Universidad Autónoma de Madrid
Departamento de Bioquímica
Facultad de Medicina

The DNA binding profile of TALE proteins in the embryo

Contrast with *in vitro* studies and functional implications

Daniel Mateos San Martín, Licenciado en Bioquímica.
Tesis doctoral dirigida por el Dr. Miguel Torres Sánchez.

CNIC, Madrid, 2013

Esta tesis doctoral ha sido realizada en el Centro Nacional de Investigaciones Cardiovasculares bajo la supervisión del Dr. Miguel Torres.

Durante la realización de esta tesis doctoral, el doctorando disfrutó de una beca de máster de la Fundación La Caixa y un Contrato para Personal Investigador de Apoyo de la Comunidad de Madrid.

Agradecimientos

Son muchas las personas que te encuentras a lo largo de una tesis. Algunos te apoyan y otros simplemente te hacen el viaje más agradable con su compañía, lo que también vale muchísimo. A todos ellos quiero darles mi agradecimiento. Algunos nombres se me quedarán en el tintero, pero no por ello me olvido de sus portadores. Creo que los que me conocen bien saben cuánto los aprecio.

Los primeros son, por supuesto, mis padres. Sin ellos no estaría donde estoy, no sería lo que soy, no sería a secas. Vosotros me instilasteis la curiosidad que me trajo aquí. Gracias. Os quiero.

A Miguel quiero reservarle también un sitio especial. He aprendido mucho en estos años. Creo que he crecido mucho, y buena parte de ello ha sido por tu ejemplo.

Ha habido muchos compañeros con los que he compartido muchísimas horas de risas y alguna de desesperación, pero pocas han sido insípidas. Todo el labo MT ha estado ahí cuando hacía falta, pero quería resaltar a Alberto, esa máquina de hacer ciencia que era un poco como un hermano mayor, a Clara con la que me pegaba por la poyata, a Laura que es como una segunda madre, un referente sólido en el que siempre te puedes apoyar cuando no tienes ni idea de por qué tu experimento no sale y que reparte ternura. A Joanna, ese cripto-pedazo de pan. A Susana, por la que tengo un cariño especial, a Bea, Irene, Félix, a todos los demás: un placer. A Carlos y Fátima gracias en especial por toda la ayuda.

Gracias a todos los de alrededor, que dan una vidilla que en otros centros no puedo imaginar que tengan: la Cañón encantadora, Teresa con su genio, Cris, Julio, Héctor, Guille, Gaetano al que ahora mismo oigo de fondo (cómo no), Vane, Juli, Marcos, Esther, Vero, Claudio… a Virginia también: sin tu atención al detalle vete a saber cómo la habría liado con las colonias. Roisin, cómo hemos sudado con el OPT, ¿eh?

Gracias a mis compañeros de máster, esos hermanos de sangre que hemos ido creciendo juntos: Edu y Alejo con los brunchcrafts y todo lo demás, Cris otra vez, Pilar, Inés, Jaime, Miguel, Lucía… Me acuerdo mucho de ese año tan especial, y del camino que hemos hecho desde entonces. ¡Flo, millones de gracias por la ayuda con la portada otra vez!

Gracias a mis queridos amigos de la carrera: hubo un momento en que me quedé algo descolgado, pero no me olvido de vosotros, y cuando nos volvemos a ver siempre es un poco igual que siempre. Yuri, compi por partida doble, Ana que se casó ya (que fuerte), Tefi de alegría contagiosa, Albertito cuyo blog leo aunque no comente mucho… os quiero.

Espero veros a todos muchas veces más en lo que me queda, y que podamos seguir compartiendo momentos.

Resumen

La Biología del Desarrollo estudia la formación de un organismo complejo con cientos de tipos celulares a partir de una única célula, el cigoto. Esto implica procesos de proliferación y diferenciación celular. Los factores de transcripción son clave en la especificación regional y diferenciación celular, delimitando qué genes se activarán en una célula concreta y confiriéndole por tanto su identidad.

Los genes Hox son una clase de factores de transcripción implicados en desarrollo particularmente importante. Están ampliamente conservados en el reino animal y forman el núcleo del sistema molecular de formación de patrón antero-posterior en todos los animales bilaterales.

Ya que los genes Hox reconocen secuencias de ADN cortas e indistintas, sus funciones de regulación transcripcional requieren de la participación de cofactores. Los cofactores Hox pertenecen a la clase TALE, un conjunto de factores de transcripción con un homeodominio divergente. En mamíferos, hay tres familias de cofactores Hox TALE: Meis, Prep y Pbx.

En esta tesis hemos abordado un análisis a escala genómica de los sitios de unión al ADN de la casi totalidad de los cofactores TALE presentes en el embrión de ratón a estadio E11.5. Meis y Prep, que se comportan casi indistinguiblemente *in vitro*, reconocen secuencias distintas en el embrión. Mientras que las secuencias unidas por Meis sugieren que actúa principalmente como cofactor Hox, Prep parece unir ADN predominantemente en forma de heterodímeros con Pbx.

No sólo sus secuencias de unión tienen aspecto muy distinto. Las demás características de sus sitios de unión son también muy distintas, sugiriendo funciones divergentes. Prep une muchos promotores y está asociado a la activación transcripcional de los genes afectados. En cambio, Meis une ADN en un patrón casi aleatorio con respecto a los sitios de inicio de transcripción. La conservación filogenética de los sitios de unión de Meis es extremadamente alta y muchos muestran marcas epigenéticas características de *enhancers*.

En los cuatro clusters Hox del ratón, Los sitios de unión de Meis, Pbx y Prep tienen un patrón denso y restringido al extremo 3', hasta el parálogo 9. Los niveles de los transcritos Hox están afectados en mutantes de pérdida de función de Meis1 y Prep1 colinealmente y en sentidos opuestos.

En conjunto, nuestros resultados dibujan un marco en el que Meis y Prep están especializados funcionalmente. Mientras que Meis actúa principalmente como cofactor Hox, Prep une Pbx y actúa sobre promotores sin participación de las proteínas Hox. A menudo tienen efectos opuestos sobre la transcripción, por ejemplo en los clusters Hox.

Summary

Developmental Biology studies the formation of a complex organism with hundreds of cell types from a single cell, the zygote. This involves both proliferation and differentiation. Transcription factors are key players in cell differentiation, selecting which genes will be active in a particular cell, thus defining its identity.

The Hox genes are a particularly important class of developmental transcription factors. They are widely conserved among animals and form the core of the molecular system for antero-posterior patterning. Hox genes require cofactors to effect their transcriptional regulatory functions. In mammals, there are three classes of TALE Hox cofactors: Meis, Prep and Pbx.

In this thesis, we have undertaken a near comprehensive ChIP-seq analysis of the genomic binding sites of the TALE Hox cofactors in the E11.5 mouse embryo. We have found that Meis and Prep, which bind DNA similarly *in vitro*, have very distinct binding preferences in the embryo. While the sequences bound by Meis suggest that is is acting mostly as a Hox cofactor, Prep seems to bind DNA predominantly in heterodimers with Pbx.

Not only do they have distinct sequence preferences, but also bind genomic sites with very different characteristics and, likely, functions. Prep, especially when in combination with Pbx, binds a high number of promoters and activates transcription. In contrast, Meis binds DNA with a nearly random distribution relative to transcription start sites. Despite being remote from promoters, the phylogenetic conservation of Meis binding sites is extremely high and many of them bear enhancer epigenetic marks in published datasets, suggesting they are highly conserved enhancer elements.

Meis, Prep and Pbx binding sites show a suggestive pattern in the four murine Hox clusters, in all of which Meis binds very densely to the 3' portion with a clear restriction to the region 3' of the Hox9 paralog. Hox genes transcript levels are affected in Meis1 and Prep1 loss-of-function models in a collinear and opposite fashion.

Taken together, our results paint a picture in which Meis and Prep are functionally specialized. While Meis is mainly a cofactor for Hox genes, binding in concert with them and Pbx to many highly conserved non coding elements, Prep binds with Pbx but no participation of Hoxes to promoters and activates the corresponding genes. They often have opposing effects on transcription, for example in the Hox clusters.

Contents

List of Figures

List of Acronyms

aa aminoacid

abd-B *Abdominal B*

ANT-C *Antennapedia* Complex

antp *antennapedia*

AP Antero-Posterior

ARE Auto-Regulatory Element

bHLH Basic Helix-Loop-Helix

bp Base-Pairs

BS Binding Site

BX-C *Bithorax* Complex

ChIP Chromatin Immunoprecipitation

CI Close Intergenic

CRM Cis Regulatory Module

C-ter Carboxi terminal

DECA Decameric Motif

DECA^{ext} Extended Decameric Motif

DV Dorso-Ventral

EMSA Electrophoretic Mobility Shift Assay

exd *extradenticle*

FI Far Intergenic

ftz *fushi tarazu*

GW Genome-Wide

HD Homeodomain

HEXA Hexameric Motif

HOM-C Homeotic complex

HSC Hematopoietic Stem Cell

HSPC Hematopoietic Stem-Progenitor Cell

hth *homothorax*

IG Intragenic

KO Knock-Out

lab *labial*

MACS Model-based Analysis of ChIP-Seq

MEF Mouse Embryonic Fibroblast

Meis^{exc} Meis-Exclusive

MEP Megakaryocyte-Erythroid Progenitor

Meis-Pbx1^{com} Meis-Pbx1 Common

Meis-Prep^{com} Meis-Prep Common

NES Nuclear Export Signal

NLS Nuclear Localization Sequence

N-ter Amino terminal

OCTA Octameric Motif

Pbx1-Prep^{com} Pbx1-Prep Common

Pbx1^{exc} Pbx1-Exclusive

PCR Polymerase Chain Reaction

PD Proximo-Distal

PICS Probabilistic Inference for Chip-Seq

PKA Protein Kinase A

PolII RNA Polymerase II

Prep^{exc} Prep-Exclusive

Prep1$^{i/i}$ *Prep1* hypomorph

PSSM Position Specific Scoring Matrix

RA Retinoic Acid

RARE Retinoic Acid Response Element

TALE Three Aminoacid Loop Extension

TF Transcription Factor

TSA Trichostatin A

TSS Transcription Start Site

TSSA Transcription Start Site-Associated

ubx *ultrabithorax*

WT Wild Type

One

Introduction

Developmental Biology deals with the process by which a single cell, the zygote, gives rise to trillions of specialized cells belonging to many different cell types that function as a coordinated whole.

In the process from zygote to adult, that single original cell must proliferate to spawn countless descendants. These descendants must also differentiate appropriately according to their position in the body.

The key to differentiation are transcription factors, proteins that bind DNA and regulate the expression of nearby genes. By restricting active genes to the specific repertoire required to differentiate a particular cell type, they are responsible for that cell's identity.

A very special class of transcription factors are responsible for translating and refining the early positional information in the embryo into a system of regional specification patterns. They are the Hox genes, and they are present in all bilaterian animals with a similar genomic organization in most of them, homologous roles in patterning and extreme sequence conservation.

1.1 The Hox genes

1.1.1 History of Hox Genes Study

William Bateson coined the term homeosis in 1894 to describe morphological variations in which a part of the body resembles another. Homeotic mutations thus affect genes important for establishing body part identity. Several such mutations were discovered during the era of classical genetics in Drosophila. Deletion of *ultrabithorax* (*ubx*), for example, results in a shift of the identity of the third thoracic segment into second thoracic, as shown by the presence of wings instead of halteres (Morata and Kerridge 1981).

After the work of many developmental geneticists culminating in Edward B. Lewis' landmark paper (Lewis 1978) the basic rules of segmental specification in Drosophila were extensively understood. There were a number of mutations known to result in anterior to posterior transformations. It was known that these mutations involved two clusters of linked genes, the *Antennapedia* Complex (ANT-C) and the *Bithorax* Complex (BX-C). It was known that loss of function mutations produce anteriorization of segmental identity while gains of function result in posteriorization.

Analysis of the phenotypes and their combinations indicated that there are two corresponding gradients in the embryo: a proximo-distal gradient in the chromosome and a BX-C gene action antero-posterior gradient such that any gene active in a segment is active in all segments posterior to it. These studies were performed

before *in situ* hybridization was developed or even the actual products of the genes affected were known.

The work of Christiane Nüsslein-Volhard and Eric F. Wieschaus expanded in this solid foundation (Nüsslein-Volhard and Wieschaus 1980). They performed an extensive screen that identified many of the genes that would occupy developmental biologists for more than a decade.

After the cloning of the *Drosophila* BX-C (Bender et al. 1983) and ANT-C (Garber, Kuroiwa, and Gehring 1983; Scott et al. 1983), it was noted that probes for *antennapedia* (*antp*) cross-hibridized with several other genes contained in them, indicating sequence similarity (McGinnis, M. S. Levine, et al. 1984). The stretch of homology spans 180 base pairs, was called the homeobox and codes for a protein domain called the Homeodomain (HD). The HD is even more conserved in sequence than the homeobox and from the very beginning was suspected to be responsible for DNA binding due to its high content in basic aminoacids (Gehring 1985).

Understanding of the homeotic selector genes and of animal development in general underwent a revolution when, shortly afterwards, it was discovered that the homeobox is present in vertebrates. In two papers published back to back, Carrasco and co-workers reported cloning of a *Xenopus* gene with homology to *antp*, *ubx* and *fushi tarazu* (*ftz*) (Carrasco et al. 1984) while McGinnis and co-workers reported the sequence of these three *Drosophila* genes and, strikingly, the presence of homologs in several animal species far removed philogenetically from insects (McGinnis, Garber, et al. 1984). In a context where no vertebrate genes controlling development were known, this must have been an earth-shattering revelation. Suddenly a system well studied in *Drosophila* potentially played a role in shaping the animals closest to us. It hinted at a deep underlying layer of similarity uniting all complex animal body plans that had been unsuspected before, and that would be confirmed and explored in the ears to come.

By the early 1990s, It was well established that the Hox clusters of vertebrates and the Homeotic complex (HOM-C) of *Drosophila* were homologous. The colinearity of gene expression domains with respect to position along the chromosome was long since described in *Drosophila*, and it had been recently reported that in vertebrates there is additional colinearity with timing of gene expression and RA sensitivity. Hox proteins were known to be able to bind DNA and suspected to be transcription factors (reviewed in M. Levine and Hoey (1988); McGinnis and Krumlauf (1992)).

A new stage in the study of Hox genes started focusing on their molecular function, through the use of homologous recombination in the mouse to generate loss of function and *in vitro* techniques to characterize protein activity. The recent understanding of the high conservation of Hox gene function and organization across

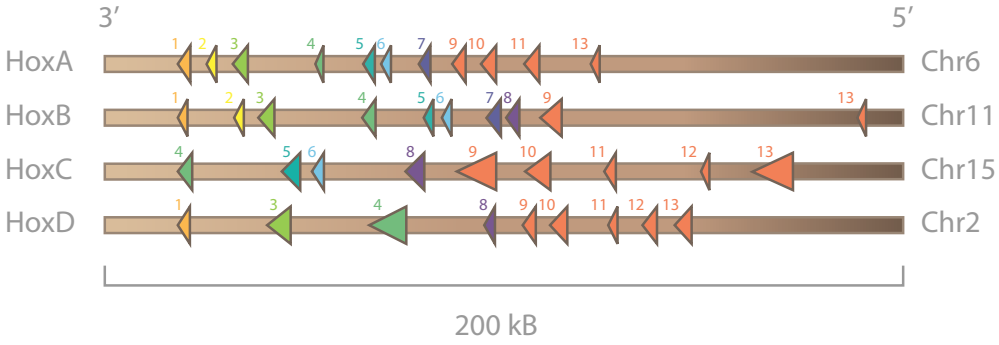

Figure 1.1: **The Organization of Hox Genes.** In mammals there are four Hox clusters, containing a total of 39 genes arranged in the same orientation in each cluster. The Hox clusters of mammals are fairly compact and devoid of non-Hox genes. Here the murine clusters are shown to scale.

most of the animal kingdom would aid by making findings in different model organisms directly translatable.

1.1.2 Cluster and Gene Structure

Drosophila homeotic genes are organized in two gene clusters, the *bithorax* (reviewed in Lewis 1978) and *antennapedia* (reviewed in Kaufman, Seeger, and Olsen 1990) complexes. The two of them form the HOM-C. The *Drosophila* "cluster" was the first to be described, but in many aspects it is quite atypical. Both components are located in the third chromosome of the fly but separated by about 10Mb, more than one third of the total length of the chromosome. They are much longer than their vertebrate counterparts. Several non-Hox-related genes are contained in the ANT-C and not all the Hox genes share the same transcriptional orientation.

In mammals there are 39 Hox genes grouped in four tightly organized clusters, each of which is homologous to the *Drosophila* HOM-C. There are 13 different sets of homologous genes (paralog groups) but no cluster contains examples of all paralogs. The only paralog groups present in all clusters are 4, 9 and 13, homologs of *Drosophila deformed* (4) and *abdominal B* (9 and 13).

For a more extensive comparative analysis of Hox clusters see Duboule (2007).

All Hox genes share a 180 bp sequence, the homeobox. The homeobox codes for a 60aa DNA binding domain, the HD, related to the helix-turn-helix DNA-binding domain of prokaryotes. It is constituted by a flexible Amino terminal (N-ter) arm and three α-helices, the third of which slots into the major groove of DNA and interacts with specific nucleotide residues. The specificity of this interaction is critically affected by aminoacid (aa) 9 of the recognition helix (Treisman et al. 1992). aas 6 and 10 also contact DNA and are very conserved across HDs. Arginines at

positions 3 and 5 of the N-ter non-helical portion of the HD also contact DNA on the core recognition site, but in the the minor groove side. Arginine 5 is almost invariant across HDs. Arginine 3 is present in Hox groups 2-8 and contributes to a preference for TAAT sites. The presence of a Lysine in position 3 is characteristic of the *Abdominal B* (*abd-B*) subfamily and probably contributes to its preference for TTAT (see 1.4 for binding specificities).

The hexapeptide/linker/N-ter of the HD is present in all Hox non-*abd-B*-like proteins (paralogs 1-8). Paralogs 9 and 10 have an alternative tryptophan-containing motif N-ter of the HD. These motifs are relevant for Hox-cofactor interactions (for more detail on Hox-cofactor interactions, see 1.3).

Transcriptional activator/repressor functions are mediated by domains other than the HD and differ between paralog groups. The balance between activation and repression is also influenced by Protein Kinase A (PKA) signalling (Saleh, Rambaldi, et al. 2000).

1.1.3 Regulation of the Hox Genes

Even before the discovery of the mouse Hox genes, it was already known that Hoxes have complex patterns of auto- and cross-regulation.

Polycomb was described as early as 1978 as a repressor needed for the correct expression of Hox genes (Lewis 1978). In addition, the *trithorax* complex is an epigenetic activator of gene expression that also plays a role in the proper expression of Hox genes.

The first diffusible molecule to be described as a regulator of Hox gene expression was Retinoic Acid (RA) (Simeone et al. 1990), which shifts their expression domains anteriorly. RA signalling is mediated by retinoid acid receptors and retinoic X receptors. These are cytoplasmic proteins that translocate to the nucleus in the presence of RA and activate gene expression by binding to Retinoic Acid Response Elements (RAREs). A number of RAREs have been described in the Hox clusters (e.g. Langston, J. R. Thompson, and Gudas 1997; Pöpperl and Featherstone 1993). RA also upregulates Pbx levels through post-transcriptional means (Knoepfler and Kamps 1997) and Meis levels.

Krox20 is a Transcription Factor (TF) that plays an important part in establishing Hox pattern in the rhombomeres of the vertebrate neural tube (Swiatek and Gridley 1993). Apart from Krox20, other regulators of Hox gene expression are Kreisler (Manzanares, Cordes, et al. 1999), Cdx1/2 (Lohnes 2003), AP-2 (Ding et al. 2013; Doerksen et al. 1996; Maconochie et al. 1999), and GATA-1 (Gosiengfiao, Horvat, and A. Thompson 2007; Vieille-Grosjean and P. Huber 1995).

1.1.4 Downstream: the Need for Cofactors

An important aspect of the effect of Hox genes in development is called phenotypic suppression (González-Reyes and Morata 1990) or posterior prevalence (Lufkin et al. 1991) and consists on the ability of more posterior (i.e. 5′) Hoxes to suppress the effect of more 3′ Hoxes. It explains the tendency of loss of function mutations to result in anteriorization while gain of function mutations usually result in posteriorization.

Hox genes are transcription factors. A simple view of their molecular effect is that they bind DNA and activate or repress the transcription of their nearby genes. However, all of the Hox HDs recognize slightly different variants of the same, very short, DNA sequence, which is found thousands of times in the genome. This begs the question of how different Hoxes recognize different, specific targets. This question baffled biologists for years until the discovery of DNA binding partners attenuated the contradiction (see Mann and Chan 1996). The Hox cofactors enhance the specifity and affinity of Hox DNA-binding, thus providing a better explanation for functional selectivity than DNA sequence recognition by Hox proteins alone can provide. We will see however that the question is still far from resolved.

1.2 History and general perspective of Hox cofactors

Pbx was identified in *Drosophila* in 1990 as *extradenticle* (*exd*) (Peifer and Wieschaus 1990) and proposed as a Hox protein cofactor based on phenotype. In parallel, it was identified as the DNA binding part of the chimeric protein produced by the t(1;19) translocation found in human pre-B cell acute lymphoblastic leukemia (Kamps et al. 1990). They were recognised as homologs in 1993 (Rauskolb, Peifer, and Wieschaus 1993).

Meis1 (Myeloid ecotropic viral insertion site 1) was identified in 1995 as a proto-oncogene locus in which the ecotropic murine leukemia virus is inserted in BXH-2 mice (Moskow et al. 1995). *homothorax* (*hth*) was identified as a *Drosophila* homolog of Meis1 required for *exd* nuclear localization (Rauskolb, Smith, et al. 1995; Rieckhof et al. 1997).

Prep1 (Pbx-regulating-protein 1) was identified in 1998 as a component of the urokinase enhancer factor 3 capable of forming a DNA binding-independent heterodimer with Pbx1 (Berthelsen, Zappavigna, Ferretti, et al. 1998; Berthelsen, Zappavigna, Mavilio, et al. 1998).

1.2.1 Philogeny: hoxes, tales

The HD is present in many genes other than Hoxes. Homologs of homeodomain genes have been found as far from vertebrates as plants. Meis, Prep and Pbx are

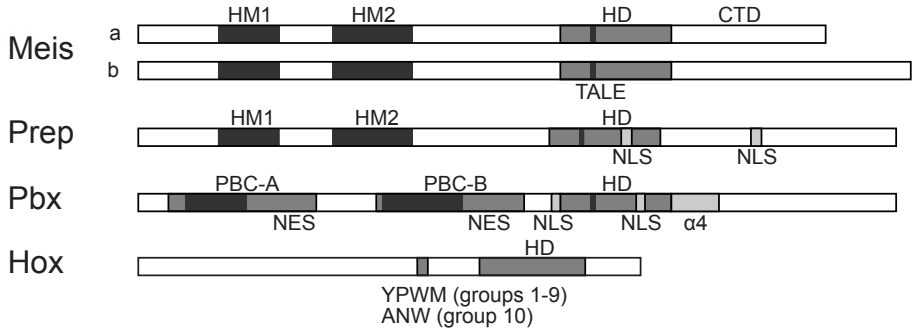

Figure 1.2: **Structural domains of the TALE and Hox proteins.**

part of a subgroup of HD proteins characterized by the presence in their HD of a Three Aminoacid Loop Extension (TALE) that gives name to the class and is situated between α-helices 1 and 2 (Bertolino et al. 1995; Burglin 1997; Moens and Selleri 2006; Mukherjee and Burglin 2007).

1.2.2 Structure of the TALE proteins

There are six distinct subgroups of TALE homodomain proteins in bilaterian genomes. Of these, only 3 are known to be Hox cofactors.

These three groups, the PREP, MEIS, and PBC classes, share homologous sequences at their N-ter ends (Burglin 1998). These sequences, called HM-1 and HM-2 in the MEIS and PREP classes and PBC-A and PBC-B in the PBC class, are exclusive to the three Hox cofactor classes despite their being apart in the TALE superclass philogeny (Mukherjee and Burglin 2007) and are responsible for much of their Hox- and cross-interactivity (see 1.3 for more details). Several Nuclear Export Signals (NESs) and Nuclear Localization Sequences (NLSs) have been reported in Pbx and in Prep. TALE protein structure is summarized in Figure 1.2.

The PBC class is the most divergent within the TALE superclass. Members of this class have a fourth α-helix 3' to the HD (Mukherjee and Burglin 2007). Also, residue 50 of the HD, the 9^{th} of the 3^{rd} helix, is a glicine in the PBC class, in contrast to the long side chain aas in Hox proteins. This residue is important for base selection by contacting the major groove of DNA.

Splicing Isoforms

Meis1 and *Meis2* are spliced to produce two major isoform families in vertebrates. These isoforms differ in their Carboxi terminal (C-ter) portion. *Meis1a*, *Meis2a* and *Meis2b* are all short-C-ter isoforms, while *Meis1b*, *Meis2c* and *Meis2d* encode a long-C-ter isoform. *Meis3* is similarly spliced. Both kinds of isoforms are responsive to Trichostatin A (TSA) (an HDAC inhibitor) and PKA but the minoritary (under 10%

in vertebrates) long-C-ter versions are much more potent transcriptional activators, both constitutively and in response to TSA and PKA (H. Huang, Rastegar, et al. 2005; Irimia et al. 2011). DNA-binding and TALE/Hox interactions of Meis are generally unaffected by splicing, since the HD and N-ter are constitutively spliced in. There are minoritary HD-less isoforms in vertebrates and insects (Irimia et al. 2011).

The existence of splicing isoforms has also been reported for *Prep*. There are at least two isoforms of *Prep2*, one of which contains the full N-ter including HM1 and HM2 but not the HD, while the other is the full-length protein (Haller et al. 2004).

1.3 Interactions Between the Hox genes and their TALE cofactors

Pbx was the first of the three cofactor classes to be shown to bind DNA cooperatively with Hoxes (Chan, Jaffe, et al. 1994). Over the years, a number of protein complexes involving two or more partners binding on the surface of DNA or independent of it have been described. The picture emerging from two decades of *in vitro* studies is complex.

