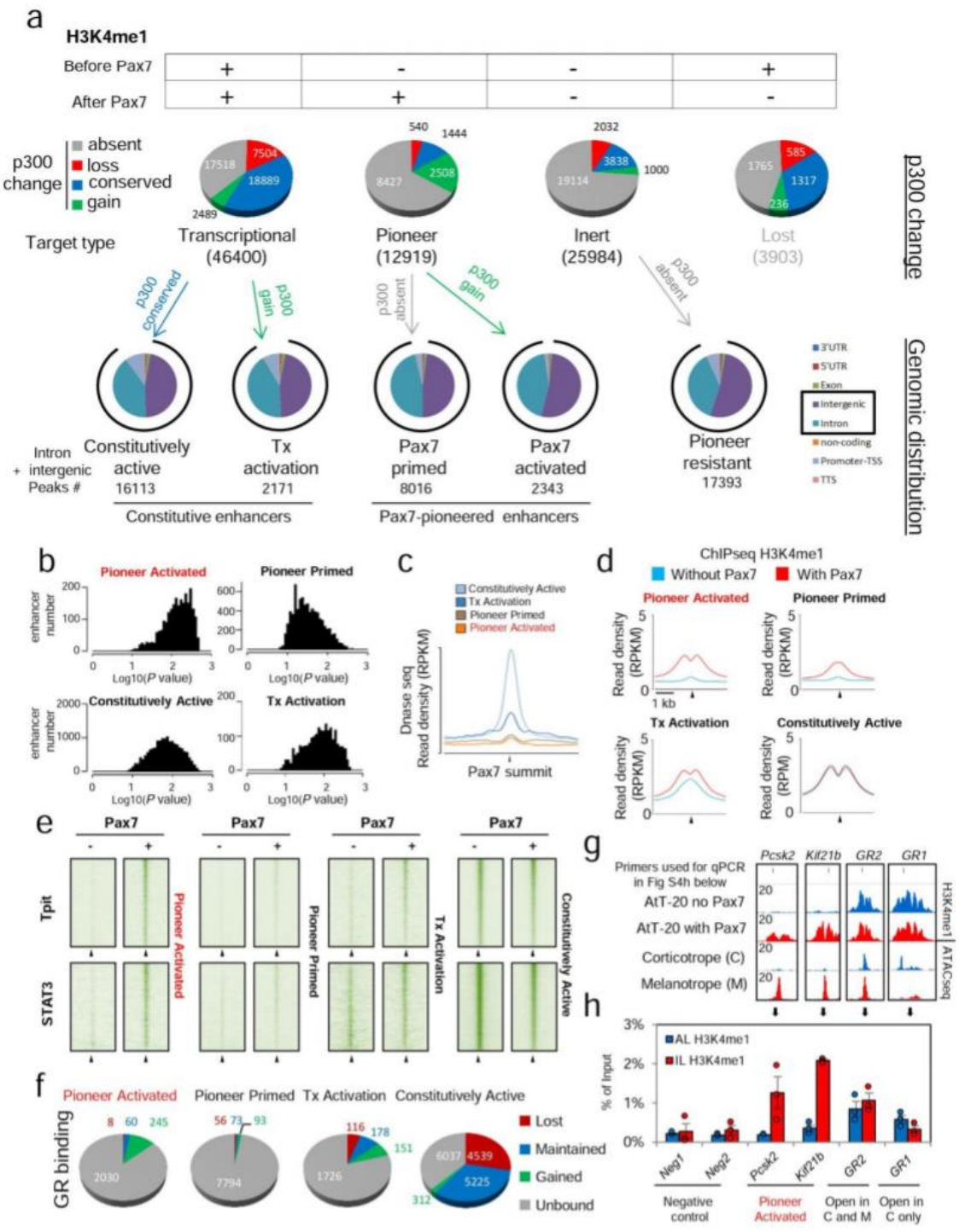
(a,b) Expression levels of the 20 most highly expressed genes in corticotropes (a) and melanotropes (b). (c) Boxplot showing the distances between cortico- (n = 558) or melano- (n = 2,891) specific DARs and the closest differentially expressed genes. Center lines show medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th to 75th percentiles; outliers are represented by dots. P values are assessed by unpaired two-sided t test. (d) Boxplot showing the level of conservation (PhastCons) of lineage specific DARs compared to random control regions. Center lines show medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th to 75th percentiles; outliers are represented by dots. P values assessed by unpaired two-sided t test. (e,f) Genome browser views of lineage-specific genes for two melanotrope (e) and two corticotrope markers (f).



Supplementary figure 3. DNA-sequence-motif searches at DARs identify Pax7 as a cell-fate regulator.

Supplementary Figure 3. DNA-sequence-motif searches at DARs identify Pax7 as a cell-fate regulator.

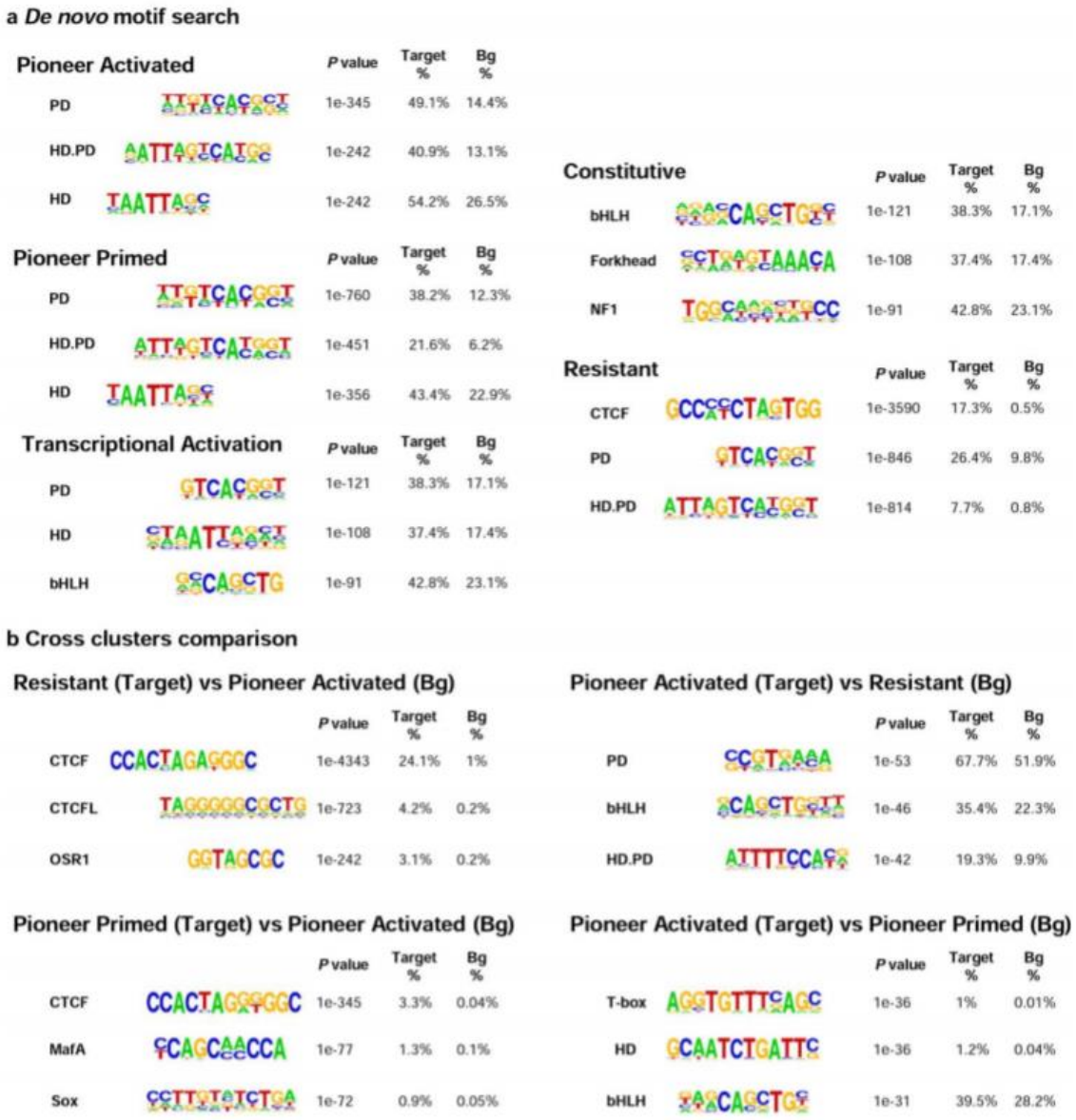
(a,b) Motif enrichments obtained using HOMER on corticotrope (a) and melanotrope (b) DARs within a 200-bp window around DARs summit. The top three de novo and known motifs are shown. The bZIP motif identified in the melano-specific DARs (b) appears unique to this subset. However, this bZIP motif is not unique to melanotrope DARs since it is also a major motif extracted from DARs shared between the two POMC lineages and it is also found in gonadotrope DARs (not shown). (c) Distribution of changes ($P<0.05$) in cortico- and melano-specific gene expression in $Pax7^{-/-}$ neuro-intermediate lobe.



Supplementary figure 4. Analysis of subsets of Pax7 genomic targets.

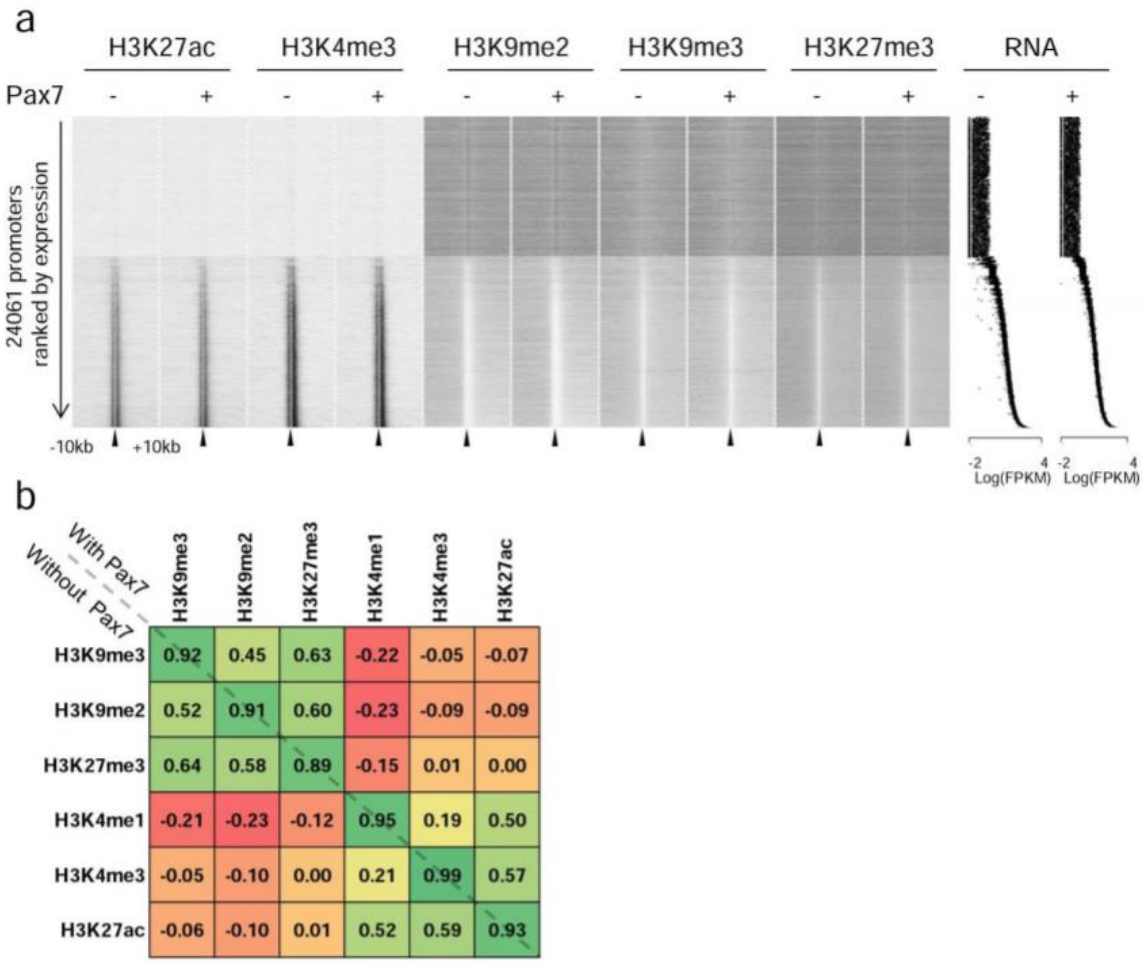
Supplementary Figure 4. Analysis of subsets of Pax7 genomic targets.

(a) Scheme for clustering subsets of Pax7 binding sites identified ($P < 10^{-5}$, derived using MACS) by ChIP-seq (n = 89,206). Pax7 peaks were associated with the presence of H3K4me1 and p300 peaks ($P < 10^{-5}$) before or after Pax7 expression (summit of H3K4me1 peak +/- 1 kb from Pax7 summit, +/- 500 bp for p300 peaks). The subsets with H3K4me1 before and after Pax7 were defined as Constitutively Active (p300 present before and after Pax7) and Pax7 Activated putative enhancers (p300 present only after Pax7). The subset that gained H3K4me1 after Pax7 binding was deemed to include putative Pioneer sites, and was further subdivided into Primed or Activated pioneer subsets depending on the accompanying gain of p300. Finally, pioneer targets being mostly at intergenic and intronic regions, we extracted for all subsets the intergenic and intronic peaks to define putative enhancers for further analyses. The number of peaks in each subset is indicated. (b) P value (derived from MACS) distribution of the four subsets (n indicated above in a) of Pax7 targets used for analyses. (c) Average DNAse hypersensitivity (GEO SRX034837) profiles for four subsets of peaks defined in a. (d) Average H3K4me1 profiles before (blue) or after (red) Pax7 at the four subsets described in Figure 3. (e) Heatmaps of Tpit and Stat3 ChIP-seq signals at the four subsets of Pax7 peaks described in Figure 3. (f) GR binding changes, measured by ChIP-seq ($P < 10^{-5}$, derived from MACS), before or after Pax7 action at the indicated subsets (n indicated above in a) of Pax7 peaks. (g) Genome browser views of the four loci used for qPCR validation in h. H3K4me1 ChIP-seq data are shown in AtT-20 cells with/without Pax7 and replicate 1 of ATAC-seq data from corticotropes (C) and melanotropes (M). (h) ChIP-qPCR for H3K4me1 at two Pax7 pioneer targets in the Kif21b and Pcsk2 loci, one putative enhancer of the GR locus open in both lineages (C and M), and another GR locus putative enhancer specifically opened in corticotropes (C) as indicated. The data shown are means +/- s.e.m. of tissue triplicates assessed by duplicate qPCR measurements. The negative control sites (Neg1 and Neg2) do not show any H3K4me1 enrichment.



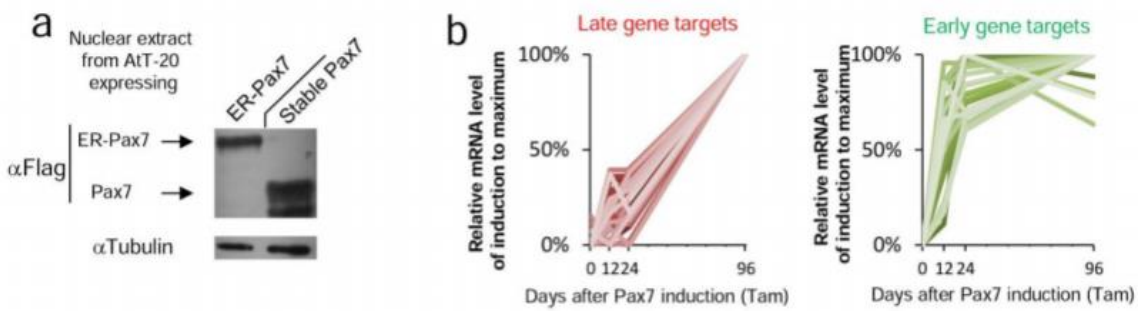
Supplementary figure 5. DNA sequence motifs identified in Pax7 subsets.

(a) De novo motif enrichments identified using HOMER at the indicated subsets (n indicated in Supplementary Fig. 4a) of Pax7 targets. The top three identified motifs are shown. (b) De novo motif enrichments identified using HOMER at the indicated subsets of Pax7 targets using the indicated subsets as background to identify sequences associated with specific Pax7 subsets. The top three identified motifs are shown.



Supplementary figure 6. Validation of chromatin-mark data produced in this study.

(a) Heatmaps of indicated ChIP-seq data around 24,061 RefSeq TSS in AtT-20 cells before or after Pax7 ranked according to expression levels derived from RNA-seq data shown at right. (b) Genome-wide Pearson correlations of ChIP-seq signals of all chromatin marks used in this study. Genomic windows of 500 bp with ≥ 20 reads were used to calculate correlations; windows with more than 500 reads were downscaled to 501 reads. The repressive histone marks (H3K9me3, H3K9me2 and H3K27me3) cluster together, while the active mark H3K27ac correlates with both H3K4me1 and H3K4me3. In all cases, the strongest correlation is obtained when comparing each dataset with or without Pax7 (>0.89).



Supplementary figure 7. Characterization of inducible ER-Pax7 cells.

(a) Western blot showing Pax7 protein expression in Tam-induced ER-Pax7 AtT-20 cells compared to stable Pax7-expressing AtT-20 cells. A representative of two independent experiments is shown. (b) Gene induction curves of ER-Pax7 gene targets assessed by RNA-seq in cells treated with Tam for 12 h, 24 h and 96 h in comparison with non-treated cells. Gene targets were separated into two induction dynamics matching those observed in Figure 5e. 100% corresponds to the maximum level of gene induction for each gene, and 0% representing the lowest level of the four time points measured. All target genes were induced at least 2-fold in stable Pax7-expressing and in ER-Pax7-expressing cells. Late targets show lower than 40% gene induction at 24 h, while early targets show more than 80% induction at 24 h.

Chapter III: Pioneer and nonpioneer cooperation drives lineage specific chromatin opening

Alexandre Mayran[1,2], Kevin Sochodolsky[1], Konstantin Khetchoumian[1], Juliette Harris[1], Yves Gauthier[1], Amandine Bemmo[1] Aurelio Balsalobre[1] and Jacques Drouin[1,2,3,4]

[1] Laboratoire de génétique moléculaire, Institut de Recherches Cliniques de Montréal (IRCM), Montreal, Québec, Canada

[2] Department of Biochemistry, McGill University, Montreal, Québec, Canada

[3] Département de Biochimie, Université de Montréal, Montreal, Québec, Canada

[4]Correspondence: jacques.drouin@ircm.qc.ca

Author contributions

A.M. and J.D. conceived the study, A.M., Au.B. and J.D. conceived and designed the experiments, A.M., K.S. and K.K. performed mouse tissue collection, A.M. performed ATACseq assays, A.M. and Y.G. performed ChIP experiments, K.S. performed immunofluorescence assays, J.H. did the pull-down assays, A.M. and Am.B. analyzed the single cell RNAseq, A.M. performed analysis of ATACseq and ChIPseq, A.M. and J.D. wrote the manuscript.

Abstract

Pioneer transcription factors are coined as having the unique property of “opening closed chromatin sites” for implementation of cell fates. We previously showed that Pax7 pioneers pituitary melanotrope cell identity through deployment of an enhancer repertoire: this allows binding of Tpit, a nonpioneer factor that drives terminal differentiation of melanotropes and the related corticotropes. Here, we used scRNAseq and chromatin accessibility profiling to define shared and cell-specific gene expression programs and chromatin landscapes in these lineages. *Pax7*- and *Tpit*-deficient pituitaries fail to deploy the melanotrope enhancer repertoire indicating that both factors are required for chromatin opening. Finally, in cells, binding of heterochromatin targets by Pax7 is independent of Tpit but Pax7-dependent chromatin opening requires Tpit. The present work shows that pioneer core properties are limited to the ability to recognize heterochromatin targets and facilitate nonpioneer binding. Chromatin opening per se may be provided through cooperation with nonpioneer factors

Introduction

Lineage specification occurs repetitively throughout the development of multicellular organisms. It leads totipotent stem cell of the early embryo to differentiate toward numerous lineages of complex organisms. Multiple specification events accumulate and incrementally decrease cell fate potency. As more decisions accumulate, cells change their program of gene expression leading to more specialized phenotypical characteristics. These decisions are typically driven by master transcription factors (TFs) that are essential to implement transcription regulatory networks. Several of these master regulators act as pioneer factors, a unique class of TFs that implement cell specific program through remodeling of the chromatin landscape. Unlike canonical TFs, they can bind “closed” heterochromatin and facilitate the binding of other TFs. For example, Foxa1 is essential for liver fate and binds regulatory sequences before gene activation (Cirillo et al., 2002; Lee et al., 2005). Ebf1 also acts as a pioneer factor during B-cell development and Neurod1 and Ascl1 (Mash1) were suggested to function as pioneer factors during neural development (Pataskar et al., 2016; Wapinski et al., 2017). Ectopic expression of pioneer factors is sufficient to drive trans-differentiation, the most extreme example being the reprogramming of fibroblasts into induced pluripotent stem cells through expression of the pioneer factors Oct4, Klf4 and Sox2 (Okita et al., 2007; Soufi et al., 2015). Similarly, C/EBPα can direct trans-differentiation of pre-B cells into macrophages (Di Tullio et al., 2011). Recently, Grainyhead (Grh) was shown to prime epithelial enhancers (Jacobs et al., 2018) by displacing nucleosomes. Whether all TFs that are critical for cell differentiation also function as pioneer factors is not really known. Further, the pioneer model implies that chromatin opening is the direct consequence of pioneer factor action and this allows nonpioneer binding to newly accessible sites. However, it has not really been assessed whether nonpioneers can play a role in pioneer-driven chromatin opening. Here, we used normal and perturbed pituitary differentiation to investigate the pioneer model and establish the specific and/or overlapping functions of the master regulators Pax7 and Tpit. The two pituitary lineages that express the hormone precursor pro-opiomelanocortin (POMC), the

melanotropes and corticotropes, both require Tpit for POMC expression and to prevent differentiation towards the gonadotrope fate (Lamolet et al., 2001; Pulichino et al., 2003). Further, both POMC lineages require Tpit to implement a secretory cell transcriptional program by activation of scaling factors for translation and secretory organellogenesis (Khetchoumian et al., 2018). While sharing the secretory POMC identity, corticotropes and melanotropes differ by their functions: indeed, these two POMC-expressing lineages control corticosteroidogenesis and pigmentation, respectively (Drouin, 2017). Melanotropes specification depends on the pioneer action of Pax7 (Budry et al., 2012) through deployment of a melanotrope enhancer repertoire (Mayran et al., 2018).

Here, we first establish that the two POMC lineages share a transcriptional program that is distinct from other pituitary cells and that in addition, they each have a unique program of gene expression. We then show that these two layers of identity (shared and lineage-specific) are reflected at the level of chromatin accessibility and that the shared POMC chromatin landscape requires Tpit. Further, the Pax7-dependent melanotrope chromatin landscape also requires Tpit suggesting that the two factors (Pax7 and Tpit) act together during melanotrope lineage specification. Finally, Pax7 and Tpit have different roles as only Pax7 has the ability for heterochromatin binding while Pax7-dependent Tpit binding is associated with chromatin opening. In summary, we propose that the essence of pioneer action may be in the ability to recognize and bind DNA sites in closed chromatin whereas cooperating nonpioneer TFs , such as Tpit, may drive chromatin opening.

Results

Diversity of pituitary cell lineages

The pituitary is a highly specialized organ where each lineage serves as a hormone producing factory. Each cell type is dedicated to the regulation of a specific endocrine organ and responds to specific signals from the hypothalamus and body. We used single cell RNAseq to decipher the transcriptional complexity of the

different pituitary lineages. For each lineage, hormone-coding mRNAs are so abundant that they appear as peaks in cDNA libraries (Figure S1A). Profiling of adult mouse male pituitary cells (Figure S1B) was achieved by plotting single cell data using t-distributed stochastic neighbor embedding (t-SNE), a common method (Maaten, 2008) that uses dimensionality reduction to cluster together cells with similar transcriptional profiles (Figure 1A). This showed that cells expressing the same pituitary hormone cluster together and that they all express Pitx1, a marker of the oral ectoderm origin of the pituitary (Drouin, 2017) (Figure S1C-G and 1B). We identified 12 clusters composed of endocrine and non-endocrine cells (Figure 1C). Cluster 1 corresponds to somatotropes as they express the growth hormone (Gh) gene. Lactotropes that produce Prolactin (Prl) are found in cluster 2. Clusters 4 and 5 correspond to melanotropes and corticotropes, respectively; both express the POMC gene, yet only melanotropes express Pcsk2. Gonadotrope that express the Lhβ gene are in cluster 8. We also detected thyrotropes as Tshβ-expressing cells; however, they did not appear as a separate cluster (Figure S1C). The pituitary stem cells that express Sox2 (Fauquier et al., 2008) are found in cluster 7. Although our study aimed at defining the transcriptome of the different pituitary lineages, we also uncovered several non-endocrine cells within the pituitary tissue that do not express Pitx1 (Figure 1B). We identified endothelial cells (cluster 9), macrophages (cluster 10), posterior pituicytes (cluster 11) and pericytes (cluster 12). Cluster 3 is fragmented in three different groups of cells that express either GH, Prolactin, POMC or Lhβ. We performed differential expression analysis between each sub-cluster 3 and its matching cell type in order to define these subsets. In all cases, cells of these subsets are specifically depleted of ribosomal proteins and enriched for mitochondrial RNA. This was shown to be an artefact of tissue dissociation (Ilicic et al., 2016) and to represent cells affected by the preparation: we excluded cluster 3 from following analyses.