1.3.1 Two-way Interactions

All three classes of TALE Hox cofactors have the nominal extension in their HDs, but only Meis and Pbx proteins have been shown to cooperatively bind DNA with Hoxes. Pbx was the earliest known Hox cofactor (Peifer and Wieschaus 1990). Its binding to Hox proteins requires its HD plus the following 15 aas and the Hox hexapeptide/linker/N-ter of the HD, which is an important determinant of DNA specificity and ability to interact with Pbx, as shown by chimeric HD experiments (Chang, Brocchieri, et al. 1996; Phelan and Featherstone 1997). The hexapeptide, also known as YPWM motif, is present in paralog groups 1 to 8 (Chang, W. F. Shen, et al. 1995; Passner et al. 1999). Paralog groups 9 and 10 can interact with Pbx through an alternative ANW motif located just 5′ to their HD (Chang, Brocchieri, et al. 1996; W. F. Shen, Rozenfeld, Lawrence, et al. 1997). Paralog groups 11 to 13 cannot bind Pbx1a despite most of them containing tryptophan residues 5′ to their HD.

Crystallographic structure determination of a Exd-Ubx-DNA complex (Passner et al. 1999) and a Pbx1-Hoxb1-DNA complex (Piper et al. 1999) showed that Pbx and the Hox component of the heterodimer bind DNA on opposite sides of the helix. The N-ter linker (20 aas long in Hoxb1 and 8 aas long in Ubx) of the Hox protein extends over the DNA so that the hexapeptide contacts Pbx directly, in an

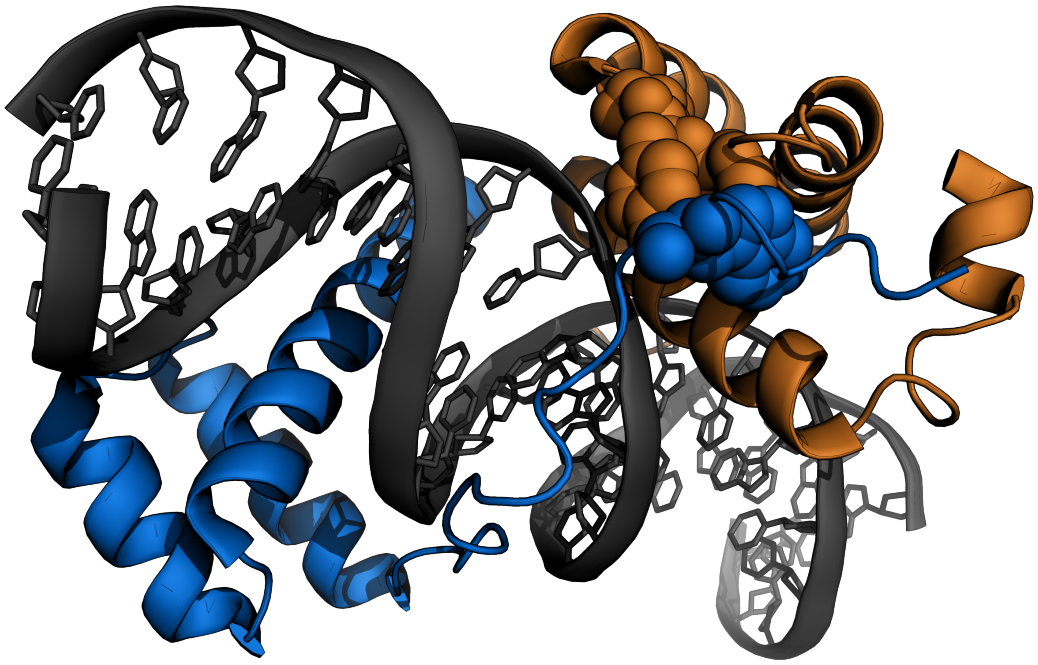

Figure 1.3: **Hox-Pbx-DNA Complex structure.** X-ray crystallographic determination of the structure of a Pbx1-Hoxa9-DNA complex, from LaRonde-LeBlanc and Wolberger 2003. The Hox protein is labelled in blue and Pbx in orange, with their hexapeptide tryptophan and TALE shown as spheres. The 3rd α-helices contact DNA directly in head-to-tail orientation. The conserved tryptophan slots into a hydrophobic pocket in Pbx partially formed by the TALE. The 4th α-helix in the divergent Pbx HD can be seen at the extreme right. The N-ter linker in Hoxa9 is ordered and contacts the minor groove of DNA, in contrast to those in the other Hox proteins crystallized, Hoxb1 and Ubx. The portions of Hox N-ter to the conserved tryptophan motif and of the Pbx N-ter to the HD were not crystallized.

hydrophobic pocket partially formed by the DNA recognition helix in the Pbx HD and by its TALE. There is also potential for the N-ter of the recognition helix of Pbx to interact with the C-ter of the Hox recognition helix, since the two helices are oriented head-to-tail and their ends are close. Later determination of an Abd-B homolog, Hoxa9, in complex with Pbx1 and DNA showed very similar arrangement of the proteins in the DNA (LaRonde-LeBlanc and Wolberger 2003). Despite Hoxa9's hexapeptide having a radically diferent backbone structure from Ubx's or Hoxb1's, the critical tryptophan residue slots into Pbx1's hydrophobic pocket in the exact same position. One of the few differences between the determined structures is that Hoxa9's N-terminal linker is ordered, unlike Ubx's or Hoxb1's.

Meis1 can bind directly Hoxes belonging to groups 9 to 13 through its C-ter portion. This interaction greatly stabilizes Meis DNA binding, which may occur

as homodimers in the absence of Hoxes or Pbx. The homeodomain of Hoxes is necessary for binding to Meis, but further aas N-ter to the Hox HD are necessary to stabilize the binding, at least in the case of Hoxa9. The part of Meis1 required for Hox binding is located somewhere within the HD or the 38 following aas, which are common to Meis1a and Meis1b. The two main Meis isoform types thus do not seem to differ in their protein or DNA binding characteristics (W. F. Shen, Montgomery, et al. 1997; T. M. Williams, M. E. Williams, and Innis 2005).

Pbx and Meis can interact independently of DNA, and *in vivo* immunopurified Pbx-containing complexes avidly bind the joint Pbx-Meis binding target (see 1.4), suggesting that most of the Pbx in the cell is complexed with Meis or Prep (Chang, Jacobs, et al. 1997; Knoepfler, Calvo, et al. 1997). This interaction is required for Pbx/Exd to translocate to the nucleus (Rieckhof et al. 1997) and for stability of the Meis/Hth protein (Abu-Shaar and Mann 1998). The Pbx Meis-interacting surface is contained within its first 88 aas, N-ter to the point of fusion in E2a-Pbx1 chimeras and within the PBC-A domain. The Meis Pbx-interacting surface is located between aas 30 and 60 (Chang, Jacobs, et al. 1997). The fact that Meis and Pbx interact through their flexible N-ter arms results in non-strict position and orientation requirements for their DNA targets, according to Jacobs, Schnabel, and Cleary (1999).

Prep binds Pbx independently of DNA through its N-ter conserved domains and the portion of Pbx missing from the E2a-Pbx1 fusion (Berthelsen, Zappavigna, Ferretti, et al. 1998), like Meis.

1.3.2 Trimeric Complexes

In addition to the heterodimeric complexes, a number of trimeric complexes have been reported. In these, generally one of the protein partners does not bind DNA.

Meis and Pbx can form three-protein complexes with at least some of the Hox proteins. Meis1 can form ternary complexes with any of the Pbx's and Hoxa9 *in vitro* independently of DNA. Addition of Meis greatly enhances the Pbx-Hoxa9 interaction, which is weak in the absence of DNA (W. F. Shen, Rozenfeld, Kwong, et al. 1999). In this study it was also shown that the three proteins colocalize *in vivo* in nuclear speckles, but surprisingly fail to activate or repress a transcriptional reporter bearing their consensus DNA target.

Meis has also been shown to form trimeric complexes with Pbx and Hoxb1. These heterotrimers bound the *Hoxb1* ARE and the *Hoxb2* r4 enhancer sequences and required intact Pbx-Hox and Meis half-sites to bind DNA, but probably not to assemble (Jacobs, Schnabel, and Cleary 1999). Formation of complexes was checked to be dependent of the heterodimerization motifs of each protein component, implying that Pbx is the hinge subunit that interacts with both Meis and Hoxb1 at the same time. Addition of the three proteins resulted in cooperative

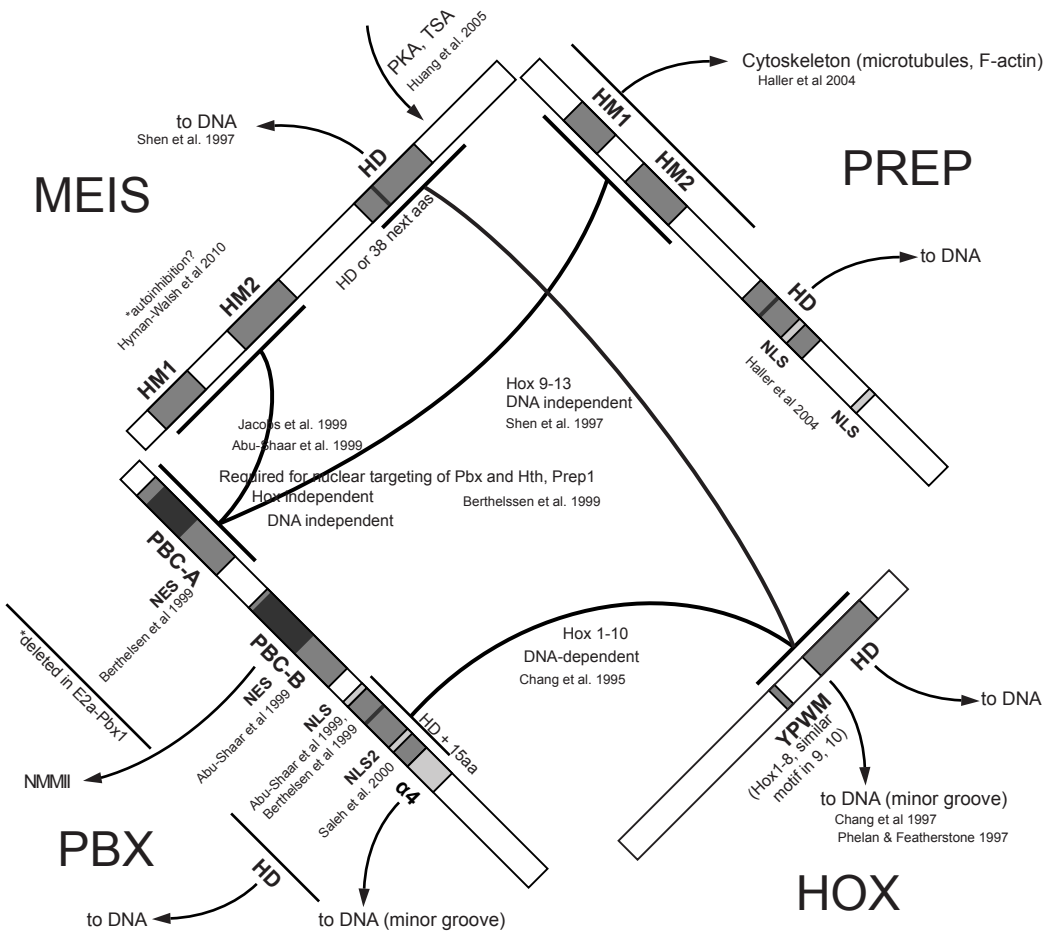

Figure 1.4: **Summary of TALE-Hox interactions.**

transcriptional activation in a transient transfection reporter assay and *in vivo* lacZ staining that depended on both Pbx-Hox and Meis recognition sequences, in contrast to W. F. Shen, Rozenfeld, Kwong, et al. (1999).

Trimeric Pbx-Meis-Hoxa9 complexes are required for Hoxa9-mediated immortalization of myeloid progenitors (Schnabel, Jacobs, and Cleary 2000).

Prep can form triple complexes with Pbx and a Hox protein as well. Co-transfection of *Prep1*, *Pbx1* and *Hoxb1* resulted in additional activation of a *Hoxb1* Auto-Regulatory Element (ARE) reporter over *Pbx1* and *Hoxb1* alone (Berthelsen, Zappavigna, Ferretti, et al. 1998). This additional activation was independent of Prep1 DNA binding ability but required its ability to bind Pbx. Prep1-Pbx1-Hoxb1 trimeric binding has been directly demonstrated *in vitro* on sequences from the rhombomeric *Hoxb1* ARE and on the *Hoxb2* r4 enhancer (Ferretti, Marshall, et al. 2000). In the case of the *Hoxb1* ARE, a non-DNA-binding version of Prep1 was surprisingly required for formation of the trimeric complex.

1.3.3 Interactions with Other Proteins

Pbx-Meis/Prep heterodimers have been shown to be able to bind DNA in concert with E2a-Myogenin and E2a-MyoD heterodimers (myogenic Basic Helix-Loop-Helix (bHLH) TFs) in the form of tetrameric complexes, as well as with Myogenin alone. Hox proteins preclude this interaction, which requires a conserved tryptophan in the bHLH TFs that is situated N-ter to its DNA-binding domain and could therefore be mediated by the same hydrophobic pocket in Pbx as Pbx-Hox interaction in a similar configuration (Knoepfler, Bergstrom, et al. 1999). Pbx-Meis interaction with MyoD has been shown to be necessary for MyoD binding to the *Myogenin* promoter, which results in chromatin remodelling and activation of *Myogenin*, suggesting recruitment of MyoD to heterochromatin by Pbx-Meis (Berkes et al. 2004).

Pbx was shown to bind the pancreatic homeodomain protein Pdx1 through a tryptophan motif similar to that of Pbx-interacting hoxes (Peers et al. 1995). Pdx1 can also form trimers with Pbx-Meis heterodimers. Pdx1 is critical for pancreas formation and participates in the differentiation of islet and acinar pancreatic cells. Expression of Pbx1 and Meis2 and trimeric association with Pdx1 is specific to acinar cells, representing a switch that changes the transcriptional effect of constitutive Pdx1 binding at the *Ela1* enhancer (Swift et al. 1998). The Pbx-Meis-Pdx1 trimer activates transcription with a bHLH heteromultimer in the *Ela1* enhancer independently of Meis DNA binding (Y. Liu, MacDonald, and Swift 2001).

The *Somatostatin* promoter has been shown to be bound by Pbx1-Prep heterodimers but not Pbx-Meis. This binding is necessary for Pdx1 to be able to activate *Somatostatin* (Goudet et al. 1999). It is interesting that in the *Somatostatin* promoter, binding of Pbx-Prep heterodimers was able to induce reporters driven by isolated components of the regulatory sequence but not by the whole sequence, which required the addition of Pdx1 for activation.

Another non-Hox homeodomain protein, Engrailed, has also been shown to interact with Pbx through a conserved tryptophan N-ter to its homeodomain, with an affinity comparable to Hoxes (Peltenburg and Murre 1996).

1.4 DNA Binding Specifities

The first homeobox proteins to be shown to be sequence-specific TFs were Ftz, En and Eve (Desplan, Theis, and O'Farrell 1988; Hoey and M. Levine 1988; Hoey, Warrior, et al. 1988). Two types of binding consensus sequence were described: TCAATTAAAT and TCAGCACCG.

Ubx Chromatin Immunoprecipitation (ChIP) showed binding to sites containing a TAAT core (A. P. Gould et al. 1990). This study also showed that binding to

sites containing very similar sequences can result in upregulation or downregulation of transcripts, so the sign of transcriptional regulation is probably not encoded in the core target sequence.

Exd was shown to bind cooperatively with Ubx to sequences in a *dpp* enhancer (Chan, Jaffe, et al. 1994). In this study, fragments of Exd and Ubx containing their HDs and C-ters were used. The reported binding sequences were ATCGAAATG and ATAAAACAA. Exd increased Ubx DNA binding by greatly reducing its dissociation rate. Surprisingly, a monoclonal antibody directed against the Ubx C-ter stabilized this interaction as well, but Antp did not produce the same effect. However, when the interaction requirements were tested in the mammalian counterparts of Ubx and Exd, the resulting picture was quite different (Chang, W. F. Shen, et al. 1995).

The vertebrate Hox HDs recognize a short DNA sequence of 4 nucleotides that in most cases contains a TAAT core (Catron, Iler, and Abate 1993; Treisman et al. 1992). The nucleotides flanking the TAAT core affect binding differentially according to the paralog group, but the magnitude of this effect is small. In addition, there is a gradient of DNA affinity such that 3′ Hox HDs have higher DNA affinities than 5′ ones (Pellerin et al. 1994).

The Pbx1 HD was shown to bind TGATTGAT (Van Dijk, Voorhoeve, and Murre 1993), but this could have been an artifact of the GST-HD fusions used. When Pbx-Hox heterodimer DNA binding was tested extensively, it was found that the slight differences in DNA selectivity shown by Hox monomers are exacerbated when binding with Pbx and the preferred target sites shift for most of the complexes. The Pbx-Hox consensus target sequence core is TGATNNAT, where the 5′ TGAT part is bound by Pbx and the 3′ NNAT is bound by the Hox protein. The Hox core nucleotides differ according to the Hox protein participating in the complex: 3′ Hoxes show highest affinity for a TGAT half-site, middle Hoxes (paralog groups 4-7) bind a half site that matches the monomeric Hox site TAAT, and 5′ Hoxes (groups 8 and higher) are very strict in their preference for a TTAT half site *vis-à-vis* other NNAT variants, but are also able to bind TTAC (Chan, Ryoo, et al. 1997; Chang, Brocchieri, et al. 1996; W. F. Shen, Rozenfeld, Lawrence, et al. 1997).

The TGATTGAT sequence has been shown to be bound by Pbx-Prep heterodimers in the context of the *Somatostatin* promoter (Goudet et al. 1999). The A/TGATAAAT sequence has been shown to be bound by trimeric Pbx-Meis-Pdx1 complexes (Y. Liu, MacDonald, and Swift 2001).

The aa residues determining sequence specificity have been explored. One of the most important is Arginine 5 of the HD, which is conserved across all Hox proteins (Phelan and Featherstone 1997).

Meis can bind DNA *in vitro* in the absence of either Pbx or Hoxes. In this case

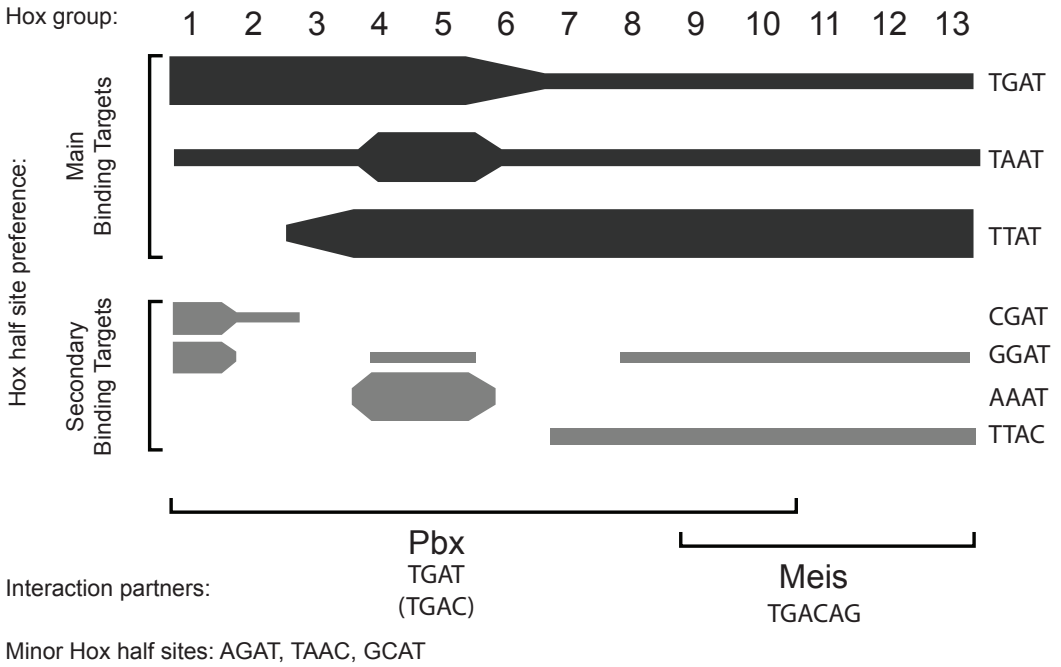

Figure 1.5: **Pbx-Hox heterodimer binding preferences.** DNA binding preferences for each paralog group of Hox proteins The width of each bar at each position represents relative affinity of that paralog group for that sequence. In black, the three major Hox half-sites. In grey, the most important secondary binding targets. Thickness of bars indicates relative affinity. Adapted from W. F. Shen, Rozenfeld, Lawrence, et al. (1997) and W. F. Shen, Montgomery, et al. (1997) to include data from Slattery et al. (2011).

it recognizes a very strict consensus sequence, TGACAG, often in multiple copies without a clear orientation or spacing (W. F. Shen, Montgomery, et al. 1997). This target sequence is very similar to that of the TGIF family of TALE proteins, and indeed Tgif binds DNA on the strand opposite to Meis and can compete with it (Yang et al. 2000). Meis monomeric binding is highly labile despite being quite intense in the steady state, which implies a rapid turnover of the bound protein.

Meis can also bind in complex with Abd-B-like Hoxes (groups 9 to 13), in which case the heterodimer recognizes a sequence containing TGACAG and the AbdB-like Hox targets TTAT or TTAC in a fixed spacing (W. F. Shen, Montgomery, et al. 1997).

When Meis binds DNA in a complex with Pbx, the preferred recognition target is the decameric sequence TGATTGACAG, containing a 5′ TGAT target and a 3′ TGACAG Meis target (Chang, Jacobs, et al. 1997; Knoepfler, Calvo, et al. 1997). Gapped configurations in which the TGAT and TGACAG half sites are spaced by 3 or 6 nucleotides have also been shown to bind Pbx-Meis heterodimers (Jacobs, Schnabel, and Cleary 1999). Several Pbx-Hox target sites have been tested for bind-

ing to Pbx-Meis, and the only ones to bind were those containing a G in position 5 of the Pbx-Hox target octamer.

Prep can bind DNA as a monomer, recognizing the same hexameric TGACAG sequence as Meis (Berthelsen, Zappavigna, Mavilio, et al. 1998). In this study, Pbx1 interaction was shown to increase DNA binding but no change in specificity was tested.

Prep can also bind DNA as an heterodimer with Pbx, in which case it recognizes either the TGACAG monomer target or a Pbx-Hox canonical sequence, TGATTGAT or TGATGGAT (Berthelsen, Zappavigna, Ferretti, et al. 1998).

Immunoprecipitation of Pbx, Meis or Hoxa9 complexes formed *in vitro* found essentially the same TGATTTAT sequence regardless of the complex component targeted (W. F. Shen, Rozenfeld, Kwong, et al. 1999). This is a consensus Pbx-Hox binding site that Meis on its own is not able to bind.

Ternary Meis-Pbx-Hox complexes are able to bind the *Hoxb2* r4 enhancer, which contains a Pbx-Hox (AGATTGAT) and a Meis monomer target site (TGACAG), both of which are necessary for trimeric binding (Jacobs, Schnabel, and Cleary 1999). The *Hoxb1* ARE also contains a Pbx-Hox (TGATGGAT) and a Meis monomer target site (TGACAG) and is likewise capable of binding by the trimeric Meis-Pbx-Hox complex, requiring both sites for binding (Jacobs, Schnabel, and Cleary 1999).

Ternary Prep1-Pbx1-Hoxb1 complexes are able to target the same Hoxb1 r3 ARE (TGATGGAT/TGACAG), but only if the Prep1 HD or the Meis/Prep target site are eliminated. In contrast, they can bind the *Hoxb2* r4 enhancer (AGATTGAT/TGACAG) with the endogenous forms of the proteins, and the elimination of the Prep1 HD in this case results in reduction of DNA binding affinity (Ferretti, Marshall, et al. 2000).

Tetrameric complexes involving Pbx, Meis or Prep, and a pair of myogenic bHLH TFs can bind a Pbx-Meis/Prep target site (TGATTGACAG) in a minimum spacing of 3bp from an E-box (CAGCTG) (Knoepfler, Bergstrom, et al. 1999).

Several relatively recent studies have applied high throughput techniques to the description of HD binding preferences. Two such studies managed to predict HD DNA binding preferences from aa sequence alone by constructing mathematical models based on affinity data for a wide selection of HDs. One tested the affinity of *in vitro* translated HDs for all possible 10-nucleotide combinations present in a protein binding microarray (Berger et al. 2008). The second used an specifically designed bacterial one hybrid system to gather their starting data, affinity of HD hybrids (Noyes et al. 2008). Although an impressive technical feat, the relevance of the described monomeric binding affinities is doubtful especially in the case of Hoxes, which change their affinity and specificity when binding DNA with their cofactors.

A later study overcame this limitation by restricting its test population to the 8 *Drosophila* Hox proteins and assessing their sequence preferences in complex with the *Drosophila* Pbx ortholog Exd and the HM domain of the *Drosophila* Meis ortholog Hth. Slattery et al. (2011) combined the SELEX technique (Tuerk and Gold 1990) with massive sequencing and a sophisticated computational analysis to infer sequence to affinity relationships for all possible 12-mers relative to each Hox-cofactor complex. This study is not only the most comprehensive exploration of Hox-Pbx DNA binding, but also the *in vitro* binding study that uses conditions most relevant to the understanding of *in vivo* binding.

Slattery et al. (2011) reaffirmed the notion that Pbx-Hox complex specificity is limited to an 8-bp long core, in which the 5′ Pbx-binding half is almost always TGAT and the 3′ Hox-dependent half has three major variants (TGAT, TAAT, TTAT) recognized preferentially by different subsets of Hoxes, but uncovered fine details slightly divergent from the classical 1996-1997 studies:

- Lab and Pb (paralog groups 1-2) recognize mainly a TGAT half-site, but also bind with comparable affinities other octamers: GGAT and CGAT for Lab, and a TGACTAAT site for Pb.

- Dfd and Scr (groups 4-5) recognize TGAT, TAAT and TTAT with similar affinities, but are also able to bind a wide variety of secondary half-sites, the most characteristic of which is AAAT.

- Antp, Ubx, Abd-A and Abd-B (groups 6-13) have a high relative affinity for TTAT, but are also able to bind the other two majoritary half-sites and a host of secondary sites, including one specific for them, TGATTGAC.

However, all these consensus sequences represent the binding site for which a particular Pbx-Hox heterodimer or Pbx-Meis-Hox trimer has the most affinity. It may be that the important factor *in vivo* is not the affinity but the selectivity of a particular site that allows it to choose between many Pbx-Hox complexes. An example is the *Hoxb1* autoregulatory enhancer. It contains several Hox-Pbx binding sites, of which one, R3, is highly specific for Hoxb1-Pbx1. In *Drosophila*, a similar sequence is present in a *labial* (*lab*) autoregulatory enhancer. Changing the sequence of this Lab-Exd target to TGATTAAT switches the specifity to Dfd-Exd (Chan, Ryoo, et al. 1997). However, the TGATGGAT original sequence is not the maximum affinity consensus for Lab-Exd (Mann and Affolter 1998; W. F. Shen, Montgomery, et al. 1997). The possibility that selectivity is more important than affinity is borne by the fact that many HDs, despite having identical highest-affinity targets, differ in their "secondary" target motifs (Berger et al. 2008).

If this is true, then *in vitro* binding studies that consider individual Hox proteins or at most combinations of a single Hox protein with different cofactors may not represent the most relevant or the majority of functional interactions taking place *in vivo*.

1.5 Expression Patterns

Prep1 and *Prep2* are ubiquitously expressed (Ferretti, Schulz, et al. 1999; Ferretti, Villaescusa, et al. 2006; Reymond et al. 2002).

Pbx1, *Pbx2*, and *Pbx3* are widely expressed, while *Pbx4* expression is restricted to both the embryonic and adult testes (Monica et al. 1991; K. Wagner et al. 2001).

Meis1 and *Meis2* expression is similar, starts around gastrulation and is regionalized (Cecconi et al. 1997; Oulad-Abdelghani et al. 1997; Tamplin et al. 2008).

1.6 Nuclear Localization of TALE proteins

The availability of TALE proteins in the cell nucleus has been shown to be regulated and to have functional relevance in both vertebrates and insects (Mercader, Leonardo, Azpiazu, et al. 1999).

The localization of Pbx/Exd is cytoplasmic in the absence of Meis/Hth and nuclear in its presence. This is due to the presence in it of at least a NES and a NLS. The NES and NLS have been mapped to aas 178-220 and 239-243 of Exd, respectively (Abu-Shaar, Ryoo, and Mann 1999). These positions fall within the PBC-B motif and the N-ter of the HD. Meis/Hth has a NLS that allows it to target NLS-deficient Pbx/Exd fragments to the nucleus (Abu-Shaar, Ryoo, and Mann 1999). Other reports have described an additional NES within PBC-A (Berthelsen, Kilstrup-Nielsen, et al. 1999) and a NLS within the 3rd helix of the HD (Saleh, H. Huang, et al. 2000). The nuclear/cytoplasmic balance can be altered by leptomycin B, a nuclear export inhibitor, suggesting the role of Meis/Hth is to retain Pbx/Exd in the nucleus (Abu-Shaar, Ryoo, and Mann 1999; Berthelsen, Kilstrup-Nielsen, et al. 1999). Two somewhat overlapping models for nuclear trafficking of Pbx/Exd have been proposed. The first considers a balance between NESs and NLSs which gets tipped by binding to NLS-containing Meis/Hth or Prep (Affolter, Marty, and Vigano 1999). The second, supported by point mutations of Pbx, implies intramolecular interactions within Pbx/Exd between its N-ter and HD that mask the NLSs and are disrupted by binding of Meis/Hth or Prep to its N-ter (Saleh, H. Huang, et al. 2000).

Another regulator of Pbx subcellular localization is Nonmuscle Myosin Heavy Chain (H. Huang, Paliouras, et al. 2003). Pbx binds it through its PBC-B domain

and *zipper* (the Nonmuscle Myosin Heavy Chain *Drosophila* homolog) mutants show increased cytoplasmic retention of Exd. When overexpressed, Zipper can even induce cytoplasmic localization of Hth, presumably through Exd.

Prep also requires co-expression with Pbx/Exd for its nuclear localization (Berthelsen, Kilstrup-Nielsen, et al. 1999). Cytoplasmic Prep2 is associated with the actin and microtubule cytoskeleton components and this cytoplasmic localization is dependent on CRM-1-mediated nuclear export and the integrity of cytoskeletal networks (Haller et al. 2004).

1.7 Functions and phenotypes

1.7.1 Knock-Out Phenotypes

Meis1 mutant mouse embryos die around E14.5 showing eye, angiogenesis and hematopoietic defects, among others (Azcoitia et al. 2005; Hisa et al. 2004). Their pervasive haemorrhages result from incomplete separation of the blood and lymphatic vasculature systems, which in turn results from failure of the megakaryocytic lineage (Carramolino et al. 2010). They also show a reduction in Hematopoietic Stem Cell (HSC) populations in fetal liver, and cardiac malformation.

Prep1 Knock-Out (KO) embryos die early in development at E7.5 due to massive epiblast p53-dependent apoptosis (Fernandez-Diaz et al. 2010). To circumvent this obstacle to embryonic Prep function study, *Prep1* hypomorph ($Prep1^{i/i}$) mice have been generated. $Prep1^{i/i}$ embryos produce around 2% of the Wild Type (WT) levels of protein and die at E17.5 with eye, angiogenesis and hematopoietic defects, apart from organ hypoplasia (Di Rosa et al. 2007; Ferretti, Villaescusa, et al. 2006).

Although the phenotypic alterations in $Prep1^{i/i}$ embryos affect the same processes altered in *Meis1* KO embryos, they have different effects. The vasculature defects are centered around the lymphatic valves in *Meis1* KO but around the capillaries in $Prep1^{i/i}$. The hematopoietic defects consist in total lack of the megakaryocytic lineage in *Meis1* KO but in $Prep1^{i/i}$ there is an incomplete maturation of erythrocytes. Those $Prep1^{i/i}$ embryos that escape embryonic lethality show defects in T-cell development (Penkov et al. 2005).

Pbx1 deletion results in both hematopoietic (DiMartino et al. 2001) and patterning (Moens and Selleri 2006; Selleri et al. 2001) defects in mice. Many of these defects phenocopy individual Hox mutants, and combined knockdowns of *Pbx1* and different Hoxes indicate genetic interaction (Moens and Selleri 2006).

Pbx1/Pdx1 double heterozygous mice have defects in pancreas formation and develop diabetes mellitus (Kim et al. 2002).

Zebrafish *Pbx* has been shown to be necessary for activation of fast muscle genes by MyoD (Maves, Waskiewicz, et al. 2007) and cardiac expression of *Hand2* (Maves,

Tyler, et al. 2009). This results support the proposed 'pioneer' function for Pbx (Berkes et al. 2004) by which it would pre-occupy closed chromatin and enable its targeting by chromatin remodelling-inducing factors, setting the range of permissible sites for these factors. Such a pioneer functions has also been proven for Pbx regarding the estrogen receptor (Magnani et al. 2011).

1.7.2 Involvement in Leukemias

While *Meis1* alone cannot induce leukemia, joint overexpression of *Meis1* and *Hoxa7* or *Hoxa9* is sufficient to induce myeloid leukemia (Nakamura et al. 1996; Schnabel, Jacobs, and Cleary 2000).

The E2a-Pbx1 fusion protein resulting from the t(1;19) translocation induces pre-B cell acute lymphoblastic leukemia (Kamps et al. 1990). This fusion protein is able to bind DNA and to interact with Hoxes, but not with Meis or Prep. The transcriptional activation domain from E2a turns it into a constitutional activator.

The portion of $Prep1^{i/i}$ embryos that survive development and reach adulthood develop spontaneous tumours or pretumoral lesions, revealing a role for *Prep1* as a tumour-suppressor gene (Longobardi, Iotti, et al. 2010).

1.7.3 Other Functions

The TALE proteins have also been implicated in limb Proximo-Distal (PD) patterning. Meis1 and Meis2 are expressed in the proximal portion of the vertebrate limb and their ectopic expression in distal limb induces nuclear translocation of normally cytoplasmic Pbx and distal to proximal transformations (Mercader, Leonardo, Azpiazu, et al. 1999). Meis is downstream of RA, which is produced by the somites and present in the limb in a PD gradient, opposing the apical ectodermal ridge-produced FGF gradient (Mercader, Leonardo, Piedra, et al. 2000; Roselló-Díez, Ros, and Torres 2011).

Meis and Pbx are involved in cardiac development, as shown by semilunar valve, ventricular septum and outflow tract defects in *Meis* and *Pbx* mutants (Stankunas et al. 2008). Recently, Meis has been shown to be expressed and recruited to enhancers during *in vitro* differentiation of cardiomyocytes, specifically at the cardiac progenitor stage and apparently in combination with GATA factors (Wamstad et al. 2012). *Meis2* knockdown in Zebrafish results in failure of cardiac tube looping (Paige et al. 2012).

Meis has also been recently implicated in mammalian neonatal heart regeneration through cardiomyocyte mitotic arrest (Mahmoud et al. 2013). Specific deletion of *Meis1* in cardiomyocytes resulted in increased postnatal cardiomyocyte proliferation while overexpression resulted in decreased proliferation. Reporter constructs showed regulation of Ink4b–Arf–Ink4a and p21 by Meis.

1.8 ChIP-seq

The ChIP assay consists of cross-linking of chromatin to preserve the non-covalent interactions between proteins and DNA for biochemical assay. It was first described by Solomon, Larsen, and Varshavsky (1988). For many years its application remained confined to focused experiments (Mardis 2007).

The first genome-wide application of the technique was achieved by microarray hibridization for determination of DNA fragment enrichment (ChIP-chip, Ren et al. 2000).

As massive sequencing has gone down in cost, its application to assaying ChIP-enriched DNA fragments has increased. The ChIP-seq technique (Robertson et al. 2007) involves deep sequencing of DNA fragments followed by bioinformatic treatment of the data to extract genomic regions bound by the factor of interest (peaks). It provides better precision and less bias than ChIP-chip (see for example Rhee and Pugh 2011) and has allowed the Genome-Wide (GW) determination of TF Binding Sites (BSs) and histone marks in an unprecedented number of cell types and tissues, for example in the ENCODE project (ENCODE Project Consortium 2004; ENCODE Project Consortium, Bernstein, et al. 2012; ENCODE Project Consortium, Birney, et al. 2007).

Two

Objectives

Further study of how the TALE proteins exert their various physiological and pathological effects is limited by sparse knowledge of their targets. DNA binding is a core aspect of the function of any TF, but in the case of the TALE proteins we only have GW data about their binding either *in vitro* or in cell lines. In this thesis, we aimed to characterize TALE protein *in vivo* DNA binding requirements as deeply as possible. We divided this aim into the following concrete objectives:

- Identify TALE protein binding sites in the genome during embryogenesis, as the necessary first step.
- Characterize the binding sites with regards to their epigenetic and conservation features.
- Characterize the DNA sequences bound by the TALE proteins *in vivo*.
- Characterize the transcriptional effect of TALE protein binding.
- Characterize the combinatorial binding of the TALE proteins.

Three

Materials & Methods

3.1 ChIP-seq

3.1.1 Chromatin Immunoprecipitation

Chromatin immunoprecipitations were performed using standard methods with anti-Prep1/2 antibody (N15, Santa Cruz Biotechnology, Santa Cruz, USA), anti-Pbx1 antibody (4342) (Cell signaling technology, Beverly, USA) and a mix of anti-Meis antibodies (K830, recognizing Meis1a, Meis2a and Meis2b isoforms, and K844, recognizing Meis1a and Meis1b isoforms; both produced at CNIC, Madrid, Spain).

A single-cell suspension was prepared from E11.5 mouse embryonic body trunks (total embryo without head, tail and legs) by crushing them against a cell strainer. Approximately $5x10^7$ cells were used for each immunoprecipitation. Cells were cross-linked in complete medium (10% FBS) containing 1% formaldehyde for 10 min, and the reaction was terminated by addition of 125 mM glycine. Fixed cells were washed three times (5 min each) in cold PBS and lysed in LB1 buffer containing 0.5% NP-40 and 0.25% triton X-100. Nuclei were then washed in LB2 buffer (containing 10mM Tris-HCl pH=8 and 200mM NaCl) to remove detergents and resuspended in LB3 buffer, containing 0.1% Na-deoxycholate and 0.5% N-laurylsarcosine. Chromatin was sonicated by 5 x 30 sec cycles at 30% of the maximum power of a Branson 450 sonicator to generate 100-400 bp chromatin fragments. After clearing by centrifugation, sonicated chromatin was incubated with antibody-bound protein A-conjugated magnetic beads (Invitrogen, Carlsbad, USA). For each IP we used 10 µg antibody. Rabbit IgG IP was performed as negative control. After overnight immunoprecipitation at 4°C the bound complexes were washed twice in WB1 (50 mM Hepes-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton-X100, 0.1% Na-doexycholate), twice in WB2 (50 mM Hepes-KOH pH 7.5, 500 mM NaCl, 1 mM EDTA, 1% Triton-X100, 0.1% Na-doexycholate) and twice in LiCl WB (10 mM Tris-Cl pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate, 1 mM EDTA). Immunoprecipitated complexes were eluted from the beads by incubating for 30 min in EB (2% SDS in TE) at 37°C. The eluted material was reverse cross-linked at 65°C overnight and incubated for 1 h at 55°C with proteinase K. The obtained material was extracted with phenol-chloroform and ethanol-precipitated. After RNAse treatment, the DNA was purified with a Polymerase Chain Reaction (PCR) purification kit (Qiagen, Netherlands). About 10 ng of immunoprecipitated DNA were processed for sequencing.

3.1.2 Sequencing and Alignment

Chromatin-immunoprecipitated DNA was sequenced using an Illumina GAII analyzer. Single-end 36bp reads (120M for Meis and 40M for Prep) were first mapped

with BWA software (H. Li and Durbin 2009) against mm9 version of the mouse genome.

3.1.3 Peak Calling

Peak-calling was performed with the Probabilistic Inference for Chip-Seq (PICS) algorithm (X. Zhang et al. 2011). The algorithm was used to identify genomic regions with a high density of reads (peaks). These regions are an indicator of the enrichment of immunoprecipitated DNA fragments for the TF of interest. PICS has been demonstrated to perform well against other peak-calling algorithms such as Model-based Analysis of ChIP-Seq (MACS) (Y. Zhang et al. 2008) and CisGenome (Ji et al. 2011).

3.2 ChIP-re-ChIP

For ChIP-reChIP the complexes were immunoprecipitated with anti-Pbx1 Ab, anti-Meis (a 1:1 mix of anti-Meis1 and anti-Meisa) Ab or rabbit IgG (negative control) and washed twice in WB1 and twice in WB2. They were then eluted from the beads by incubating for 30 min at 37 °C in reChiP elution buffer (2% SDS in TE supplemented with 15 mM DTT). The obtained material was diluted 20-fold in ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris–HCl, pH 8.0, 167 mM NaCl) and used for the second IP with anti-Pbx1, anti-Prep1/2, anti-Meis antibodies or rabbit IgG. The second immunoprecipitation was performed as described above.

3.3 Antero-Posterior ChIP

For comparison of Meis site occupancy between positions along the anterior to posterior axis of the embryo, E11.5 embryos were dissected according to the diagram in Figure 4.23. ChIP was performed as described on nuclear extracts from each of the portion and the resulting chromatin de-crosslinked for PCR testing.

3.3.1 Primers

Proportionality of product amount to input concentration was tested along a range of input DNA amount and number of cycles. 30 to 35 PCR cycles were chosen as the optimal proportionality point and used for occupancy testing. Some primers were found not to produce a satisfactory proportionality between input and output and results discarded.

3.3.2 PCR Conditions

Primers were designed for all Meis peaks in the Hox clusters using primer-BLAST (`http://www.ncbi.nlm.nih.gov/tools/primer-blast/`)and targeting amplification conditions of 60°C and product length around 100pb.

3.4 RNA extraction

For RNA-seq, total RNA was purified from whole E11.5 mouse embryos (*Prep1*$^{i/i}$, Meis1 KO and wt control littermates of each). mRNA was purified and the library for Illumina chromatin sequencing prepared according to the Illumina recommendations.

3.4.1 Sequencing, Alignment and Fold Change estimation

mRNA samples were sequenced using the paired-end 50bp protocol. Reads (7M per sample) were mapped and transcript expression estimated using RSEM (B. Li and Dewey 2011). This program aligns the reads against a set of predefined transcripts (in our case mouse ensemble 63 genebuild) and uses an expectation maximization algorithm to assign reads probabilistically to one of the isoforms of a given gene. The quantification results from RSEM were then analyzed with the Bioconductor package DESeq (Anders and W. Huber 2010), which fits a negative binomial distribution to estimate technical and biological variability.

3.5 EMSA

Nuclear extracts were isolated from cells prepared from E11.5 mouse embryonic body trunks as described (Longobardi and Blasi 2003). Briefly, cells were washed twice with cold PBS, collected in 500 µl cold buffer A (10 mM HEPES, pH 7.9, 10 mM KCl, 1.5 mM MgCl2, 0.5 mM DTT, proteinase inhibitor cocktail), left for 10 min on ice, and lysed by adding Triton X-100 to a final concentration of 0.3%. Nuclear extracts were prepared by resuspending pelleted nuclei in 100 µl buffer C (20 mm HEPES, pH 7.9, 25% glycerol (v/v), 0.42 m NaCl, 1.5 mm MgCl2, 0.5 mM DTT, 0.5 mm EDTA, proteinase inhibitor cocktail) for 30 min on ice. The extract was cleared by centrifugation. EMSA reactions were performed in 20 µl reaction mix containing 10 mM Tris-HCl (pH 7.6), 0.5 mM EDTA 0.5, 0.5 mM EGTA, 5% glycerol buffer, 80 mM NaCl, 1µg poly-dIdC, 1 mM DTT, 5 µg nuclear extract and 30,000 cpm of the 32P-labeled probe. When indicated, we added unlabeled probe in 50X excess or specific antibodies (4 µg) against Prep1/2 (N15, Santa Cruz Biotechnology, Santa Cruz, USA), Meis (1:1 mix of K830 and K844) or Pbx1 (P20, Santa Cruz Biotechnology, Santa Cruz, USA). Reactions were carried out for 30 min at

room temperature, and the complexes resolved by 5% non-denaturing PAGE. The gel was dried and scanned in a phosphorimager. The following double-stranded oligonucleotides were used, in addition to mutant versions as indicated in figures and text:

Meis Peak 586 – HEXA:
5′-AGGGAAGAGCCTGACAGATGACAGTTTCGAAAAA-3′
Meis Peak 313 – OCTA:
5′-CAAATAACTGATTGATTGCGGTCGAGGCACATTG-3′
DECA: 5′-GGCCTCGTGATTGACAGGCTCGCCG-3′
Meis Peak 1 – OCTA + HEXA:
5′-ATGCTGTGACAGTGATAAATGACGGTGCAGAA-3′
Meis Peak 595:
5′-AATTTAAGCAGTGATGAATGAGCTCGGCTG-3
Meis Peak 9:
5′-TTTGGGTGACAAAGATGAATGGTCTATTGT-3′
Meis Peak 54:
5′-GACAACCTCGCCTGTGATTGACCCCTGGAGTGG-3′
Prep Peak 1440:
5′-AGGGGCCCGTGATTGACAGGCTGAACTACAGACT-3′
Meis Peak 857:
5′-TTGCTGACAACTGACTGATAAATTATTTTCTGCT-3′

3.6 Bioinformatics

3.6.1 Motif Discovery, FIMO

For the identified peaks, *de novo* motif discovery was run to identify consensus sequences enriched in the selected regions versus the whole genome using rGADEM (L. Li 2009). While similar in its model to previous de novo motif finder algorithms such as MEME (Bailey, N. Williams, et al. 2006), GADEM performs better in large ChIP-seq experiments. Interestingly, rGADEM can identify dimer motifs located close to each other. The algorithm reports probability weight matrices (PWM) summarizing the consensus sequences identified in the peaks. (iii) We identified, interpreted and annotated de novo motifs using MotIV (motif identification and validation). MotIV compares the PWM of the identified de novo motifs with those in the JASPAR TF database (`http://jaspar.genereg.net`) and provides a significance level for the similarity. MotIV can furthermore report the distribution of distances between pairs of motifs and the percentage of co-localized motifs within the same peak, which might suggest cooperation between TFs. Steps (ii) and (iii)

of the pipeline were run for the peaks identified for each TF separately and also for peaks found for two or more TFs.

Individual instances of the core motifs within all peaks were searched with a local install of the FIMO (Find Individual Motif Ocurrences) program from the MEME suite (Bailey, Boden, et al. 2009) with default parameters (individual p-value cutoff 10^{-4}). The resulting files were parsed and the data added to the peak tables using a custom Python script.

3.6.2 Exploratory Data Analysis

Peak overlapping, correlation with RNA-seq data and conservation data aggregation were performed on the Galaxy bioinformatics platform (Blankenberg et al. 2010; Goecks et al. 2010). The set of peak coordinates for each factor was intersected with the sets for the other factors, the nearest transcript expression data were added to the tables, and the average conservation value (measured as the PhastCons 30-way vertebrate score (Siepel et al. 2005) was aggregated over the length of each peak with the aggregate genomic scores tool.

GC percentage count and literal motif match finding were performed using custom Python scripts (`http://www.python.org/`) run over the fasta sequence of the peaks.

Genomic profiling of peaks was performed using custom Python scripts and the CEAS (Cis-regulatory Element Annotation System) tool (Shin et al. 2009) within the Cistrome package (T. Liu et al. 2011). Peak association with regulated genes was estimated by calculating peak density per megabase in all Ensembl v63 nuclear genes, their promoters and close intergenic regions and comparing with the corresponding values for genes up- or down-regulated in Meis1 KO and *Prep1*$^{i/i}$ embryos.

3.7 Statistical Testing

3.7.1 Random Expectations

For Octameric Motif (OCTA) motif variants, we considered the expected number of instances of a motif containing i G/Cs and j A/Ts to be $I_{variant} = Nx\left(\frac{P(GC)}{2}\right)^{i} x\left(\frac{P(AT)}{2}\right)^{j}$, where I is the number of instances expected, N is the number of bases in the collection and GC and AT are the GC and AT fractions in the collection.

For the number of co-regulated genes, we considered a random expectation the intersection of each of 4 pairs two independent sets.

3.7.2 Tests Used

We estimated the significance of OCTA motif variants with a Poisson test in which λ equals the random expected number of instances.

Overlap between TALE factors and Hoxes was tested with χ^2 tests of independence applied to the relevant contingency tables, as was co-regulation of misregulated genes in MEis1 and Prep1 deficient embryos.

Intron-exon ratio of TALE peaks versus GW intron/exon ratio was evaluated with an exact binomial test.

Four

Results

4.1 Characterization of TALE protein ChIP-seq Peaks

4.1.1 Immunoprecipitation and Sequencing

ChIP was performed in the dissected trunks of embryonic stage E11.5 C57BL/6 mice. The trunk and anterior head was selected as input material because this is the main domain of Meis and Hox expression. ~120 million reads were mapped for Meis and ~40 million each for Prep and Pbx1 ChIP-seqs.

The antibodies used recognize most of the TALE protein variants found in the embryo. For Meis, we used an antibody raised against the N-ter of Meis1 that recognizes both short and long C-ter isoforms of Meis1 combined with an antibody raised against the C-ter portion of Meis that recognizes the short isoforms of both Meis1 and Meis2 (Meis1a, Meis2a and Meis2b). For Prep1, we used an antibody that recognizes both Prep1 and Prep2. For Pbx, we used an antibody that only recognizes Pbx1, so our study is not comprehensive for Pbx proteins.

4.1.2 Peak Calling

We initially attempted to call peaks from read data with the widely used MACS algorithm (Y. Zhang et al. 2008). However, precision was very low, i.e. peaks were about 2000pb wide. Visual inspection of the read profile revealed that, even when the enrichment was clearly delimited and an obvious peak was visible, the peak regions called by MACS extended far beyond the summit and even the base of the peak. The mean peak profile for the highest enriched peaks is generated by MACS in the process of defining an "archetypal" peak, which is later used as a reference to score read accumulations. This profile, in our case, included obviously extraneous adjacent regions where the read density was at genomic background levels. No combination of parameters could refine the peak calling. We speculate that this glitch might be due to the read numbers used in contemporary deep sequencing experiments, much higher than those common when MACS was designed.

We then used the PICS algorithm (X. Zhang et al. 2011) for calling peaks from read data. PICS produced much more precise peak calls than MACS. These were used for all following analyses.

4.1.3 Overlap of Meis, Prep and Pbx1 peaks

5686 peaks were found for Meis, 3331 for Prep, and 3504 for Pbx1, for a total of 10326 non-redundant genomic regions. The overlaps between the three sets are represented in Figure 4.1. The first conclusion from this analysis is that Meis overlaps much less with either Prep or Pbx1 than Prep and Pbx1 with each other. 4837 of 5686 (85%) Meis peaks are Meis-Exclusive ($Meis^{exc}$). Of the remaining 15%, 186

Figure 4.1: **Overlap of Meis, Prep, and Pbx1 peaks.** Venn diagram of the number of Meis^{exc},Prep^{exc}, Pbx1^{exc}, Meis-Prep^{com}, Meis-Pbx1^{com} and Meis-Prep^{com} peaks. Areas of the circles are proportional to total peak numbers for Meis (red), Prep (blue) and Pbx1 (green) peaks. Intersection areas are roughly proportional to the number of Meis-Prep^{com}, Meis-Pbx1^{com}, Meis-Prep^{com} and triple peaks.

(3.3%) are Meis-Prep Common (Meis-Prep^{com}), 444 (7.8%) were Meis-Pbx1 Common (Meis-Pbx1^{com}), and 222 (3.9%) are bound by all three factors.

Prep and Pbx1 peaks overlap much more with each other than with Meis. Of 3331 Prep peaks, 1900 (57%) are Prep^{exc}, 186 (5.6%) are Meis-Prep^{com}, 1023 (30.7%) are Pbx1-Prep Common (Pbx1-Prep^{com}) and the 222 triple peaks represent the remaining 6.7%. For the 3504 Pbx1 peaks, 1821 (52%) are Pbx1^{exc}, 444 (12.7%) are Meis-Pbx1^{com}, 1023 (29.2%) are Pbx1-Prep^{com} and triple peaks represent 6.3%.

Triple peak Chip-reChip

Triple peaks showed the most obvious immediate interest, since they might represent Cis Regulatory Modules (CRMs) integrating inputs from all three factors. The experimental result that three different factors are able to bind DNA on the same genomic site can have two interpretations. The factors might bind simultaneously in the form of a complex or to very close BSs on the same CRM. The other possibility is that, since our input chromatin is a nuclear extract from many different cell types, the factors might bind alternatively, maybe competing for the same target sequences.