We then compared the transcriptomes of the different pituitary lineages by performing differential expression analysis between clusters 1, 2, 4, 5, 7 and 8. Genes with two-fold differential expression between each cluster (p value <0.05 and minimum 0.2 UMI in at least one pooled cluster) are shown as a heatmap

(Figure 1D). The Sox2+ stem cell niche is the most transcriptionally divergent from other pituitary lineages based on correlation clustering analyses. The two Pit1-dependent lineages, the lactotropes and somatotropes, are the most different compared to the two POMC lineages and gonadotropes that cluster together. Within the latter group, corticotropes and melanotropes are more transcriptionally correlated together than with gonadotropes. Thus, the two POMC lineages, melanotropes and corticotropes, have both a shared and a specific transcriptional program.

Lineage-specific chromatin landscapes

We next aimed to identify cis-regulatory elements that regulate the transcriptional identity of pituitary lineages. We used ATACseq (Buenrostro et al., 2013) to identify putative regulatory elements that are accessible in each lineage. We complemented our previously published datasets of purified melanotrope and corticotrope ATACseq with accessibility profiles for gonadotropes and anterior lobe (AL) cells. Gonadotropes were FACS-purified from transgenic pituitaries expressing the LHβ-Cerulean transgene (Budry et al., 2011). As control, we isolated the remaining AL cells that are mostly composed of a combination of Pit1-dependent somatotropes and lactotropes. Interestingly, we found that the promoters of hormone genes *POMC*, *αGSU*, *Gh* and *Prl* as well as lineage specifiers *Tpit* (*Tbx19*), *Pax7*, *SF-1* (*Nr5a1*) and *Pit1* show lineage-specific accessibility (Figure 2A and S2A). However, the *Pcsk2* promoter is accessible in all pituitary lineages but its numerous distal accessible sites (putative enhancers) are only accessible in melanotropes. Globally, we identified 98926 open chromatin regions across the pituitary lineages (Figure 2B). Segregation of lineage-specific accessibility yielded 33451 regions opened in all lineages, 14025 regions opened in a combination of three lineages, 20374 in two lineages and finally 31076 opened in only one lineage. Thus, there are regions specifically accessible in melanotropes, corticotropes, gonadotropes or in the AL. In accordance with the close transcriptional correlation between melanotropes and corticotropes, we also found shared regions accessible in both melanotropes and corticotropes (POMC-

specific) but closed in gonadotropes and AL. Finally, we identified 13130 pituitary-specific sites that are closed in embryonic stem cells, as well as a set of 20321 regions accessible in both pituitary and ES cells (Figure S2B). These ubiquitous peaks are for the majority (58%) composed of promoter elements (Figure S2C) while regions opened in two or in only one pituitary lineage are mostly distal elements (94% and 96%, respectively). This reinforces the idea that promoter accessibility is established early during differentiation. Further, this suggests that lineage specific opening of promoters tends to be an exception and may be involved in restricting appropriate expression of critical genes such as hormone-coding genes and lineage specifiers.

Comparison of the spatial relationship between the POMC-specific transcriptional program (Figure 2C) and lineage-specific accessibility (Figure 2D) shows that POMC-specific gene promoters tend to be closer to both POMC- and melano-specific open regions (Figure 2E) and are enriched for the motif of the lineage specifier Tpit (Figure 2F). Melanotrope genes are also close to their lineage-specific open chromatin regions and their lineage specific chromatin landscape is enriched for the motif of their lineage specifier Pax7 (Figure 2G-J). Similar relationships were found for the Pit1-dependent lineages, the corticotropes and the gonadotropes (Figure S2D-H).

Pax7 is required for phenotypical features of melanotropes

The two POMC lineages have shared and specific transcriptional programs and open chromatin landscapes. Their most obvious similarity is expression of the hormone precursor POMC as well as expression of the terminal differentiation factor Tpit. However, POMC expression varies between these two lineages with low (blue) and high expression (red) cells in scRNAseq analyses (Figure 3A). Indeed, POMC is more highly expressed in Pax7-expressing melanotropes compared to GR (nr3c1)-expressing corticotropes as shown in t-SNE and differential expression Volcano plots (Figure 3A,B). Analysis of the POMC-EGFP (Lavoie et al., 2008) transgenic pituitaries (Figure 3C) also showed high and low-expressing cells. Indeed, IL (melanotropes) transgenic pituitary cells have much

higher EGFP levels compared to most EGFP-positive corticotrope cells of the AL (Figure 3D). This indicates that the transgene expressed in both melanotropes and corticotropes is subject to cell-type specific regulation and can be used as a surrogate phenotypical readout. In a *Pax7-/-* background, melanotropes express the EGFP transgene at the same level as in AL corticotropes (Figure 3D, E). Melanotropes are also typically larger and have a more complex organelle content than corticotropes as assessed by forward and side scatter distributions in FACS profiles and this is also lost in Pax7-deficient mice (Figure S3A, B). Thus, all three melanotrope features (high POMC expression, large cell size and granularity) are fully dependent on Pax7 and virtually all melanotropes switch to a corticotrope phenotype in *Pax7-/-* mice (Figure3F). However, the proportion of total EGFP-positive cells in AL or IL is not affected by loss of Pax7 (Figue 3G). Thus, Pax7 implements melanotrope but not the shared POMC cell identity which is under the control of Tpit (Pulichino et al., 2003).

Pax7 and Tpit are required for opening cognate lineage-specific enhancer landscape

We next sought to uncouple the specific roles of Pax7 and Tpit for implementation of the shared and melanotrope-specific chromatin landscapes. To do so, we performed ATACseq on IL cells from mice of genetic backgrounds lacking one or both alleles of *Pax7* and/or *Tpit*. We used *Pax7* and *Tpit* double heterozygote littermates as control; these show minimal differences compared to wild-type animals (Fig S4). We found that melanotrope-specific peaks (Figure 2B) are not accessible in *Pax7-/-;Tpit*+/- IL whereas they are present in wild-type or double heterozygotes (Fig 4A) in agreement with previous data (Mayran et al., 2018) that showed Pax7 requirement for accessibility of melanotrope regulatory modules. Also, we found that Tpit is required for the open status of POMC-specific open chromatin (Figure 4B). This suggests that Tpit is involved in the pioneering process of the shared POMC lineage enhancers. For example, POMC gene expression critically relies on Tpit and accordingly, both its promoter and enhancer accessibility strongly depend on Tpit but not on Pax7 (Figure 4C). This is also true for the promoter of POMC-specific gene Tnxb and the Tpit-dependent enhancer

(Khetchoumian et al., 2018) of the Creb3l2 gene (Figure 4C). Thus, Tpit is required for opening the shared POMC regulatory modules while Pax7 is only required for melanotrope enhancers.

Tpit is required for Pax7-dependent chromatin opening

In order to define and compare the Pax7 and Tpit-dependent ATACseq landscapes, we performed differential enrichment analysis ($p<0.05$ and fold changes >2) and found 16024 sites altered in *Tpit-/-* IL (Figure 5A). Most changes are decreased accessibility (12573 sites) and these show greater quantitative differences compared to sites with increased accessibility (3451 sites). Analysis of these Tpit-dependent sites revealed two subsets, one dependent and one independent of Pax7 (Figure 5B). In both Tpit-dependent subsets, chromatin accessibility in *Pax7-/-;Tpit-/-* IL is the same as in *Tpit-/-*. In order to ascertain the reliability of these analyses, we performed qPCR analyses of ATAC samples at index loci for dependence on Pax7, Tpit or both (Figure S5). Principal component analysis using all Tpit-regulated sites shows that most of the variance between samples (77%) is explained by component 1 (Fig 5C). Consistent with the heatmaps of Figure 5B, wild-type and *Pax7+/-;Tpit+/-* cluster together while *Pax7* knockout are between wild-type and *Tpit* knockout samples. Thus, a subset of Tpit chromatin targets are also dependent on Pax7. We then focused on Pax7-dependent chromatin access to assess whether co-dependency on Tpit and Pax7 is a general feature of Pax7-dependent sites. We found 7057 Pax7-dependent sites (Fig 5D): most changes are decreased accessibility (6112 sites) and these are greater effects compared to increased sites (945 sites). Strikingly, and unlike Tpit-dependent accessibility, virtually all Pax7-dependent sites are also dependent on Tpit (Figure 5E). Accordingly, principal component analysis of Pax7-dependent sites showed that Pax7 knockout samples cluster with *Tpit* knockout and Pax7 on component 1 (79% of variance). This indicates that *in vivo,* Tpit is absolutely required for the establishment of the Pax7-dependent chromatin landscape. Thus in both cases, *Pax7* and *Tpit* double knockout samples cluster closely with the *Tpit*

single knockout (Fig 5C, F) showing that loss of both factors does not have a greater impact on accessibility than the loss of Tpit alone.

To assess the biological basis of this co-dependency, we focused on the *Pcsk2* gene, encoding the PC2 protein, a hallmark of melanotrope identity. We previously showed strong Pax7 dependency for chromatin opening of an upstream enhancer (Budry et al., 2012) together with multiple distal elements that are ATACseq sensitive in a lineage-specific manner. Interestingly, this melanotrope-specific chromatin accessibility is confined within a topologically associated domain (TAD) and stops at the border of this TAD (Figure 5G). In accordance with the role of Pax7 for *Pcsk2* expression in melanotropes, we found that this TAD-wide chromatin opening doesn't occur in *Pax7-/-* animals. Consistent with our previous observation that Pax7-dependent access also depends on Tpit, chromatin opening within this TAD does not take place in *Tpit-/-* and in *Pax7-/-;Tpit-/-* animals. Thus, Tpit is required for opening of the *Pcsk2* TAD. It is noteworthy that *Pcsk2* is also expressed, albeit at low levels, in gonadotropes (Figure 5H). This gonadotrope pattern of *Pcsk2* expression is also found in the *Tpit-/-* IL cells that have switched fate (Budry et al., 2012); these cells occupy the dorsal side of the *Tpit-/-* mutant IL (Figure 5I). In contrast, the ventral side of the same *Tpit-/-* IL harbors Pax7-positive cells that fail to express Pcsk2 (Figure 5I). This confirms that Pax7 functionally requires Tpit to establish melanotrope identity.

Productive Pax7 pioneer action in Tpit-positive cells

In order to evaluate the chromatin binding properties of Pax7 and Tpit, we compared their binding properties to those of a pituitary nonpioneer factor Pitx1 and to two pioneer factors, Neurod1 that is expressed in pituitary corticotropes (Lamolet et al., 2004; Poulin et al., 1997), and the reprogramming factor Sox2 (Okita et al., 2007). The heatmaps of ATACseq signals for all binding sites of these five TFs in AtT20 cells reveal striking differences between the profiles of Tpit and Pitx1, compared to those of the pioneers, Neurod1, Pax7 and Sox2 (Figure 6A). Whereas Tpit and Pitx1 mostly (~ 95% of all sites) bind to sites that are open (ie. have a detectable ATACseq signal), a significant proportion of the pioneer binding

sites have little or no ATACseq; thus for Pax7 and Neurod1, ~ 30 % of binding sites are in inaccessible chromatin whereas this is true for ~ 60% of all Sox2 binding sites. The ability to bind closed chromatin (ATAC-negative) appears to be a common property of the three pioneer factors in contrast to the non-pioneers Tpit and Pitx1.

In order to reveal the contribution of Tpit to chromatin pioneering by Pax7, we compared the impact of Pax7 expression in two different pituitary cell lines, one that expresses Tpit (AtT20) and one that does not (αT3). In agreement with expression data, the Tpit locus exhibits ATACseq signals in AtT20 cells but not in αT3 cells (Figure 6B) whereas the locus for the gonadotrope-specific regulator SF1 presents ATACseq signals in αT3 but not in AtT20 cells (Figure 6C).

We looked at the chromatin accessibility by ATACseq before/after Pax7 expression in both lineages and found that in AtT-20 cells, Pax7 binding leads to opening of 6607 regions. In contrast, only 661 sites are opened following Pax7 expression in αT3 cells and this opening is of greater magnitude in AtT-20 cells (Figure 6D). Thus, expression of Pax7 in αT3 cells does not lead to efficient chromatin opening. We have previously identified steps of pioneering as follow: initial weak binding to closed enhancers (less than 30 minutes), stabilization of pioneer binding followed by chromatin opening (Mayran et al., 2018). Failure to perform any of these three steps will impede Pax7 pioneering. To identify which specific step limits Pax7 pioneering in αT3 cells, we compared Pax7 binding in AtT-20 and αT3 cells (Figure S6A). Pax7 has similar ability to bind closed heterochromatin in both AtT-20 and αT3 cells (Figure 6A and S6A,B). We extracted all Pax7 sites bound to closed chromatin in AtT-20 and in αT3 (Fig 6e). In both AtT-20 and αT3, there are sites with strong and low binding signals. This suggest that Pax7 is able to bind strongly to closed chromatin in both cell context. However in AtT-20 cells, Pax7 binding strength correlates (r^2=0.41) with accessibility after Pax7 expression whereas in αT3 cells, there is no correlation (r^2=0.09) between Pax7 binding and post-Pax7 chromatin opening (Fig 6E). We identified a subset of heterochromatin Pax7-bound sites based on their ability for Tpit binding after Pax7

expression; these sites are not bound by Tpit in absence of Pax7 (Figure 6F) and become accessible in presence of Tpit and Pax7 (Figure 6G). It is noteworthy that only these newly accessible sites contain the Tpit DNA binding motif (Figure 6H). This suggests that Pax7-dependent Tpit binding also depends on the Tpit DNA motif. Notwithstanding, Pax7 and Tpit interact directly in vitro as shown in pull-down assays (Figure 6I). The interaction between the two factors may also contribute to their cooperation. This suggests that Pax7`s ability to bind strongly to closed chromatin is not dependent on cell specific factors and that stable binding is not sufficient to drive chromatin opening. During melanotrope differentiation, this pioneer-dependent chromatin opening also requires Tpit.

Discussion

Pioneer factors are coined as “factors that can open closed chromatin”. This label implied that pioneers were expected to directly provide this ability. The present work shows that there can be division of labor between pioneer and nonpioneer factors: in the case of Pax7 and Tpit, the former recognizes and engages pioneering sites and the latter provides the chromatin opening ability. Indeed, we show that Tpit is required for chromatin opening at sites that must first be pioneered by Pax7, and we did not find evidence for the reverse. While we envision that this is made possible by Tpit-dependent recruitment of chromatin remodeling machineries, the finding clearly circumscribes the unique features of the Pax7 pioneer function. Namely, the unique aspect of pioneer action is the ability to bind DNA sites within heterochromatin that are inaccessible to probing by techniques such as ATACseq; this appears to be a property shared with other pioneers such as Sox2, NeuroD1 (Figure 6A) in contrast to nonpioneer factors such as Tpit or Pitx1.

For Pax7, we have shown that this initial binding to closed heterochromatin (Figure 7A) is rapid (within 30 min, Figure 7B) and that the first detectable change in chromatin structure at these binding sites is revealed by stabilization (within 24 h, Figure 7C) of Pax7 binding (Mayran et al., 2018). This stabilization precedes chromatin opening (revealed by ATACseq). Here, we find that strong Pax7 binding

occurs even in absence of Tpit (ie in αT3 cells, Figure 6E): this excludes Tpit`s involvement in Pax7 stabilisation of binding. In contrast, the next step of pioneer action, namely chromatin opening, requires Tpit as it does not occur in αT3 cells. The recruitment of Tpit at a subset of Pax7 heterochromatin binding sites (Figure 7D) that have a Tpit DNA binding site in the vicinity (within a few hundred bp) results in chromatin opening (Figure 7E). It would thus be the combined interaction of Pax7 with Tpit together with the latter's ability to bind its DNA site exposed through the initial action of Pax7 that altogether would lead to completion of pioneer action with Tpit bringing in the chromatin remodeling ability (Figure 7E). With these interdependent functions, the cooperation between pioneer and nonpioneer factors provides robust stringency and a fail-safe mechanism for triggering chromatin opening at a very specific subset of Pax7 sites.

The requirement for DNA binding sites for both Pax7 and Tpit at the Pax7 pioneered enhancers is reminiscent of the action of other pioneers where two different pioneers cooperate for chromatin opening. For example, the pluripotency factors cooperate with each other to facilitate the binding of other and binding of multiple pioneers is required at some sites to drive chromatin opening (Chronis et al., 2017; Soufi et al., 2015). Also, there is cooperation between FoxA1 and GATA4 for chromatin binding (Donaghey et al., 2018). Foxa1 also facilitates the binding of the estrogen receptor by opening chromatin (Hurtado et al., 2011; Laganiere et al., 2005). Reciprocally, another study showed that steroid receptors can facilitate the binding of Foxa1 (Swinstead et al., 2016). This suggest that in the case of steroid receptors and Foxa1, both can play the role of pioneer at specific subsets of their targets. To our knowledge, the required participation of a nonpioneer for pioneer-driven chromatin opening has not been shown so far.

Prior data indicated that pioneer binding precedes gene activation, for example the binding of FoxA at liver targets precedes their activation (Cirillo et al., 1998). More recently, Grainy head binding at pioneer sites was shown to determine chromatin opening (ATAC signal) but not enhancer activity (Jacobs et al., 2018).

The required cooperation between pioneer and nonpioneer factor for implementation of a specific genetic program ensures specificity of action. Indeed as Pax7 is also involved in cell fate specification in muscle and neural tissues in addition to pituitary, its interdependent action with Tpit in the pituitary ensures that Pax7 access to pituitary-unrelated heterochromatin sites may not result in chromatin opening in absence of Tpit. This essential cooperation provides robustness for lineage specification by preventing mis-activation of inappropriate gene regulatory networks. This limitation to the capacity of pioneer factors would allows specific combinations of pioneers and nonpioneers to activate different regulatory networks and explain the wide variability in targets depending on the same pioneer factor in different contexts.

Acknowledgments

We are grateful to Odile Neyret for NGS analyses, Eric Massicotte for FACS sorting, Dominic Fillon for microscopy, Simone Terouz for histology preparations, Dimitar Dimitrov for mouse management and to Evelyne Joyal for her expert secretarial assistance. Analyses were performed on the server of compute Canada. This work was supported by grants to J.D. from the Canadian Institutes of Health Research.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Methods

Mice, tissues, and cell culture.

Mice genotypes and strains used in this study. Foe single cell RNAseq, a 4 month old male C57Bl/6 mouse pituitary was used; cells were dissociated as described below. For ATACseq, FACS-purified pituitary cells are isolated from 3-5 month old adult LH-cerulean (Budry et al., 2011) or POMC EGFP (Lavoie et al., 2008) C57Bl/6 transgenic pituitaries. ATACseq was performed in duplicates for each genotypes of *Pax7* (Mansouri and Gruss, 1998) and *Tpit* (Pulichino et al., 2003) knockouts. Each replicate used a pool of four dissected intermediate pituitaries from 8-20 days old mice in mixed Balb/c and 129sv backgrounds. FACS analyses of intermediate and anterior pituitaries used 15-20 days-old *Pax7* knockout mice harboring the *POMC-EGFP* transgene in a mixed 129sv and Balb/c background. All animal experimentation was approved by the IRCM Animal Ethics Committee in accordance with Canadian regulations.

Cell lines used in this study. AtT-20 cells (obtained from the late E. Herbert in 1981 and subsequently maintained in our laboratory, with yearly negative mycoplasma tests) and αT3 cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and antibiotics (penicillin/streptomycin).

Generation of stable Neo, Pax7 and Sox2 expressing cells. Expression vectors constructed in the pLNCX2 vector were described previously (Budry et al., 2012). Retroviruses were packed using the EcoPack 2-293 cells (Clontech, Mountain View, CA) and infections were performed as described (Budry et al., 2012). Selection of retrovirus-infected cell populations was achieved with 400 µg/ml Geneticin (Gibco, 11811-031). Resistant colonies were pooled to generate retrovirus-infected populations of more than 1,000 independent colonies.

Pituitary intermediate and anterior lobe cell dissociation

After dissection, mouse pituitary intermediate or anterior lobes were dissociated as described (Budry et al., 2011). Briefly, dissected pituitaries were separated into intermediate and anterior lobes and kept during the dissections in 300µl of dissection buffer (DMEM, 10% FBS, HEPES 10mM and DNase 100 U/ml). Anterior lobes were cut in pieces using a scalpel to facilitate tissue dissociation and digested at 37°C using 5mg/ml Trypsin for 10 minutes. We then added 2mM EDTA and incubated 5 minutes more. 10% FBS was then added to stop the dissociation and samples were centrifuged and then resuspended into 150µl of PBS 1X, 0.1% BSA, 10mM HEPES and for FACS analyses.

3' end single cell RNAseq

Dissociated pituitary cells were diluted at 500 cells per µl and processed using Chromium Single Cell 3′ v2 Reagent (10x Genomics, Pleasanton, CA) following the manufacturer recommendation. Briefly, cells were passed on the channel and 9269 cells were recovered. Pituitary cells were partitioned into gel beads in emulsion for cell lysis and barcoded with oligo-dT priming and reverse transcribed. cDNA library was amplified fragmented and size selected. Samples were controlled at multiple steps during the procedure by running on BioAnalyzer. Libraries were sequenced on Hiseq 4000 with 100bp paired-end reads.

Purification of pituitary lineages by FACS

Dissociated anterior pituitary cells from 62 LH-cerulean mice (Budry et al., 2011) were sorted using FACSAria instrument (BD) and the gate used to define cerulean positive versus negative cells were defined by first assessing auto fluorescence of WT mice of the same strain, C57/Bl6. Cerulean positive and negative cells constituted the gonadotrope and anterior lobe (AL) samples, respectively, that were used in this study.

FACS analyses

Dissociated anterior or intermediate pituitaries from *WT* or *Pax7-KO* mice crossed with the POMC-EGFP transgene were analysed using the FACSCalibur cell analyser (BD Bioscience). EGFP levels were quantified together with forward

scatter (FSC) as an indicator of cell size and side scatter (SSC) as an indicator of granularity for both high and low EGFP-expressing cells. Experiments were repeated on 3 *Pax7* knockout and 5 wild-type litter mates in a mixed C57/Bl6 and 129sv background.

Immunohistofluorescence

Immunohistofluorescence was performed on PFA-fixed paraffin sections as described (Bilodeau et al., 2009). Briefly, 5 days old *Tpit* knockout and WT pituitaries were dissected, fixed in 4% PFA, embedded in paraffin and cut into 5 µm thick sections. The following antibodies used for immunohistofluorescence: Pax7 (DSHB AB_528428), PC2 (a gift of Dr Nabil Seidah, IRCM, Montreal).

ATACseq

All ATACseq samples were processed as previously described (Mayran et al., 2018). Briefly, 50 000 cells were washed in PBS and incubated on ice for 30 minute in a hypotonic cell lysis buffer (0.1% $^{w}/_{v}$ sodium citrate tribasic dehydrate and 0.1% $^{v}/_{v}$ Triton X100) and centrifuged (5 minutes at 2000g at 4°C). Cells were then incubated 30 minutes on ice in cell lysis buffer (10mM Tris-HCl, pH7.4, 10mM NaCl, 3mM $MgCl_2$, 0.1% $^{v}/_{v}$ IGEPAL CA-630. After centrifugation (5 minutes at 2000g at 4°C), the nuclei pellets were resuspended in transposase Master Mix (1.25 µl 10x TD buffer, 5 µl H_2O and 6.5 µl of Tn5: Illumina Nextera Kit; FC-121-1031) and incubated for 30 minutes at 37°C. Samples were purified using the DCC purification columns (Zymo). The eluted DNA was barcoded for multiplexing of samples using Nextera barcodes and PCR-enriched using the Phusion kit. Libraries were recovered with GeneRead Purification columns. Samples were then evaluated by qPCR to test enrichments and sequenced on Illumina Hiseq 2500 with 50bp or 125bp paired-end reads according to Illumina's recommendation.