To distinguish between the two possibilities, we performed ChIP-re-ChIP experiments. In these, the output of a ChIP is used as input for a further round of ChIP with antibodies against a different protein. In this way, the resulting DNA

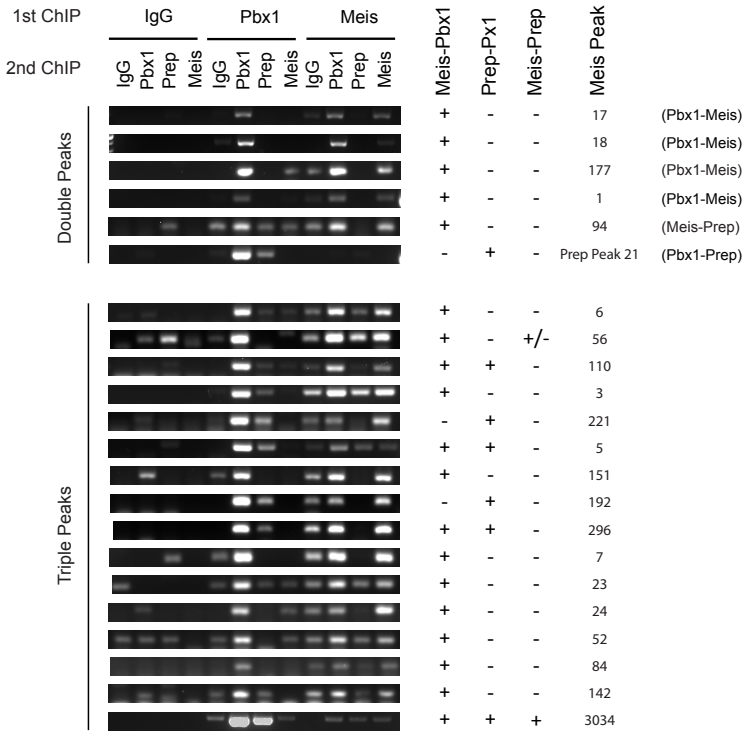

Figure 4.2: **ChIP-re-ChIP experiments.** Factors were subjected to succesive round of ChIP with antibodies with different specificities, as indicated. Unspecific antibody (IgG) was used as a negative control. Purified DNA was amplified in a rate-limiting PCR to obtain an amount of product roughly proportional to the peak DNA present. Band intensities over IgG band were considered as positive signals.

will be enriched specifically in fragments bound by the two proteins simultaneously. Results are shown in Figure 4.2.

We tested 6 double peaks and 16 triple peaks. All of them showed simultaneous binding by at least two factors. The most frequent co-binding combinations we saw were either Pbx1+Prep (10/17) or Pbx1+Meis (17/21). Of the 16 triple peaks we tested, we saw Meis+Prep co-binding in only 2 (Meis peaks 56 and 3034).

These results suggest that Prep and Meis very rarely bind to the same genomic site at the same time. Therefore, the 222 triple peaks probably represent instances of competition between Meis and Prep for binding to Pbx1 and DNA.

4.1.4 Position of Bindings Sites Relative to Transcription Start Sites

The first step in the genomic characterization of peaks was to find the TSS nearest to each peak. The gene set we used for this and all following gene-centric analyses was Ensembl63, the latest version available at the time of mapping.

Distance between peak summits and TSSs was calculated using the TSS as position 0 and direction of transcription as + direction, in order to differentiate upstream and downstream peaks. Results are represented in Figure 4.3. We found a clear concentration of peaks near TSSs for both Prep and Pbx1. Meis peaks, in contrast, show very little concentration near TSSs, and a much higher proportion of them is located at extreme distances from TSSs.

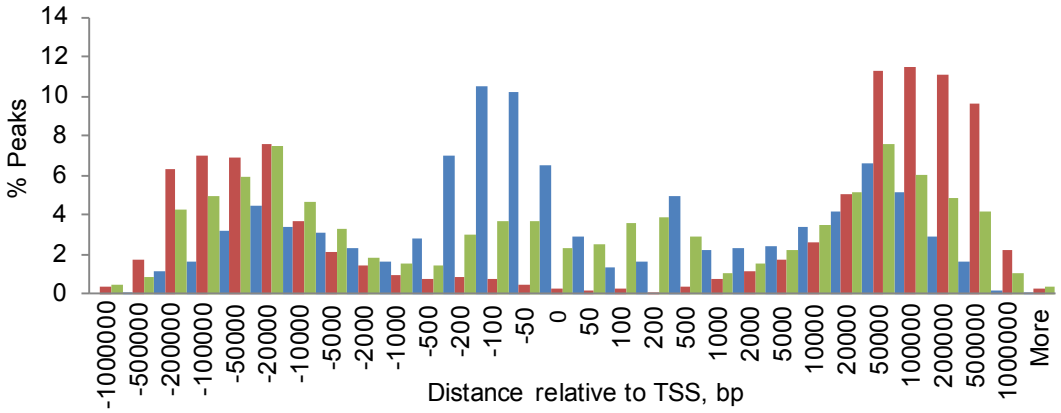

Figure 4.3: **Distance From Peaks to TSSs.** The percentage of peaks in each category over total factor peaks is represented in the Y axis. X axis categories are labelled with the right limit of the interval, in bp. Red, Meis; Blue, Prep; Green, Pbx1.

4.1.5 TSSA, IG, CI, FI

In order to understand better the interplay between peak overlap and distance to transcriptional units, we classified peaks into Transcription Start Site-Associated (TSSA), Intragenic (IG), Close Intergenic (CI) or Far Intergenic (FI). The criteria for the definition of TSSA were based on the results represented in Figure 4.3, in which the enrichment in Prep peaks near TSSs is very clearly delimited between -500 and +100 bp. Therefore, peaks located between -500 and +100 were considered to be TSSA. Of the rest, those within the gene body were considered IG. We subdivided intergenic peaks into CI (those outside the transcriptional unit but within 20kb of the nearest TSS) and FI (further than 20kb away from the nearest TSS) because of the difficulty of assigning potential transcriptional targets to peaks that are at extreme distances from the nearest TSS.

The results of this classification are represented in Figure 4.4. There are obvious differences specially in the proportion of TSSA peaks bound by different combinations of factors. Meis and Pbx1 bind TSSs very rarely, either alone or in combination. In contrast, over 30% of Prep^{exc} peaks are located near a TSS, a proportion that rises to 71.5% when bound in combination with Pbx1 (Pbx1-Prep^{com} peaks).

Meis-Prep^{com} peaks show a distribution intermediate between those of Meis^{exc} and Prep^{exc} peaks.

Intergenic peaks form the majority of Meis^{exc} peaks: 45.7% of Meis^{exc} peaks are FI. Pbx1^{exc} and Meis-Pbx1^{com} peaks show a profile very similar to that of Meis^{exc} peaks. Triple peaks have a flat distribution in these four categories.

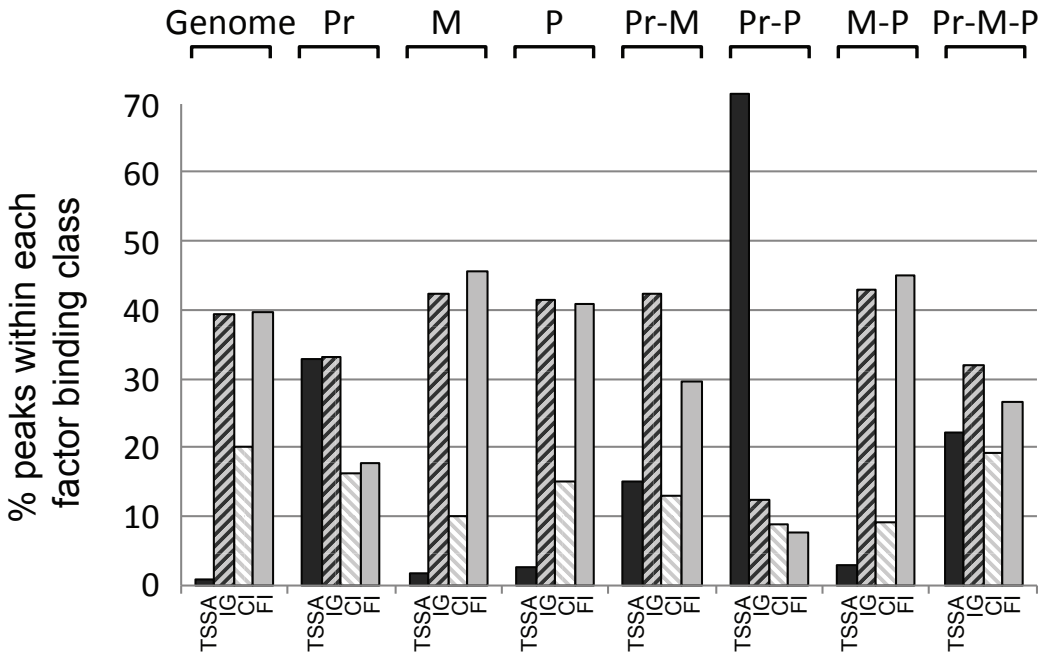

Figure 4.4: **Genomic location classes within each factor-binding profile category.** The peaks for each combination of binding factors were classified into TSSA, IG, CI or FI categories. The resulting profile is shown here. ALL = total non-redundant genomic sites. Pr = Prep^{exc}, M = Meis^{exc}, P = Pbx1^{exc}, Pr-M = Meis-Prep^{com}, Pr-P = Pbx1-Prep^{com}, M-P = Meis-Pbx1^{com}, Pr-M-P = triple peaks.

Intron/Exon Ratio of Meis Peaks is skewed

We considered the IG peaks. 39.3% of the mouse genome is covered by introns while 2% is covered by exons, which results in an intron/exon ratio of 19.65. The corresponding percentages for Meis (39.9% versus 0.5%), Prep (24.9% versus 1.6%) and Pbx1 (31.4% versus 1.6%) work out to ratios of 79.8, 15.56 and 19.63, respectively. The difference was highly significant for Meis (p-value $< 10^{-20}$) but not for Prep or Pbx1 (both p-values over 10^{-1}).

4.1.6 Conservation Profiling

We analyzed the evolutionary conservation profile of peaks. For this analysis, the peak summit was used as position 0 and the phastCons (Siepel et al. 2005) score was averaged for equivalent positions of all peaks of each subset. PhastCons is a measure of phylogenetic conservation that varies between 0 and 1 and roughly represents the probability that a given base is a conserved element. PhastCons is estimated from a hidden Markov model based on multiple sequence alignments and is available precomputed in several genomic databases. The version we used was 30-vertebrate phastCons.

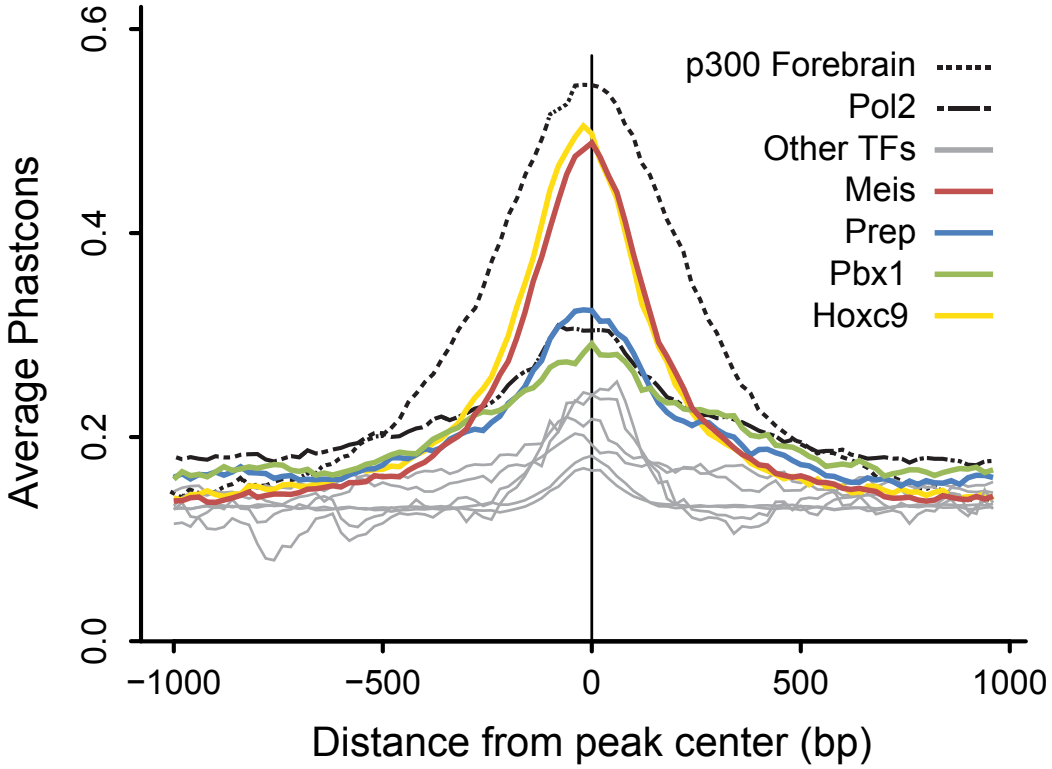

Figure 4.5: **Conservation Profile of Peaks.** PhastCons values at each position from summit were averaged across all peaks in a subset and plotted. Peaks for Meis, Prep, and Pbx1 are from this study. Peaks for Hoxc9 are from Jung et al. 2010. Peaks for p300 in mouse embryonic forebrain were extracted from Blow et al. 2010. Peaks for PolII are from Mahony et al. 2011. Other TFs in grey are CEBPA and HNF4A (Schmidt et al. 2010), RAR (Mahony et al. 2011) and FLAG-tagged Meis1 and Hoxa9 in a leukemia cell line (Y. Huang et al. 2012).

The resulting profile shows that Prep and Pbx1 peaks have a high conservation value near the summit that tails off to genomic average values. The maximum conservation value and the profile are very similar to TSSs marked by RNA Polymerase II (PolII) peaks from Mahony et al. (2011), which together with the previously mentioned high proportion of TSSA peaks within Prep and Pbx1 suggests a significant portion of these may be promoters.

Meis peaks have a much higher conservation profile, approaching that of genomic regions bound by p300 in the embryonic forebrain. These p300 peaks represent the most conserved enhancers (Blow et al. 2010). The only other TF peaks approaching this level of conservation that we have observed are, notably, those for Hoxc9 in the embryonic spinal cord (Jung et al. 2010). Thus, genomic loci bound by Meis are far away from TSSs but highly conserved, which strongly suggests functional relevance.

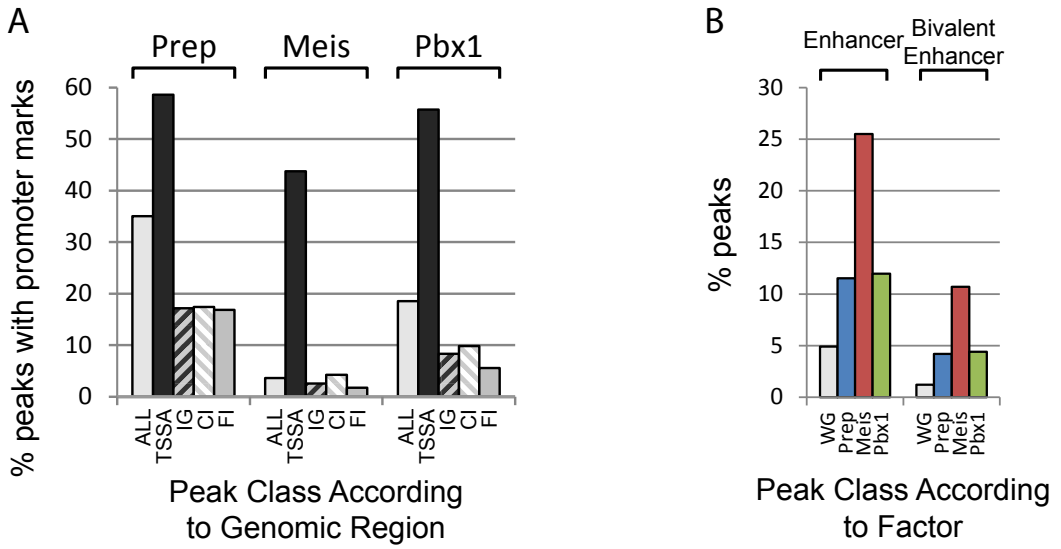

Figure 4.6: **Histone Marks in Peak Subsets.** A) percentage of peaks in each genomic binding class that show promoter (RNApolII$^+$; H3K4Me3$^+$) epigenetic marks. B) percentage of peaks for each factor that show enhancer (H3K4Me1$^+$; H3K4Me3$^-$) or bivalent enhancer (H3K4Me1$^+$; H3K27Ac$^+$) histone marks.

4.1.7 Histone Marks

In order to study the correlation between the peak sets identified and known functional genomic regulatory elements, we checked their association with histone post-translational modifications. Different histone marks are associated with active and inactive promoters and enhancers, and can therefore be used to identify functional elements (Mikkelsen et al. 2007). In particular, RNApolII$^+$ H3K4Me3$^+$ marks promoters, H3K4Me1$^+$ H3K4Me3$^-$ marks enhancers and H3K4Me1$^+$ H3K27Ac$^+$ marks bivalent (poised) enhancers (Creyghton et al. 2010).

We took advantage of the extensive, high-quality epigenetic data generated by the Ren lab at LICR for the ENCODE project (Y. Shen et al. 2012). We chose the source tissue that most closely resembled our sample, Mouse Embryonic Fibroblasts (MEFs).

For all three factor peak sets, TSSA peaks showed a high level of overlap with promoter epigenetic marks (between 43.8 and 58.6% of all TSSA peaks versus 0.38% of the genome tagged RNApolII$^+$ H3K4Me3$^+$). Only Prep non-TSSA peaks showed high association with promoter marks (around 17% for all non-TSSA categories). Meis non-TSSA peaks had low levels of coincidence (between 1.7 and 4.2%) and Pbx1 non-TSSA peaks had levels intermediate between those of Meis and Prep (between 5.6 and 9.8%) (Figure 4.6a). Thus we conclude that TSSA peaks likely represent *bona fide* promoters. In addition, many non-TSSA Prep peaks could in fact be either undescribed or cryptic promoters or simply promoters not present in

the gene catalogue we used.

In contrast with promoter marks, enhancer (H3K4Me1$^+$ H3K4Me3$^-$) and bivalent enhancer (H3K4Me1$^+$ H3K27Ac$^+$) marks are much more associated with Meis peaks than with Prep or Pbx1 peaks. 25.5% of Meis peaks bear enhancer marks.

4.2 Sequence analysis

Beyond the genomic location of the BSs of our factors of interest, we sought to characterize the sequences they select *in vivo*, and then contrast the preferences of each factor and compare them to their *in vitro* behaviour.

4.2.1 Motif discovery

We performed motif discovery on the peak sequences. In short, motif discovery is a way to discover short sequences that are more frequent in a sequence dataset than would be expected by chance. We used the rGADEM software, an implementation of the GADEM (L. Li 2009) algorithm in the R statistical computing language that integrates smoothly with PICS and produces a detailed and readable output.

We ran rGADEM with default parameters on the sequence of non-overlapping subsets of peaks defined by their factor combination: Meis^{exc}, Prep^{exc}, Pbx1^{exc}, Meis-Prep^{com}, Meis-Pbx1^{com}, Meis-Prep^{com} and triple peaks, separately. The results are shown in Figure 4.7.

4.2.2 Core Versus Accessory Motifs

The representation of motif distances to the peak summit shows a fairly clear distinction between two different motif classes: "core" motifs that tend to be located near the summit of the peak and therefore shown a unimodal distance distribution, and "accessory" motifs that are depleted near the summit of the peak and thus show a bimodal distribution.

Core motifs are likely to be *bona fide* factor recognition sequences. The fact that they tend to be located at or close to the maximum confidence point for factor binding, the summit, suggests that they are directly responsible for this binding.

In contrast, accessory motifs may be interpreted in various ways. Accessory motifs might represent BSs for cofactors that collaborate in the binding. Alternatively, they could represent local DNA compositional biases that facilitate the binding of the factors without interacting with them, for example by imparting a certain curvature or groove topology to the stretch of DNA. They might also be simply artifacts: if a subset of peaks is located preferentially in genomic regions with particular DNA sequence characteristics, one would expect those characteris-

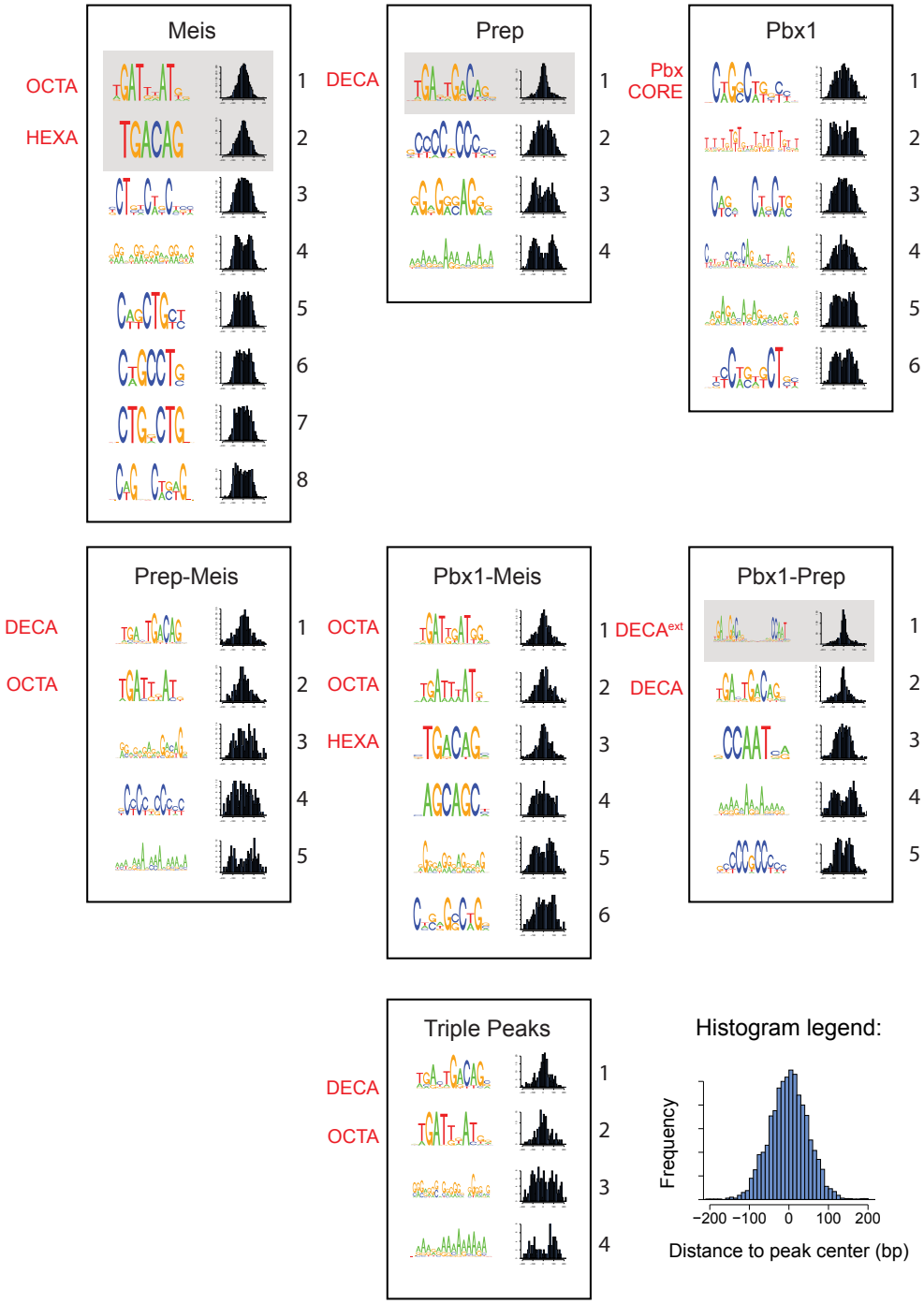

Figure 4.7: **All Motifs Found in Peak subsets.** Sequence logos for every motif found in each non-overlapping peak subset. "Core" motifs labelled in red. The logos resulting from the motif definitions (PSSMs) which we selected as canonical and used for directed motif search are highlighted in grey. The histograms to the right of the sequence logo show the distribution of motif positions with respect to the summit of the peak.

tics to be reflected in the results of motif discovery, although they might not have a direct effect on factor binding.

4.2.3 Core Motifs

Most of the core motifs we found are immediately recognizable from the literature on Hox cofactor DNA binding preferences and were identified in several peak subsets. The motif we call Hexameric Motif (HEXA) corresponds to the canonical monomeric Meis/Prep BS and is detected in the Meis^{exc} and Meis-Pbx1^{com} subsets, but surprisingly not in the Prep^{exc} or Meis-Prep^{com} subsets. The Octameric Motif (OCTA) fits the well-known Pbx-Hox heterodimeric BS and is present in all subsets with Meis binding, even in those peaks in which we have not found evidence of Pbx binding, such as Meis^{exc} and Meis-Prep^{com}. The Decameric Motif (DECA) clearly resembles a heterodimeric Pbx-Meis or Pbx-Prep BS with a 5′ TGAT Pbx part and a 3′ Prep/Meis TGACAG part. We found an additional previously unreported bipartite motif that contains a DECA and a CCAAT sequence at a fixed distance, which we have called Extended Decameric Motif (DECA^{ext}). DECA^{ext} only appears in motif discovery on Pbx1-Prep^{com} peaks. In Pbx1^{exc} peaks we only found one low complexity core motif that did not match any found in the literature for Hox or TALE proteins, but might be related to accessory motifs found in other subsets (see below). Because of its poor definition, we discarded it from further analysis.

4.2.4 Accessory Motifs

Several intriguing sequences appear as accessory motifs. All subsets show variants of a poorly defined sequence composed of Gs and As in no particular order: motif 4 in Meis^{exc}, 3 and 4 in Prep^{exc}, 5 in Pbx1^{exc}, 3 in Meis-Pbx1^{com}, 5 in Meis-Pbx1^{com}, 4 in Pbx1-Prep^{com}, 3 and 4 in triple peaks (Figure 4.7). It is interesting to note that this represents DNA stretches in which one strand is exclusively populated by purines and the complementary strand is only populated by pyrimidines.