ChIPseq

ChIPseq were performed as previously described (Langlais et al., 2012). At least 3 immuno-precipitations were pooled per ChIP experiments. Library and flow cells were prepared by the IRCM Molecular Biology Core Facility according to Illumina's recommendations and sequenced on Illumina Hiseq 2500. The following

antibodies were used for ChIPseq: FlagM2 (Sigma F3165), Neurod1 (Poulin et al., 2000), Sox2 (Ab59776, Abcam).

Pull-down assay

MBP fusion proteins coupled to maltose amylose beads were produced as described (Bilodeau et al., 2006). ^{35}S labeled proteins were synthesized *in vitro* using the TNT T7 Quick for PCR DNA kit (Promega, L5540). Labeled proteins were incubated with MBP-tagged proteins in $TNEN_{50}$ (50mM Tris pH7.5, 5mM EDTA, 50mM Nacl, 0.1% NP-40) with 1mM PMSF and 2% BSA for 4 hours at 4°C. Beads were washed three times with 1ml $TNEN_{125}$. Bound proteins were resolved by SDS-Page and visualized by autoradiography.

Data Analyses

Single Cell RNAseq analyses. Using Cell Ranger v2.1.1 (10X Genomics), reads were aligned on the mm10 mouse reference genome using the default parameters of Cell Ranger to generate unique molecular identifier counts for each genes across the 9269 cells that were profiled. We obtained an average of 25220 reads per cell and we detected 1807 median genes per cell. Using the Cell Ranger pipeline with default parameters, we generated a gene-barcode matrix, principal component analysis and dimensionality reduction using the t-SNE algorithm. Unbiased clustering of single cells was performed using Cell Ranger which combines K-means clustering and graph-based clustering uncovering 12 clusters. We used Loupe Cell Browser (10X Genomics) to visualize the t-SNE plot with colored cell according to their assigned cluster or colored by gene UMI. Loupe Cell Browser was also used to perform local differential expression analyses of clusters 1, 2, 4, 5, 7, 8. Cluster corresponds to 4 islands of cells that each express a different hormone gene, Gh, Prolactin, Lh and POMC. Differential analysis of the four sub-clusters against their matching hormone-expressing cell cluster showed that mitochondrial RNAs are down-regulated in each case and this is an indicator of low quality cells (Ilicic et al., 2016). Thus to avoid confounding effects, cluster 3 was not included in all further analyses.

ATACseq. Reads were trimmed, if required, to obtain a read length of 50bp and aligned to the mm10 mouse reference genome using Bowtie v2.3.1 (Langmead et al., 2009) with the following parameters: --fr --no-mixed --no-unal. Sam files were converted into tag directories using HOMER v4.9.1 (Heinz et al., 2010) and into bam files using Samtools v1.4.1 (Li et al., 2009) view function. Tag directories were used to generate the normalized BigWig files with Homer using the command makeUCSCfile with the parameters: -fsize 1e20 -res 5 -fragLength 100. Peaks were identified by comparing each sample replicate to sequenced input DNA from pituitary using MACS v2.1.1.20160309 (Zhang et al., 2008) callpeak function using the parameters: -f BAMPE --bw 250 -g mm --mfold 10 30 -p 1e-5. Peaks with an associated pvalue less than 10^{-5} were kept. First we compared ATACseq profiles of purified pituitary cells: melanotrope (2 replicates), corticotropes (2 replicates) gonadotropes (1 replicate) and whole AL (1 replicate). Peaks from all datasets from purified pituitary cells were merge using HOMER v4.9.1 mergePeaks tool to obtain a file with all unique positions from the ATACseq datasets. This list was clustered by k-means in 2 clusters for each samples giving the 16 combinations of ATAC clustering as represented in a heatmap in Figure 2B. Peaks from all datasets from the various genotypes of *Pax7* and *Tpit* knockout ILs were merged together using HOMER v4.9.1 mergePeaks tool to obtain a file with all unique IL positions from all ATACseq datasets. ATACseq signals were quantified in these different datasets using the analyzeRepeats.pl HOMER command and differential accessibility analyses was performed using getDiffExpression.pl with default parameters which uses Deseq2. Peaks showing a differential p-value less than 0.05 and a fold change of 2 fold or more were considered differentially accessible.

ChIPseq. We mapped ChIPseq reads on the mouse genome assembly mm10 by using Bowtie v1.1.2 with the following settings: bowtie -t -p 4 --trim5 1 --best mm10 –S. Sam files were converted into tag directories using HOMER v4.9.1 and into bam files using Samtools v1.4.1 view function. Peaks were identified by comparing each sample to its control (IP Flag for Pax7, IP IgG for others) using MACS v2.1.1.20160309 callpeak function using the parameters: --bw 250 -g mm --mfold 10 30 -p 1e-5. Peaks with an associated p value less than 10^{-5} were kept.

Data presentation. Heatmaps and average profiles were generated using Easeq (Lerdrup et al., 2016), We used IGV (Thorvaldsdottir et al., 2013) to visualize the BigWig files on the genome. Principal component analysis , clustered Heatmap associated with dendrograms from Figures 1D,5C,F were generated using ClustVis (Metsalu and Vilo, 2015).

References

Bilodeau, S., Roussel-Gervais, A., and Drouin, J. (2009). Distinct developmental roles of cell cycle inhibitors p57Kip2 and p27Kip1 distinguish pituitary progenitor cell cycle exit from cell cycle re-entry of differentiated cells. Mol Cell Biol *29*, 1895-1908.

Bilodeau, S., Vallette-Kasic, S., Gauthier, Y., Figarella-Branger, D., Brue, T., Berthelet, F., Lacroix, A., Batista, D., Stratakis, C., Hanson, J.*, et al.* (2006). Role of Brg1 and HDAC2 in GR trans-repression of the pituitary POMC gene and misexpression in Cushing disease. Genes Dev *20*, 2871-2886.

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.P., Tanay, A.*, et al.* (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell *171*, 557-572.e524.

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 Dev *26*, 2299-2310.

Budry, L., Lafont, C., El Yandouzi, T., Chauvet, N., Conéjero, G., Drouin, J., and Mollard, P. (2011). Related pituitary cell lineages develop into interdigitated 3D cell networks. Proc Natl Acad Sci USA *108*, 12515-12520.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature methods *10*, 1213-1218.

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell *168*, 442-459.e420.

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.

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.

Di Tullio, A., Vu Manh, T.P., Schubert, A., Castellano, G., Mansson, R., and Graf, T. (2011). CCAAT/enhancer binding protein alpha (C/EBP(alpha))-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation. Proc Natl Acad Sci U S A *108*, 17016-17021.

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.

Drouin, J. (2017). Chap 1 - Pituitary Development. In The Pituitary (Fourth Edition), S. Melmed, ed. (Academic Press), pp. 3-22.

Fauquier, T., Rizzoti, K., Dattani, M., Lovell-Badge, R., and 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 USA *105*, 2907-2912.

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.

Hurtado, A., Holmes, K.A., Ross-Innes, C.S., Schmidt, D., and Carroll, J.S. (2011). FOXA1 is a key determinant of estrogen receptor function and endocrine response. Nat Genet *43*, 27-33.

Ilicic, T., Kim, J.K., Kolodziejczyk, A.A., Bagger, F.O., McCarthy, D.J., Marioni, J.C., and Teichmann, S.A. (2016). Classification of low quality cells from single-cell RNA-seq data. Genome Biol *17*, 29.

Jacobs, J., Atkins, M., Davie, K., Imrichova, H., Romanelli, L., Christiaens, V., Hulselmans, G., Potier, D., Wouters, J., Taskiran, II*, et al.* (2018). The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. Nat Genet *50*, 1011-1020.

Khetchoumian, K., Balsalobre, A., Mayran, A., Christian, H., Chénard, V., St-Pierre, J., and Drouin, J. (2018). Cell-autonomous transcriptional mechanism for enhancement of translation capacity in secretory cells BioRxiv.

Laganiere, J., Deblois, G., Lefebvre, C., Bataille, A.R., Robert, F., and Giguere, V. (2005). From the Cover: Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response. Proc Natl Acad Sci USA *102*, 11651-11656.

Lamolet, B., Poulin, G., Chu, K., Guillemot, F., Tsai, M.J., and Drouin, J. (2004). Tpit-independent function of NeuroD1(BETA2) in pituitary corticotroph differentiation. Mol Endocrinol *18*, 995-1003.

Lamolet, B., Pulichino, A.M., Lamonerie, T., Gauthier, Y., Brue, T., Enjalbert, A., and Drouin, J. (2001). A pituitary cell-restricted T-box factor, Tpit, activates POMC transcription in cooperation with Pitx homeoproteins. Cell *104*, 849-859.

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.

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.

Lavoie, P.L., Budry, L., Balsalobre, A., and Drouin, J. (2008). Developmental Dependence on NurRE and EboxNeuro for Expression of Pituitary POMC. Mol Endocrinol *22*, 1647-1657.

Lee, C.S., Friedman, J.R., Fulmer, J.T., and Kaestner, K.H. (2005). The initiation of liver development is dependent on Foxa transcription factors. Nature *435*, 944-947.

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.

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.

Maaten, v.d. (2008). Visualizing Data using t-SNE. Journal of Machine Learning Research JMLR *9*, 2579-2605.

Mansouri, A., and Gruss, P. (1998). Pax3 and Pax7 are expressed in commissural neurons and restrict ventral neuronal identity in the spinal cord. Mech Dev *78*, 171-178.

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

Metsalu, T., and Vilo, J. (2015). ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. Nucleic Acids Res *43*, W566-570.

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature *448*, 313-317.

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.

Poulin, G., Lebel, M., Chamberland, M., Paradis, F.W., and Drouin, J. (2000). Specific protein:protein interaction between basic Helix-Loop-Helix transcription factors and homeoproteins of the Pitx family. Mol Cell Biol *20*, 4826-4837.

Poulin, G., Turgeon, B., and Drouin, J. (1997). NeuroD1/BETA2 contributes to cell-specific transcription of the POMC gene. Mol Cell Biol *17*, 6673-6682.

Pulichino, A.M., Vallette-Kasic, S., Tsai, J.P., Couture, C., Gauthier, Y., and Drouin, J. (2003). Tpit determines alternate fates during pituitary cell differentiation. Genes Dev *17*, 738-747.

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.

Swinstead, E.E., Miranda, T.B., Paakinaho, V., Baek, S., Goldstein, I., Hawkins, M., Karpova, T.S., Ball, D., Mazza, D., Lavis, L.D.*, et al.* (2016). Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. Cell *165*, 593-605.

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.

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.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, 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.

Figures

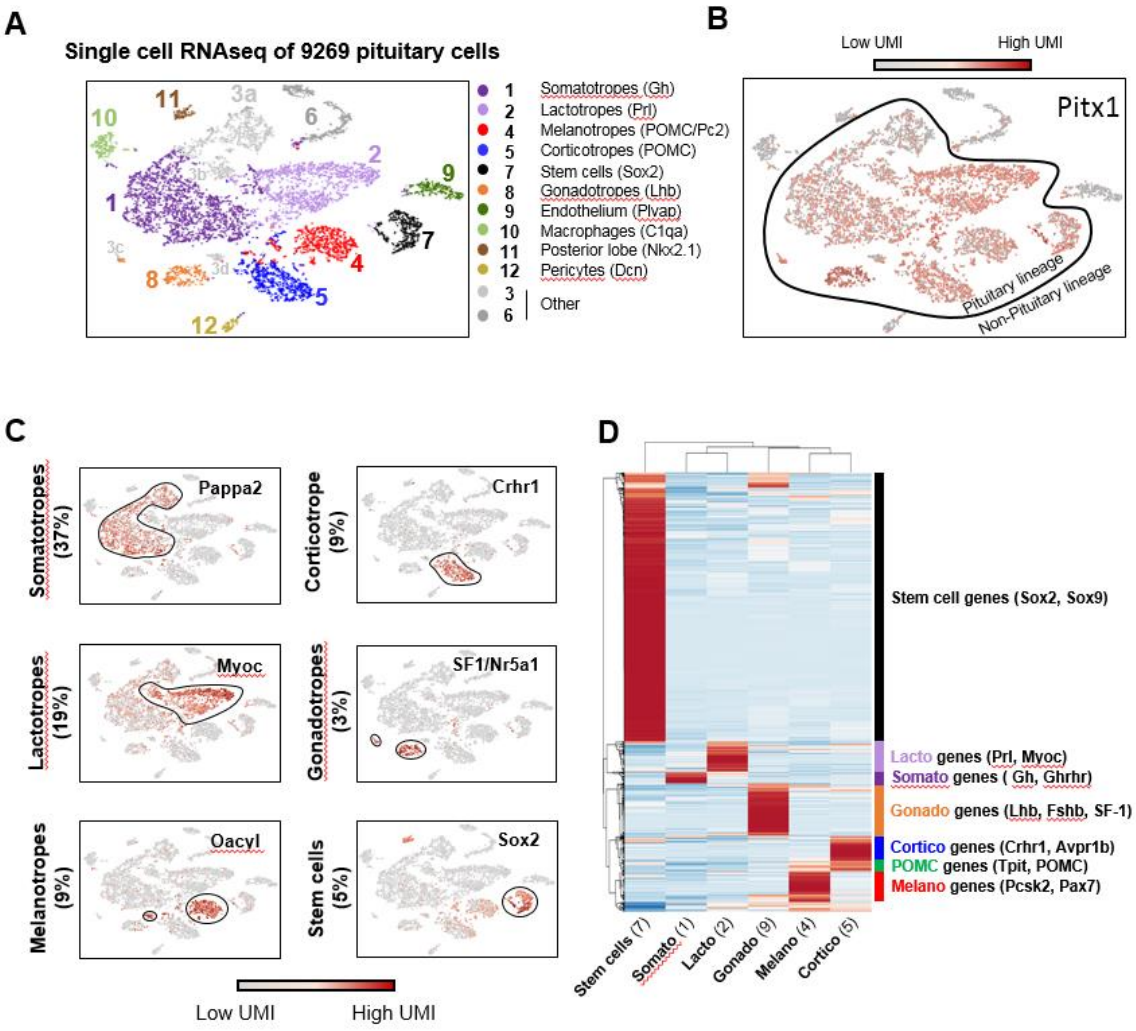


Figure 1. Transcriptional complexity of the pituitary adult gland

Figure 1. Transcriptional complexity of the pituitary adult gland

A. t-SNE map (t-distributed stochastic neighbor embedding) plot of the 9269 profiled pituitary cells colored by the 12 clusters identified using unsupervised k-means clustering. Cluster identification included expression of the hallmarks gene(s) indicated between parenthesis and the markers shown in Figure 1C.
B. t-SNE map showing color-coded Pitx1 expression.
C. t-SNE map showing color-coded expression of indicated markers for the major pituitary lineages.
D. Heatmap showing normalized expression of the 1000 most differentially expressed genes (p value<0.05, FC>2 and minimum number of UMI >0.3) in clusters representing the different endocrine and progenitor cells (cluster 1, 2, 4, 5, 7, 8). Rows are centered; unit variance scaling is applied to rows. Both rows and columns are clustered using correlation distance and average linkage.

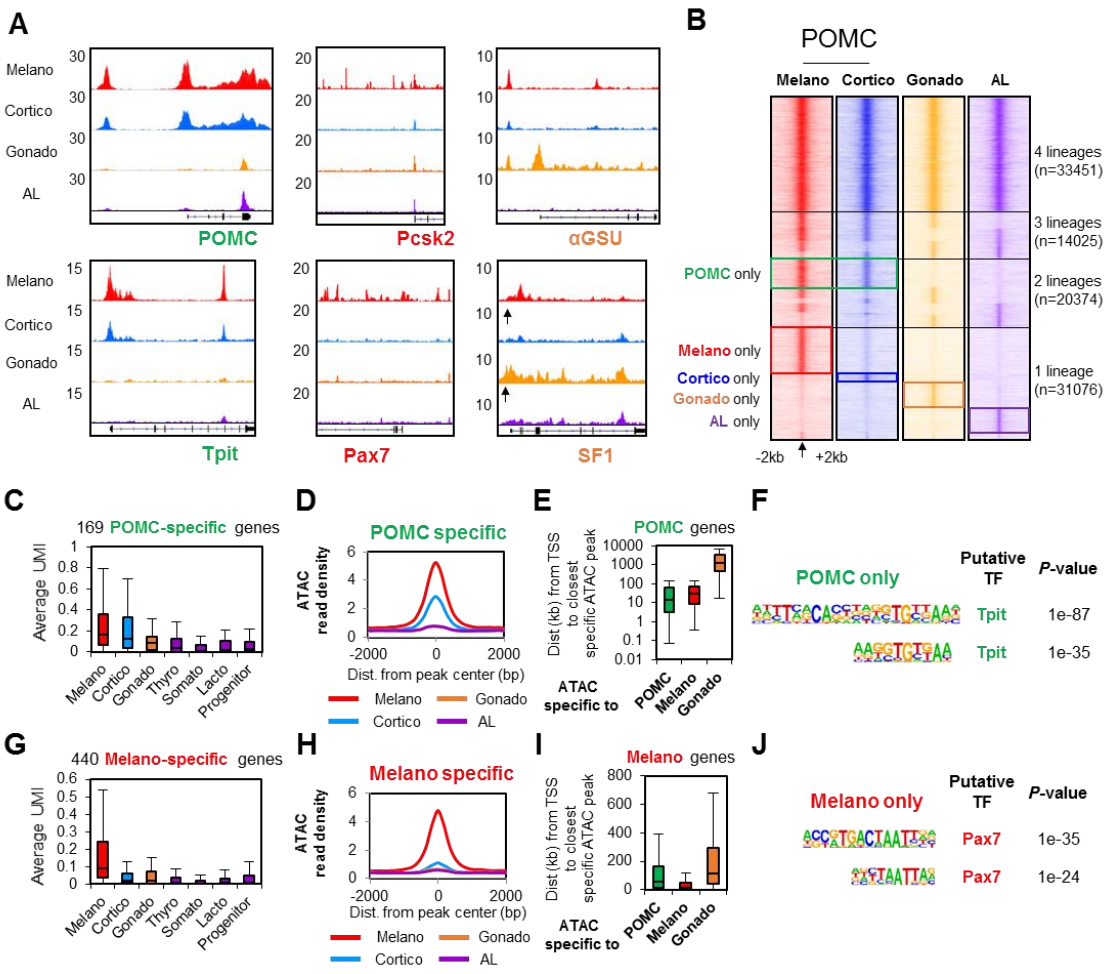


Figure 2. Lineage-specific chromatin access reveals lineage regulators

Figure 2. Lineage-specific chromatin access reveals lineage regulators

A. Genome browser view (IGV) of ATACseq profiles at genes marking pituitary lineages: shared POMC markers (green), melanotrope (red) or gonadotrope (orange). The *SF1* promoter is indicated by an arrow.
B. Heatmaps showing ATACseq signals (RPKM) across the different pituitary lineages in a 4kb window around the ATACseq peak center indicated by an arrow. Colored boxes indicate peaks specifically enriched in the indicated lineage.
C. Boxplot showing expression (UMI) of gene markers of the POMC lineages across the different lineages. Center lines show medians; box limits indicate the twenty-fifth and seventy-fifth percentiles; whiskers extend to 1.5 times the interquartile range from the twenty-fifth to seventy-fifth percentiles.
D. Average profiles of ATACseq signals at POMC-specific ATACseq peaks.
E. Box-plot of distances between the TSS of POMC-specific genes and the closest ATACseq peak in POMC, melanotropes and gonadotrope cells. Box plot features as in C.
F. Motif enriched (assessed by HOMER) under POMC-specific ATACseq peaks and not found in other subsets.
G. Box-plot showing expression (UMI) of melanotrope gene markers across the different lineages. Box plot features as in C.
H. Average profiles of ATACseq signals at melanotrope specific ATACseq peaks.
I. Box-plot of distances between the TSS of melanotrope-specific genes and the closest ATACseq peak in POMC, melanotrope and gonadotrope cells. Box plot features as in C.
J. Motif enriched (assessed by HOMER) under melanotrope specific ATACseq peaks and not found in other subsets.

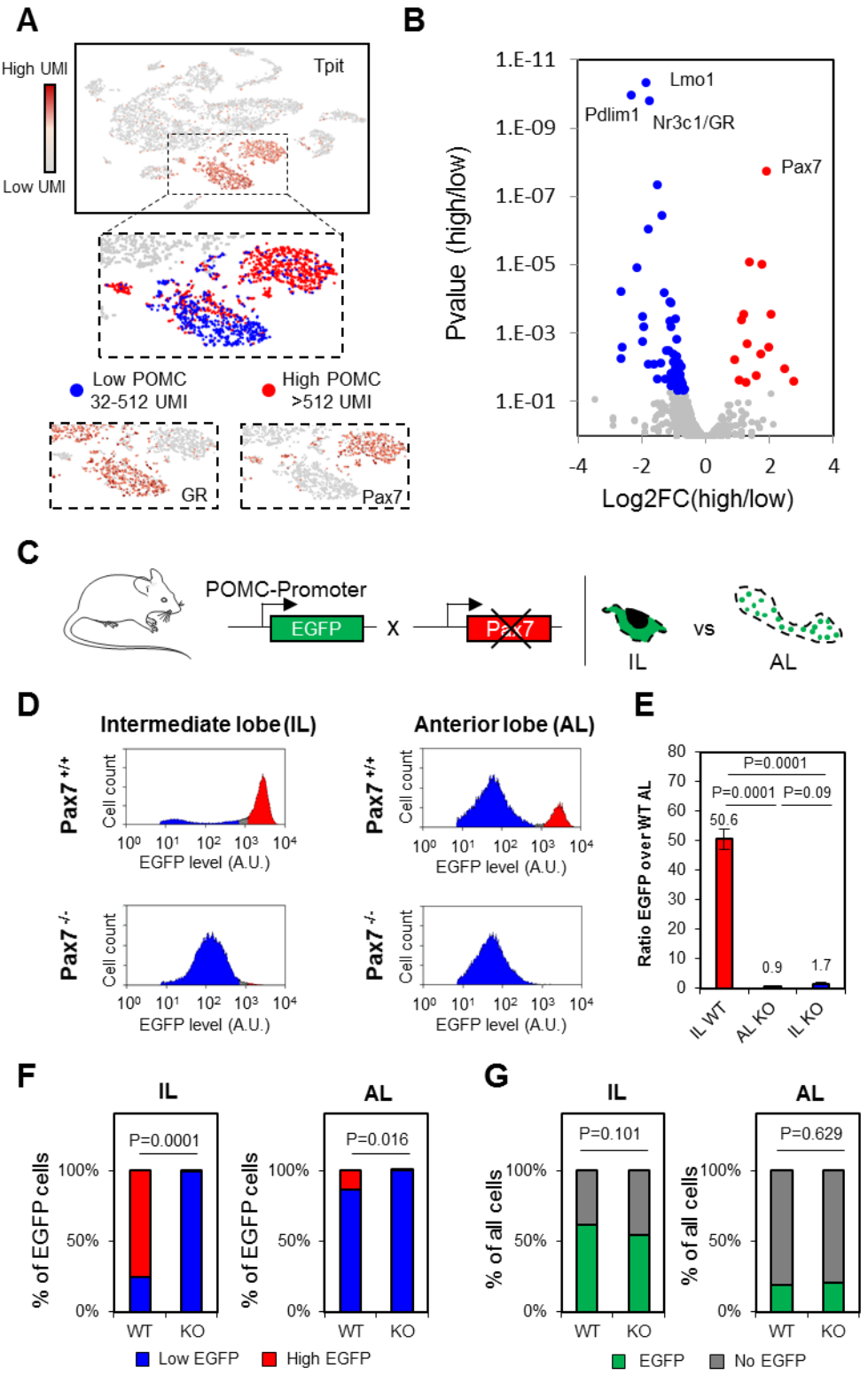


Figure 3. Pax7 implements melanotrope features onto a shared POMC cell identity

Figure 3. Pax7 implements melanotrope features onto a shared POMC cell identity

A. t-SNE map of adult pituitary cells colored for expression of the POMC lineage regulator Tpit (Tbx19). Enlarged panels showing two clusters of POMC cells with high and low POMC levels, melanotropes expressing Pax7 and corticotropes expressing GR.
B. Volcano plot showing differential transcription factor gene expression (p value vs log2 FC) between the high versus low POMC expressing cells. GR and Pax7 are highlighted.
C. Experimental scheme to assess Pax7 dependence of melanotrope phenotypical features. Transgenic mice expressing POMC-EGFP were crossed into *Pax7-/-* mice and compared to WT. The two pituitary lobes were dissected for each genotype and analyzed by FACS.
D. Representative FACS profiles showing cell populations with different POMC-EGFP transgene levels in intermediate (IL) and anterior lobes (AL) of *WT* (n=5) and *Pax7 KO* (n=3) pituitaries.
E. Bar graph showing ratios of EGFP signals in *WT* IL, *Pax7 KO* AL, *Pax7 KO* IL compared to *WT* AL. The analyses included 5 *WT* and 3 *Pax7 KO* replicates. P values were computed using unpaired two-sided t-test.
F. Bar graph showing the proportion of high versus low EGFP expressing cells based on fluorescence signals detected by FACS. P values were computed using unpaired two-sided t-test.
G. Bar graph showing the proportion of EGFP expressing and non-expressing cells based on fluorescence signals detected by FACS. P values were computed using unpaired two-sided t-test.

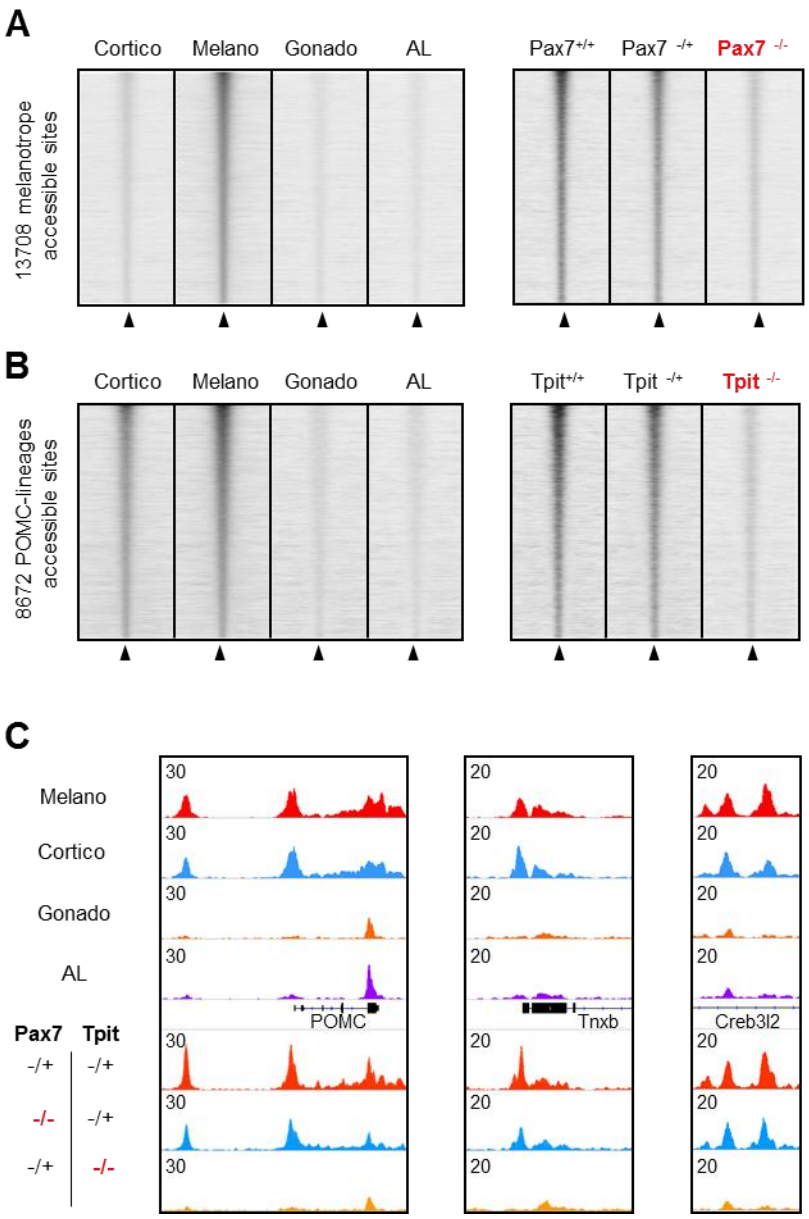


Figure 4. Pax7 and Tpit are required for opening cognate enhancer landscapes

Figure 4. Pax7 and Tpit are required for opening cognate enhancer landscapes

A. Read density heatmaps showing ATACseq signals (RPKM) across the different pituitary lineages in a 4kb window centered at melanotrope specific ATAC peaks (left panel). Right panel shows corresponding ATACseq heatmaps in the ILs of *WT*, *Pax7*$^{+/-}$*;Tpit*$^{+/-}$ (labeled Pax7$^{+/-}$) and *Pax7*$^{-/-}$*;Tpit*$^{+/-}$ (labeled Pax7$^{-/-}$) mice.
B. Read density heatmaps showing ATACseq signals (RPKM) across the different pituitary lineages in a 4kb window centered at POMC-specific ATAC peaks (left panel). Right panel shows corresponding ATACseq heatmaps in the intermediate lobe of *WT*, *Pax7*$^{+/-}$*;Tpit*$^{+/-}$ (labeled Pax7$^{+/-}$) and *Pax7*$^{-/-}$*;Tpit*$^{+/-}$ (labeled Pax7$^{-/-}$) mice.
C. Genome browser view (IGV) of ATACseq profiles at POMC-specific ATACseq peaks in the different pituitary lineages and in ILs of *Pax7*$^{+/-}$*;Tpit*$^{+/-}$*, Pax7*$^{-/-}$*;Tpit*$^{+/-}$*and Pax7*$^{+/-}$*;Tpit*$^{-/-}$ mice.

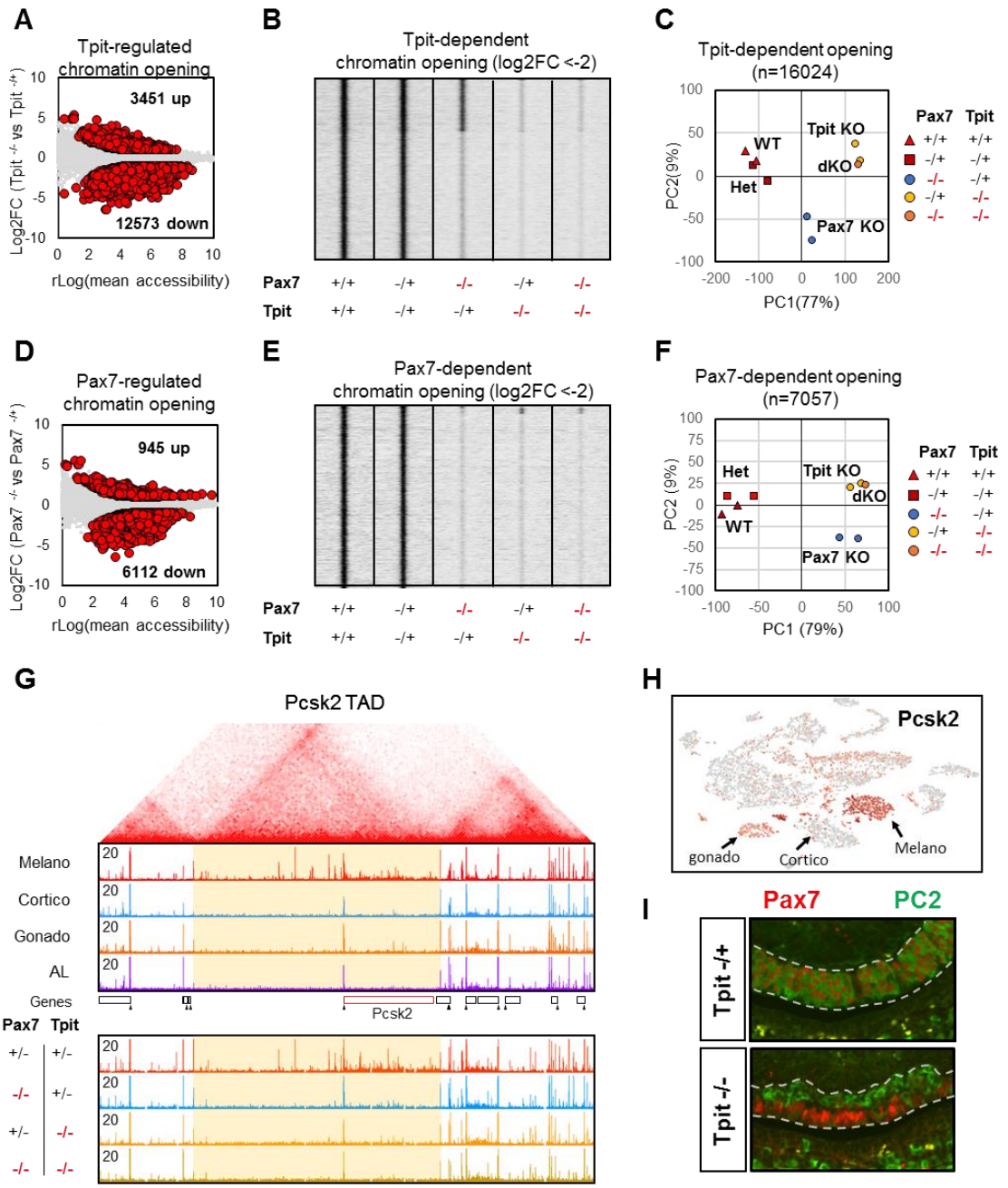


Figure 5. Pax7 dependent chromatin landscape requires Tpit.

Figure 5. Pax7 dependent chromatin landscape requires Tpit.

A. Dispersion plot showing average ATACseq rlog values (assessed by Deseq2) over the log2 fold changes of *Tpit* heterozygote versus *Tpit* knockout IL at all accessible regions. Differentially accessible regions (p-value <0.05 and log2 FC > +/- 1 as computed by Deseq2) are shown as red circles.
B. Read density heatmaps showing ATACseq signals at Tpit-dependent chromatin opening (log2 FC<-2) in the indicated mouse genotypes.
C. Principal component analysis of the ATAC signals at all 16024 Tpit regulated chromatin opening across the tested genotypes.
D. Dispersion plot showing average ATACseq rlog values (assessed by Deseq2) over the log2 fold changes of *Pax7* heterozygote versus *Pax7* knockout IL at all accessible regions. Differentially accessible regions (p value <0.05 and log2 FC > +/- 1 as computed by Deseq2) are shown as red circles.
E. Read density heatmap showing ATACseq signals at Pax7-dependent chromatin opening (log2 FC<-2) in the indicated mouse genotypes.
F. Principal component analysis of the ATAC signals at all 7058 Pax7 regulated chromatin opening across the tested genotypes.
G. Hi-C interaction map (top) from mouse ES cells (Bonev et al., 2017) around the *Pcsk2* locus showing the boundaries of the *Pcsk2* TAD. Genome browser views (bottom) of the ATACseq profiles in purified pituitary cells and ILs of the indicated genotypes at the corresponding genome location.
H. t-SNE map colored for single cell *Pcsk2* expression showing no Pcsk2 expression in corticotropes, highest expression in melanotropes and weak expression in gonadotropes.
I. Co-staining immunofluorescence for Pax7 (red) and PC2 (green) of *Tpit* heterozygote and *Tpit* knockout pituitaries from 5 days postnatal mice.

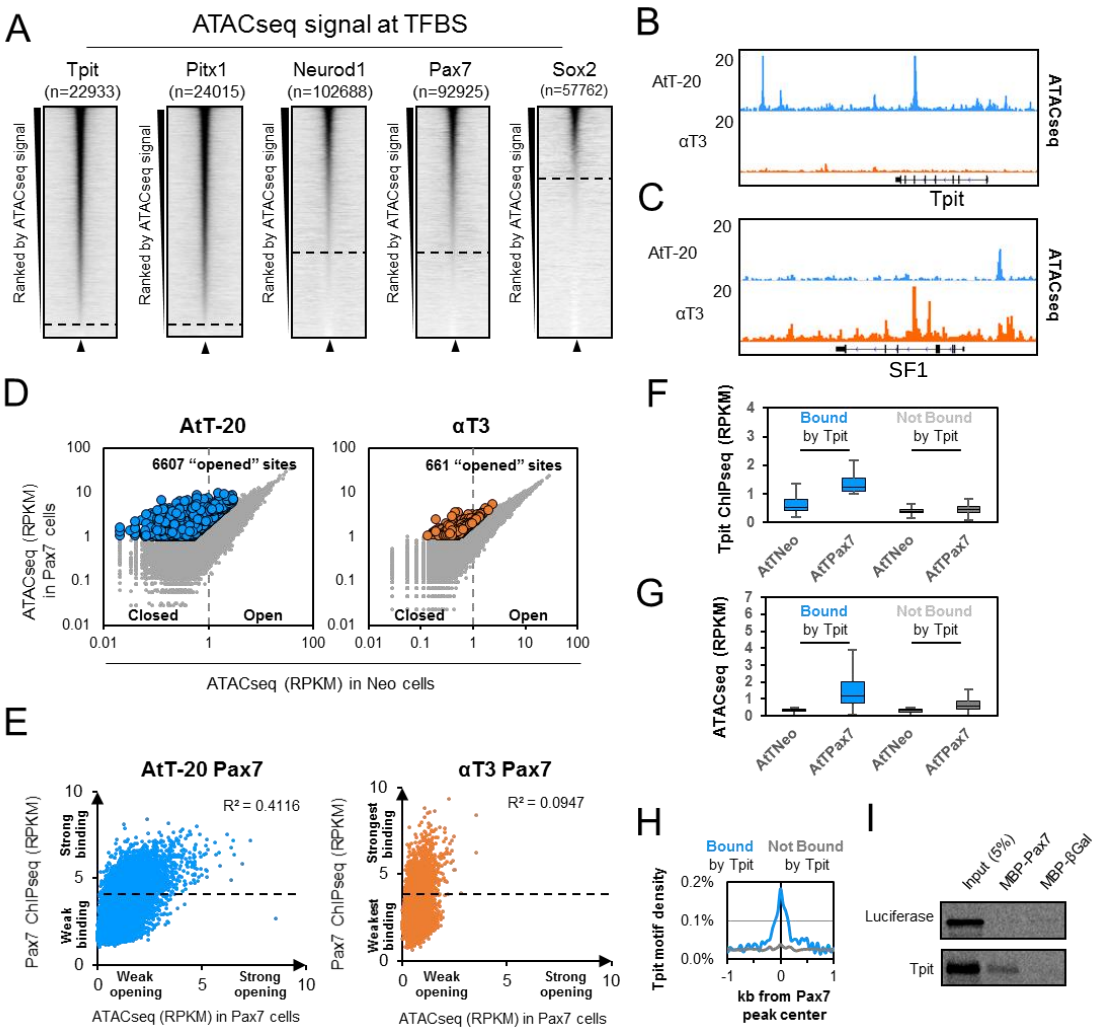


Figure 6. Pax7 binding on closed chromatin is only productive in Tpit-expressing cells.

Figure 6. Pax7 binding on closed chromatin is only productive in Tpit-expressing cells.

A. Read density heatmaps of ATACseq signal density in a 4kb window centered on binding sites for the indicated factors. The heatmaps are ranked by their decreasing ATACseq central (200bp) read densities.

B,C. Genome browser views of ATACseq profiles in AtT-20 and αT3-cells at the Tpit (B) and SF1 (C) loci.

D. Dispersion plots of central (200 bp) ATACseq read densities in Neo (x-axis) versus Pax7 (y-axis) expressing AtT-20 (left) and αT3 cells (right) at all Pax7 binding sites in the indicated cell lines. Colored dots represent sites with significantly stronger signals after Pax7 expression.

E. Dispersion plots of ATACseq read densities in Pax7-expressing cells (x-axis) over Pax7 ChIPseq read densities (y-axis) in AtT-20 (left) and αT3 cells (right) at Pax7 binding sites with no ATACseq signal before Pax7 expression.

F,G. Boxplots of Tpit ChIPseq (F) and ATACseq (G) read densities in Neo and Pax7 expressing AtT-20 cells at Pax7 sites no ATACseq signal before Pax7 expression subdivided into Tpit-bound (>1 RPKM in Pax7 expressing cells, blue) and not bound by Tpit (<1 RPKM in Pax7 expressing cells, grey). Center lines show medians; box limits indicate the twenty-fifth and seventy-fifth percentiles; whiskers extend to 1.5 times the interquartile range from the twenty-fifth to seventy-fifth percentiles.

H. Tpit motif density at Pax7 sites no ATACseq signal before Pax7 expression subdivided into Tpit-bound (>1 RPKM in Pax7 expressing cells, blue) and not bound by Tpit (<1 RPKM in Pax7 expressing cells, grey).

I. Pull-down assay of in vitro translated Tpit interaction with MBP-Pax7 but not with MBP-βGal.

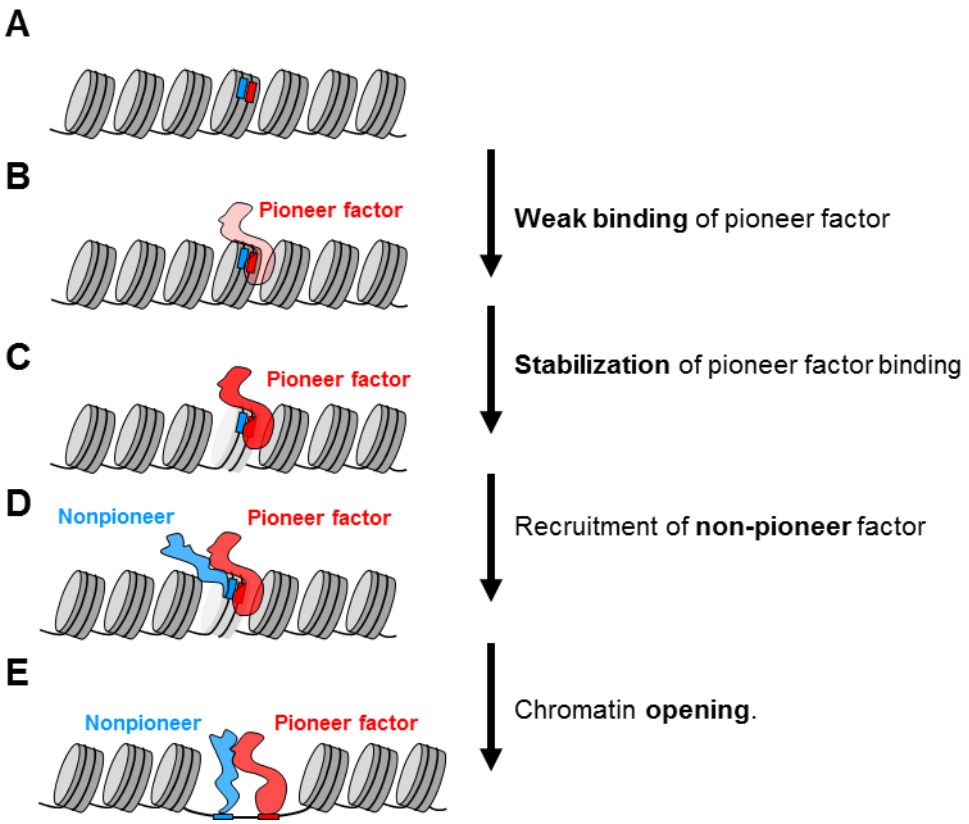


Figure 7. Stepwise model of pioneer and nonpioneer cooperation for chromatin opening.

Figure 7. Stepwise model of pioneer and nonpioneer cooperation for chromatin opening.

A. Inactive nucleosomal heterochromatin with binding sites for Pax7 (red) and Tpit (blue).
B. Weak binding of pioneer factor Pax7 to heterochromatin site.
C. Stabilized Pax7 binding with altered heterochromatin.
D. Recruitment of nonpioneer Tpit through interaction with Pax7 and its DNA binding site.
E. Accessible chromatin with both factors bound at enhancer.

Supplementary Figures

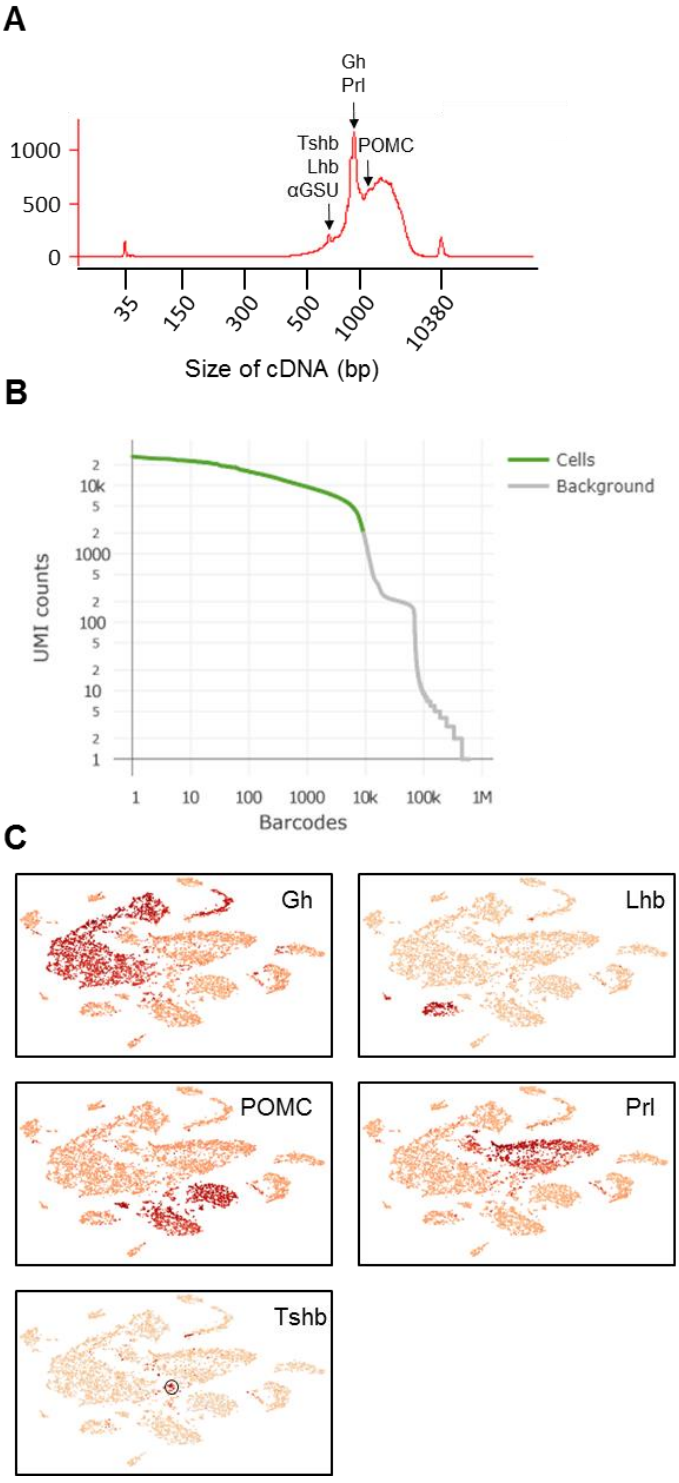


Figure S1. Validation of single cell RNAseq data.

Figure S1. Validation of single cell RNAseq data.

(A) Bioanalyzer profile showing size distribution of cDNAs from dissected pituitary cells.

(B) Single cell RNAseq plot showing barcode numbers (x-axis) over the number of UMI per barcode (y-axis). The threshold (colored) used for selecting the 9269 cells analysed.

(C) t-SNE map showing color-coded expression of hormone genes Gh, Lhb, POMC, Prl and Tshb.