Variants of a CTGnCTG sequence appear in many of the subsets. In Meis^{exc} peaks, motifs 3 and 5 to 8 all fit this consensus, as do motifs 1, 3 and 6 from Pbx1^{exc} peaks. In the Meis-Pbx1^{com} subset, motif 6 and the reverse complement of motif 4 could represent variants of this consensus. This motif was found in a previous report (Tijssen et al. 2011) and described as a "motif resembling an E-box". E-boxes are binding sites for bHLH proteins, and the myogenic bHLH proteins have been reported to be able to bind DNA with Pbx-Meis heterodimers (Berkes et al. 2004; Knoepfler, Bergstrom, et al. 1999).

Another of the accessory motifs appeared in several subsets and has suggestive connections in the literature. A CCCgCCC motif appears in the Prep^{exc}, Meis-Prep^{com}

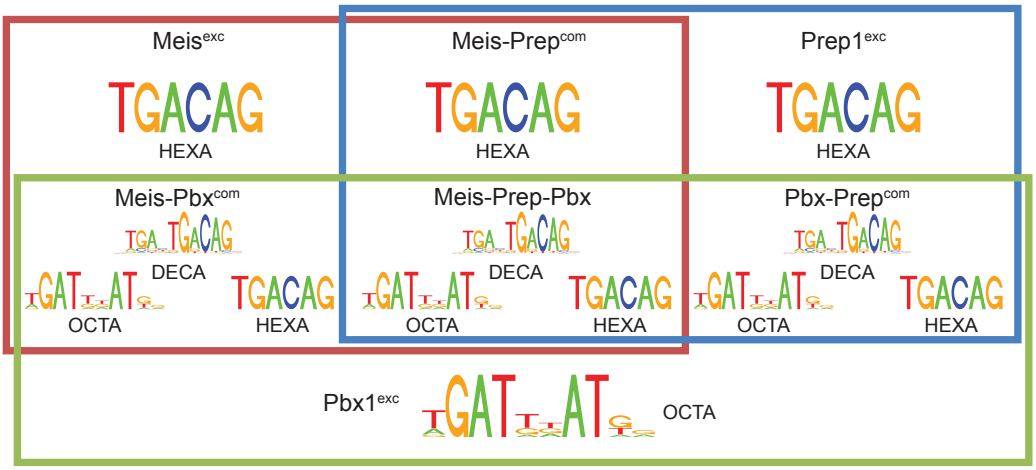

Figure 4.8: **"Core" Motifs Expected in Peak subsets.**

and Pbx1-Prep^{com} subsets. It is very similar to the well known Sp1 binding site (GGGGnGGGG, Briggs et al. 1986).

4.2.5 Comparison with Expected Distribution of Motifs

Going forward, we focused on the core motifs. Their distribution across the subsets defined by TALE protein binding was unexpected and reveals previously unnoticed specificities in TALE protein *in vivo* site selection.

Our *a priori* expectation for the distribution of motifs, based on the extensive literature on *in vitro* binding of the TALE factors, is represented in Figure 4.8. We expected the HEXA motif to be prevalent in all subsets in which no Pbx1 was detected: Meis^{exc}, Prep^{exc}, and Meis-Prep^{com}. We expected the OCTA motif to be present in all subsets with Pbx1 binding, and the DECA motif to be present in the intersection of Meis/Prep with Pbx1. These expected motifs could have included additionally all TGAT-containing motifs in Meis or Prep subsets, if many of those peaks not bound by Pbx1 were instead bound by other Pbx proteins.

However, the actual results (summarised in Figure 4.9) diverged from our expectations. We found that the OCTA motif is present in all Meis-positive subsets, included those that are Pbx1-negative. The DECA motif, conversely, is present in all Prep-positive subsets, even in those that show no evidence of Pbx1 binding. HEXA appears in both Meis^{exc} and Meis-Pbx1^{com}, but not in any Prep-positive subset.

4.2.6 Motif Frequencies

Motif discovery is a form of undirected search. Once having the motif definitions, we performed a directed search using the FIMO program from the MEME suite.

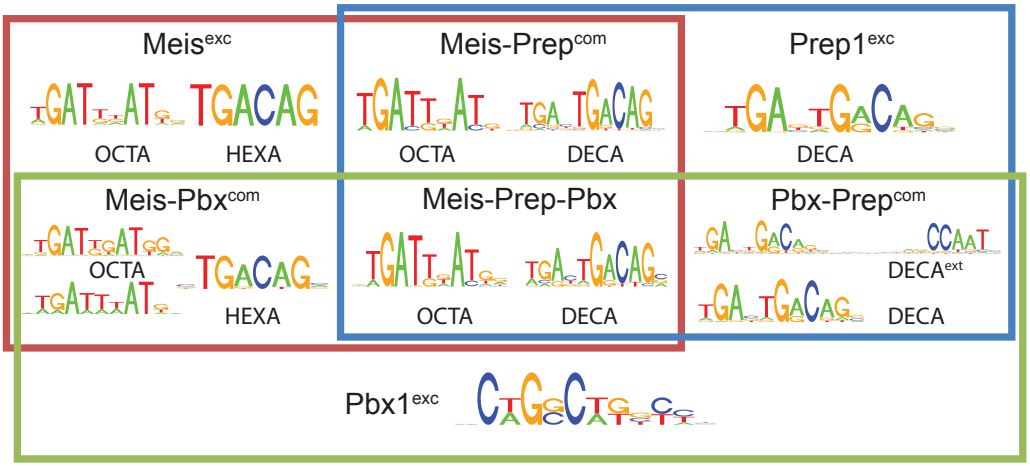

Figure 4.9: **"Core" Motifs Found in Peak subsets.** Sequence logos (Schneider and Stephens 1990) for each core motif found in each peak subset. In a sequence logo, the height of a letter is proportional to its frequency at that position and the height of the total stack at each position is proportional to the invariance at that position in the original alignment.

Directed search can reveal the presence of motifs in subsets were they are present but not frequent enough to be detected by motif discovery. It also allows us to quantify and compare motif frequencies.

The results are represented in Figure 4.10. HEXA is most frequent in triple peaks, but it is not the most frequent core motif in any peak subset. OCTA is the most frequent core motif in Meis^{exc} peaks, in triple peaks and in Meis-Pbx1^{com} peaks, where it is present in over 50% of the peaks. The special abundance of OCTA motif in Meis-Pbx1^{com} reinforces the idea that these sites may represent Pbx-Hox binding sequences. DECA and DECA^{ext} are very frequent in all subsets where Prep is involved. This is especially marked in Pbx1-Prep^{com} peaks, where HEXA and OCTA are present in under 10% of peaks but DECA^{ext} is present in over 70% of the peaks. In general, Meis peaks tend to be OCTA-positive and to a lesser extent HEXA-positive and Prep peaks tend to be DECA- or DECA^{ext}-positive, and this tendencies are exacerbated by co-binding with Pbx1.

4.2.7 EMSA on Core Motifs

To determine whether the identified core motifs are indeed capable of being bound by the TALE proteins tested, we performed Electrophoretic Mobility Shift Assay (EMSA) experiments on peak sequences containing each motif. In EMSA, the electrophoretic mobility of a radioactively labelled DNA probe is measured in the absence and presence of purified proteins or a nuclear extract. A change in mobility

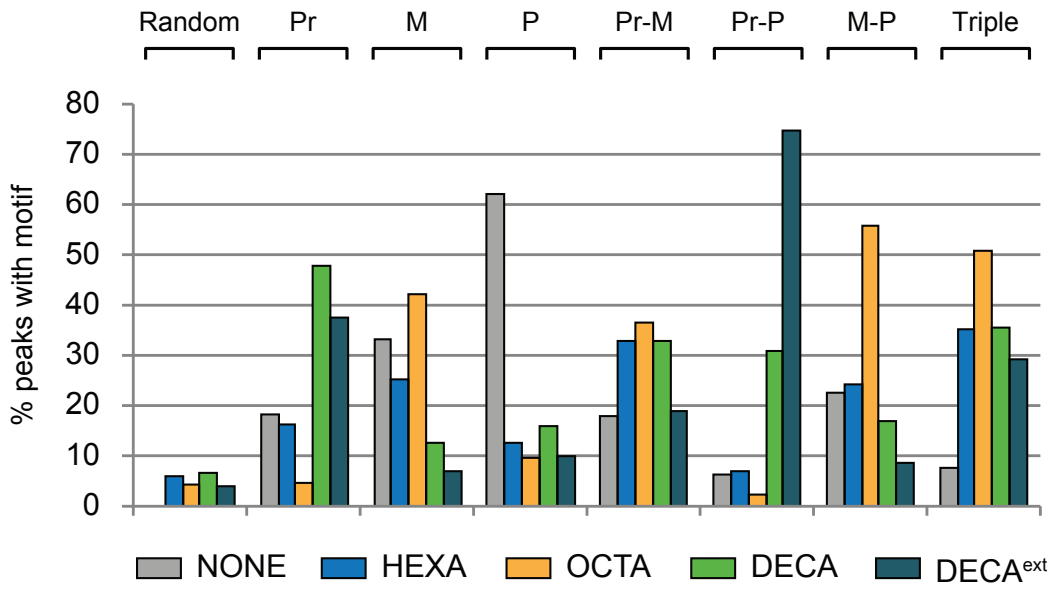

Figure 4.10: **Core Motif Frequency in Peak Subsets.** Percentage of peaks in each factor binding category that contain each core motif. Random represents average values in sets of 5686 sequences with length distribution equal to Meis peaks. Pr = Prep^{exc}, M = Meis^{exc}, P = Pbx1^{exc}, Pr-M = Meis-Prep^{com}, Pr-P = Pbx1-Prep^{com}, M-P = Meis-Pbx1^{com}.

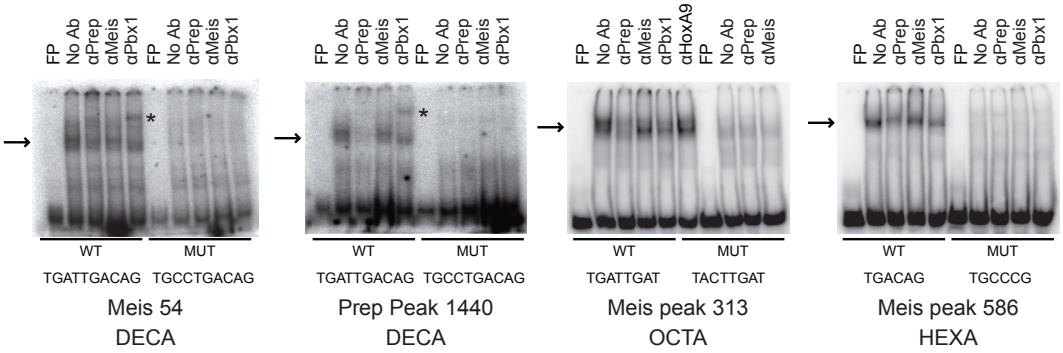

Figure 4.11: **EMSA Experiments on Core Motifs.** A) EMSA on the core motifs HEXA, OCTA, and DECA. Oligonucleotide sequences were taken from peaks containing the aforementioned motifs, and mutated versions of these were use to control specificity of binding. Arrows indicate complex bands and asterisks indicate super-shifted bands.

indicates protein binding to the probe. If an undefined protein mixture such as a nuclear extract is used, super-shift experiments can be performed in which a specific antibody will further shift or inhibit the complex if it recognizes its target.

The result of EMSA experiments on core motifs is shown in Figure 4.11. EMSA on probes from selected peaks showed that while Prep and Pbx can bind any of the core motifs identified, Meis can bind the HEXA and OCTA sequences but can only

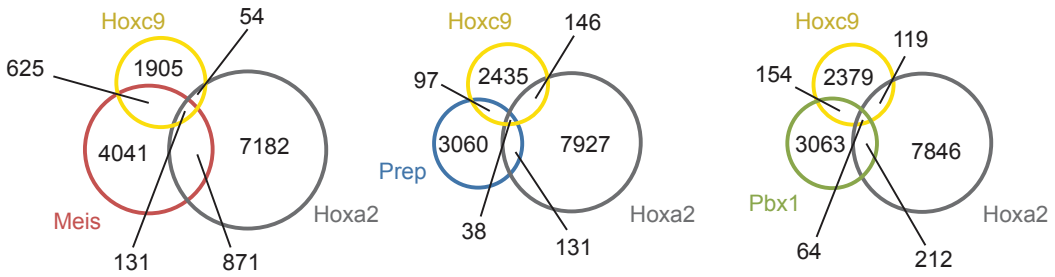

Figure 4.12: **Hox Peak Overlap.** Overlap between Meis (red), Prep (blue) or Pbx1 (green) peaks and Hoxa2 (grey) or Hoxc9 (yellow) peaks. Areas of the circles are proportional to total peak numbers. Intersection areas are roughly proportional to the number of common peaks.

weakly bind to the DECA sequence.

4.2.8 Hox Peak Overlap

Only two ChIP-seq vertebrate Hox protein binding datasets have been published that use antibodies against the endogenous proteins rather than tagged versions of them (Donaldson et al. 2012; Jung et al. 2010). These studies characterized binding of Hoxc9 in E13.5 embryonic mouse spinal cord (Jung et al. 2010) and of Hoxa2 in E11.5 mouse embryo branchial arches (Donaldson et al. 2012).

Since the motifs found in our peaks suggest a pre-eminent role for Meis as a Hox cofactor, we took advantage of the fact that the samples used are comparable to ours to contrast our datasets to the ones published. The results are summarised in Figure 4.12.

We found that, of the TALE factors we tested, Meis is the one that overlaps the most with known Hox GW binding site datasets. 28.9% of our Meis peaks overlap with either Hoxa2 peaks, Hoxc9 peaks, or both. The figures for Prep and Pbx1 are 8.1% and 12.6%, respectively. Moreover, the majority of Prep-Hox and Pbx1-Hox peaks are also Meis peaks: 185/271 (68.3%) for Prep and 348/440 (79%) for Pbx1.

Surprisingly, Meis peaks bearing the OCTA motif are not much more likely to overlap with Hox peaks than those that do not contain it: 842/2467 (34.1%) versus 801/3219 (24.9%), although this difference is significant (χ^2 p-value $= 2.46x10^{-14}$). In contrast, Pbx1 peaks with the OCTA motif are much more likely to overlap Hoxes: 245/554 (44.2%) versus 195/2950 (6.6%) (χ^2 p-value $= 1.04x10^{-132}$). The same is true of Prep peaks: 116/286 (40.6%) OCTA-positive peaks overlap Hoxes, versus 155/3045 (5.1%) that do not (χ^2 p-value $= 1.04x10^{-97}$).

These data support the notion that most Meis peaks represent actual targets of Hox binding and suggests that a subset of these could represent Meis-Hox interac-

tions in the absence of Pbx. Since binding sites of only two of the 39 Hox proteins in the mouse embryo cover almost 30% of Meis peaks despite the samples not being exactly equal, it is reasonable to hypothesize that additional Hox GW DNA-binding datasets will eventually reveal most Meis sites as Hox DNA binding sites. We have not seen a clear association of Meis binding sites with Hoxc9 rather than Hoxa2, despite the fact that Meis can bind Hoxc9 but not Hoxa2 directly. In addition to this, OCTA is more abundant than HEXA in Meis peaks, suggesting that Meis binding to Hox through Pbx interaction without itself binding DNA is the most frequent mode of meis interaction with Hoxes.

They also suggest that the most common cofactor for Hox proteins in the embryo is actually Meis, rather than Pbx. This is surprising since only Hox proteins in paralog groups 9 to 13 have been described to bind Meis directly, so other Hox proteins would depend on Pbx for Meis complex formation.

4.2.9 OCTA Motif Variants

The OCTA motif, which we hypothesize represents binding sites for Hox proteins, is an eight-nucleotide sequence with high variability in positions 5 and 6. Some of the 16 possible matches have been shown to be able to bind Hoxes. Specific variants are known to be preferentially bound *in vitro* by 3′ Hoxes (TGATTGAT), middle Hoxes (TGATTAAT) or more 5′ Hoxes (TGATTTAT) (W. F. Shen, Rozenfeld, Lawrence, et al. 1997; Slattery et al. 2011).

To study the correlation between TALE protein binding and OCTA motif variants, we quantified the representation of each of the 16 possible combinations of nucleotides at positions 5 and 6. To do this, we calculated the expected number of occurrences based on the nucleotide composition of the peak subset. Over-representation is thus the ratio of actual motif instances found by literal string matching to this expected number, for each OCTA variant and peak subset. Results are represented in Figure 4.13.

We see a clear pattern of over-representation. Most OCTA motif variants that have been described to be bound by Hoxes are over-represented in the non-redundant list of all peaks. Exceptions to this rule are WGATAGAT, WGATGCAT, and WGATCGAT. In contrast, all variants not described to be bound by Hoxes are not significantly over-represented except WGATGAAT.

We do not see or see very little enrichment over random expectation of OCTA variants in the subsets in which Meis is not involved (Prep^{exc}, Pbx1^{exc}, $\text{Pbx1-Prep}^{\text{com}}$). The only clear exception to this is the WGATTGAT variant in $\text{Pbx1-Prep}^{\text{com}}$ peaks. This lack of OCTA enrichment is especially surprising for Pbx1^{exc} peaks, which we expected would represent majoritarily Pbx-Hox binding.

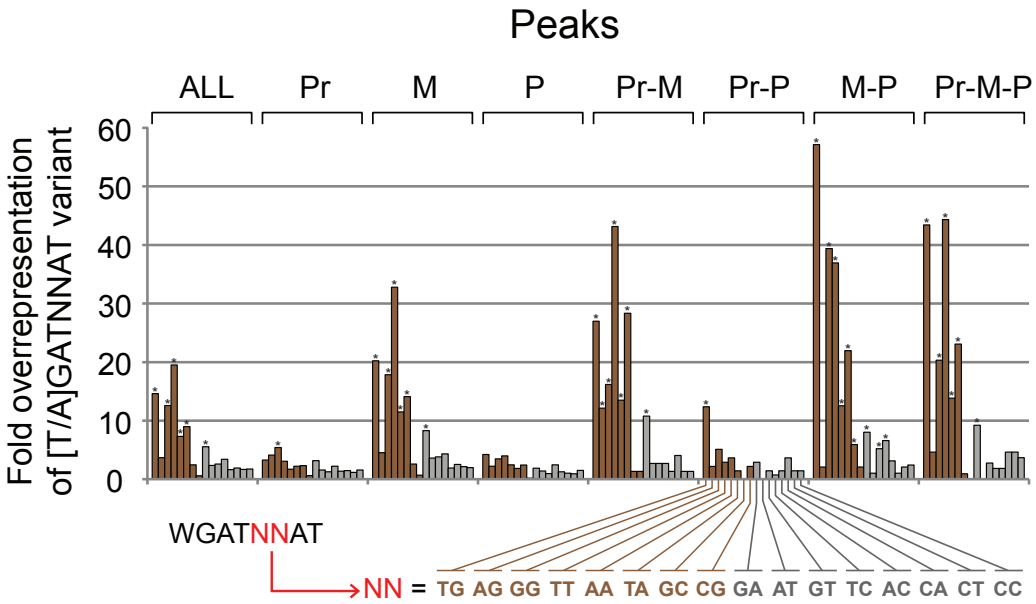

Figure 4.13: **OCTA Motif Variants.** Fold overrepresentation of A/TGATNNAT variants in peak subsets. Variants known to be bound by Hoxes are marked in brown, while those not known to be bound by Hoxes are marked in grey. Asterisks mark values of overrepresentation over 5 with Bonferroni adjusted p-values under 10^{-3}.

In contrast, all subsets in which Meis is involved show marked enrichment of at least some OCTA variants. In Meis^{exc} peaks, the most enriched variant is WGATTTAT, which appears over 30 times more frequently than would be expected by chance and is described in the literature as the preferred binding sequence for Hox paralog groups 6 to 10 in combination with Pbx. After WGATTTAT, the next OCTA variants most enriched in Meis^{exc} peaks are WGATTGAT and WGATGGAT. WGATTGAT is the preferred binding target for Hox paralog groups 1 to 5 in combination with Pbx *in vitro*. WGATGGAT has been described as a binding site in vivo in the *Hoxb1* rhombomere 4 ARE repeat 3. Also enriched in Meis^{exc} peaks are WGATAAAT, WGATTAAT (preferred target for paralogs 3 to 7) and WGATGAAT.

Meis-Prep^{com} peaks show a pattern of enrichment similar to Meis^{exc} peaks, but with higher levels of over-representation. The exception is the WGATAGAT variant, which is over-represented in Meis-Prep^{com} but not Meis^{exc} peaks.

The pattern of OCTA over-representation in Meis-Pbx1^{com} peaks is markedly different from that of Meis^{exc} peaks. General levels of enrichment are also higher. In these, the most enriched variant is WGATTGAT, appearing over 55 times more frequently than would be expected by chance. The next most enriched variants are WGATGGAT and WGATTTAT, and then WGATTAAT. In this subset, WGATAGAT is not significantly enriched. A number of undescribed variants are enriched at

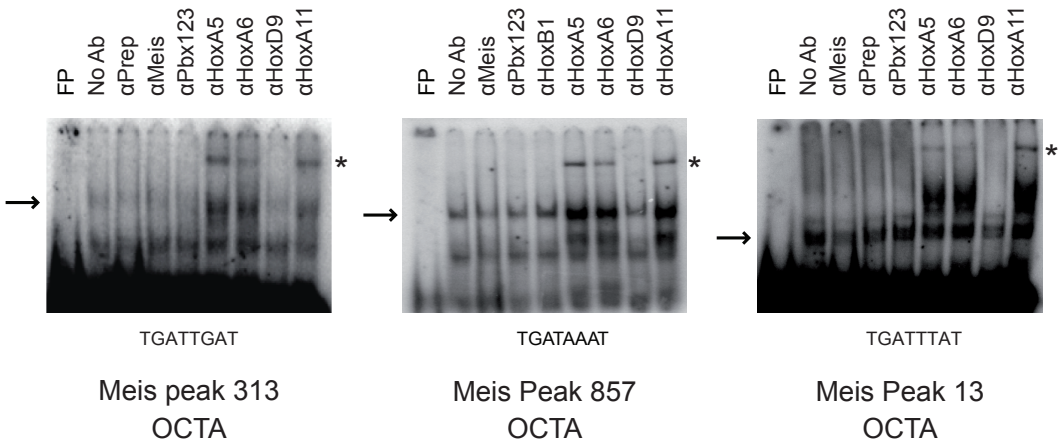

Figure 4.14: **EMSA experiments on OCTA motif variants.** EMSA was performed on OCTA variants with anti-Hox antibodies. Arrows indicate complex bands and asterisks indicate super-shifted bands.

moderate levels that are not over-represented in any other subset.

The over-representation pattern and levels in triple peaks are intermediate between those of Meis-Pbx1^{com} and Meis-Prep^{com} peaks.

As expected, EMSA experiments on OCTA-containing peaks with anti-Hox antibodies (shown in Figure 4.14) confirmed that these OCTA-containing peaks are capable of Hox binding *in vitro*, and the relative affinities of the different Hox paralog groups tested matched well the OCTA variant preferences described in the literature.

These results strongly suggest that many Meis peaks are Hox targets *in vivo*. If they are indeed, it is difficult to reach conclusions on the preferred paralog groups for formation of Pbx1-Meis-Hox complexes following Slattery et al. (2011), since the OCTA variants most overrepresented in Meis peaks are those that are strongly bound by a wide variety of paralog groups. The only exception to this is TGATGGAT, which is strongly associated with group 1 and is highly overrepresented in Meis peaks, particularly in Meis-Pbx1^{com} peaks. However, the relative frequency of specific Pbx-Hox target sequences in our data might be explained by the relative abundance of Hox proteins in the input sample.

4.2.10 EMSA on TGATGAAT

The OCTA variant TGATGAAT has never been reported to bind Hox proteins or Pbx-Hox heterodimers. Its overrepresentation in Meis^{exc}, Meis-Prep^{com}, Meis-Pbx1^{com} and triple peak subsets suggested it could be a novel target for Pbx-Hox heterodimers. We decided to test this possibility by EMSA.

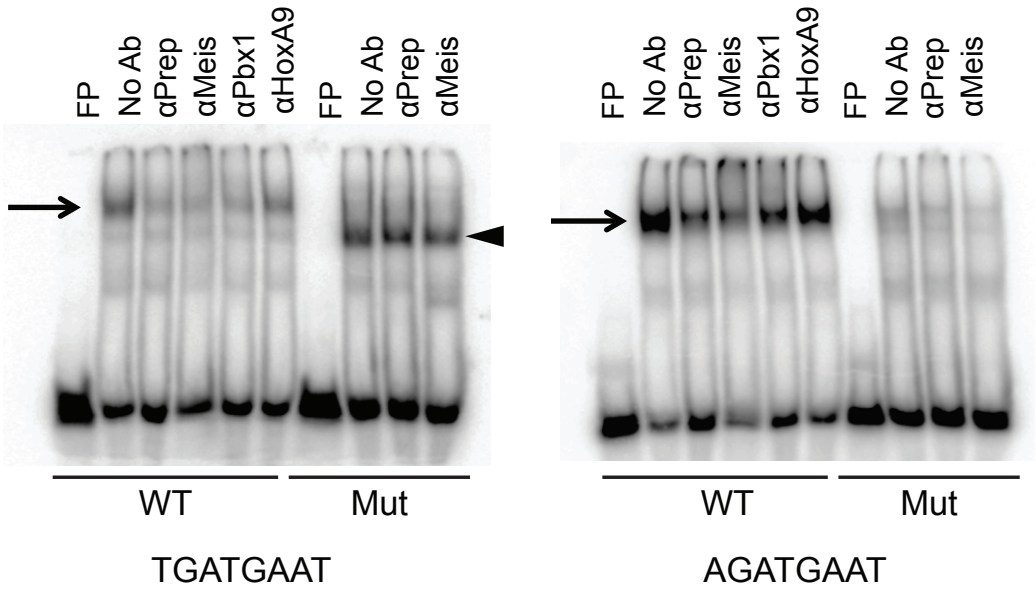

Figure 4.15: **EMSA on the novel OCTA variant TGATGAAT.** EMSA was performed on oligonucleotides containing the novel OCTA variant TGATGAAT and a mutated version, using antibodies for TALE proteins and the Hox proteins Hoxa9. Arrows indicate specific complexes that were inhibited by addition of anti-TALE antibodies. Arrowhead indicates a non-specific complex that forms on the mutated version of the oligonucleotide on the left.