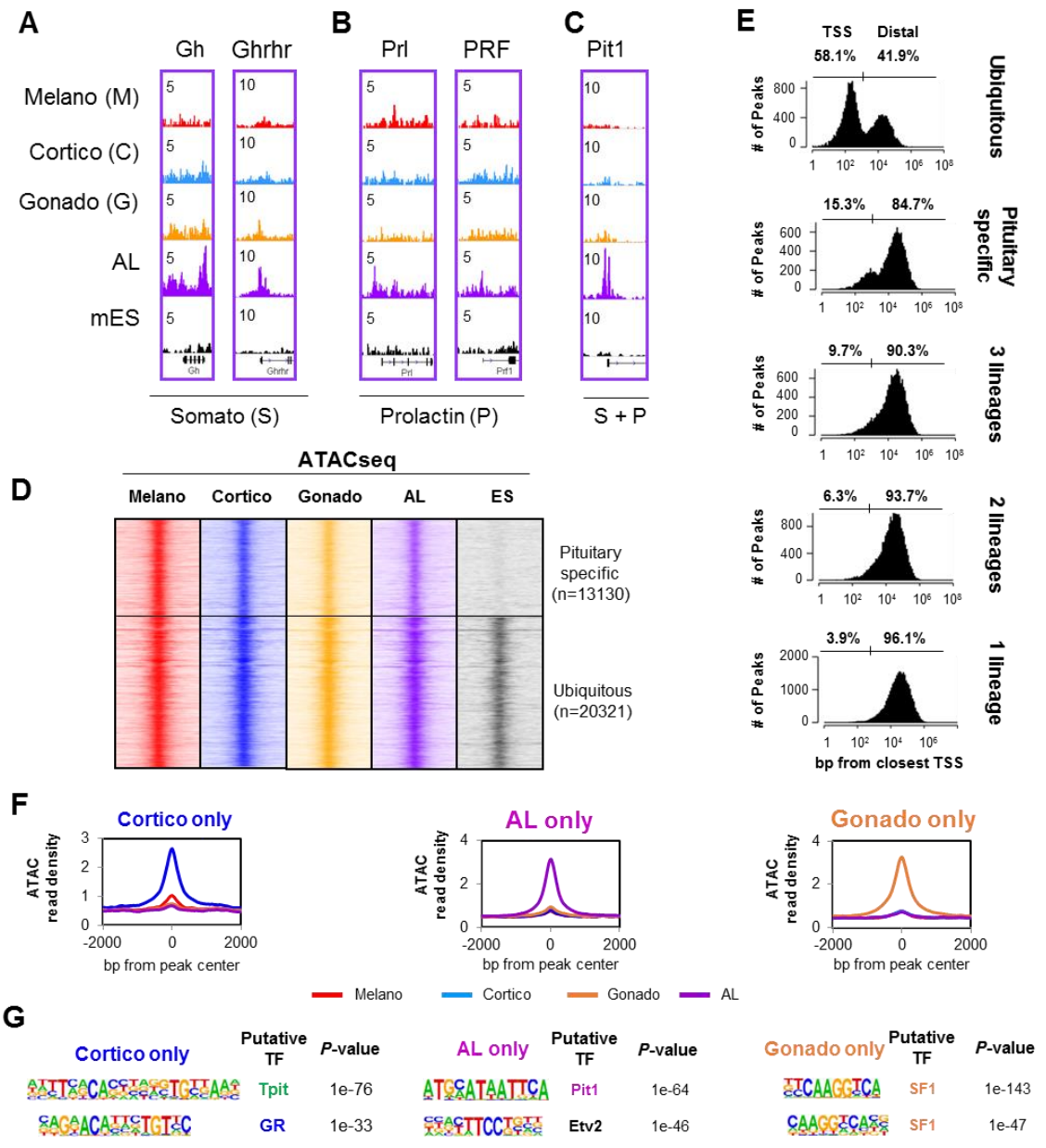


Figure S2. Chromatin landscapes in other pituitary lineages.

Figure S2. Chromatin landscapes in other pituitary lineages.

(A-C) Genome browser views (IGV) of ATACseq profiles from the indicated pituitary lineages at genes marking the identity of somatotropes (A), lactotropes (B) and both (C).

(D) Read density heatmaps showing ATACseq signals at sites of accessibility common to all four pituitary lineages clustered by ATACseq signals in mouse embryonic stem cells (GSE64058).

(E) Genomic distribution of the distances between the indicated category of ATACseq peaks and the closest TSS.

(F) Average profiles of ATACseq signals for the four indicated lineages at cortico-(blue), anterior lobe (purple) and gonado- (orange) specific ATACseq sites.

(G) Motif enrichments (assessed by HOMER) under the indicated category of ATACseq peaks.

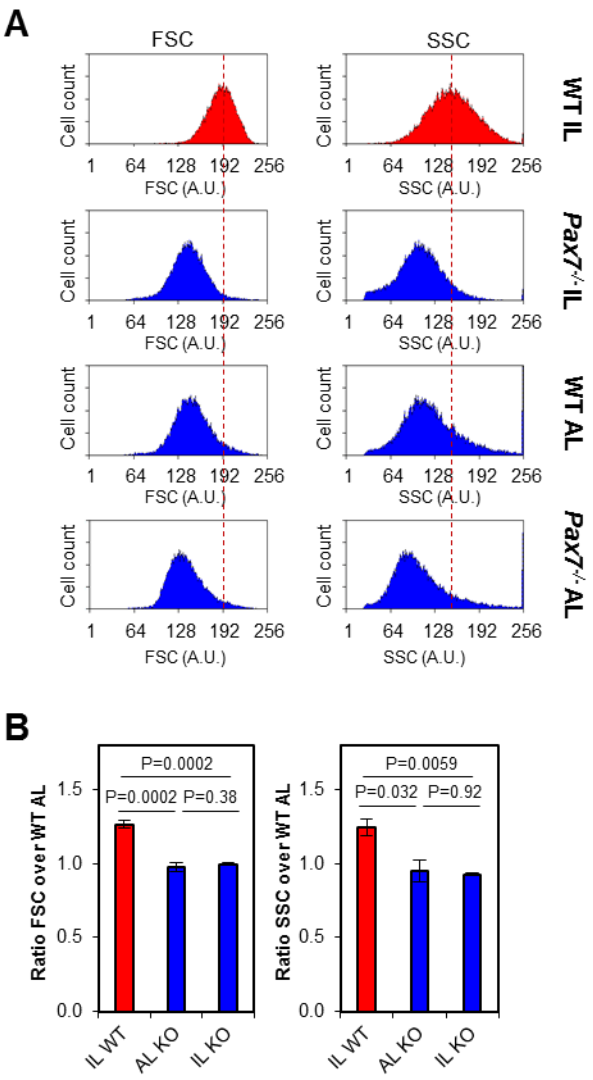


Figure S3. FACS analyses of wild-type (WT) and *Pax7-/- POMC-EGFP* pituitary cells.

(A) Representative FACS profiles showing forward (left) and side (right) scatter for cells from wild type and *Pax7*$^{-/-}$ IL and AL.

(B) Bar graphs showing the ratios of forward (right) and side (right) scatter for EGFP positive cells from wild type (n=5) and *Pax7*$^{-/-}$ (n=3) IL and AL.

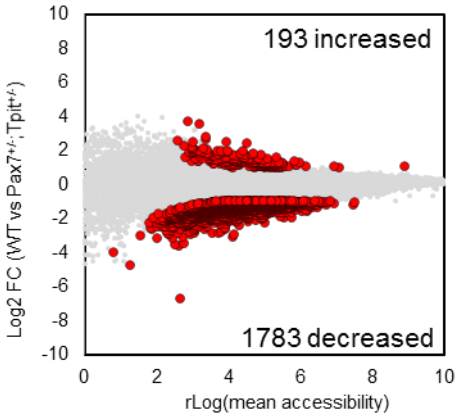


Figure S4. Dispersion plot comparing ATACseq signals in WT and *Pax7*+/-;*Tpit*+/- IL cells.

Dispersion plot showing rLog values of accessibility (ATACseq, x-axis) in wild type and *Pax7*$^{+/-}$;*Tpit*$^{+/-}$ IL over log2 fold changes in wild type versus *Pax7*$^{+/-}$;*Tpit*$^{+/-}$ IL. Red circles identify the differentially accessible regions (p value<0.05, Log2 FC >+/-1).

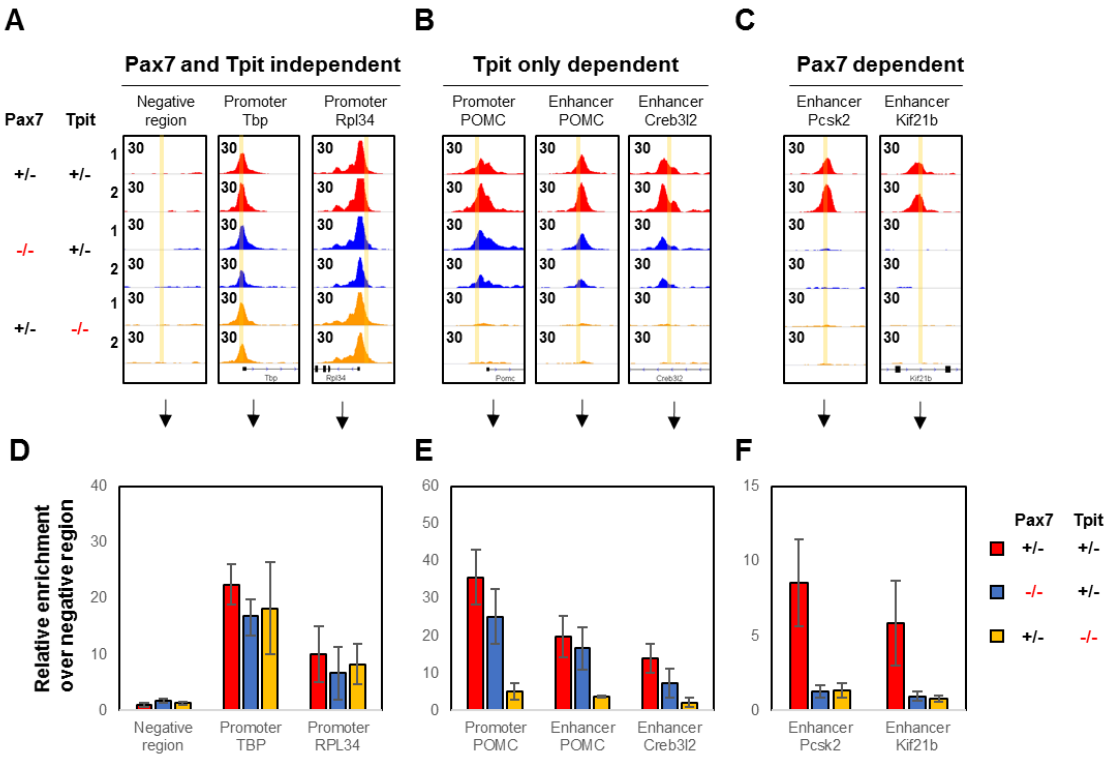


Figure S5. Q-PCR validation of Pax7 and/or Tpit-dependent accessibility (ATACseq).

(A-C) Genome browser views (IGV) of ATACseq profiles from the indicated genotypes at unaffected (A), Tpit-only dependent (B) and Pax7 dependent (C) sites. Regions amplified in the qPCR measurements of Figure S5D-F are highlighted in yellow.

(D-F) Relative enrichments over a negative region, measured by qPCR of ATACseq libraries at unaffected (D), Tpit-only dependent (E) and Pax7 dependent (F) sites.

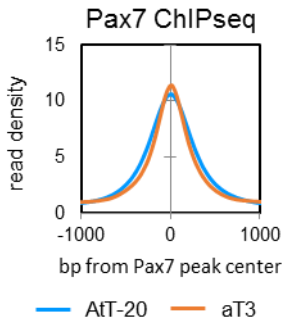


Figure S6. Comparison of Pax7 ChIPseq data in AtT-20 and αT3 cells.

Average profiles of Pax7 ChIPseq in AtT-20 (blue) and αT3 (orange) cells at the best 2000 Pax7 peaks from each lineage showing similar enrichments in both ChIPseqs.

Chapter IV: Discussion

1. Transcriptional and chromatin landscape diversity within the pituitary.

The pituitary has been an unending source of discoveries from physiology or endocrinology to developmental biology and more recently epigenetics. Traditional genetic analyses and molecular biology approaches led to the discovery of Pitx and Tpit as critical regulator of pituitary differentiation and POMC transcription [44, 46, 75]. The advent of the genomic era, lead us to move from single gene candidate approaches to unbiased whole genome analyses. Micro-array or RNAseq analyses identified several regulators involved in pituitary development. For example, these have led to the discovery of ETV1 as a regulator of POMC transcription [76], Pax7 as the selector gene for melanotrope identity [45]. These approaches also identified the role of Tpit as a driver of secretory and translation pathways that transform fetal differentiated cells into hormone-producing factories [77]. However, the anterior pituitary is a heterogeneous tissue; it is composed of five hormone-expressing lineages as well as a stem cell niche and various non-endocrine cells. This makes it challenging to perform biochemical assessments of the cell-specific consequences of disturbed pituitary function and development. Investigating the cell-specific transcription program of the pituitary, relied on the use of transgenic animal expressing a reporter engineered with cell-specific expression.

Studying rare lineages that lack well-defined markers is technically challenging. For example, the posterior lobe is composed of the axonal projection of hypothalamic neurons as well as poorly defined pituicytes. In chapter III, we presented a single cell RNAseq (scRNAseq) analysis performed on an adult male mouse pituitary. This revealed a cluster of posterior pituicytes and showed its associated transcriptome (Chapter III, Fig. 1a). The identification of this transcriptome unexpectedly helped us better define the scope of Pax7 dependence for the melanotrope lineage. Indeed, we had previously shown that Pax7 is required for the expression of the melanotrope gene program in intermediate lobe cells [45]. Yet, some genes (e.g. *Fndc3c1*, *Gldc*, *Gpr50*) that

were identified by micro-array analysis as melanotrope specific were not mis-regulated in *Pax7* knockout mice. This suggested that Pax7 is not the only determinant of the melanotrope fate. We now know that, these genes are not expressed in melanotropes (Appendix 1) but in pituicytes of the posterior lobe which probably contaminated the melanotropes during FACS-sorting of intermediate lobe cells.

Another long-standing question on pituitary development and renewal concerns the pituitary stem cell niche. This niche is characterized by expression of the transcription factor Sox2 as well as Sox9 and it was shown that these cells participate in the renewal of the adult pituitary cells [78, 79]. We do not know what triggers the exit from a stem cell state and the entry towards differentiation. The transition between stem cell and hormone expressing differentiation is likely occurring in a short time period and this limits our ability to study this process. Using scRNAseq, we were able to define a very rich transcriptome of pituitary stem cells and numerous stem cell markers are not expressed in differentiated cells (Chapter III, Fig. 1a, b). Interestingly, a subset of stem cells express endocrine cell genes (Appendix 2a, b). Sub-clustering of the stem cell population in these two categories: hormone-expressing and non-hormone expressing cells (Appendix 2c) showed that most of the stem cell program is still expressed while gene associated with differentiated endocrine cells are turned on (Appendix 2d). This suggest that the differentiation program is turned on before the cell leave the stem cell state and possibly gene associated with stem cell identity are only turned off once the new identity is fully established. However, we did not find differentiating-stem cells that express multiple hormones. This indicates that each of these cells may already be specified toward one endocrine lineage identity. Finally, principal component analyses support the interpretation that these cells may be transitioning towards differentiation (appendix 2e). Further analyses of these transitioning cells should give us insights into the molecular mechanisms underlying the exit from the stem cell state towards differentiation. So far, we only exploited these datasets to define the transcriptomes of the different pituitary lineages. This is already useful for many purposes; for example, the identification of surface markers of each lineage allows

easy purification by FACS for future studies. Clearly, a lot more information remains to be extracted from this experiment. This short discussion illustrates how age-old questions become amenable to novel insights when new technologies are applied to complex biological systems.

2. Chromatin landscape of pituitary differentiation

In the beginning of the work presented here, the chromatin landscape of pituitary cells was unknown. Indeed, the small size and limited materials available from the pituitary made it challenging to perform experiments such as chromatin immunoprecipitations on normal cells. Thus, tissue like the pituitary could not have been investigated in large-scale project such as, ENCODE. However, new techniques such as Assay for Transposase of Accessible Chromatin followed by high throughput sequencing (ATACseq) now provides a snapshot of the epigenome using very few cells. Chapter II and III used ATACseq on FACS-purified pituitary cells to define the open chromatin landscape in different lineages: corticotrope and melanotrope in chapter II and gonadotrope and anterior lobe (mostly somatotropes and lactotropes) in chapter III. This defined differentially accessible regions (DARs) for each cell type. The DNA motifs found under each DARs revealed lineage regulators without prior knowledge. For instance, and as validation, in Chapter III, we identified the DNA motif for binding of SF1, the transcription factor required for gonadotrope identity [70, 71], under gonadotropes DARs. Under regions of open chromatin specific to the anterior lobe, we found the motif of Pit1 that is required for the differentiation of both somatotropes, lactotropes and thyrotropes [58]. This suggests that SF1 and Pit1 could participate in a pioneer factor paradigm during lineage specification towards their cognate lineages. This database of open chromatin regions is also useful to predict enhancers that may be active in one or another lineage. This would be particularly useful to generate reporter mice for the different lineages.

Comparisons of the open chromatin landscapes of corticotropes and melanotropes cells in chapter II [80] revealed that the melanotrope-specific enhancer repertoire is five times greater than the number of regions with corticotrope-specific

accessibility. However, the transcriptional programs of each of these two cells is equally rich, with 243 corticotrope-enriched and 261 melanotrope-enriched genes. This suggest that the corticotropes chromatin landscape may be a default configuration on top of which Pax7 deploys an enhancer repertoire driving melanotrope specification. Yet, it is still not understood how Pax7 exerts the observed extinction of the corticotrope program. In Chapter III, we investigated the consequences of Tpit loss on the chromatin landscape of POMC cells. Interestingly, we found that regions of chromatin accessibility shared between the two POMC lineages requires Tpit. These regions are near genes shared by the two lineages as well as genes specific to corticotropes (data not shown). This suggests that the corticotrope-specific program requires Tpit and has an open chromatin landscape in both corticotropes and melanotropes but that an unclear mechanism modulates gene expression. One possible scenario may be that corticotrope-specific transcription factors act on these previously established enhancers. Several transcription factors are only active in corticotropes for example the glucocorticoid receptor (GR) expression is only repressed in melanotropes and thus specific to corticotropes. Accordingly, several known GR targets are specifically expressed in corticotropes (Fkbp5, Rasd1) [81, 82]. This suggest that GR is a good candidate as regulator of a portion of the cortico-specific gene expression program.

3. Mechanism of pioneer-driven chromatin opening

Pioneer transcription factors are critical to drive lineage transitions. The paradigm of pioneer action relies on the ability to access closed chromatin and trigger chromatin opening to allow the binding of classical transcription factors. Many pioneer factors have been characterized by their ability to directly bind nucleosomes [83]. Although these analyses rely on biochemical experiments that may somewhat differ from the situation inside a nucleus, they likely reflect a unique ability of pioneer factors that is not shared by other transcription factors. The work presented here illustrates the unique properties of the pioneer factor Pax7 and allowed identification of multiple steps of pioneer-driven chromatin opening [80].

We found that regions opened by Pax7 acquire *bona fide* histone modifications found at active enhancers. Pax7 pioneer targets gain the H3K4me1 mark and this is associated with the recruitment of Ash2, a protein of the MLL (Trithorax) complex that is responsible for deposition of this mark (Chapter II Fig. 3). However, some sites resist Pax7-driven chromatin opening and histone modification deposition (Chapter II Fig. 4). These resistant sites are marked by higher levels of histone 3 lysine 9 trimethyl (H3K9me3) but have similar levels of H3K9me2 compared to pioneered enhancers. Interestingly, these sites are still bound by Pax7, which suggests that H3K9me3 does not impair Pax7 binding. On the other hand, H3K9me2 is similarly high in pioneer and resistant targets. This may reflect that H3K9me2 and H3K9me3 underlie two distinct (hetero)chromatin environments, one where Pax7 drives chromatin remodeling and one where it does not. Thus, the chromatin environment may be important to define the scope of pioneer factors.

Our analyses of the different steps of Pax7-driven pioneering provide several clues on mechanisms that impose limits to Pax7 remodeling (Chapter II Fig. 5). Indeed, Pax7 binding to closed chromatin (both resistant and sensitive to remodeling) is rapid (less than 30 minutes) but initially weak. After 24 hours, Pax7 binding to pioneered sites is stabilized but this does not occur at resistant sites. The initial binding of Pax7 in the first 30 minutes following activation suggests that Pax7 is very efficient at “scanning” closed chromatin. In resistant sites, Pax7 fails to stabilize and chromatin opening does not occur. This showed that Pax7 initial binding is not sufficient to drive the remaining of the pioneer process, as the stabilization of Pax7 seems to be required. This probably reflects the fact that Pax7 stabilization requires some unknown change to chromatin structure. One possibility is that di- or tri-methylation of histone H3 directly prevents Pax7 stabilization and that loss of this mark is required for Pax7 stabilization. Indeed, at pioneered sites, H3K9me2 is lost after Pax7 remodeling; however, H3K9me3 remains at resistant sites. H3K9me2 may thus limit Pax7 stabilization until it is removed by an unknown mechanism in the first 24 hours, thus allowing Pax7 stabilization. However, at resistant sites, Pax7 being unable to trigger the removal of H3K9me3, these sites resist Pax7 remodeling. Another explanation could be

that proteins associated with the two different chromatin environments (H3K9me2 and H3K9me3) may also limit Pax7 ability to stabilize on its targets. For example, it was recently shown that the HP1α protein, which is associated with H3K9me3 and constitutive heterochromatin, drives phase separation of bound DNA [84]. Pax7 is still able to access resistant sites, it is thus unlikely that Pax7 binding is hindered by these phase separated droplets; however, it is possible that phase separation segregates some Pax7 partners outside of its reach preventing Pax7 stabilization and the following chromatin remodeling step.

In addition, we showed that Pax7 pioneering leads to loss of DNA methylation and this is associated with the implementation of a long term epigenetic memory assessed by chromatin accessibility and transcription factor binding for more than 20 days after Pax7 withdrawal. We propose that this reprogramming stability relies on the loss of DNA methylation observed in Pax7-expressing cells. Indeed as mentioned in the introduction, DNA methylation is one of the most stable epigenetic mark as it is present on the DNA itself and maintained during replication. The ability of Pax7 to remodel the chromatin at methylated DNA target sites is already uncommon, but its ability to drive loss of DNA methylation is particularly interesting. There are two known mechanisms of DNA demethylation, the first one is active and mediated by the Tet enzymes, while the second one may be passive through inhibition of DNA methylation maintenance during replication, which would dilute DNA methylation at each round of replication. Given the slow timescale of Pax7 chromatin opening (more than three days), loss of CpG methylation likely follows the same kinetics. It is thus unlikely that Tet-mediated de-methylation occurs at Pax7 sites as these rely on direct enzymatic modifications of the ^{5me}CpG and as such, we may expect Tet-dependent loss of CpG methylation to be somewhat rapid. Thus, Pax7-driven loss of CpG methylation may arise from a failure to maintain CpG methylation during replication. Interestingly when we purified Pax7-associated partners through RIME [85], (data not shown), two strong Pax7-interacting partners are DNMT1 as well as its regulator protein UHRF1[86]. This suggests that Pax7, DNMT1 and UHRF1 can interact and further supports the hypothesis of a passive mode of Pax7-driven loss of CpG methylation.

In chapter III, we show that *in vivo,* Tpit is required for Pax7-driven chromatin opening but not for Pax7 binding to closed chromatin This demonstrates that during pioneer factor action, a non-pioneer factor may not be as passive as previously thought. Indeed, Tpit is unable to bind closed chromatin in the absence of Pax7 action, thus Tpit is not a pioneer factor. In the absence of Tpit expression, Pax7 can still strongly bind to closed chromatin. This suggests that Tpit is either required for the step of chromatin opening or for a currently unknown step occurring before chromatin opening but after Pax7 binding stabilization. Purification of Tpit-interacting partners through RIME experiments identified several ATP-dependent nucleosome remodelers of the SWI-SNF family of proteins as partners of Tpit (data not shown). This argues for a direct role of Tpit and its interacting partners during the chromatin opening step. In future work, it will be crucial to assess the precise role of Tpit, its interactors and the dynamics of their recruitments following Pax7 activation. Also, stable Pax7 expression in combination with an inducible Tpit may be a useful tool to uncover the dynamics of Tpit function for chromatin remodeling during Pax7 pioneering.

Inherently, pioneer factors represent a risk that could potentially initiate disease conditions. Indeed as factors that can target closed chromatin, mis-expression of a pioneer factor may lead to long-term activation of inappropriate regulatory networks. And indeed, pioneer factors are frequently mutated or over-expressed in cancers [87]. However, there are several limits to pioneer factor capabilities. We already discussed above the possible role of the histone modification H3K9me3 for limiting Pax7 capabilities. This may prevent the activation of the Pax7-dependent myogenic targets in the pituitary. In another system, H3K9me3 was found to prevent the binding of the pioneer factors Oct4, Sox2 and Klf4 [88] suggesting that this mark may limit pioneer factor competence through several mechanisms. Another strategy may consist in preventing pioneer mis-expression. This can be achieved through promoter occlusion in lineages where they are not supposed to be expressed. As shown in chapter II and chapter III, cell specific accessibility occurs mostly at enhancers while promoter sequences tend to be ubiquitously accessible. Yet, gene encoding lineage regulators such as *Pax7*, *Tpit*,

SF1 or *Pit1* each have lineage-specific promoter accessibility i.e. their promoters are occluded in the non-relevant lineages. Finally, another strategy was identified in chapter III and discussed above; it consists in a separation of tasks and requisite cooperation between a pioneer that targets closed chromatin, Pax7, and a nonpioneer, Tpit, that provides opening of the chromatin. In the context of development, this can have several advantages; it provides robustness to the system by requiring co-expression of two or more factors in order to activate certain programs/genes. In addition, this can provide a way to control the timing of activation of different subsets of targets through the expression of different cooperating factors at different times. Understanding the limits and capabilities of transcription factors to drive cell fate transitions is critical for control of these processes.