As shown in Figure 4.15, the antibodies against Pbx1, Prep and Meis all interfered with complex formation on the WT sequence. Anti-Hoxa9 antibody interfered weakly but clearly with the formation of the complex on both oligonucleotide probes, showing the sequence is indeed a target for at least some Hox proteins. No specific complex formed on the mutated versions of the sequence.

4.3 Transcriptional targets

In order to identify correlations between TALE factor binding and transcriptional activity in the embryo, we obtained RNA from E11.5 *Meis1* KO embryos, *Prep1*$^{i/i}$ embryos and WT littermates.

We found 855 up-regulated and 631 down-regulated genes in *Prep1*$^{i/i}$ embryos, and 210 up-regulated and 198 down-regulated genes in *Meis1* KO embryos.

4.3.1 Co-Regulation

To estimate the level of co-regulation between Meis and Prep proteins, we compared the number of genes that are mis-regulated in both kinds of defective embryos, either in the same direction or in oposite directions.

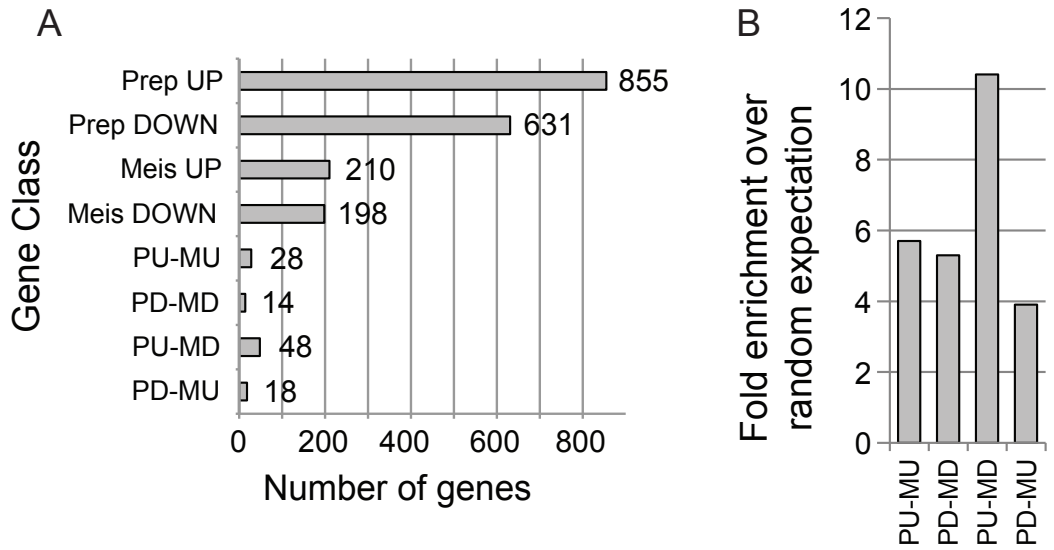

Figure 4.16: **Misregulated Genes in Meis1/Prep1-Deficient Embryos.** A) number of genes up- or down-regulated in *Meis1* KO and *Prep1*$^{i/i}$ embryos, and coincidence between sets. B) over-representation of co-regulated genes over random expectation.

The chance that a given gene is misregulated in *Meis1*-deficient embryos is about 5 times higher if it is misregulated in *Prep1*-deficient embryos and vice-versa, as shown in Figure 4.16. All categories of co-regulation were over-represented compared to random expectation. The most over-represented category, with 10 times more genes than would be expected if the gene sets were independent, is composed of genes that are up-regulated in *Prep1*$^{i/i}$ embryos and down-regulated in *Meis1* KO embryos, i.e., genes activated by *Meis1* and repressed by *Prep1*.

4.3.2 Gene-Peak Correlation

We tried to establish how nearby binding events could be affecting the expressing of misregulated genes. In order to do that, we calculated the Meis and Prep peak density in TSSA, IG and CI regions near misregulated genes and compared it to the corresponding GW densities. The ratios are represented in Figure 4.17.

We found some significant correlations between TALE factor binding near genes and misregulation of those genes in TALE factor mutants. Binding of Meis near the TSS of a gene is associated to the gene being up-regulated in the *Meis1* KO, i.e., repressed by Meis in normal conditions. Surprisingly, binding of Prep to IG or CI regions is also associated to up-regulation in *Meis1* KO, although the effect is smaller in magnitude.

Genes down-regulated in the *Prep1*$^{i/i}$ embryo have a higher than average probability to show Prep peaks in their promoter, and Meis or Prep peaks within the

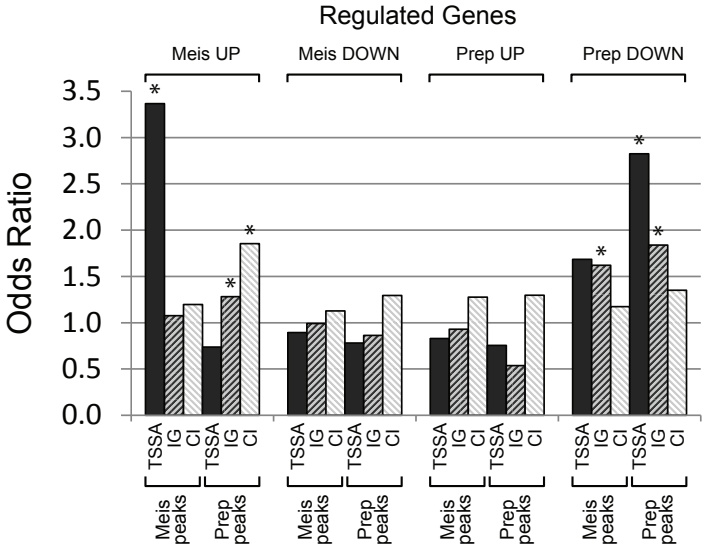

Figure 4.17: **Correlation Between Misregulation and TALE Binding.** Ratio of Meis and Prep peak density in misregulated gene-associated TSSA, IG and CI regions to the corresponding GW averages. Asterisks mark p-value $< 10^{-3}$.

transcriptional unit.

We could not find association between binding in any region and down-regulation in *Meis1* KO or up-regulation in *Prep1*$^{i/i}$. It has to be noted that we did not knockout the full complement of *Meis* nor *Prep* genes in the embryo, so the weak correlations between protein binding and transcriptional effect we report could be due to redundancy with the rest of Meis/Prep factors. It could also mean that the relationship is more indirect than a simple binding-activation/repression model, or that we could not link the protein binding events to their actual target genes.

4.3.3 Gene Ontology

Since Prep or Meis binding in the TSSA or IG regions of a gene might be associated to its transcriptional control by these factors, we performed Gene Ontology analysis on all genes bound by Meis or Prep to gain insight into their general functions.

For both Meis- and Prep-bound genes, the most significant category was regulation of transcription. That supports the view that the TALE proteins are master regulators in the sense that they regulate the transcription of other regulators (TFs) rather than effector genes.

Most other over-represented categories were consistent with known roles of *Meis1* (Nervous System Development, Heart Development, Embryonic Limb Morphogenesis, Blood Vessel Development) or *Prep1* (DNA Metabolic Process, Cell Cycle, In Utero Embryonic Development) and generally suggested a framework in which *Meis1* has many functions in specific developmental programs while *Prep1* is more involved in basal cellular processes.

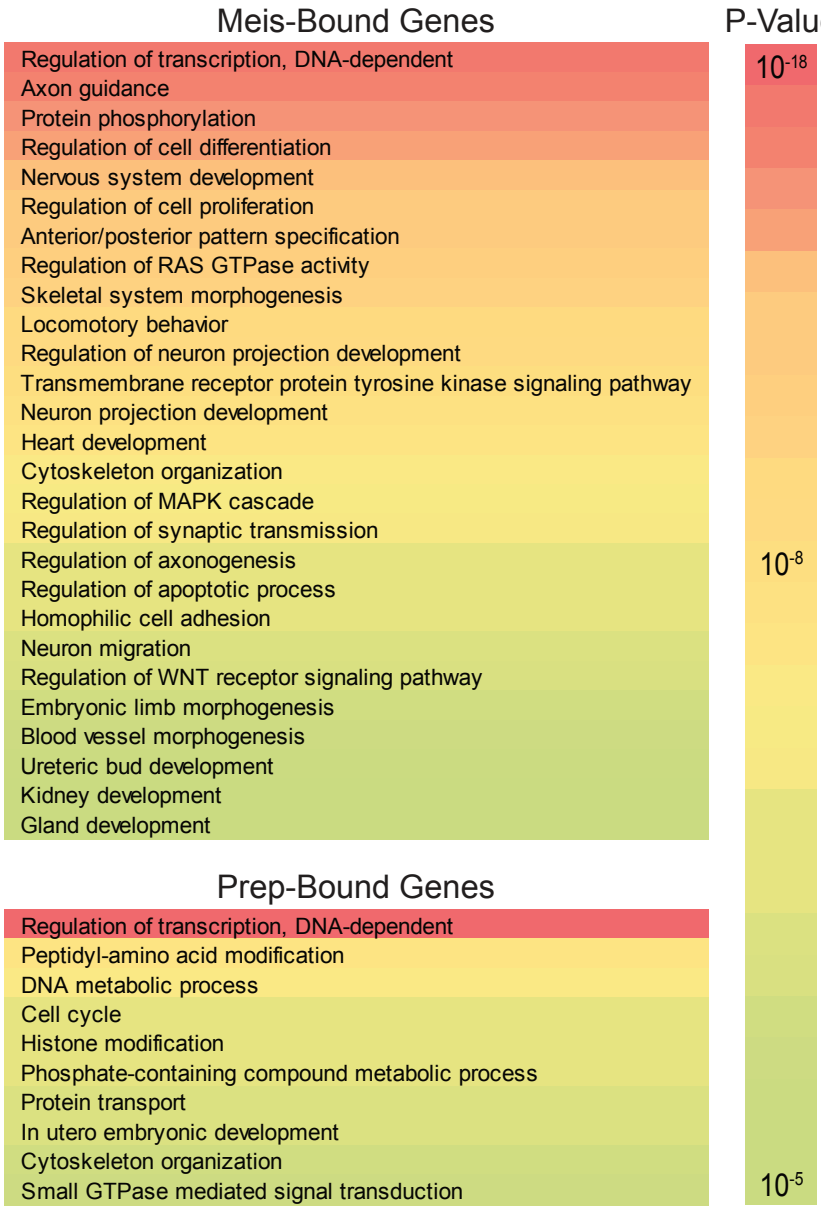

Figure 4.18: **Gene Ontology Categories Over-Represented in TALE-bound genes.** We restricted our search to the Biological Process ontology. We hand-selected the categories that were informative, i.e., specific and non-redundant. P-value cutoff was 10^{-5}. Categories sorted by significance.

4.4 Hox Clusters

The Hox clusters are a genomic region of particular interest for the study of TALE factor binding.

4.4.1 Distribution in the Hox clusters

We studied the read distribution in the Hox clusters (Figure 4.19). In all 4 Hox clusters, Meis read density is much higher than in the surrounding regions, and shows a concentration near the 3′ end that extends to paralog 9 but no further.

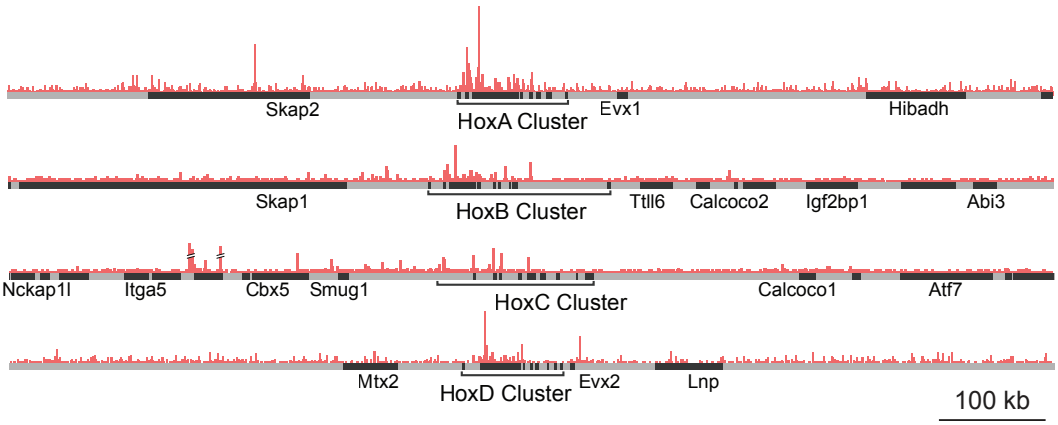

Figure 4.19: **Read Density in the Hox Clusters.** Meis ChIP-seq read density in the four mouse Hox Clusters, represented as bar graphs.

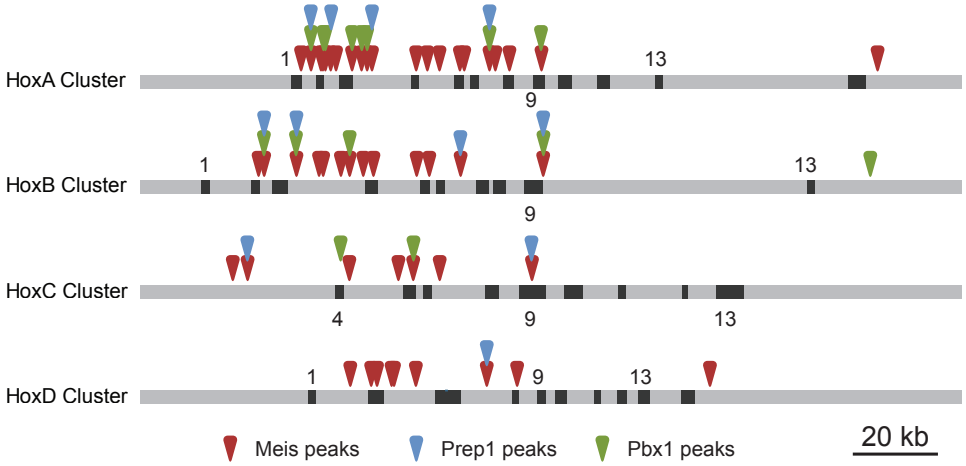

Figure 4.20: **TALE Factor ChIP Peaks in the Hox Clusters.** Arrowheads represent peaks, black boxes represent transcriptional units. We have represented the short isoform of each paralog 3 for easier orientation.

4.4.2 Peaks in the Hox Clusters

Plotting Meis, Prep and Pbx1 peaks in the Hox clusters (Figure 4.20) reveals a clear restriction of TALE protein binding events in our sample to the region 3′ to paralog 9 in every cluster.

We found abundant Meis peaks in the Hox clusters (20 in HoxA, 13 in HoxB, 7 in HoxC and 8 in HoxD) and many fewer Prep or Pbx1 peaks (4, 4, 2, and 1 for Prep; 8, 4, 2 and 0 for Pbx1). The HoxA cluster is clearly more bound by the TALE Hox cofactors than the rest. All but one Prep and Pbx1 peaks in the Hox clusters overlap with Meis.

It is notable that the border of Meis binding versus non-binding is located pre-

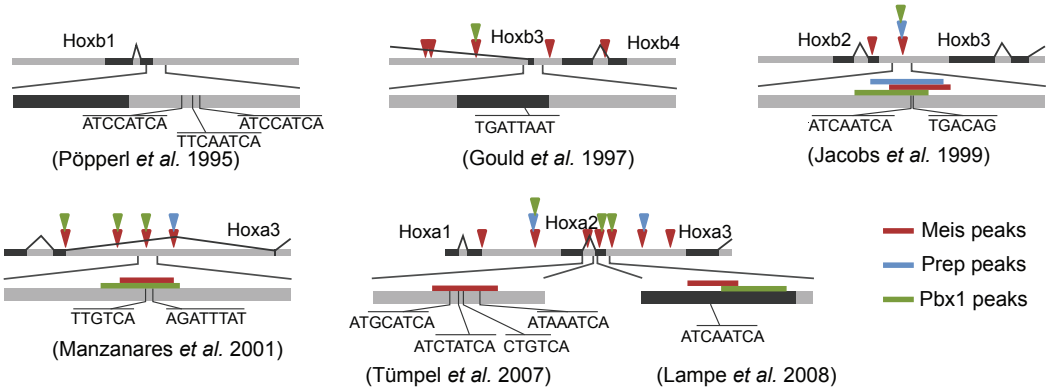

Figure 4.21: **Previously Described Hox Regulatory Sites.** In each panel, the top section represents the genomic neighbourhood (15kb), with peak positions marked by arrowheads as in Figure 4.20. The bottom section shows a close-up (1kb) on the described regulatory site, with motif instance sequences marked at their position and peaks represented as solid bars covering their full span. Black bars represent exons, gray bars represent introns or intergenic regions.

cisely at the 9th paralog. It is one of the three only paralog groups conserved in all four murine Hox clusters (the others being 4 and 13) and marks the transition between non-*abd-B*-related Hoxes and *abd-B*-related Hoxes.

4.4.3 Hox Regulatory Sites from the Literature

Several auto- and cross-regulatory TALE protein BSs have been described in the Hox clusters (A. Gould et al. 1997; Jacobs, Schnabel, and Cleary 1999; Lampe et al. 2008; Manzanares, Bel-Vialar, et al. 2001; Pöpperl, Bienz, et al. 1995; Tümpel et al. 2007).

Of these 6 described regulatory sites, 4 show Meis peaks. 3 of the 4 also show Pbx1 binding, and one of these three shows a Meis-Prep-Pbx1 triple peak. All of them contain OCTA motifs. In three of the six cases HEXA motifs are present at a short distance to the OCTA sequence.

4.4.4 Hox RNAseq

The subdivision of the Hox clusters regarding Meis is not restricted to DNA binding. We plotted transcriptional variations in Hox genes in the absence of *Meis1* or *Prep1*, from our RNAseq data.

While most of the transcriptional changes were too subtle to be significant individually, the pattern is consistent across clusters and significant (p-value < 0.05). In the absence of *Meis1*, 3′ Hoxes tend to be overexpressed while 5′ Hoxes tend to be repressed. The opposite pattern can be seen when considering the expression of

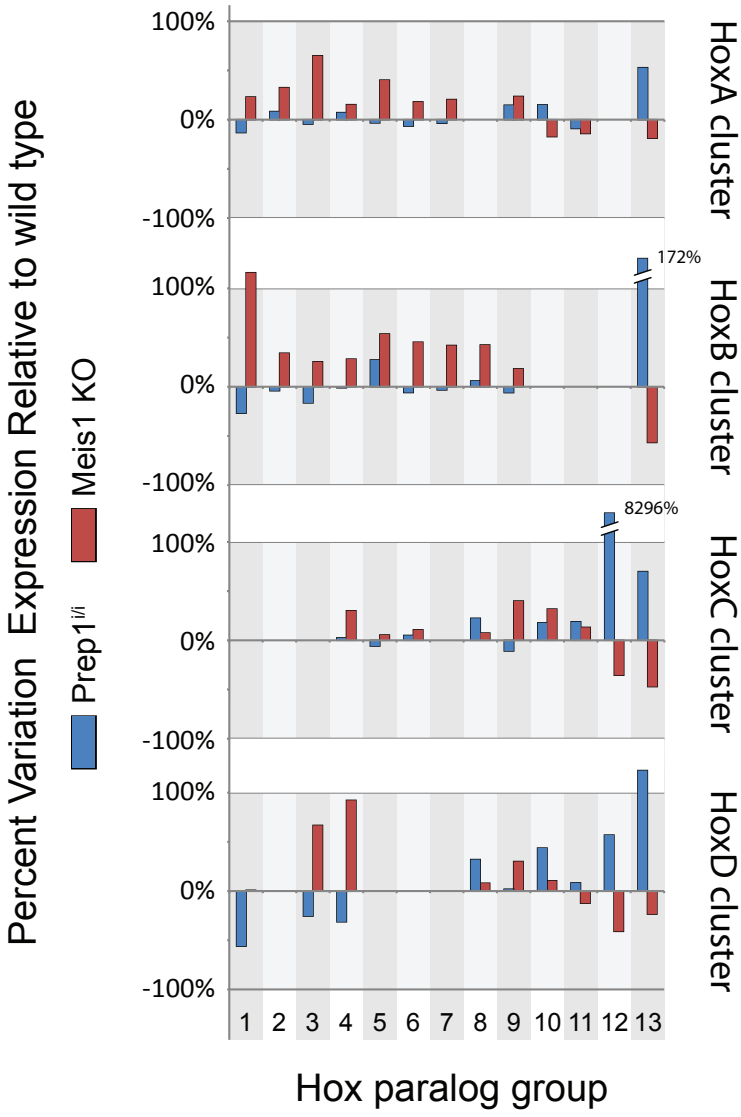

Figure 4.22: **Differential transcriptional response of the 3′ and 5′ halves of the Hox clusters.** *Meis1* KO (red) or *Prep1*$^{i/i}$ (blue) Hox transcripts abundance relative to WT littermate controls, as percentages.

Hox genes in *Prep1*$^{i/i}$ embryos: 3′ Hoxes are generally very slightly repressed while 5′ Hoxes are markedly overexpressed.

This pattern is surprising in at least two ways: Meis is more co-expressed with 3′ Hoxes than with 5′ Hoxes, so our expectation was that its lack would inhibit 3′ gene expression while maybe derepressing 5′ Hoxes. At the same time, the largest effect we detect in the transcription of the Hox genes is the overexpression of 5′ Hoxes in the absence of Prep1, a region where neither Prep nor Meis bind.

4.4.5 AP ChIP

The Meis peaks in a given Hox cluster could be occupied by Meis constitutively or they could be occupied alternatively in different cells. Since our initial approach can not distinguish between these two possibilities, we performed ChIP in embryo

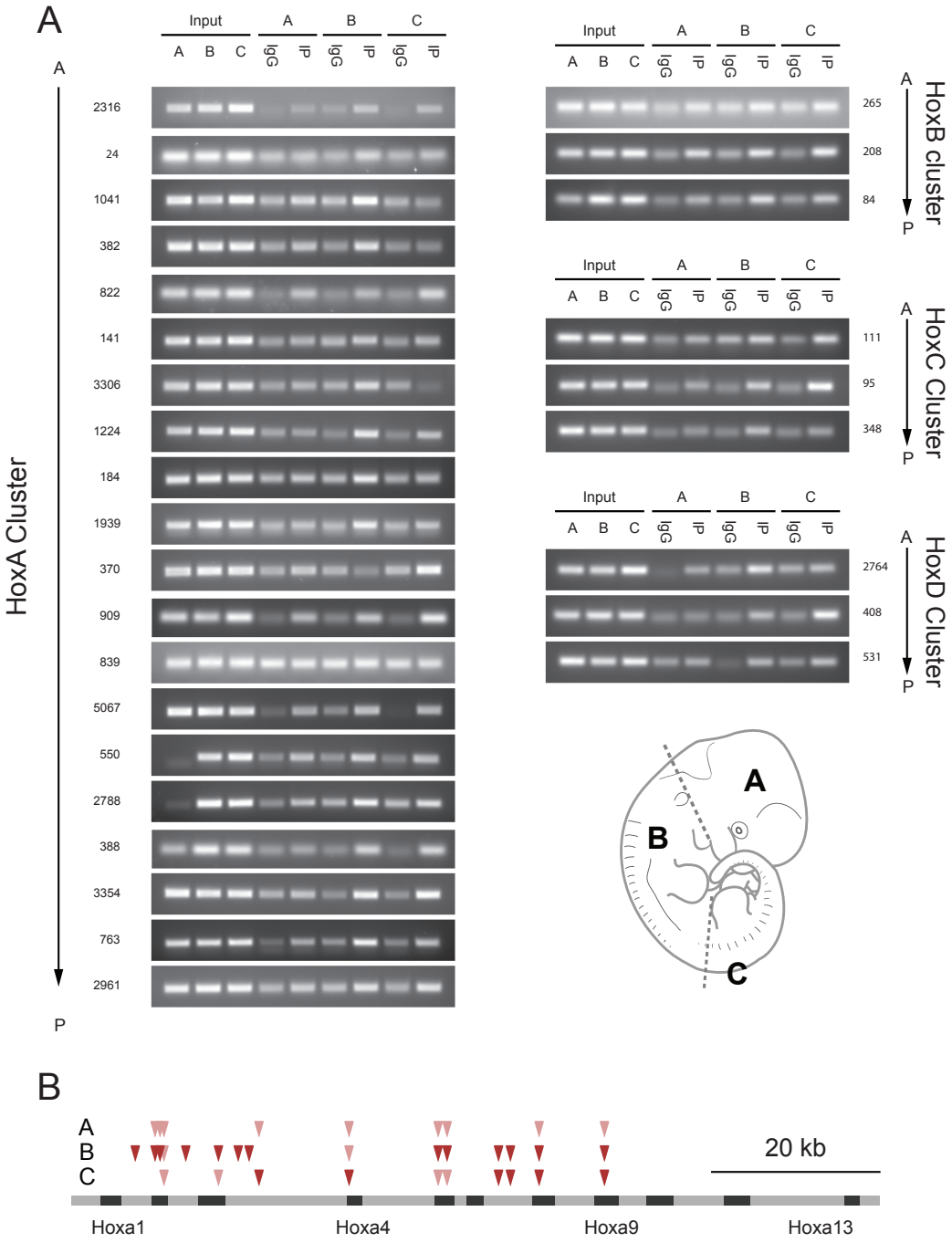

Figure 4.23: **ChIP on Antero-Posterior Embryo Fragments.** A) Band intensity of rate-limited PCR on the Meis ChIP products with peak primers. Numbers denote Meis peak number. For each peak, the left three lanes are PCR products from the three inputs at the same DNA concentration. The right six lanes are PCR products of a fixed amount of control (IgG) or Meis ChIP DNA. B) Summary of the results in A for the HoxA cluster. An absent arrowhead indicates no binding detected, a light-colored arrowhead indicates positive but not predominant binding compared to other embryo regions, and a dark-colored arrowhead indicates predominant binding.