4. General conclusion

In summary, the work presented in chapter II and chapter III define the diversity of transcriptional, and chromatin landscapes of the pituitary. This is a resource for the identification of lineage specific regulatory elements, to identify cell fate markers, lineage transition or cell type heterogeneity.

In chapter II, we dissected the molecular mechanism of pioneer-driven chromatin opening by Pax7 and identified key steps in this process, which were unknown before this work. Namely, we uncovered that pioneer factor binding occurs very rapidly in the first 30 minutes; in the following 24 hours, pioneer binding is stabilized and progressive chromatin opening spans over more than three days. We also showed that Pax7 action leads to the loss of DNA methylation and this is strikingly associated with the implementation of a long-term epigenetic memory that lasts numerous cell divisions.

Finally, in chapter III, we provide evidence that a key property of pioneer factor action had been incorrectly assumed. The assumption being that nonpioneer factors are binding to region previously opened by pioneer factor. Here we show that pioneer-driven chromatin remodeling may be a cooperative process between a pioneer Pax7 and a non-pioneer factor Tpit. This restricts the core property of a pioneer factor such as Pax7 as the ability to target closed chromatin and the immediate local change in chromatin structure. In contrast, chromatin opening may constitute a separate task that requires a cooperation between a pioneer like Pax7 with a nonpioneer factor such as Tpit.

References

1. Ruiz-Trillo, I., et al., *The origins of multicellularity: a multi-taxon genome initiative.* Trends Genet, 2007. **23**(3): p. 113-8.
2. Rosenberg, A.B., et al., *Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding.* Science, 2018. **360**(6385): p. 176-182.
3. Vaquerizas, J.M., et al., *A census of human transcription factors: function, expression and evolution.* Nat Rev Genet, 2009. **10**(4): p. 252-63.
4. Lambert, S.A., et al., *The Human Transcription Factors.* Cell, 2018. **172**(4): p. 650-665.
5. Payvar, F., et al., *Purified glucocorticoid receptors bind selectively in vitro to a cloned DNA fragment whose transcription is regulated by glucocorticoids in vivo.* Proc. Natl. Acad. Sci. USA, 1981. **78**: p. 6628-6632.
6. Webster, N.J.G., et al., *The hormone-binding domains of the estrogen and glucocorticoid receptors contain an inducible transcription activation function.* Cell, 1988. **54**: p. 199-207.
7. Weintraub, H., et al., *Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD.* Proc. Natl. Acad. Sci. USA, 1989. **86**: p. 5434-5438.
8. Sekiya, S. and A. Suzuki, *Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors.* Nature, 2011. **475**(7356): p. 390-3.
9. Lee, C.S., et al., *The initiation of liver development is dependent on Foxa transcription factors.* Nature, 2005. **435**(7044): p. 944-7.
10. Rudnicki, M.A., et al., *MyoD or Myf-5 is required for the formation of skeletal muscle.* Cell, 1993. **75**(7): p. 1351-9.
11. Luger, K., M.L. Dechassa, and D.J. Tremethick, *New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?* Nat Rev Mol Cell Biol, 2012. **13**(7): p. 436-47.
12. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 A resolution.* Nature, 1997. **389**(6648): p. 251-60.
13. Robinson, P.J. and D. Rhodes, *Structure of the '30 nm' chromatin fibre: a key role for the linker histone.* Curr Opin Struct Biol, 2006. **16**(3): p. 336-43.
14. Marsden, M.P. and U.K. Laemmli, *Metaphase chromosome structure: evidence for a radial loop model.* Cell, 1979. **17**(4): p. 849-58.
15. Shlyueva, D., G. Stampfel, and A. Stark, *Transcriptional enhancers: from properties to genome-wide predictions.* Nat Rev Genet, 2014. **15**(4): p. 272-86.
16. Jin, C., et al., *H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions.* Nat Genet, 2009. **41**(8): p. 941-5.
17. Jin, C. and G. Felsenfeld, *Nucleosome stability mediated by histone variants H3.3 and H2A.Z.* Genes Dev, 2007. **21**(12): p. 1519-29.
18. Jenuwein, T. and C.D. Allis, *Translating the histone code.* Science, 2001. **293**(5532): p. 1074-80.
19. Calo, E. and J. Wysocka, *Modification of enhancer chromatin: what, how, and why?* Mol. Cell, 2013. **49**(5): p. 825-837.
20. Bernstein, B.E., et al., *A bivalent chromatin structure marks key developmental genes in embryonic stem cells.* Cell, 2006. **125**(2): p. 315-326.

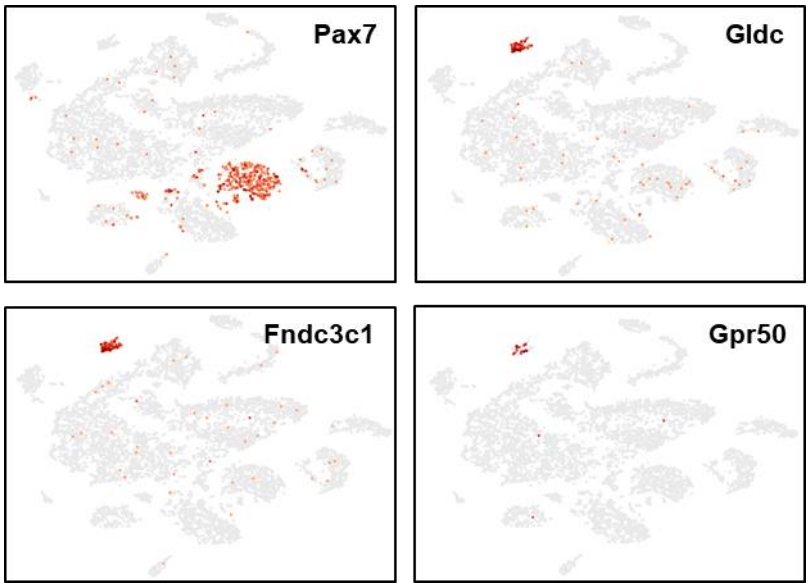
21. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Res, 2011. **21**(3): p. 381-95.
22. Teves, S.S., et al., *A dynamic mode of mitotic bookmarking by transcription factors.* Elife, 2016. **5**.
23. Reinberg, D. and L.D. Vales, *Chromatin domains rich in inheritance.* Science, 2018. **361**(6397): p. 33-34.
24. Li, E., T.H. Bestor, and R. Jaenisch, *Targeted mutation of the DNA methyltransferase gene results in embryonic lethality.* Cell, 1992. **69**(6): p. 915-26.
25. Razin, A. and H. Cedar, *Distribution of 5-methylcytosine in chromatin.* Proc Natl Acad Sci U S A, 1977. **74**(7): p. 2725-8.
26. Smith, Z.D. and A. Meissner, *DNA methylation: roles in mammalian development.* Nat Rev Genet, 2013. **14**(3): p. 204-20.
27. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.* Cell, 1999. **99**(3): p. 247-57.
28. Bestor, T.H. and V.M. Ingram, *Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA.* Proc Natl Acad Sci U S A, 1983. **80**(18): p. 5559-63.
29. Pastor, W.A., L. Aravind, and A. Rao, *TETonic shift: biological roles of TET proteins in DNA demethylation and transcription.* Nat Rev Mol Cell Biol, 2013. **14**(6): p. 341-56.
30. Baran, Y., et al., *The landscape of genomic imprinting across diverse adult human tissues.* Genome Res, 2015. **25**(7): p. 927-36.
31. Drouin, J., *Chap 1 - Pituitary Development*, in *The Pituitary (Fourth Edition)*, S. Melmed, Editor. 2017, Academic Press. p. 3-22.
32. Harris, G.W., *Neural control of pituitary gland.* Physiol. Rev, 1948. **28**: p. 139-179.
33. Harris, G.W., *The function of the pituitary stalk.* Bull. Johns. Hopkins. Hosp, 1955. **97**(5): p. 358-375.
34. Couly, G.F. and N.M. Le Douarin, *Mapping of the early neural primordium in quail-chick chimeras. II. The prosencephalic neural plate and neural folds: implications for the genesis of cephalic human congenital abnormalities.* Dev. Biol, 1987. **120**: p. 198-214.
35. Rizzoti, K., et al., *SOX3 is required during the formation of the hypothalamo-pituitary axis.* Nat. Genet, 2004. **36**(3): p. 247-255.
36. Suga, H., et al., *Self-formation of functional adenohypophysis in three-dimensional culture.* Nature, 2011. **480**(7375): p. 57-62.
37. Davis, S.W., A.H. Mortensen, and S.A. Camper, *Birthdating studies reshape models for pituitary gland cell specification.* Dev Biol, 2011. **352**(2): p. 215-227.
38. Brownstein, M.J., J.T. Russel, and H. Gainer, *Synthesis, transport and release of posterior pituitary hormones.* Science, 1980. **207**(4429): p. 373-378.
39. Gimpl, G. and F. Fahrenholz, *The oxytocin receptor system: structure, function, and regulation.* Physiol. Rev, 2001. **81**(2): p. 629-683.
40. Rivier, C. and W. Vale, *Modulation of stress-induced ACTH release by corticotropin-releasing factor, catecholamines and vasopressin.* Nature, 1983. **305**: p. 325-327.
41. Corstens, G.J., et al., *Analysis of Xenopus melanotrope cell size and POMC-gene expression.* Ann. N. Y. Acad. Sci, 2005. **1040**: p. 269-272.
42. Roubos, E.W., et al., *Plasticity of melanotrope cell regulations in Xenopus laevis.* Eur. J. Neurosci, 2010. **32**(12): p. 2082-2086.
43. Celio, M.R., et al., *'Proopiocortin fragments' in normal human adult pituitary. Distribution and ultrastructural characterization of immunoreactive cells.* Acta endocrino (Copenh.), 1980. **95**(1): p. 27-40.

44. Lamolet, B., et al., *A pituitary cell-restricted T box factor, Tpit, activates POMC transcription in cooperation with Pitx homeoproteins.* Cell, 2001. **104**(6): p. 849-59.
45. 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. 2299-2310.
46. Pulichino, A.M., et al., *Tpit determines alternate fates during pituitary cell differentiation.* Genes Dev, 2003. **17**(6): p. 738-747.
47. Papadimitriou, A. and K.N. Priftis, *Regulation of the hypothalamic-pituitary-adrenal axis.* Neuroimmunomodulation, 2009. **16**(5): p. 265-271.
48. Philips, A., et al., *Novel dimeric Nur77 signaling mechanisms in endocrine and lymphoid cells.* Mol. Cell. Biol, 1997. **17**(10): p. 5946-5951.
49. Philips, A., et al., *Antagonism between Nur77 and glucocorticoid receptor for control of transcription.* Mol. Cell. Biol, 1997. **17**(10): p. 5952-5959.
50. Pulichino, A.M., et al., *Human and mouse Tpit gene mutations cause early onset pituitary ACTH deficiency.* Genes Dev, 2003. **17**(6): p. 711-716.
51. Lavoie, P.L., et al., *Developmental Dependence on NurRE and EboxNeuro for Expression of Pituitary POMC.* Mol. Endocrinol, 2008. **22**(7): p. 1647-1657.
52. Lamolet, B., et al., *Tpit-independent function of NeuroD1(BETA2) in pituitary corticotroph differentiation.* Mol. Endocrinol, 2004. **18**(4): p. 995-1003.
53. Zhang, F., et al., *Enhancer-bound LDB1 regulates a corticotrope promoter-pausing repression program.* Proc Natl Acad Sci U S A, 2015. **112**(5): p. 1380-5.
54. Kato, Y., et al., *Regulation of human growth hormone secretion and its disorders.* Intern Med, 2002. **41**(1): p. 7-13.
55. Mangalam, H.J., et al., *A pituitary POU domain protein, Pit-1, activates both growth hormone and prolactin promoters transcriptionally.* Genes Dev, 1989. **3**: p. 946-958.
56. Tatsumi, K., et al., *Cretinism with combined hormone deficiency caused by a mutation in the PIT1 gene.* Nat. Genet, 1992. **1**(1): p. 56-58.
57. Nica, G., et al., *Zebrafish pit1 mutants lack three pituitary cell types and develop severe dwarfism.* Mol. Endocrinol, 2004. **18**(5): p. 1196-1209.
58. Li, S., et al., *Dwarf locus mutants lacking three pituitary cell types result from mutations in the POU-domain gene pit-1.* Nature, 1990. **347**: p. 528-533.
59. Sornson, M.W., et al., *Pituitary lineage determination by the Prophet of Pit-1 homeodomain factor defective in Ames dwarfism.* Nature, 1996. **384**(6607): p. 327-333.
60. Wu, W., et al., *Mutations in PROP1 cause familial combined pituitary hormone deficiency.* Nat. Genet, 1998. **18**(2): p. 147-149.
61. Cogan, J.D., et al., *The PROP1 2-base pair deletion is a common cause of combined pituitary hormone deficiency.* J. Clin. Endocrinol. Metab, 1998. **83**(9): p. 3346-3349.
62. Zhu, X., et al., *Sustained Notch signaling in progenitors is required for sequential emergence of distinct cell lineages during organogenesis.* Genes Dev, 2006. **20**(19): p. 2739-53.
63. Hinuma, S., et al., *A prolactin-releasing peptide in the brain.* Nature, 1998. **393**(6682): p. 272-6.
64. Saiardi, A., et al., *Antiproliferative role of dopamine: loss of D2 receptors causes hormonal dysfunction and pituitary hyperplasia.* Neuron, 1997. **19**(1): p. 115-126.
65. Nowakowski, B.E. and R.A. Maurer, *Multiple Pit-1-binding sites facilitate estrogen responsiveness of the prolactin gene.* Mol. Endocrinol, 1994. **8**(12): p. 1742-1749.
66. Scully, K.M., et al., *Role of estrogen receptor-alpha in the anterior pituitary gland.* Mol. Endocrinol, 1997. **11**(6): p. 674-681.

67. Lin, S.C., et al., *Pituitary ontogeny of the Snell dwarf mouse reveals Pit-1-independent and Pit-1-dependent origins of the thyrotrope.* Development, 1994. **120**(3): p. 515-22.
68. Charles, M.A., et al., *Pituitary-specific Gata2 knockout: effects on gonadotrope and thyrotrope function.* Mol. Endocrinol, 2006. **20**(6): p. 1366-1377.
69. Bliss, S.P., et al., *GnRH signaling, the gonadotrope and endocrine control of fertility.* Front Neuroendocrinol, 2010. **31**(3): p. 322-40.
70. Ingraham, H.A., et al., *The nuclear receptor steroidogenic factor 1 acts at multiple levels of the reproductive axis.* Genes Dev, 1994. **8**: p. 2302-2312.
71. Zhao, L., et al., *Steroidogenic factor 1 (SF1) is essential for pituitary gonadotrope function.* Development, 2001. **128**(2): p. 147-54.
72. Luo, X., Y. Ikeda, and K.L. Parker, *A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation.* Cell, 1994. **77**(4): p. 481-490.
73. Ikeda, Y., et al., *The nuclear receptor steroidogenic factor 1 is essential for the formation of the ventromedial hypothalamic nucleus.* Mol. Endocrinol, 1995. **9**(4): p. 478-486.
74. Szeto, D.P., et al., *P-OTX: a PIT-1-interacting homeodomain factor expressed during anterior pituitary gland development.* Proc. Natl. Acad. Sci. USA, 1996. **93**(15): p. 7706-7710.
75. Lamonerie, T., et al., *PTX1, a bicoid-related homeobox transcription factor involved in transcription of pro-opiomelanocortin (POMC) gene.* Genes Dev, 1996. **10**(10): p. 1284-1295.
76. Budry, L., et al., *The Ets Factor Etv1 Interacts with Tpit Protein for Pituitary Pro-opiomelanocortin (POMC) Gene Transcription.* J Biol Chem, 2011. **286**(28): p. 25387-25396.
77. Khetchoumian, K., et al., *Cell-autonomous transcriptional mechanism for enhancement of translation capacity in secretory cells* BioRxiv, 2018.
78. Fauquier, T., et al., *SOX2-expressing progenitor cells generate all of the major cell types in the adult mouse pituitary gland.* Proc. Natl. Acad. Sci. USA, 2008. **105**(8): p. 2907-2912.
79. Rizzoti, K., H. Akiyama, and R. Lovell-Badge, *Mobilized adult pituitary stem cells contribute to endocrine regeneration in response to physiological demand.* Cell Stem Cell, 2013. **13**(4): p. 419-32.
80. 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
81. Kemppainen, R.J. and E.N. Behrend, *Dexamethasone rapidly induces a novel ras superfamily member-related gene in AtT-20 cells.* J Biol Chem, 1998. **273**(6): p. 3129-31.
82. Binder, E.B., *The role of FKBP5, a co-chaperone of the glucocorticoid receptor in the pathogenesis and therapy of affective and anxiety disorders.* Psychoneuroendocrinology, 2009. **34 Suppl 1**: p. S186-95.
83. Soufi, A., et al., *Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming.* Cell, 2015. **161**(3): p. 555-68.
84. Larson, A.G., et al., *Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin.* Nature, 2017. **547**(7662): p. 236-240.
85. Mohammed, H., et al., *Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor.* Cell Rep, 2013. **3**(2): p. 342-9.
86. Bostick, M., et al., *UHRF1 plays a role in maintaining DNA methylation in mammalian cells.* Science, 2007. **317**(5845): p. 1760-4.
87. Katoh, M., et al., *Cancer genetics and genomics of human FOX family genes.* Cancer Lett, 2013. **328**(2): p. 198-206.

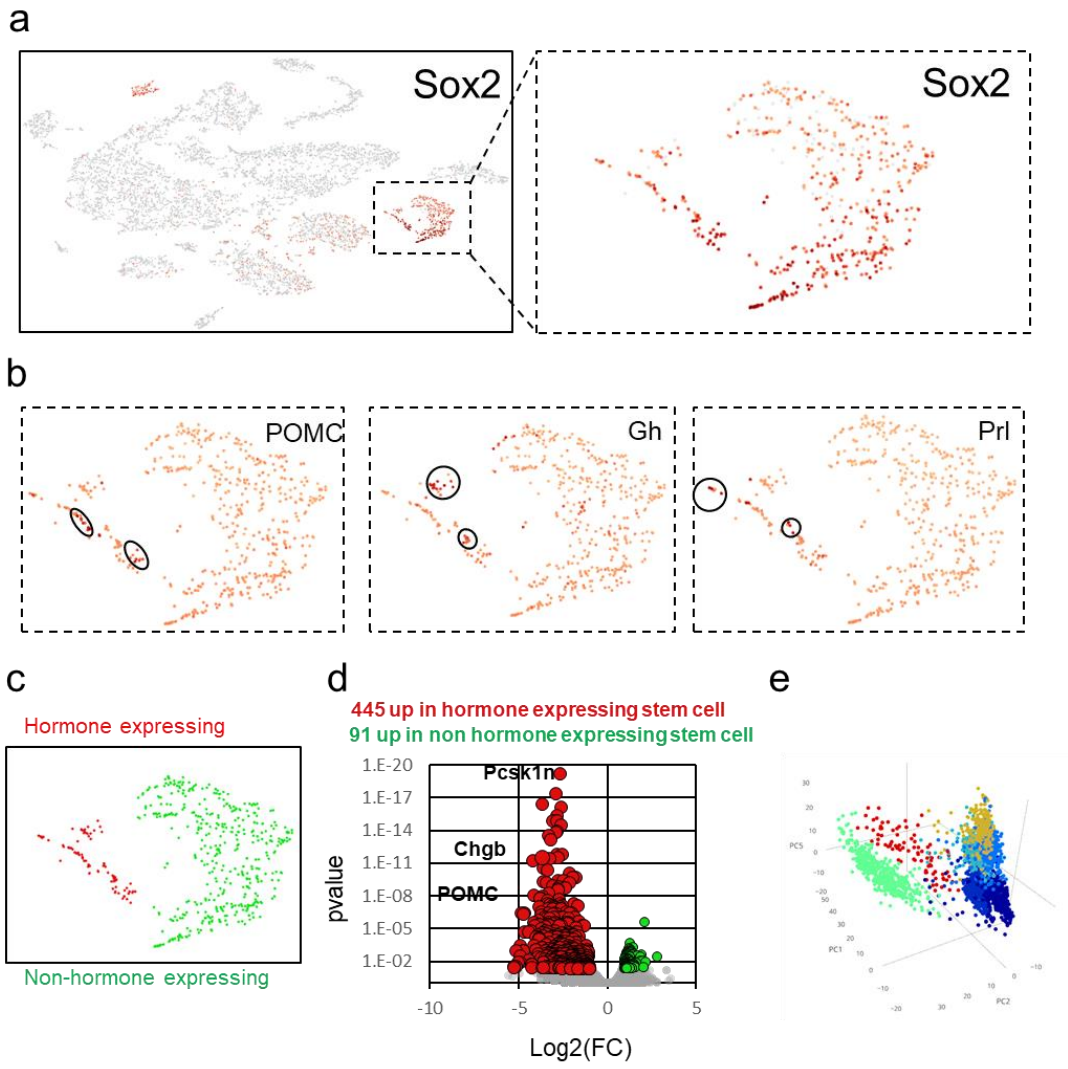
88. Soufi, A., G. Donahue, and K.S. Zaret, *Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome.* Cell, 2012. **151**(5): p. 994-1004.

Appendix



Appendix 1: Gldc, Fndc3c1 and Gpr50 are not expressed in melanotropes.

t-SNE map showing the expression of the indicated genes from scRNAseq



Appendix 2: Pituitary stem cells transitioning towards differentiation from scRNAseq.

(a) t-SNE map showing the expression of Sox2 in the stem cell population
(b) t-SNE map showing the expression of POMC, GH Prl in a subset of the stem cell population
(c) Subclustering of the stem cell population into hormone expressing (red) and non-hormone expressing (blue)
(d) Volcano plot showing differential gene expression (Log2FC) over pvalue in the hormone-expressing versus the non-hormone expressing pituitary stem cell niche
(e) 3D principal component analysis showing component 1, 2 and 5 for the differentiated hormone expressing cell in blue and yellow while the stem cell

population is in green and the stem cell population that express hormones is in red

Appendix 3: The pentadactyl state relies on a pioneer activity of the HOXA13 and HOXD13 transcription factors (Desanlis et al.)

The pentadactyl state relies on a pioneer activity of the HOXA13 and HOXD13 transcription factors

Ines Desanlis[1,#], Yacine Kherdjemil[1,@,#], Alexandre Mayran[2,3,#] , Claudia Gentile[1], Rushikesh Sheth[4], Rolf Zeller[4], Jacques Drouin[2,3] and Marie Kmita[1,5,6*]

[1]*Genetics and Development Research Unit, Institut de Recherches Cliniques de Montréal, Montreal, Quebec, Canada.*

[2]*Department of Biochemistry, McGill University, Montreal, Quebec, Canada.*

[3]*Molecular Genetics Research Unit, Institut de Recherches Cliniques de Montréal, Montreal, Quebec, Canada.*

[4]*Developmental Genetics, Department Biomedicine, University of Basel, 4058 Basel, Switzerland*

[5]*Département de Médecine (Programme de Biologie Moléculaire), Université de Montréal, Montreal, Quebec, Canada*

[6]*Department of Experimental Medicine, McGill University, Montreal, Quebec, Canada*

[#]Equal contributions

* Corresponding author: Marie Kmita

Email: Marie.Kmita@ircm.qc.ca

Tel: (1) 514-987-5749

@ Present address: EMBL Heidelberg, Meyerhofstrasse 1, D-69117 Heidelberg, Germany.

AUTHOR CONTRIBUTIONS

Y.K., A.M. and M.K. conceived the study. I.D., Y.K., A.M., designed and conducted the experiments. C.G. contributed to figures 3 and 4b. RS performed the experiment for figure 4d. I.D., Y.K., A.M., M.K. analyzed and interpreted the data. M.K. wrote the paper together with A.M., I.D., and Y.K. All authors commented on the manuscript.

ABSTRACT

Pioneer factors are transcription factors that are able to recognize their target sites even when concealed in "closed" chromatin, eventually eliciting the switch to an accessible chromatin state, permissive to other transcription factors (TFs) and the transcriptional machinery[1,2]. As such, pioneer factors play a key role in switching cell fate. Here, we provide evidence that HOXA13 and HOXD13 (HOX13 hereafter), two transcription factors of the Hox family of developmental genes, act as pioneer factors in the developing limb. We show that Hox13 function is mandatory for switching a series of target loci to an accessible chromatin state, allowing the binding of other transcription factors. These target loci include an enhancer previously identified as essential for the pentadactyl state, providing evidence that the pioneer activity of HOX13 is key for digit patterning. Based on the data reported here and previous studies[3,4], we propose that during the fin-to-limb transition, the implementation of digit-specific enhancer elements required an ancestral pioneer function of the HOX13 TFs.

MAIN TEXT

The *Hox* family of developmental genes, which encode transcription factors (TFs), have pivotal roles in organogenesis and patterning. Their differential expression along the main body axis establishes positional information that instructs cells about their fate, ultimately generating morphological diversity[5-7]. How the various HOX TFs activate distinct genetic programs has thus become of major interest, notably because diverse mutations as well as improper expression are often associated with severe morphological defects. Moreover, increasing evidence suggest that changes in *Hox* regulation contributed to morphological adaptation/diversity during evolution. Intriguingly, the various HOX TFs appear to have similar DNA-binding properties[8]. Several studies uncovered the relevance of co-factors, notably the PBX/Exd family of TFs, in increasing DNA-binding specificity. Yet, analyses based on primary mesenchymal limb bud cells transfected with the subset of Hox proteins known for their patterning function in developing limbs, revealed that these TFs cluster into only two subgroups based on their genome-wide binding[9]. Strikingly, one of the two subgroups is composed of HOXA11, HOXD11 and HOXD13 despite the fact that HOXA11/HOXD11 are required for the zeugopod (forearm) developmental program[10] and HOXD13, together with HOXA13, is mandatory for digit development[11] (Fig. 1a). This finding could explain how *Hoxa11* is able to trigger the formation of extra-digits when ectopically expressed in distal limb[3]. In turn, it raises the possibility that *Hoxa11* functional specificity varies depending on the environmental context, i.e. whether its expression is restricted to the presumptive zeugopod (as in wild type embryos) or ectopically located in distal limbs. Yet, in absence of *Hox13* function (*Hoxa13-/-;Hoxd13-/-*), even though *Hoxa11* becomes ectopically expressed in distal cells, it fails to trigger digit formation[4, 11,12]. While this suggests that *Hoxa11*-dependent formation of extra-digits requires *Hox13* function, it raises the possibility that HOX13 TFs, directly or indirectly, impact HOXA11 binding specificity.

We first analyzed the genome-wide binding of HOXA11 in wild type limb buds at E11.5 by performing Chromatin immunoprecipitation followed by high throughput sequencing (ChIP-seq) for HOXA11 and compared the data to the ChIP-seq data previously reported for HOXA13[4] (Fig.1; supplementary Fig. 1). We found the vast majority of HOXA11 ChIP-seq peaks are located at genomic regions distinct from transcriptional start site (TSS; Fig. 1b), reminiscent of the HOX13 genome-wide binding in distal limbs[4]. A significant proportion (65%) of HOXA11-bound loci in proximal limb, are also occupied by HOXA13 in distal limbs, signifying that only 35% of the HOXA11 targets (6333 targets) are not "shared" with HOXA13 (Fig. 1c). Similarly, 43% of the HOXA13-bound loci (9179 targets) are specific to HOXA13 while 57% are bound by HOXA11 in the proximal limb (Fig. 1c). GREAT[13] analysis revealed that common and specific sites are associated with limb development, consistent with the function of both TFs (Fig. 1d). Interestingly, *de novo* motif search uncovered distinct motifs for common *versus* specific targets (Fig. 1e, f). These HOXA13 specific sites are also occupied by HOXD13 suggesting that this binding specificity is shared between the HOX13 paralogs (Fig. 1g). As HOXA11 and HOXA13 are expressed in distinct cells in the developing limb, we next tested whether the cellular environment could contribute to HOXA11 and HOXA13 binding specificity. To this aim, we analyzed the HOXA11 binding profile in mutant limb buds in which *Hoxa11* is ectopically expressed in distal limbs[3] (*Prrx1:Cre;Rosa26*$^{Hoxa11/Hoxa11}$; referred to as *RA11KI* hereafter, Fig. 2a). Compared to the wild type HOXA11 binding profile (*Hoxa11* expression restricted to the presumptive zeugopod), we observed ectopic binding of HOXA11 in *RA11KI* limbs (Fig. 2b, c). Interestingly, distally expressed HOXA11 binds to loci specific to HOXA13 in the wild type context (Fig. 2d), which could explain how distal expression of HOXA11 triggers the formation of extra-digits in the *RA11KI* limbs[3] while it is unable to compensate for the absence of HOX13 in *Hox13-/-* limbs[4,11]. Interestingly, HOXA11-bound loci that are specific to the *RA11KI* limb buds are also bound by HOXA13 and enriched for the DNA motif associated with HOXA13 specific sites (supplementary Fig. 2a, b; Fig. 1f). This indicates that the binding specificity of HOXA11 and HOXA13 does not rely on distinct DNA motifs. Rather,

our findings suggest that the primary driver for the distinct HOXA11 and HOXA13 binding patterns in wild type limb is cell type dependent.

Pioneer factors have emerged as key TFs for modifying cell fate by triggering chromatin accessibility at target loci, allowing for the binding of other TFs and the transcriptional machinery[1]. Previous analysis of the impact of HOX13 on gene regulation revealed that the transition from the early to the distal/late limb developmental program relies on HOX13 function and it was proposed that this switch in developmental program could be mediated by a pioneer activity of the HOX13 TFs[4]. We thus hypothesized that ectopic HOXA11 binding at loci normally bound specifically by HOX13 could be associated with a pioneer activity of HOX13. To test this hypothesis, we first performed ATAC-seq (Assay for Transposase Accessible Chromatin followed by high-throughput sequencing[14]) on wild type and *Hox13-/-* distal limb buds (Fig. 3a; supplementary Fig. 3a). We identified changes in chromatin accessibility (4573 increases and 3034 decreases) (Fig. 3b), a subset of which coincides with loci bound by HOXA13 in wild type distal limb. Importantly, those are restricted to loci losing accessibility in *Hox13-/-,* suggesting that chromatin accessibility at a subset of HOX13 targets directly relies on HOX13 (Fig. 3c). In contrast, genomic regions with increased accessibility in *Hox13-/-* distal limbs are distinct from the wild type HOX13-bound loci, excluding a direct contribution of HOX13 in the establishment of closed chromatin states (Fig. 3c). We then analyzed the binding specificity of HOXA11 and HOX13 at these sites. While the binding of HOXA11 and HOXA13 correlate (R^2=0.65) at sites with unchanged accessibility, HOX13-dependent accessibility shows an unambiguous preference for HOXA13-bound loci and a low correlation with HOXA11 (R^2=0.23) (Fig. 3d). However, we found that distal *Hoxa11* expression (in *RA11KI* limbs) leads to ectopic HOXA11 binding compared to wild type (Fig. 3e; supplementary Fig. 3b). This suggests that, in *RA11KI* limbs, in which *Hoxa11* is expressed in the *Hox13* domain, HOXA11 becomes capable of binding genomic regions made accessible by HOX13.

Hox13 was previously shown to bind the regulatory landscape controlling *HoxD* expression[4,15]. We thus looked at chromatin accessibility at this locus. Consistent with the centromeric regulatory landscape (C-DOM) controlling *HoxD* expression in distal limb and the telomeric landscape (T-DOM) triggering *HoxD* expression in the proximal limb[16], we found that chromatin accessibility is significantly higher within the T-DOM in proximal limb cells and within the C-DOM in distal cells (Fig. 3f). Interestingly, these differences in accessibility are dissolved in *Hox13-/-* distal limb, with increased accessibility at wild type HOXA11 targets and decreased accessibility at HOX13 targets (Fig. 3f). These results suggest that HOX13 binding within the C-DOM promotes the accessibility of the distal limb enhancers while the distal gain of *Hoxa11* expression in *Hox13-/-* limbs results in increased accessibility within the T-DOM regulatory landscape. Based on the existence of loci bound by HOX13 TFs which require HOX13 function to become accessible, as well as the evidence that the HOX13-dependent chromatin accessibility allows for the binding of at least another TF, we propose that HOX13 TFs behave as pioneer factors during limb development.

Interestingly, the enhancer that we previously identified, which drives *Hoxa11* antisense transcription and consequently blocks *Hoxa11* expression in distal limbs[3], belongs to the subset of HOX13 targets identified in this study, whose chromatin accessibility relies on HOX13 function (Fig. 4). The function of this enhancer is mandatory for the pentadactyl state (five digits), and as such was proposed to be a key component in the mechanism underlying the transition from polydactyly (more than 5 digits) in ancestral tetrapods to the pentadactyl state of modern tetrapods[3]. Consistent with previous results, analysis of RNA-seq data in wild type and *Hox13-/-* distal limbs shows the specific loss of the antisense transcript overlapping with *Hoxa11* exon 1 (Fig. 4a). In both wild type proximal limb and *Hox13-/-* distal limb, this enhancer is in an inaccessible chromatin state (Fig. 4b, middle tracks) indicating that its activity requires HOX13 function. Interestingly, we found that this enhancer is one of the loci ectopically bound by HOXA11 in *RA11KI* limbs (Fig. 4b, bottom track). These data thus suggest that HOX13 TFs are mandatory for the chromatin accessibility of this enhancer and in turn, allows

for HOXA11 binding when HOXA11 is expressed distally. To assess whether chromatin accessibility at this enhancer is also dependent on HOX13 TFs in other tetrapod species, ATAC-seq data from early and distal late wing buds from chick embryos were compared. This comparison was used as a proxy for distal specific chromatin opening, as *Hox13* expression occurs exclusively in distal limb at the stage the presumptive autopod is forming. ATAC-seq data from chick reveal that the enhancer located in the *Hoxa11* intron is in a closed chromatin state at early stage and becomes subsequently accessible in the presumptive autopod, where *Hox13* is expressed (Fig. 4d). Although these ATAC-seq data do not provide a direct proof of HOX13 pioneer activity in chick, based on the data obtained from wild type and *Hox13-/-* mouse distal limb buds, it is reasonable to assume that the HOX13-dependent accessibility of the enhancer driving *Hoxa11* antisense transcription is conserved among tetrapods. Interestingly, transgenic *zebrafish* carrying the mouse enhancer for *Hoxa11* antisense transcription, provided evidence that this mouse enhancer is activated in the *hox13* domain of the transgenic fin[3]. Therefore, it is likely that the pioneer activity uncovered for the mouse HOX13 TFs is actually conserved in all vertebrates.

In summary, our genomic analyses of the HOXA13 and HOXD13 reveal that these transcription factors act during limb development as pioneer factors. Whether this functional characteristic is specific to HOX13 or a common feature of the HOX family remains to be investigated. Importantly, the finding that the HOX13 pioneer function is required for the activity of the digit-specific enhancer, whose function is mandatory for the pentadactyl state of modern tetrapods indicate that HOX13 pioneer function was likely instrumental during the fin-to-limb transition. Finally, the data presented in this work suggest that HOX13 pioneer function is conserved in tetrapods and most likely in all vertebrates. Based on these results and previous reports[3,4], we propose that, during the fin-to-limb transition, the implementation of digit-specific enhancers (at least a subset of them) required an ancestral pioneer function of the HOX13 TFs.

ACKNOWLEDGEMENTS

We are particularly grateful to lab members for insightful discussions and sharing reagents. This work was supported by the Canadian Institute for Health Research (MOP-115127) to M.K. Bioinformatics analyses were enabled in part by support provided by Calcul Quebec (www.calculquebec.ca) and Compute Canada (www.computecanada.ca). Y.K. was supported by a fellowship from the Molecular Biology program of the Université de Montréal and the IRCM fellowship Michel-Bélanger. A.M. was supported by the IRCM challenge fellowship, C.G. was supported by the Jacques Gauthier IRCM fellowship.

AUTHOR INFORMATION

The authors declare no competing financial interests.
Correspondence and requests for materials should be addressed to: marie.kmita@ircm.qc.ca

METHODS (ONLINE-ONLY)

Mouse lines

Hoxa13null (Hoxa13^{Str}), *Hoxd13null (Hoxd13^{lacZ}), Rosa^{Hoxa11}* mouse lines were previously described[3,11,18]. All mice were maintained in mixed background (C57BL/6 X 129). Noon of the day of the vaginal plug was considered as E0.5. Mice and embryos were genotyped by PCR using genomic DNA extracted from tail biopsy specimens and yolk sacs, respectively. Mice work at the Institut de Recherches Cliniques de Montréal (IRCM) was reviewed and approved by the IRCM animal care committee (protocols 2015-14 and 2017-10).

Chromatin Immunoprecipitation and Sequencing

HOXA11 ChIP was performed in forelimb buds of CD1 (wild type) and *Prrx1Cre; Rosa$^{Hoxa11/Hoxa11}$ (RA11KI)* mice at E11.5 in the same conditions as previously described for HOX13 ChIP[4]. Chromatin was cross-linked using a combination of

disuccinimidyl glutarate (DSG) and formaldehyde and sonicated using Fisher Scientific, Model 100 sonic dismembrator to obtain fragments between 100-600 bp. Protein A and Protein G Dynabeads (Invitrogen) were incubated for 6 hours at 4°C with 5ug HOXA11 (SAB1304728, Sigma) antibody. The chromatin was coupled to the beads overnight at 4°C. The immunoprecipitated samples were then sequentially washed in low salt (1% Triton, 0,1% SDS, 150 mM NaCl, 20 mM Tris (pH8), 2 mM EDTA), high salt (1% Triton, 0,1% SDS, 500 mM NaCl, 20 mM Tris pH8, 2 mM EDTA), LiCl (1% NP-40, 250 mM LiCl, 10 mM Tris (pH8), 1 mM EDTA) and TE buffer (50 mM NaCl, 10 mM Tris (pH8), 1 mM EDTA). The DNA was then purified on QIAquick columns (Qiagen). Library and flow cells were prepared by the IRCM Molecular Biology Core Facility according to Illumina's recommendations and sequenced on Illumina Hiseq 2500 in a 50 cycles paired-end configuration.

ATAC-seq

Dissection of proximal and distal of forelimb buds from wild type embryos and distal forelimb buds of *Hox13*$^{-/-}$ embryos[4] were performed at E11.5. All samples for ATAC-seq were processed as previously described[2]. Briefly, 50 000 cells were washed in PBS and incubated on ice for 30 minutes in a hypotonic cell lysis buffer (0.1% $^{w}/_{v}$ sodium citrate tribasic dehydrate and 0.1% $^{v}/_{v}$ Triton X100) centrifugated (5 minutes at 2000g at 4°C). Cells were then incubated 30 minutes on ice in cell lysis buffer (10mM Tris-HCl, pH7.4, 10mM NaCl, 3mM $MgCl_2$, 0.1% $^{v}/_{v}$ IGEPAL CA-630. After centrifugation (5 minutes at 2000g at 4°C) the resulting pellet of nuclei was re-suspended in transposase Master Mix (1.25 µl 10x TD buffer, 5 µl H_2O and 6.5 µl of Tn5: Illumina Nextera Kit; FC-121-1031) and incubated for 30 minutes at 37°C. Samples were purified using MinElute PCR purification column (Qiagen). The eluted DNA was enriched and barcoded for multiplexing of samples using Nextera barcodes by PCR using Phusion kit. The library was recovered with GeneRead Purification columns. Samples were then evaluated by qPCR to test for the enrichment of open regions and sequenced on Illumina Hiseq 2500 with 50bp or paired end reads, according to Illumina's recommendation.

RNA Preparation and Sequencing

Dissections of proximal and distal forelimb buds were performed at E11.5 as described above. The dissected limb buds were stored at -80 in Qiagen RNAlater until genotyping was performed. RNA was extracted from two independent embryos and performed in biological duplicate using RNAeasy Plus mini kit (Qiagen 74134). mRNA enrichment, library preparation and flow-cell preparation for sequencing were performed by the IRCM Molecular Biology Core Facility according to Illumina's recommendations. Sequencing was done on a HiSeq 2500 instrument with a paired-end 50 cycles protocol.

ChIP-seq and ATAC-seq Data analysis

ChIP-seq and ATAC–seq reads were aligned to the mm10 genome using bowtie v.2.3.1 with the following settings: bowtie2 -p 8 --fr --no-mixed --no-unal –x. Sam files were converted into tag directories using HOMER v4.9.1 and into bam files using Samtools v1.4.1 view function. Peaks were identified by comparing each sample to its input using MACS v2.1.1.20160309 callpeak function using the parameters: --bw 250 -g mm --mfold 10 30 -p 1e-5. Peaks with an associated pvalue less than 10^{-5} were kept. Heatmaps and average profiles were generated using the Easeq software[19]. ChIP-seq and ATAC-seq data were visualized on the IGV software[20] using BigWig files generated using the makeUCSCfile HOMER command. For ATACseq differential analysis, peaks from ATAC datasets were merged using HOMER v4.9.1 mergePeaks tool to obtain a file with all the unique position from all the ATAC-seq datasets. ATAC-seq signals were quantified in these different datasets using the analyzeRepeats.pl HOMER command and differential accessibility analyses were performed using getDiffExpression.pl with default parameters, which uses Deseq2 to perform differential analysis. Peaks showing differential accessibility of more than two folds and a pvalue smaller than 0.05 were considered differentially accessible.

Motifs analysis

The HOMER findMotifsGenome command was used to perform de novo analysis within 200-bp windows around the center of summit of peaks against background sequences generated by HOMER that matches GC content. In all cases the shown motifs were the top scoring motif based on p-value.

RNA-seq Data analysis

Strand specific paired-end reads were aligned to the mm10 reference genome using TopHat2[21] with the parameters --rg-library “L” --rg-platform “ILLUMINA” --rg-platform-unit “X” --rg-id “run#many” --no-novel-juncs --library-type fr-firststrand -p 12. The resulted Bam files were converted to tagDirectory using HOMER and BigWig were produced using the makeUCSCfile HOMER command.

Protein extraction and western-blot

Nuclear extracts were performed using pooled forelimb and hindlimb buds at E11.5 from wild type and *Hoxa11*-/- embryos. Western blot was performed using the anti-HOXA11 antibody (1:500) (SAB1304728, Sigma) and the anti-H3 antibody (1:3000) (Ab1791, Abcam) was used as loading control.

Accession numbers

The ChIP-seq data for HOX13 in distal limb and RNAseq for wild type and *Hox13*$^{-/-}$ distal limb buds[4] at E11.5 are available from the NCBI Gene Expression Omnibus repository under the accession numbers GSE81356. Chicken ChIPseq data of Hoxa13 and Hoxa11[9] were obtained from the accession numbers GSE86088.

Study approval

All mice experiments described in this article were approved by the Animal Care Committee of the Institut de Recherches Cliniques de Montréal (protocols # 2014-14 and 2017-10).

REFERENCES

1 Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* **25**, 2227-2241, doi:10.1101/gad.176826.111 (2011).

2 Mayran, A. *et al.* Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. *Nat Genet* **50**, 259-269, doi:10.1038/s41588-017-0035-2 (2018).

3 Kherdjemil, Y. *et al.* Evolution of Hoxa11 regulation in vertebrates is linked to the pentadactyl state. *Nature* **539**, 89-92, doi:10.1038/nature19813 (2016).

4 Sheth, R. *et al.* Distal Limb Patterning Requires Modulation of cis-Regulatory Activities by HOX13. *Cell Rep* **17**, 2913-2926, doi:10.1016/j.celrep.2016.11.039 (2016).

5 Kmita, M. & Duboule, D. Organizing Axes in Time and Space; 25 Years of Colinear Tinkering. *Science* **301**, 331-333 (2003).

6 Krumlauf, R. Hox genes in vertebrate development. *Cell* **78**, 191-201, doi:https://doi.org/10.1016/0092-8674(94)90290-9 (1994).

7 Mallo, M., Wellik, D. M. & Deschamps, J. Hox genes and regional patterning of the vertebrate body plan. *Dev Biol* **344**, 7-15, doi:10.1016/j.ydbio.2010.04.024 (2010).

8 Saurin, A. J., Delfini, M. C., Maurel-Zaffran, C. & Graba, Y. The Generic Facet of Hox Protein Function. *Trends Genet*, doi:10.1016/j.tig.2018.08.006 (2018).

9 Jerkovic, I. *et al.* Genome-Wide Binding of Posterior HOXA/D Transcription Factors Reveals Subgrouping and Association with CTCF. *PLoS Genet* **13**, e1006567, doi:10.1371/journal.pgen.1006567 (2017).

10 Davis, P. D., Witte, P. D., Hsieh-Li, M. H., Potter, S. S. & Capecchi, R. M. Absence of radius and ulna in mice lacking hoxa-11 and hoxd-11. *Nature* **375** (1995).

11 Fromental-Ramain, C. *et al.* Hoxa-13 and Hoxd-13 play a crucial role in the patterning of the limb autopod. *Development* (1996).

12 Woltering, J. M., Noordermeer, D., Leleu, M. & Duboule, D. Conservation and divergence of regulatory strategies at Hox Loci and the origin of tetrapod digits. *PLoS Biol* **12**, e1001773, doi:10.1371/journal.pbio.1001773 (2014).

13 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501, doi:10.1038/nbt.1630 (2010).

14 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218, doi:10.1038/nmeth.2688 (2013).

15 Beccari, L. *et al.* A role for HOX13 proteins in the regulatory switch between TADs at the HoxD locus. *Genes Dev* **30**, 1172-1186, doi:10.1101/gad.281055.116 (2016).

16 Andrey, G. *et al.* A Switch Between Topological Domains Underlies *HoxD* Genes Collinearity in Mouse Limbs. *Science* **340**, doi:10.1126/science.1234167 (2013).

17 Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572.e524, doi:10.1016/j.cell.2017.09.043 (2017).

18 Kmita, M., Kondo, T. & Duboule, D. Targeted inversion of a polar silencer within the HoxD complex re-allocates domains of enhancer sharing. *Nature Genetics* **26**, 451, doi:10.1038/82593 (2000).

19 Lerdrup, M., Johansen, J. V., Agrawal-Singh, S. & Hansen, K. An interactive environment for agile analysis and visualization of ChIP-sequencing data. *Nat Struct Mol Biol* **23**, 349-357, doi:10.1038/nsmb.3180 (2016).

20 Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192, doi:10.1093/bib/bbs017 (2013).

21 Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**, R36, doi:10.1186/gb-2013-14-4-r36 (2013).

FIGURE

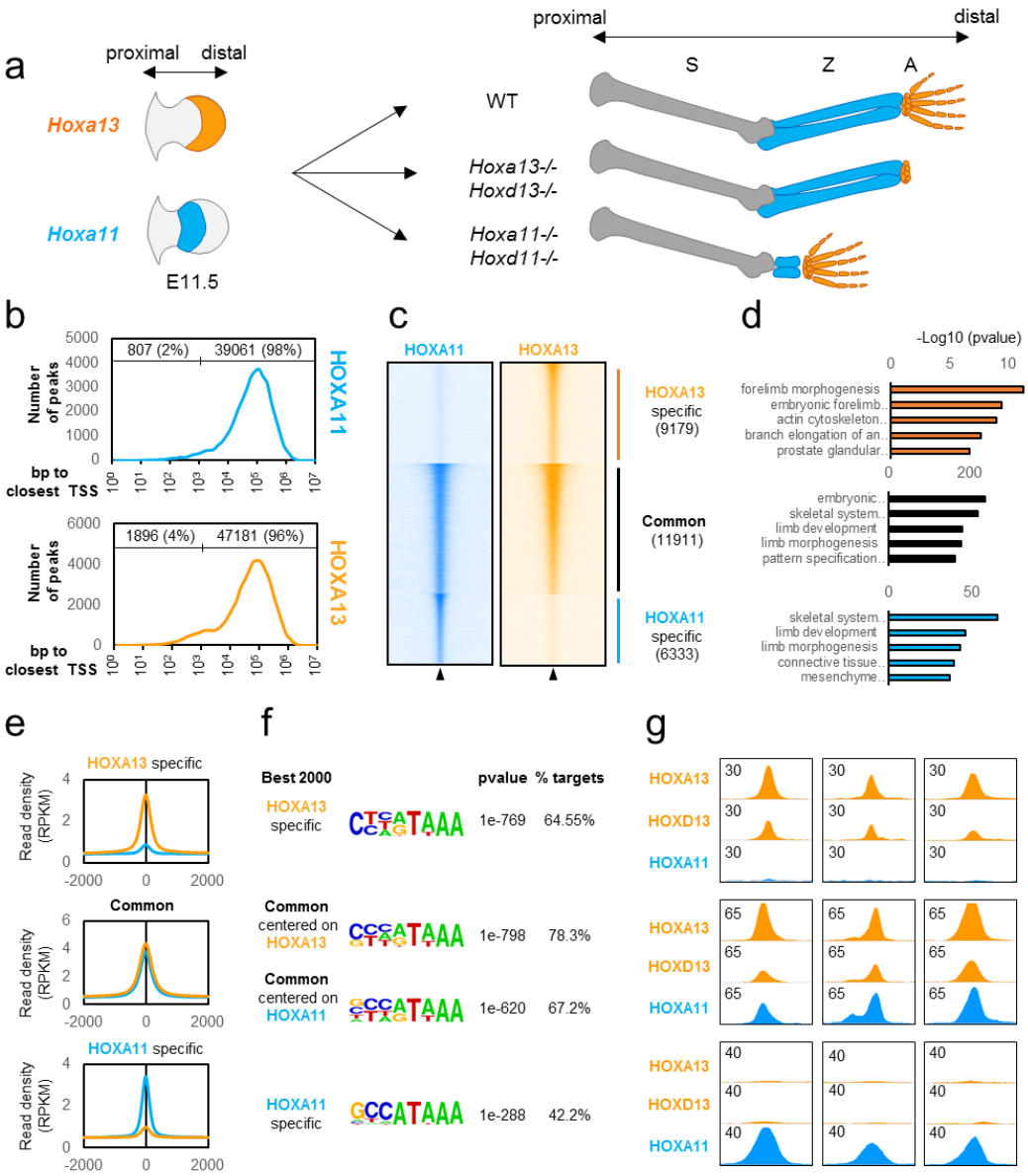


Figure 1. HOXA13 and HOXA11 bind specific limb cis-regulatory modules.