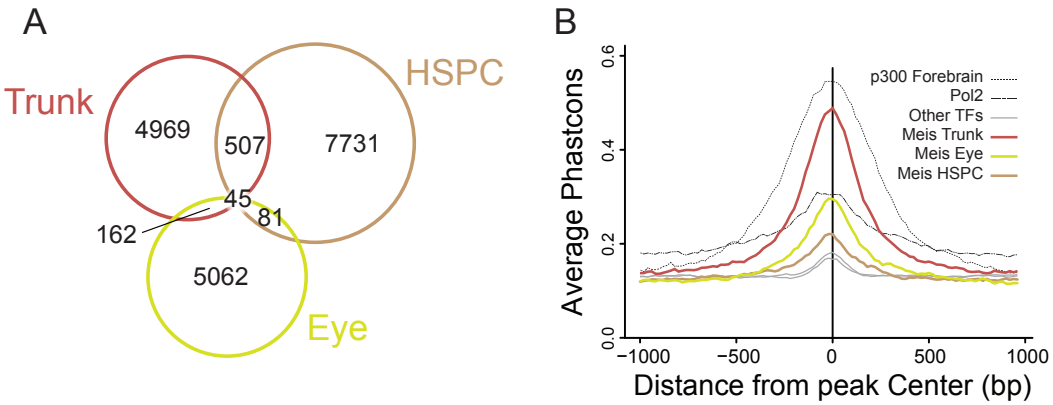

Figure 4.24: **Meis Binding in Different Contexts.** A) Venn diagram of the overlap between Meis peaks in Meis trunk peaks, Meis eye peaks provided by Dr Paola Bovolenta, and Meis HSPC peaks from Wilson et al. (2010). B) Conservation profile of Meis peaks from different contexts

fragments. We dissected embryos along the Antero-Posterior (AP) axis and performed Meis and unspecific IgG ChIP on the resulting nuclear extracts.

The resulting DNA was subjected to PCR using primers designed for amplification of products centered around each of the Meis peaks in the HoxA cluster and some peaks in other clusters, spanning approximately 100bp around each summit. The PCR conditions were chosen so that the amount of product was roughly proportional to the input. The results are shown in Figure 4.23.

We discarded those peaks for which the IgG bands were of uneven intensity and compared the intensity of the bands for the rest of the peaks. As can be seen in Figure 4.23b, Meis peaks occupancy is limited in the anteriormost part of the embryo, maximal in the middle section and skewed towards 5' peaks in the posteriormost part of the embryo.

4.5 Meis Contexts

We are interested in what determines transcription factor binding. In order to gain insight on the binding preferences of Meis in different contexts, we took advantage of a published ChIP-seq dataset for Meis in HSPCs (Wilson et al. 2010) and a ChIP-seq dataset for Meis in the developing (E12.5) eye kindly provided by Dr Paola Bovolenta.

We first checked the overlap of Meis BSs in different contexts, shown in Figure 4.24a. A surprisingly low (45/18087, or 0.25% of all non-redundant genomic regions) number of peaks are present in all three data sets, indicating that Meis has a GW pattern of binding that is highly context-specific. The conservation profile of Meis peaks, shown in Figure 4.24b, Is also different in these three tissues – neither

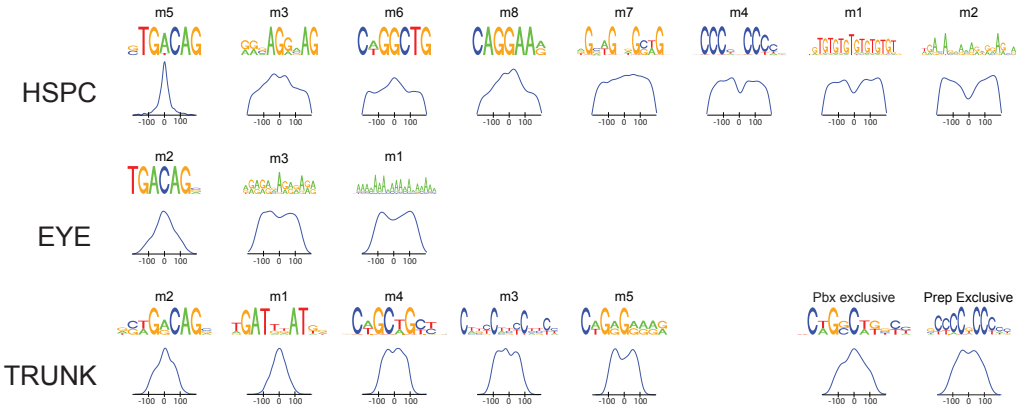

Figure 4.25: **Motif Discovery on the Sequences of Meis Peaks from different contexts.** Motif logos and histograms for motif position within peak for Meis trunk peaks, Meis eye peaks provided by Dr Paola Bovolenta, and Meis HSPC peaks from Wilson et al. (2010).

in the developing eye nor in HSPCs do we see the extreme conservation of binding locations from the embryonic trunk.

When we compared the locations of peaks relative to TSSs, we found further differences. Meis eye peaks, like trunk Meis peaks, are seldom located in promoters (75/5361, or 1.4% of all Meis eye peaks). In contrast, HSPC Meis peaks are clearly enriched in promoters, although not to the degree of Prep^{exc} peaks: 1037/8386 (12.4%) of Meis HSPC peaks are TSSA. This parallel between Meis in the devoping eye and trunk with contrast to Meis in HSPCs extends, intriguingly, to the imbalance of IG peaks between exon and intron peaks: while the Meis eye peaks are excluded from exons like the trunk peaks (intron/exon ratio of 37.2% versus 0.6%, i.e. 62.06, compared to a Meis trunk ratio 79.8), HSPC peaks have an intron/exon ratio of 25.86 (1.7% versus 43.8%), which is close to the genomic average of 19.65.

We then performed an unsupervised motif discovery as with the subsets of Meis, Prep and Pbx1 trunk peaks. We used the same parameters as earlier, but for trunk peaks we used the whole set. The results are in Figure 4.25. In both HSPC and eye peaks, the main core motif is HEXA. In the eye, which is a Hox-free environment, it is not surprising that Meis does not bind OCTA sequences, but no sign of DECA is found. DECA could be expected if it was bound by Pbx-Meis heterodimers in the absence of Hoxes.

In HSPCs, there is not obvious binding to OCTA sequences despite the fact that several Hoxes are expressed at high levels. There are also noteworthy accessory motifs. In Figure 4.25 m6, m4 and m9 from HSPCs are old acquaintances; m6 is very similar to the Pbx1^{exc} subset core motif, which as mentioned above is also present in

several variants in Meis^{exc} peaks, and has appeared in several hematopoietic ChIP-seq reports such as Tijssen et al. (2011). Its profile in Meis HSPC peaks is interesting because it seems a hybrid of that which would be characteristic of accessory and core motifs.

HSPC m4 is similar to the CCCgCCC motif we detected in all single- or double-factor trunk peak subsets with Prep participation. This motif is similar to the binding site of Sp1 and Klf4, both of which have functions in hematopoiesis.

Five

Discussion

The results presented here represent the first near-comprehensive GW view of TALE family proteins BSs in the vertebrate embryo. As such, they deepen our understanding of the target sequences and modes of action of these Hox cofactors, and raise interesting questions that will have to be addressed in further studies.

5.1 Meis vs Prep

To date, the general picture revealed by *in vitro* studies of Meis and Prep molecular function has revealed very few differences between them (see for example Moens and Selleri 2006, where they are considered interchangeably). Both can bind Pbx directly (Knoepfler, Calvo, et al. 1997; W. F. Shen, Montgomery, et al. 1997), both recognize very similar sequences (Chang, Jacobs, et al. 1997; Knoepfler, Calvo, et al. 1997), both can participate in trimeric complexes with Pbx and Hox proteins (Berthelsen, Zappavigna, Ferretti, et al. 1998; W. F. Shen, Rozenfeld, Kwong, et al. 1999), and both can activate the transcription of promoters containing their target sequences (Berthelsen, Zappavigna, Ferretti, et al. 1998; Jacobs, Schnabel, and Cleary 1999). Only two clear differences at the molecular level have been established: the first is that Meis is resposive to TSA and PKA signalling (H. Huang, Rastegar, et al. 2005) while Prep is not. The second is that Meis is able to bind some Hox proteins directly without the participation of Pbx, while Prep is not (T. M. Williams, M. E. Williams, and Innis 2005).

However, their general biological functions are quite different. The *Meis1* mutant has defects in hematopoietiesis and organogenesis. Prep, in contrast, is critical at a much earlier stage than Meis in functions apparently more basal, although the *Prep1* hypomorphic mutant survives to E17.5, when it dies with a phenotype similar but distinct to that of the *Meis1* mutant (Di Rosa et al. 2007; Ferretti, Villaescusa, et al. 2006). One partial explanation of this functional difference is that while the *Prep* genes are expressed ubiquitously, *Meis* genes have restricted and dynamic expression domains. However, if they were completely interchangeable the expression of Meis in territories that already express Prep should be redundant. This is clearly not the case. Indeed, regarding their involvement in cancer, Meis and Prep actually have effects in opposite directions: *Meis1* is a proto-oncogene that can induce leukemias by over-expression, while *Prep1* is a tumor suppressor gene (Iotti et al. 2011; Longobardi, Iotti, et al. 2010; Moskow et al. 1995; Thorsteinsdottir et al. 2001; Wong et al. 2007).

The work described here provides another foundation for explaining the functional differences between Meis and Prep. The near-comprehensive catalogue of TALE factor binding sites will be a useful resource in itself for us and other groups when considering mechanisms and mediators of Meis, Prep and Pbx effects.

The *in vivo* binding sequence preferences we have described will guide the understanding of TALE protein control of expression of downstream genes. A DECA or DECA^{ext} sequence found to act in a promoter should be suspected primarily of being bound by a Pbx-Prep complex. Functional OCTA sequences located in enhancers should be considered as likely candidates for regulation by Meis and Hoxes, with or without the participation of Pbx.

Although only a few of the peaks we describe are triple Meis-Prep-Pbx binding sites, they could represent the points of input into jointly regulated systems. Meis and Prep could be collaborating or antagonizing each other. The fact that in almost all of the triple peaks we have assayed there is no simultaneous binding of Meis and Prep means that, in most cases, Meis and Prep compete for binding to the same sequences. In practice that would mean Meis displaces the ubiquitous and pre-existing Prep in some or all of the cells in which it is expressed.

5.2 Binding Preferences

Despite their very similar binding behaviour *in vitro*, Prep and Meis have shown radically different binding sequence preferences *in vivo* in our study.

This study is the first detailed view and comparison of TALE protein *in vivo* BSs. Only a few target sites were known previously. These have been invariably been regulatory sites located near genes. In contrast, our unbiased GW search has revealed striking differences in TALE protein promoter and enhancer occupancy.

Meis binds thousands of extremely conserved gene-remote regions, a high proportion of which show enhancer chromatin marks in comparable cells. It seems reasonable to conclude that many of these sites are enhancers, at least some of which are inactive but "poised". Poised enhancers have been suggested to represent the range of responses available to a cell (Creyghton et al. 2010) and could represent a developmental path restriction by the TALE system which would be later narrowed down by the participation in TALE-Hox complexes of subsets of Hox proteins available in different cells.

Along this line, the Pbx-Meis heterodimer has been proposed to function as a "pioneer" factor that can bind heterochromatin and recruit MyoD and the estrogen receptor to some of their targets, such as *Myogenin* or *M-cadherin* (Berkes et al. 2004; Magnani et al. 2011). It must be noted though that our ChIP-seq shows binding of Pbx and Prep at the *M-cadherin* promoter, but not binding of Meis. *M-cadherin* is expressed at E11.5 in the proximal limb. Meis could be indirectly responsible for this restriction, even if it does not bind the *M-cadherin* promoter at that stage, by making Pbx available at the nucleus for Pbx-Prep complex formation. It could also bind heterochromatin in complex with Pbx and be replaced by Prep afterwards.

In remarkable contrast with Meis-Pbx1^{com} peaks, Pbx1-Prep^{com} peaks are very strongly associated with promoters. This preference is driven by Prep, since Prep^{exc} but not Pbx1^{exc} peaks show strong association with promoters.

The transcriptional effects of Meis and Prep are often in opposing directions on the same gene, despite the fact that they rarely bind the same cis-regulatory sequences.

While *in vitro* both Meis and Prep show the ability to bind DNA with Pbx as heterodimers, we have found that, *in vivo*, Prep is much more likely to do so than Meis. Less than 15% of Meis peaks show a DECA motif, the target for this kind of heterodimers. Conversely, while both Meis and Prep are able to form triple complexes with Pbx and Hoxes *in vitro*, only Meis seems to do so *in vivo*: the proportion of Prep^{exc} and Pbx1-Prep^{com} peaks that show the OCTA motif is under 5%.

This, together with the clear divide in binding location between Meis and Prep peaks, paints a picture in which Meis and Prep are functionally specialized, with Meis participating almost exclusively in the Pbx-Hox system by binding extremely conserved enhancers in TSS-remote positions and Prep binding promoters in concert with Pbx but independently of Hoxes. It is remarkable that no other study has detected such a clear division of labour between Meis and Prep.

5.3 TALE Proteins as Hox cofactors

The surprising finding that the most frequent sequence motif in Meis peaks is OCTA, the Pbx-Hox target site, rather than HEXA, the Meis monomer binding site, or DECA, the Pbx-Meis/Prep target, suggests that most Meis binding sites are trimeric Pbx-Hox-Meis complex binding sites in which Meis does not contact DNA directly. The large overlap between Meis peaks and published Hox binding sites reinforces the view that Meis is an important Hox cofactor. *In vitro* experiments show Hox proteins can bind OCTA-containing Meis peak sequences.

If most Meis peaks with the OCTA motif are indeed Hox BSs, this work represents the first GW catalog of *in vivo* BSs from several different Hox paralog groups. Since to date only two ChIP-seq datasets for vertebrate Hox proteins have been published (Donaldson et al. 2012; Jung et al. 2010) and very few of their downstream genes are known, it could represent an extremely valuable asset for understanding the missing link between Hox protein expression and control of morphogenesis.

5.3.1 OCTA Variant Preferences

The detailed analysis of OCTA motif variants reveals preferences in the variable dinucleotide that match what we know about Pbx-Hox preferences and the ability

of Pbx and Meis to bind subsets of the Hox proteins.

In Meis^{exc} peaks, the most overrepresented variant is TGATTTAT. This OCTA variant is the preferred binding target for *abd-B*-related Hoxes, which are precisely those that can bind Meis without the participation of Pbx. In contrast, the most common variant in Meis-Pbx1^{com} peaks is TGATTGAT, which is the preferred variant of 3′ Hoxes. These can not bind Meis directly, so Meis has to rely on indirect binding to Hoxes via Pbx. However, as mentioned earlier, it is hard to extract conclusions on the relative participation of different Hox paralog groups in Meis-Pbx1-Hox complexes because the abundance of Hoxes in our sample is uneven.

We have found an OCTA variant, TGATGAAT, that had not been described to be bound by Hoxes. Its overrepresentation in several TALE peak subsets suggested it could represent a novel Hox binding target. We have confirmed this by EMSA experiments.

5.4 Restricted Binding in the Hox Clusters

It was unexpected that the TALE Hox cofactors we have inspected would have a pattern of binding in the Hox clusters that is tightly restricted to their 3′ part.

Additionally, Meis peaks extend past the 3′, but not 5′, end of all Hox clusters and into the adjacent gene deserts. These have been shown to contain cis-regulatory elements and to establish long-range chromatin interactions with the Hox clusters themselves (Andrey et al. 2013; Noordermeer et al. 2011). In all Hox clusters, the range of these regulatory interactions coincides with the range of Meis peaks extending past the end of the clusters. Consistently, we find no Meis peaks in the gene desert 5′ to the HoxD cluster that is known to contain many regulatory elements important for the expression of HoxD genes in the digits (Montavon et al. 2011), where Meis is not expressed. This "regulatory archipelago" establishes long-range contacts with the 5′ region of the HoxD cluster up to *Hoxd9*, the 3′-most HoxD gene expressed in the digits.

Still, there is an unexplained aspect of TALE binding in the Hox clusters. The transcriptional effect of Prep deficiency is strongest in the 5′ part of the clusters, where neither Meis nor Prep bind. Silent Hox clusters form discrete inactive chromatin compartments, and upon activation they adopt a chromatin conformation separated in an active and an inactive compartment (Noordermeer et al. 2011). The boundary between the topologically associated domains evolves over time, providing a mechanistic basis for temporal colinearity. This proposed mechanism also provides a possible explanation for Meis and Prep effects on the transcription of parts of the Hox cluster to which they do not bind: they could be influencing this colinearity mechanism by serving as anchors between the 3′ part of the clusters and

its associated gene desert.

5.5 Additional Binding Partners

We have discovered four new sequences not shown previously to be involved in TALE protein binding.

The sequence CWGSCWG is the only Pbx core motif we found. It appears in the Meis^{exc} and Pbx1^{exc} subsets and a similar one appears in Meis-Pbx1^{com} peaks. It is likely to represent the binding sequence for an unknown Pbx or Pbx-Meis cofactor. It is interesting that highly similar sequences have been reported previously as unidentified motifs in several hematopoietic factor ChIP-seq experiments: An Scl/Tal1 ChIP-seq in immature erythocytes (Kassouf, Hughes, et al. 2010), a 10-factor ChIP-seq in the hematopoietic progenitor cell line HPC-7 (Wilson et al. 2010), and a GW analysis of Gata1, Gata2, Runx1, Fli1, and Scl/Tal1 binding in primary human megakaryocytes (Tijssen et al. 2011). The only common factor to these three studies is Scl/Tal1. Scl/Tal1 is a bHLH protein and thus binds an E-box. Although this novel motif resembles an E-box, the binding preferences of Scl/Tal1 are well characterized and do not include this variant, in which there is an additional nucleotide between the two halves of a canonical E-box.

Scl/Tal1 can interact with Gata1 to specify erythroid cells (Lahlil et al. 2004) and we have found an Scl/Tal1-related motif but not any motif resembling Gata1. Considering that Meis is necessary for megakaryocyte generation in the mouse embryo, it is tempting to speculate that Scl might be recruited by Gata1 in erythroid cells (Anguita et al. 2004) and Meis1 in megakaryocytes. These two cell types share a precursor cell, the Megakaryocyte-Erythroid Progenitor (MEP) (Doré and Crispino 2011), so this could represent a protein recruitment-based developmental switch. *Scl/Tal1* mutants die at 9.5 due to complete lack of blood, but deletion in adult mice shows it is necessary for maintenance only of the erythroid and megakaryocytic lineages (Mikkola et al. 2003), which points to dual roles in embryonic HSCs and MEPs. It is remarkable that a DNA-binding deficient form of Scl rescues the early hematopoiesis-related lethality and allows the embryo to progress to E14.5, when it dies with incomplete maturation of erythrocytes (Kassouf, Chagraoui, et al. 2008; Porcher et al. 1999). This probably means some of its functions are mediated by cofactors that allow it to bind DNA independently of its own DNA-binding domain. Scl is also expressed in the neural tube in a Dorso-Ventral (DV)-restricted manner (Muroyama et al. 2005), as is Meis.

The CCCgCCC sequence is an accessory motif found in all peaks subsets in which Prep is present. It exactly matches the GC-box, a promoter component that is the binding site for Sp1 (Kriwacki et al. 1992) and its related factors Sp3, Sp4 and

the KLF family. The defining characteristic of the Sp/Klf family is a C-ter DNA-binding domain containing three C2H2 Zinc finger motifs. A short motif N-ter to the Zinc fingers, called the Buttonhead (btd) box, is present in Sp proteins but not in Klfs (Suske, Bruford, and Philipsen 2005). It does not contain a tryptophan. It is remarkable that Sp genes have a conserved genomic organization: they are linked two by two and 3 of these 4 syntenyc pairs are linked to Hox genes. Additionaly, 5 of the 9 Sp genes in the mouse genome have TALE factor peaks in their promoters, most often Pbx1-Prep^{com} peaks. The coincidence of Pbx1-Prep with an Sp target sequence and possible control of Sp genes by Prep and Pbx suggests these two families might act as a previously unrecognised functional module. Deeper study would be needed to ascertain the nature of the relationship, if it exists. However, a computational analysis of human promoters found the motif for Sp1 (Hartmann et al. 2013), so its presence in Prep subsets could be a byproduct of their high proportion in promoters.

The CCAAT sequence appears in Pbx1-Prep^{com} peaks. It is a known component of promoters and enhancers, bound by NF-Y and the CEBP proteins. As with the GC-box, it is not surprising to find a promoter-related sequence in a subset of peaks that are overwhelmingly associated with promoters, and indeed this sequence also appears in a general motif search in promoters (e.g. see Hartmann et al. 2013). However, in this case it is clear that there is a specific interaction between the Pbx1-Prep heterodimer and a CCAAT-binding factor, since DECA and the CCAAT box appear at a fixed distance within the DECA^{ext} motif. The CCAAT box is present in many promoters of genes overexpressed in cancer (Dolfini and Mantovani 2013). The NFY heterotrimer might itself be transcriptionally regulated by the TALE Hox cofactors, since the *Nfyc* gene (coding for the γ subunit) has Meis and Prep peaks with several DECA^{ext} motifs within its first intron. In fact, the Prep peaks actually overlap the alignment point for the TSS of a rat isoform of *Nfyc* that has not yet been described to exist in the mouse, and show LICR MEF promoter (RNApolII^{+} H3K4Me3^{+}) marks. Very obviously, this fact suggests the existence of a previously unrecognized isoform of *Nfyc* in the mouse and its specific regulation by the TALE proteins through differential promoter recruitment. *Nfyc* is downregulated under the significance threshold in the *Prep1*$^{i/i}$ embryo, but this does not rule out isoform balance effects or redundant control by Prep1/2. This is an interesting avenue to pursue regarding the search for mediators of the involvement in cell proliferation and cancer of Meis and Prep.

The poly-purine motif is interesting. It might be the binding site for an unknown partner of Meis, but if this was the case it is hard to explain why it does not seem to represent any specific sequence requirement other than a separation of purines and pyrimidines in complementary strands. It seems more likely that

it represents a sequence pattern that does not bind a particular protein but results in structural features favouring Meis binding. Polypurine tracts have been shown to be able to form DNA triple helical structures *in vitro* (Mirkin et al. 1987; Mooren et al. 1990), to be implicated in splicing of many genes including *Fgfr2* (Carstens, E. J. Wagner, and Garcia-Blanco 2000), and are a known component of retroviruses (Rausch and Le Grice 2004).

Polypurine/polypyrimidine tracts with mirror sequence repeats favour the adoption by DNA of a triple helix conformation (DNA-H). In this conformation, two strands pair in a standard double helix and the third strand pairs with the bases in one of them through Hoogsteen hydrogen bonds. This conformational change is slow but stable and results in distortions that can trigger recombination and DNA repair mechanisms, which can result in genomic instability (G. Wang and Vasquez 2004). Long polypurine/polypyrimidine tracts are associated with genes involved in development and morphogenesis, among other functions, and may have an evolutionary role by promoting variability (Bacolla et al. 2006). There might be a relationship between this mutagenicity and the oncogenic ability of Meis. A more exotic but suggestive possibility is that, if indeed this polypurine motif represents stretches capable of forming a triple helix, it could be a mechanism to anchor long-range chromatin loops between enhancers and promoters. The pairing of three of the four strands involved would leave the fourth available for transcription.

5.6 Meis Contexts

The comparison of Meis binding sites in the trunk, eye and HSPCs provides an interesting frame in which to evaluate the significance of some aspects of Meis trunk binding sites. It is clear that Meis eye and HSPC binding sites are more similar in general to each other than to trunk binding sites, despite trunk and HSPC peaks overlapping the most of the three possible pairings: both core motif composition and conservation values point to trunk binding sites being apart from the other two sets. One aspect that could be related to this is that compared to those tissues, our sample was much more heterogeneous. This must have biased the detection in favour of binding sites that are relatively invariant across tissues, even if Meis affinity for them in individual cells is lower than for tissue-specific binding sites.

Regarding motifs, the lack of OCTA in eye peaks is to be expected, since there is no Hox expression in the eye, but for HSPC peaks it is puzzling. According to the Gene Expression Commons (`http://gexc.stanford.edu`), a repository of hematopoietic gene expression patterns, several Hox genes are expressed in HSPCs: at least *Hoxc6*, *Hoxa2*, 3, 5 and 9 and *Hoxb2* to 5 are expressed at this stage in hematopoietic development. We don't find any reason why Meis should not bind

DNA in concert with them as we understand it does in the trunk. One possible explanation is the aforementioned bias inherent to our sample: if Meis-Hox binding has lower affinity than Meis monomeric binding but is more invariant across tissues, it could have stayed under the detection threshold in Wilson et al. (2010). However, there is no evidence supporting this explanation and in fact the DNA binding affinity of Hox-Meis heterodimers is higher than that of Meis monomers. One other difference is that in HSPCs *Meis1* is the only expressed gene of the family, but again there is no reason to suspect that in our sample Meis2 is exclusively responsible for co-Hox binding.

Another aspect of the comparison that should be pointed out is that we do not find DECA binding in the eye. Since Pbx is present there, there is no obvious reason why Meis should not bind in concert with it. Maybe heterodimeric Pbx-Meis binding is something that happens *in vitro* but is just not common *in vivo*, just as heterotrimeric Prep-Pbx-Hox binding seems to be.

The appearance in HSPC peaks of the accessory motif CWGSCWG supports the hypothesis that it is the recognition sequence for an hematopoietic transcription factor. The presence of CCCgCCC in a set that is not predominantly TSSA argues against it appearing in Prep peaks purely because of their high association with promoters

The polypurine motif present in all TALE trunk peak subsets appears again in HSPCs (m2) and eye (m3) peaks, which suggests the possibility of it being a general requirement for Meis binding, since it seems to be the only constant pattern in Meis binding sites across tissues apart from the HEXA motif.

One conclusion to be drawn from the fact that Meis binding sites in different tissues are almost non-overlapping is Meis binding is not at all constitutive but highly context-dependent. The fact that they share many of the same sequence motifs implies that presence of a binding target sequence is insufficient to determine binding *in vivo*. This is important because to date, we do not know what determines transcription factor binding beyond DNA sequence: all known TFs have recognition sequences that appear many thousands of times in the genome but they only bind a subset of them.

The selection of actual targets from a pool of adequate candidates might be based on chromatin accessibility, interaction with additional cofactors, or sequence features that might affect fine double helix structural features, such as minor groove width. A recent study of the regulatory activity of short sequences from *in vivo* binding sites has shown that, in the case of the homeodomain factor Crx, high GC content is as powerful a predictor of cis-regulatory potential as the presence of high quality recognition sites for Crx (White et al. 2013). Since this study was carried out with reporter constructs, chromatin accessibility was not a factor and therefore can

not be the only differentiator between bound and not bound potential binding sites. In the case of Meis, it is clear that the sequence patterns described in our work are not sufficient to explain GW patterns of binding and their change across tissues. Maybe the key to predicting tissue-specific binding site occupancy lies in the large scale chromatin state. Alternatively, it could be dependent on subtler, more local sequence patterns that our current methods are unable to characterize. Time will tell.