Figure 1. HOXA13 and HOXA11 bind specific limb cis-regulatory modules.

(a) Schematics illustrating the restricted expression of *Hoxa11* and *Hoxa13* in limb buds as well as the limb phenotype associated with the loss of *Hoxa/d11* and *Hoxa/d13*. (b) Distance of the HOXA11 (blue) and HOXA13 (orange) bound loci to the closest TSS of all HOXA11 and HOXA13 peaks (p-value < 10^{-5}) in E11.5 forelimb buds. (c) Heatmap showing ChIP-seq read density of HOXA11 (blue) and HOXA13 (orange) at all peaks for HOXA11 and HOXA13 with a p-value <10^{-20}, in a 4kb window. Peaks are ranked based on p-value and according to the specificity of binding. (d) GREAT[13] analysis of each category of HOXA11 and HOXA13 binding sites. The five top enriched biological process terms are shown. Binomial raw p-value as computed by GREAT are shown. (e) Average profile showing ChIP-seq signal (RPKM) of HOXA11 (blue) and HOXA13 (orange) at the three categories depicted in Fig. 1b. The x-axis represents the distance (bp) to peak summit. (f) Top scoring motif found by HOMER *de novo* motif analysis in a 200bp window around the 2000 best peaks of the three categories depicted in fig. 1b. Common HOXA11 and HOXA13 binding sites were analyzed using HOXA11 peak summit and HOXA13 peak summit. (g) Genome browser view (IGV) showing HOXA13, HOXD13 and HOXA11 binding (reads per million reads) at prototype loci of the three categories of HOXA11 and HOXA13 depicted in fig. 1b.

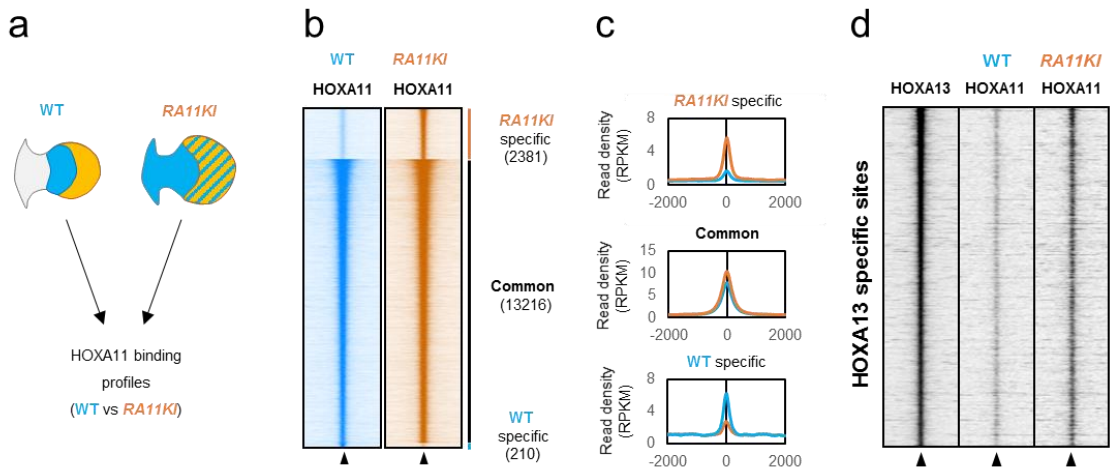


Figure 2. Loss of HOXA11-HOX13 specificity following ectopic *Hoxa11* expression.

Figure 2. Loss of HOXA11-HOX13 specificity following ectopic *Hoxa11* expression.

(a) Schematics showing the expression pattern of *Hoxa11* (blue) and *Hoxa13* (orange) in mouse limb bud from wild type and in *Prrx1Cre; Rosa*$^{Hoxa11/Hoxa11}$ mouse (*RA11KI)*. These were used to assess binding of HOXA11 after ectopic expression. (b) Heatmap showing ChIPseq read density of HOXA11 in wild type (blue) and in *RA11KI* (brown) from wild type and *RA11KI* E11.5 forelimb buds, at all HOXA11 peaks with a p-value $<10^{-20}$, in a 4kb window. Peaks are ranked based on p-value and according to binding specificity. (c) Average profile showing ChIP-seq signal (RPKM) for HOXA11 in wild type (blue) and *RA11KI* (brown) at the three categories depicted in Fig. 2b. The x-axis represents the distance (bp) to peak summit. (d) Heatmap showing ChIP-seq read density of HOXA13 in wild type and HOXA11 in wild type and *RA11KI* E11.5 forelimb buds at HOXA13 specific peaks with a p-value $<10^{-20}$, in a 4kb window. Peaks are ranked based on p-value of HOXA13 binding.

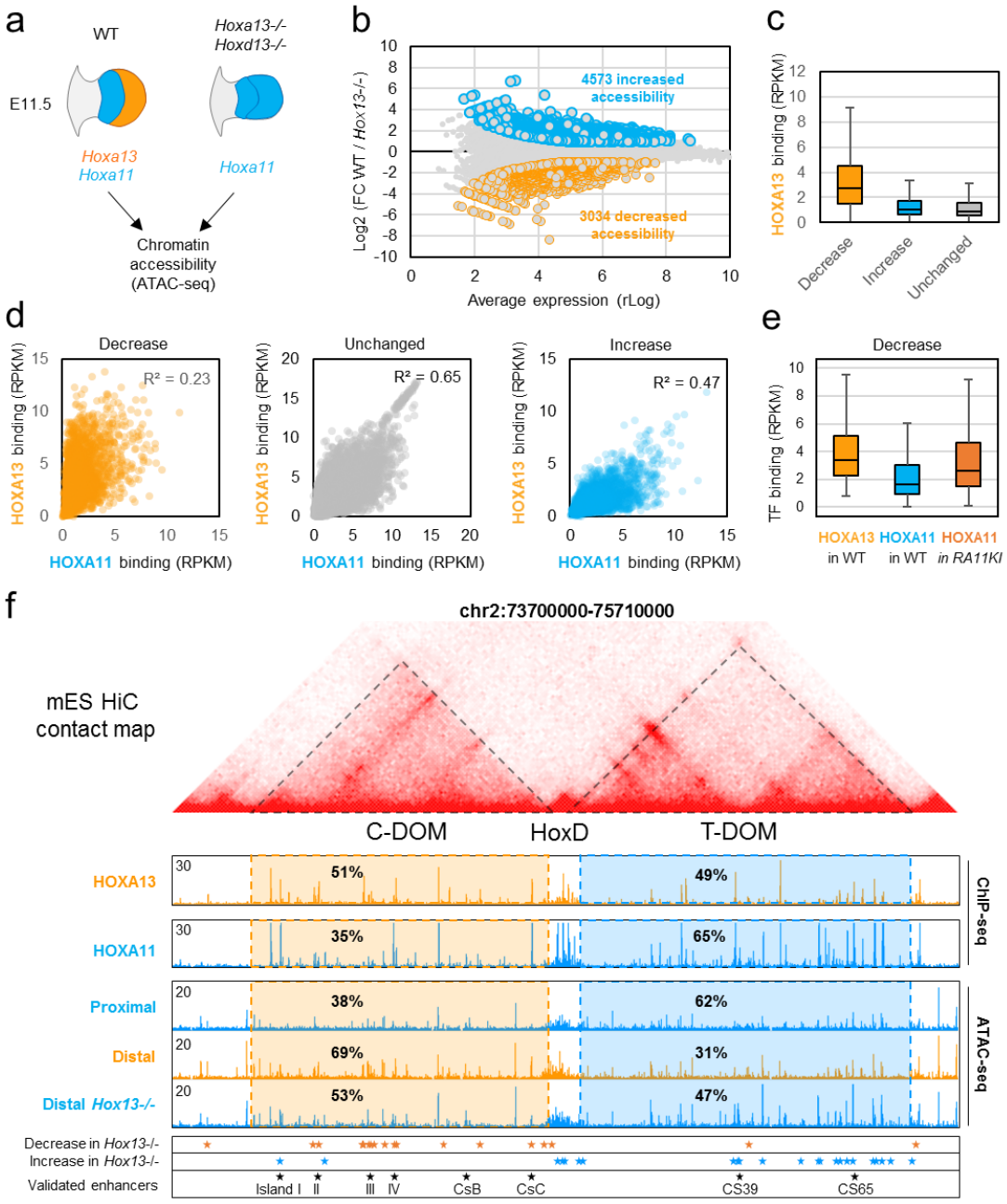


Figure 3. HOXA13 specific binding is linked to HOX13-dependent chromatin accessibility.

Figure 3. HOXA13 specific binding is linked to HOX13-dependent chromatin accessibility.

(a) Schematics illustrating *Hoxa11* and *Hoxa13* expression in wild type and *Hoxa13-/-*; *Hoxd13-/-* (*Hox13-/-* hereafter) E11.5 limb buds. These were used to assess HOX13 regulated accessibility by ATAC-seq. (b) Dispersion plot showing average expression (Deseq2 computed rLog, x-axis) over log2 fold change in wild type versus *Hox13*-/-limb buds (y-axis) at all ATAC-seq peaks. Differentially accessible ATAC-seq peaks (p-value <0.05, fold change > +/- 2) are shown in blue for KO-enriched and in red for wild type-enriched ATAC signals. (c) Box plot showing level of HOXA13 binding (RPKM) at the three categories of ATAC-seq peaks. Center lines show medians; box limits indicate the twenty-fifth and seventy-fifth percentiles; whiskers extend to 1.5 times the interquartile range from the twenty-fifth to seventy-fifth percentiles. Outliers are not shown. (d) Correlation plots showing the level (RPKM) of HOXA11 (x-axis) over HOXA13 (y-axis) at the three indicated categories of ATAC-seq peaks. Correlation coefficient were computed using excel. (e) Box plot showing HOXA13 and HoxA11 binding in wild type and *RA11KI* limb bud (RPKM) at the three categories of ATAC-seq peaks. Center lines show medians; box limits indicate the twenty-fifth and seventy-fifth percentiles; whiskers extend to 1.5 times the interquartile range from the twenty-fifth to seventy-fifth percentiles. Outliers are not shown. (f) HiC contact map (top) of mES cells[17] at the HoxD cluster and its surrounding two topological domains, C-DOM and T-DOM. Genome browser view (IGV, bottom) showing the intensity (reads per millions) of HOXA13 and HOXA11 ChIP-seq, and ATAC-seq signals in proximal and distal wild type limb bud as well as in *Hox13*-/- distal limb bud. The orange and blue boxes highlight the C-DOM and T-DOM landscapes, respectively. The percentage values correspond to the proportion of reads found at the peaks for each dataset in the indicated topological domains. Differentially accessible sites are shown with orange and blue stars for ATAC-seq peaks enriched in wild type or in *Hox13-/-* respectively. Validated limb enhancers are shown as a black star.

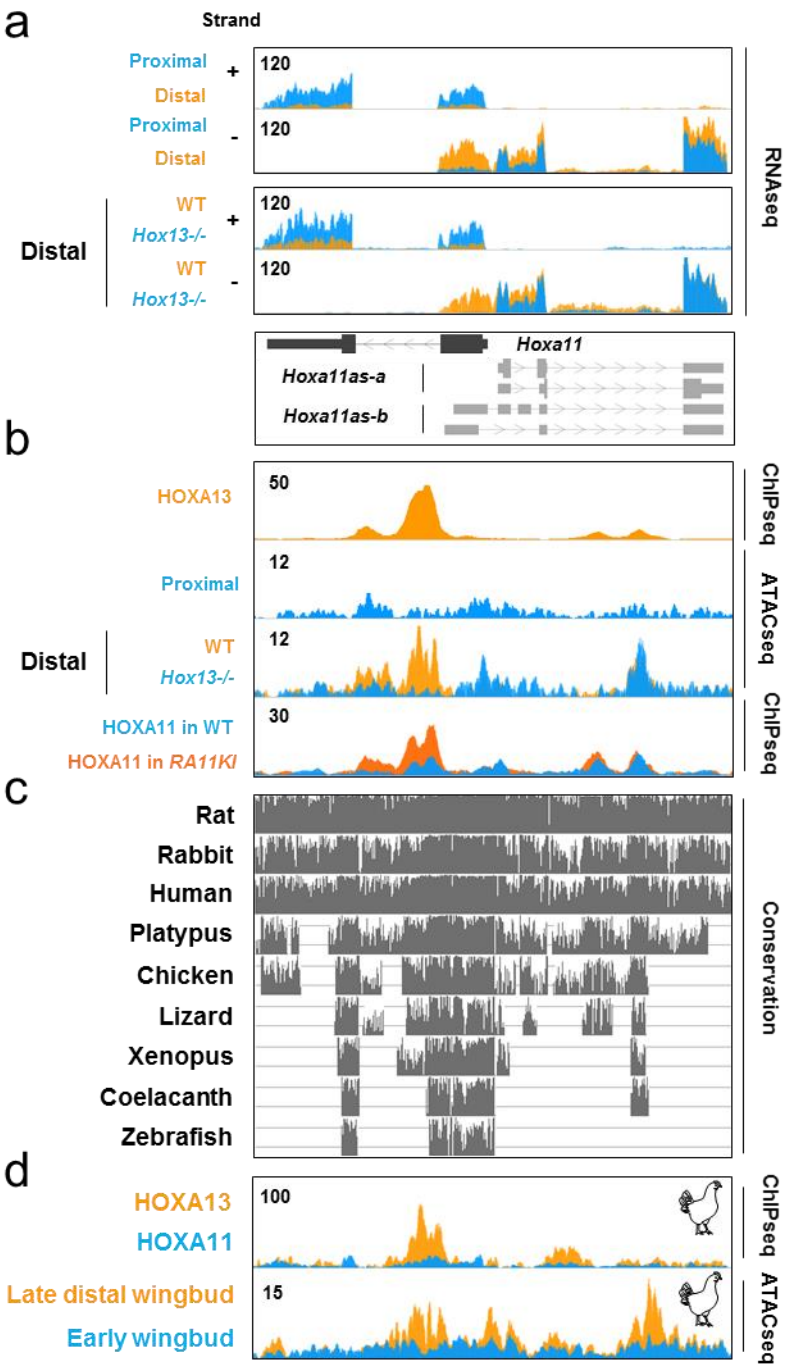
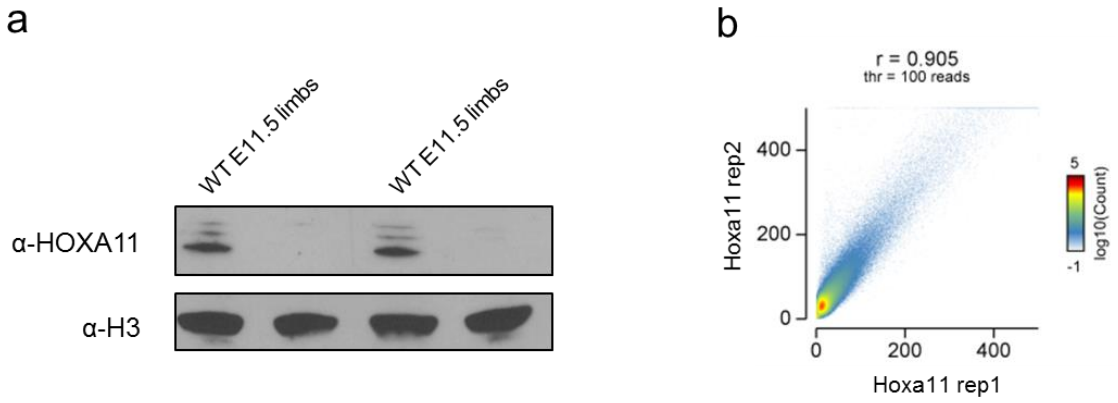


Figure 4. HOX13-dependent opening of the enhancer controlling *Hoxa11* antisense transcription in tetrapods.

Figure 4: HOX13-dependent opening of the enhancer controlling *Hoxa11* antisense transcription in tetrapods.

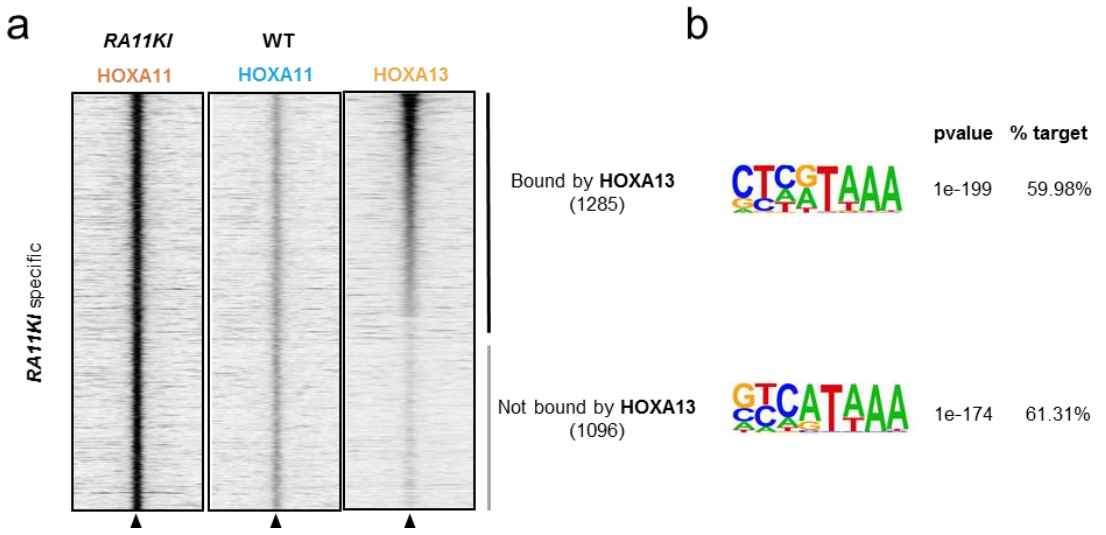
(a) Genome browser view (IGV) of overlaid strand specific RNA-seq at the *Hoxa11* locus in the indicated tissues showing the mutually exclusive expression of *Hoxa11* and the antisense transcripts overlapping with *Hoxa11* exon1. (b) Genome browser view (IGV) of the ChIP-seq and ATAC-seq data from mouse E11.5 limb bud at the *Hoxa11* locus. (c) UCSC tracks at the same genomic locus showing sequence conservation between mouse and both tetrapod and non-tetrapod species. (d) Genome browser view (IGV) of HoxA13 and HoxA11 ChIP-seq data[9], primary chicken mesenchymal limb progenitor cells) and ATAC-seq from chicken wing bud at early stage (HH20, prior to *Hox13* transcriptional activation) and late (HH26/27) distal wing bud (*Hox13*-expressing domain) at the chicken genomic region corresponding to the mouse coordinates in Fig. 4a.

SUPPLEMENTARY FIGURES



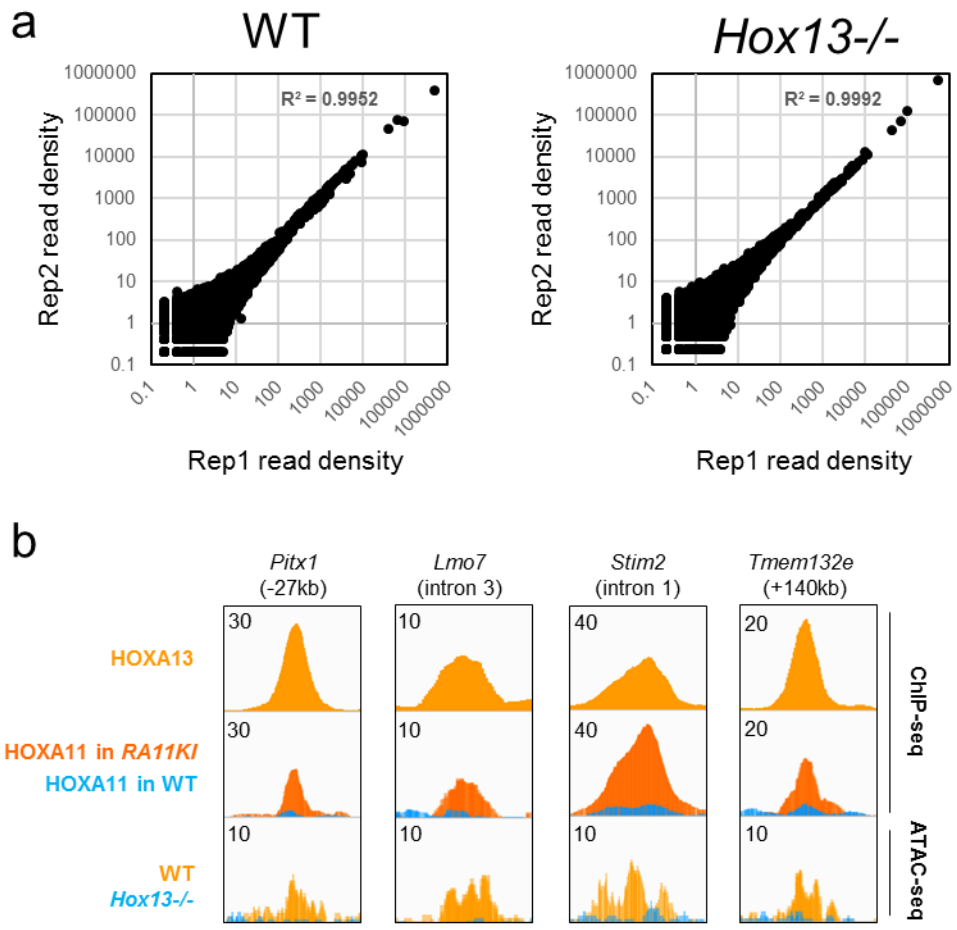
Supplementary figure 1. Validation of HOXA11 ChIPseq in E11.5 mouse limb bud.

(a) Western blot of nuclear extracts from E11.5 wild type and *Hoxa11*-/- limb buds. Two replicates are shown. Antibody against HOXA11 (top panel) and Histone H3 (bottom panel) were used. (b) Genome wide correlation of 500bp windows of the two HOXA11 ChIP-seq replicates signal. Pearson correlation is shown. Windows with more than 500 reads are shown as 501 reads.



Supplementary figure 2. Association between ectopic HOXA11 binding sites and HOXA13 specific sites.

(a) Heatmap showing ChIP-seq read density of HOXA11 in wild type and *RA11KI,* and HOXA13 from E11.5 forelimb buds at *RA11KI* specific peaks with a p-value $<10^{-20}$, in a 4kb window. Peaks are categorized based on their colocalization with a HOXA13 peak and are ranked by HOXA13 binding. (b) *De novo* motif analyses using HOMER showing the top scoring motif in a 200bp window for *RA11KI* specific peaks bound by HOXA13 (pvalue$<10^{-20}$) or not.



Supplementary figure 3. ATACseq analyses in HOX13-/- compared to wild type mice.

(a) Correlation between ATACseq signal in replicate 1 and replicate 2 of wild type (left panel) and *Hox13-/-* (right panel). Correlation coefficient was computed using Excel. Values are reads per kilobases per million reads. (b) Genome browser view (IGV) of the indicated ChIP-seq and ATAC-seq data from mouse E11.5 limb bud at the mentioned loci.