5.7 Stories from Particular Peaks

In addition to the systematic GW and Hox-focused analysis we have presented so far, a number of interesting TALE protein binding patterns have caught our eye during the exploration of the dataset, apart from the *Nfyc* case described above.

Id2 (*Inhibitor of differentiation 2*) has many TALE binding sites in its vicinity, including a Pbx1-Prep^{com} peak in its promoter. *Id2* represses myogenesis and it is in turn repressed by *Rp58* (also known as *Zfp238*), a direct target of MyoD, to promote myogenesis (Yokoyama et al. 2009). *Rp58* has a Pbx1-Prep^{com} peak 2kb upstream of its TSS plus an hematopoietic Meis peak from Wilson et al. (2010). It is suggestive that this system has already been shown to be connected with the TALE proteins (see introduction), which suggests a relatively widespread intertwining of TALE function and MyoD.

Dnm3 (*Dynamin 3*) is a member of a family of enzymes that have a role in membrane dynamics. The *Dnm3* gene has a high density of Meis peaks and Hoxa2 peaks from Donaldson et al. (2012), several of which overlap Meis peaks. *Dnm3* was originally identified as a transcript upregulated during megakaryocyte development and that could stimulate megakaryocyte development when overexpressed in CD34+ cells (Reems et al. 2008). Later, it was shown that *Dnm3* knockdown results in reduced megakaryocyte progenitor amplification and reduced megakaryocyte cytoplasmic enlargement (W. Wang et al. 2011). Strikingly, this report also showed interaction of Dnm3 with Nonmuscle myosin IIA, one of several nonmuscle myosin heavy chains in mammals. Nonmuscle myosin heavy chain IIB interacts with Pbx and when overexpressed *zipper*, its Drosophila homolog, induces cytoplasmic retention of Hth (see Introduction). The three nonmuscle myosin heavy chain genes in mammals (*Myh9*, *Myh10* and *Myh14*) all have Pbx1-Prep^{com} peaks in their promoters. This connections suggest that *Dnm3* may be a mediator of Meis's role in megakaryocyte development, which could be worth pursuing.

Other mediators of the requirement of Meis for megakaryocyte development may be *Ets1* and *Fli-1*. These are linked in the chromosome and have Meis peaks nearby. They have been shown to stimulate transcription of *Pf4* (Okada et al.

2011), a megakaryocyte-specific gene that itself has hematopoietic Meis peaks from Wilson et al. (2010) in its promoter and is under 5kb from *Ppbp*, which is also megakaryocyte-specific and has a Meis peak from our study in its promoter.

These few examples illustrate the potential of the data we have generated to give insight into unexpected functions and relationships of the TALE genes and to generate hypotheses about mediators of TALE protein function.

Six

Conclusions

- Meis and Prep have largely non-overlapping genomic binding sequences. Pbx1 binding overlaps much more with Prep than with Meis.
- While Prep binds mainly at TSSs, specially in concert with Pbx, Meis BSs are extremely conserved and distributed throughout the genome and do not show a particular relationship with transcriptional units, with the exception that exons are less likely to contain Meis BSs than introns. A large proportion of Meis BSs are enhancers.
- The sequences at sites bound by Meis, Prep and Pbx1 suggest Prep tends to bind as a heterodimer with Pbx, without participation of Hox proteins. In contrast, Meis tends to bind DNA in concert with Hoxes, probably as a non-DNA-binding partner in trimeric Meis-Hox-Pbx complexes.
- Meis and Prep regulate some of the same genes through different cis-regulatory sequences. In many cases, they regulate transcription of a given gene in opposite directions.
- TALE factor binding landscape subdivides Hox clusters in two domains. DNA binding in the clusters is dominated by Meis peaks and is restricted to the part of each cluster 3′ to the 9^{th} paralog.
- The transcriptional effect on the Hox clusters is colinear with chromosomic position– in opposite directions between Meis and Prep and between the 5′ and 3′ parts of each cluster. Surprisingly, the largest effect is that seen in *Prep1*$^{i/i}$ embryos on the 5′ part of the clusters, in which we detect no binding of TALE proteins.
- At least some Hox proteins are able to bind the novel sequence TGATGAAT.

Seven

Conclusiones

- Los sitios de unión al ADN de Meis y Prep apenas solapan. Los sitios de unión de Pbx1 solapan mucho más con los de Prep que con los de Meis.
- En contraste con Prep, que se une al ADN mayoritariamente en promotores especialmente en combinación con Pbx1, los sitios de unión de Meis están distribuidos por el genoma sin una relación general con unidades transcripcionales. Cabe el matiz de que los exones están deplecionados de picos Meis. Una proporción importante de picos Meis son enhancers.
- Las secuencias de bases que encontramos en sitios de unión de Meis, Prep y Pbx1 sugieren que Prep une ADN casi siempre en complejo con Pbx, sin participación de proteínas Hox. En cambio, Meis tiende a unir ADN en complejo con Hoxes, probablemente como miembro de complejos Pbx-Meis-Hox pero sin unir ADN.
- Meis y Prep regulan algunos de los mismos genes a través de secuencias reguladoras en *cis* distintas. En muchos casos, regulan la transcripción en sentidos opuestos.
- El paisaje de unión de proteínas TALE divide los clusters Hox en dos dominios. La unión de proteínas TALE a los clusters está dominada por Meis y está restringida a la parte de cada cluster situada 3' del parálogo 9.
- El efecto transcripcional de Meis y Prep en los genes Hox es colineal con la posición cromosómica de los genes, y en sentidos opuestos entre Meis y Prep. Sorprendentemente, el efecto más marcado lo observamos en la parte 5' de los clusters, en la que no detectamos unión de proteínas TALE.
- Al menos algunas proteínas Hox son capaces de unir la secuencia de unión TGATGAAT, no descrita previamente.

Bibliography

ABU-SHAAR, M. and MANN, R. S. (1998). "Generation of multiple antagonistic domains along the proximodistal axis during Drosophila leg development." *Development* 125.19, pp. 3821–3830.

ABU-SHAAR, M., RYOO, H. D., and MANN, R. S. (1999). "Control of the nuclear localization of Extradenticle by competing nuclear import and export signals." *Genes Dev* 13.8, pp. 935–945.

AFFOLTER, M., MARTY, T., and VIGANO, M. A. (1999). "Balancing import and export in development." *Genes Dev* 13.8, pp. 913–915.

ANDERS, S. and HUBER, W. (2010). "Differential expression analysis for sequence count data." *Genome Biol* 11.10, R106.

ANDREY, G., MONTAVON, T., MASCREZ, B., GONZALEZ, F., NOORDERMEER, D., LELEU, M., TRONO, D., SPITZ, F., and DUBOULE, D. (2013). "A switch between topological domains underlies HoxD genes collinearity in mouse limbs." *Science* 340.6137, p. 1234167.

ANGUITA, E., HUGHES, J., HEYWORTH, C., BLOBEL, G. A., WOOD, W. G., and HIGGS, D. R. (2004). "Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2." *EMBO J* 23.14, pp. 2841–2852.

AZCOITIA, V., ARACIL, M., MARTÍNEZ-A, C., and TORRES, M. (2005). "The homeodomain protein Meis1 is essential for definitive hematopoiesis and vascular patterning in the mouse embryo." *Dev Biol* 280.2, pp. 307–320.

BACOLLA, A., COLLINS, J. R., GOLD, B., CHUZHANOVA, N., YI, M., STEPHENS, R. M., STEFANOV, S., OLSH, A., JAKUPCIAK, J. P., DEAN, M., LEMPICKI, R. A., COOPER, D. N., and WELLS, R. D. (2006). "Long homopurine*homopyrimidine sequences are characteristic of genes expressed in brain and the pseudoautosomal region." *Nucleic Acids Res* 34.9, pp. 2663–2675.

BAILEY, T. L., BODEN, M., BUSKE, F. A., FRITH, M., GRANT, C. E., CLEMENTI, L., REN, J., LI, W. W., and NOBLE, W. S. (2009). "MEME SUITE: tools for motif discovery and searching." *Nucleic Acids Res* 37.Web Server issue, W202–W208.

BAILEY, T. L., WILLIAMS, N., MISLEH, C., and LI, W. W. (2006). "MEME: discovering and analyzing DNA and protein sequence motifs." *Nucleic Acids Res* 34.Web Server issue, W369–W373.

BENDER, W., AKAM, M., KARCH, F., BEACHY, P. A., PEIFER, M., SPIERER, P., LEWIS, E. B., and HOGNESS, D. S. (1983). "Molecular Genetics of the Bithorax Complex in Drosophila melanogaster." *Science* 221.4605, pp. 23–29.

BERGER, M. F. et al. (2008). "Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences." *Cell* 133.7, pp. 1266–1276.

BERKES, C. A., BERGSTROM, D. A., PENN, B. H., SEAVER, K. J., KNOEPFLER, P. S., and TAPSCOTT, S. J. (2004). "Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential." *Mol Cell* 14.4, pp. 465–477.

BERTHELSEN, J., KILSTRUP-NIELSEN, C., BLASI, F., MAVILIO, F., and ZAPPAVIGNA, V. (1999). "The subcellular localization of PBX1 and EXD proteins depends on nuclear import and export signals and is modulated by association with PREP1 and HTH." *Genes Dev* 13.8, pp. 946–953.

BERTHELSEN, J., ZAPPAVIGNA, V., FERRETTI, E., MAVILIO, F., and BLASI, F. (1998). "The novel homeoprotein Prep1 modulates Pbx-Hox protein cooperativity". *The EMBO journal* 17.5, pp. 1434–45.

BERTHELSEN, J., ZAPPAVIGNA, V., MAVILIO, F., and BLASI, F. (1998). "Prep1, a novel functional partner of Pbx proteins". *The EMBO journal* 17.5, pp. 1423–33.

BERTOLINO, E., REIMUND, B., WILDT-PERINIC, D., and CLERC, R. G. (1995). "A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif." *J Biol Chem* 270.52, pp. 31178–31188.

BLANKENBERG, D., VON KUSTER, G., CORAOR, N., ANANDA, G., LAZARUS, R., MANGAN, M., NEKRUTENKO, A., and TAYLOR, J. (2010). "Galaxy: a web-based genome analysis tool for experimentalists." *Curr Protoc Mol Biol* Chapter 19, pages.

BLOW, M. J. et al. (2010). "ChIP-Seq identification of weakly conserved heart enhancers." *Nat Genet* 42.9, pp. 806–810.

BRIGGS, M. R., KADONAGA, J. T., BELL, S. P., and TJIAN, R. (1986). "Purification and biochemical characterization of the promoter-specific transcription factor, Sp1." *Science* 234.4772, pp. 47–52.

BURGLIN, T. R. (1997). "Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals". *Nucleic acids research* 25.21, pp. 4173–80.

— (1998). "The PBC domain contains a MEINOX domain: coevolution of Hox and TALE homeobox genes?" *Development genes and evolution* 208.2, pp. 113–6.

CARRAMOLINO, L., FUENTES, J., GARCIA-ANDRES, C., AZCOITIA, V., RIETHMACHER, D., and TORRES, M. (2010). "Platelets play an essential role in separating the blood and lymphatic vasculatures during embryonic angiogenesis". *Circulation research* 106.7, pp. 1197–201.

CARRASCO, A. E., MCGINNIS, W., GEHRING, W. J., and DE ROBERTIS, E. M. (1984). "Cloning of an X. laevis gene expressed during early embryogenesis coding for a peptide region homologous to Drosophila homeotic genes." *Cell* 37.2, pp. 409–414.

CARSTENS, R. P., WAGNER, E. J., and GARCIA-BLANCO, M. A. (2000). "An intronic splicing silencer causes skipping of the IIIb exon of fibroblast growth factor receptor 2 through involvement of polypyrimidine tract binding protein." *Mol Cell Biol* 20.19, pp. 7388–7400.

CATRON, K. M., ILER, N., and ABATE, C. (1993). "Nucleotides flanking a conserved TAAT core dictate the DNA binding specificity of three murine homeodomain proteins." *Mol Cell Biol* 13.4, pp. 2354–2365.

CECCONI, F., PROETZEL, G., ALVAREZ-BOLADO, G., JAY, D., and GRUSS, P. (1997). "Expression of Meis2, a Knotted-related murine homeobox gene, indicates a role in the differentiation of the forebrain and the somitic mesoderm." *Dev Dyn* 210.2, pp. 184–190.

CHAN, S. K., JAFFE, L., CAPOVILLA, M., BOTAS, J., and MANN, R. S. (1994). "The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein." *Cell* 78.4, pp. 603–615.

CHAN, S. K., RYOO, H. D., GOULD, A., KRUMLAUF, R., and MANN, R. S. (1997). "Switching the in vivo specificity of a minimal Hox-responsive element." *Development* 124.10, pp. 2007–2014.

CHANG, C. P., BROCCHIERI, L., SHEN, W. F., LARGMAN, C., and CLEARY, M. L. (1996). "Pbx modulation of Hox homeodomain amino-terminal arms establishes different DNA-binding specificities across the Hox locus." *Mol Cell Biol* 16.4, pp. 1734–1745.

CHANG, C. P., JACOBS, Y., NAKAMURA, T., JENKINS, N. A., COPELAND, N. G., and CLEARY, M. L. (1997). "Meis proteins are major in vivo DNA binding partners for wild-type but not chimeric Pbx proteins." *Mol Cell Biol* 17.10, pp. 5679–5687.

CHANG, C. P., SHEN, W. F., ROZENFELD, S., LAWRENCE, H. J., LARGMAN, C., and CLEARY, M. L. (1995). "Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins." *Genes Dev* 9.6, pp. 663–674.

CREYGHTON, M. P., CHENG, A. W., WELSTEAD, G. G., KOOISTRA, T., CAREY, B. W., STEINE, E. J., HANNA, J., LODATO, M. A., FRAMPTON, G. M., SHARP, P. A., BOYER, L. A., YOUNG, R. A., and JAENISCH, R. (2010). "Histone H3K27ac

separates active from poised enhancers and predicts developmental state." *Proc Natl Acad Sci U S A* 107.50, pp. 21931–21936.

DESPLAN, C., THEIS, J., and O'FARRELL, P. H. (1988). "The sequence specificity of homeodomain-DNA interaction." *Cell* 54.7, pp. 1081–1090.

DI ROSA, P., VILLAESCUSA, J. C., LONGOBARDI, E., IOTTI, G., FERRETTI, E., DIAZ, V. M., MICCIO, A., FERRARI, G., and BLASI, F. (2007). "The homeodomain transcription factor Prep1 (pKnox1) is required for hematopoietic stem and progenitor cell activity." *Dev Biol* 311.2, pp. 324–334.

DIMARTINO, J. F., SELLERI, L., TRAVER, D., FIRPO, M. T., RHEE, J., WARNKE, R., O'GORMAN, S., WEISSMAN, I. L., and CLEARY, M. L. (2001). "The Hox cofactor and proto-oncogene Pbx1 is required for maintenance of definitive hematopoiesis in the fetal liver". *Blood* 98.3, pp. 618–26.

DING, X., YANG, Z., ZHOU, F., WANG, F., LI, X., CHEN, C., LI, X., HU, X., XIANG, S., and ZHANG, J. (2013). "Transcription factor AP-2α regulates acute myeloid leukemia cell proliferation by influencing Hoxa gene expression." *Int J Biochem Cell Biol* 45.8, pp. 1647–1656.

DOERKSEN, L. F., BHATTACHARYA, A., KANNAN, P., PRATT, D., and TAINSKY, M. A. (1996). "Functional interaction between a RARE and an AP-2 binding site in the regulation of the human HOX A4 gene promoter." *Nucleic Acids Res* 24.14, pp. 2849–2856.

DOLFINI, D. and MANTOVANI, R. (2013). "Targeting the Y/CCAAT box in cancer: YB-1 (YBX1) or NF-Y?" *Cell Death Differ* 20.5, pp. 676–685.

DONALDSON, I. J., AMIN, S., HENSMAN, J. J., KUTEJOVA, E., RATTRAY, M., LAWRENCE, N., HAYES, A., WARD, C. M., and BOBOLA, N. (2012). "Genome-wide occupancy links Hoxa2 to Wnt-beta-catenin signaling in mouse embryonic development." *Nucleic Acids Res* 40.9, pp. 3990–4001.

DORÉ, L. C. and CRISPINO, J. D. (2011). "Transcription factor networks in erythroid cell and megakaryocyte development." *Blood* 118.2, pp. 231–239.

DUBOULE, D. (2007). "The rise and fall of Hox gene clusters." *Development* 134.14, pp. 2549–2560.

ENCODE PROJECT CONSORTIUM (2004). "The ENCODE (ENCyclopedia Of DNA Elements) Project." *Science* 306.5696, pp. 636–640.

ENCODE PROJECT CONSORTIUM, BERNSTEIN, B. E., BIRNEY, E., DUNHAM, I., GREEN, E. D., GUNTER, C., and SNYDER, M. (2012). "An integrated encyclopedia of DNA elements in the human genome." *Nature* 489.7414, pp. 57–74.

ENCODE PROJECT CONSORTIUM, BIRNEY, E., et al. (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." *Nature* 447.7146, pp. 799–816.

FERNANDEZ-DIAZ, L. C., LAURENT, A., GIRASOLI, S., TURCO, M., LONGOBARDI, E., IOTTI, G., JENKINS, N. A., FIORENZA, M. T., COPELAND, N. G., and BLASI, F. (2010). "The absence of Prep1 causes p53-dependent apoptosis of mouse pluripotent epiblast cells." *Development* 137.20, pp. 3393–3403.

FERRETTI, E., MARSHALL, H., PÖPPERL, H., MACONOCHIE, M., KRUMLAUF, R., and BLASI, F. (2000). "Segmental expression of Hoxb2 in r4 requires two separate sites that integrate cooperative interactions between Prep1, Pbx and Hox proteins." *Development* 127.1, pp. 155–166.

FERRETTI, E., SCHULZ, H., TALARICO, D., BLASI, F., and BERTHELSEN, J. (1999). "The PBX-regulating protein PREP1 is present in different PBX-complexed forms in mouse." *Mech Dev* 83.1-2, pp. 53–64.

FERRETTI, E., VILLAESCUSA, J. C., DI ROSA, P., FERNANDEZ-DIAZ, L. C., LONGOBARDI, E., MAZZIERI, R., MICCIO, A., MICALI, N., SELLERI, L., FERRARI, G., and BLASI, F. (2006). "Hypomorphic mutation of the TALE gene Prep1 (pKnox1) causes a major reduction of Pbx and Meis proteins and a pleiotropic embryonic phenotype". *Molecular and cellular biology* 26.15, pp. 5650–62.

GARBER, R. L., KUROIWA, A., and GEHRING, W. J. (1983). "Genomic and cDNA clones of the homeotic locus Antennapedia in Drosophila." *EMBO J* 2.11, pp. 2027–2036.

GEHRING, W. J. (1985). "The homeo box: a key to the understanding of development?" *Cell* 40.1, pp. 3–5.

GOECKS, J., NEKRUTENKO, A., TAYLOR, J., and TEAM, G. (2010). "Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences." *Genome Biol* 11.8, R86.

GONZÁLEZ-REYES, A. and MORATA, G. (1990). "The developmental effect of overexpressing a Ubx product in Drosophila embryos is dependent on its interactions with other homeotic products." *Cell* 61.3, pp. 515–522.

GOSIENGFIAO, Y., HORVAT, R., and THOMPSON, A. (2007). "Transcription factors GATA-1 and Fli-1 regulate human HOXA10 expression in megakaryocytic cells." *DNA Cell Biol* 26.8, pp. 577–587.

GOUDET, G., DELHALLE, S., BIEMAR, F., MARTIAL, J. A., and PEERS, B. (1999). "Functional and cooperative interactions between the homeodomain PDX1, Pbx, and Prep1 factors on the somatostatin promoter." *J Biol Chem* 274.7, pp. 4067–4073.

GOULD, A. P., BROOKMAN, J. J., STRUTT, D. I., and WHITE, R. A. (1990). "Targets of homeotic gene control in Drosophila." *Nature* 348.6299, pp. 308–312.

GOULD, A., MORRISON, A., SPROAT, G., WHITE, R. A., and KRUMLAUF, R. (1997). "Positive cross-regulation and enhancer sharing: two mechanisms for specifying overlapping Hox expression patterns." *Genes Dev* 11.7, pp. 900–913.

HALLER, K., RAMBALDI, I., DANIELS, E., and FEATHERSTONE, M. (2004). "Subcellular localization of multiple PREP2 isoforms is regulated by actin, tubulin, and nuclear export." *J Biol Chem* 279.47, pp. 49384–49394.

HARTMANN, H., GUTHÖHRLEIN, E. W., SIEBERT, M., LUEHR, S., and SÖDING, J. (2013). "P-value-based regulatory motif discovery using positional weight matrices." *Genome Res* 23.1, pp. 181–194.

HISA, T., SPENCE, S. E., RACHEL, R. A., FUJITA, M., NAKAMURA, T., WARD, J. M., DEVOR-HENNEMAN, D. E., SAIKI, Y., KUTSUNA, H., TESSAROLLO, L., JENKINS, N. A., and COPELAND, N. G. (2004). "Hematopoietic, angiogenic and eye defects in Meis1 mutant animals." *EMBO J* 23.2, pp. 450–459.

HOEY, T. and LEVINE, M. (1988). "Divergent homeo box proteins recognize similar DNA sequences in Drosophila." *Nature* 332.6167, pp. 858–861.

HOEY, T., WARRIOR, R., MANAK, J., and LEVINE, M. (1988). "DNA-binding activities of the Drosophila melanogaster even-skipped protein are mediated by its homeo domain and influenced by protein context." *Mol Cell Biol* 8.11, pp. 4598–4607.

HUANG, H., PALIOURAS, M., RAMBALDI, I., LASKO, P., and FEATHERSTONE, M. (2003). "Nonmuscle myosin promotes cytoplasmic localization of PBX." *Mol Cell Biol* 23.10, pp. 3636–3645.

HUANG, H., RASTEGAR, M., BODNER, C., GOH, S.-L., RAMBALDI, I., and FEATHERSTONE, M. (2005). "MEIS C termini harbor transcriptional activation domains that respond to cell signaling." *J Biol Chem* 280.11, pp. 10119–10127.

HUANG, Y. et al. (2012). "Identification and characterization of Hoxa9 binding sites in hematopoietic cells". *Blood* 119.2, pp. 388–98.

IOTTI, G., LONGOBARDI, E., MASELLA, S., DARDAEI, L., DE SANTIS, F., MICALI, N., and BLASI, F. (2011). "Homeodomain transcription factor and tumor suppressor Prep1 is required to maintain genomic stability." *Proc Natl Acad Sci U S A* 108.29, E314–E322.

IRIMIA, M., MAESO, I., BURGUERA, D., HIDALGO-SÁNCHEZ, M., PUELLES, L., ROY, S. W., GARCIA-FERNÀNDEZ, J., and FERRAN, J. L. (2011). "Contrasting 5′ and 3′ evolutionary histories and frequent evolutionary convergence in Meis/hth gene structures." *Genome Biol Evol* 3, pp. 551–564.

JACOBS, Y., SCHNABEL, C. A., and CLEARY, M. L. (1999). "Trimeric association of Hox and TALE homeodomain proteins mediates Hoxb2 hindbrain enhancer activity." *Mol Cell Biol* 19.7, pp. 5134–5142.

JI, H., JIANG, H., MA, W., and WONG, W. H. (2011). "Using CisGenome to analyze ChIP-chip and ChIP-seq data." *Curr Protoc Bioinformatics* Chapter 2, Unit2.13.

JUNG, H., LACOMBE, J., MAZZONI, E. O., LIEM JR, K. F., GRINSTEIN, J., MAHONY, S., MUKHOPADHYAY, D., GIFFORD, D. K., YOUNG, R. A., ANDERSON, K. V.,

WICHTERLE, H., and DASEN, J. S. (2010). "Global control of motor neuron topography mediated by the repressive actions of a single hox gene." *Neuron* 67.5, pp. 781–796.

KAMPS, M. P., MURRE, C., SUN, X. H., and BALTIMORE, D. (1990). "A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B ALL". *Cell* 60.4, pp. 547–55.

KASSOUF, M. T., CHAGRAOUI, H., VYAS, P., and PORCHER, C. (2008). "Differential use of SCL/TAL-1 DNA-binding domain in developmental hematopoiesis." *Blood* 112.4, pp. 1056–1067.

KASSOUF, M. T., HUGHES, J. R., TAYLOR, S., MCGOWAN, S. J., SONEJI, S., GREEN, A. L., VYAS, P., and PORCHER, C. (2010). "Genome-wide identification of TAL1's functional targets: insights into its mechanisms of action in primary erythroid cells." *Genome Res* 20.8, pp. 1064–1083.

KAUFMAN, T. C., SEEGER, M. A., and OLSEN, G. (1990). "Molecular and genetic organization of the antennapedia gene complex of Drosophila melanogaster". *Advances in genetics* 27, pp. 309–62.

KIM, S. K., SELLERI, L., LEE, J. S., ZHANG, A. Y., GU, X., JACOBS, Y., and CLEARY, M. L. (2002). "Pbx1 inactivation disrupts pancreas development and in Ipf1-deficient mice promotes diabetes mellitus." *Nat Genet* 30.4, pp. 430–435.

KNOEPFLER, P. S., BERGSTROM, D. A., UETSUKI, T., DAC-KORYTKO, I., SUN, Y. H., WRIGHT, W. E., TAPSCOTT, S. J., and KAMPS, M. P. (1999). "A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with pbx-Meis1/Prep1." *Nucleic Acids Res* 27.18, pp. 3752–3761.

KNOEPFLER, P. S., CALVO, K. R., CHEN, H., ANTONARAKIS, S. E., and KAMPS, M. P. (1997). "Meis1 and pKnox1 bind DNA cooperatively with Pbx1 utilizing an interaction surface disrupted in oncoprotein E2a-Pbx1." *Proc Natl Acad Sci U S A* 94.26, pp. 14553–14558.

KNOEPFLER, P. S. and KAMPS, M. P. (1997). "The Pbx family of proteins is strongly upregulated by a post-transcriptional mechanism during retinoic acid-induced differentiation of P19 embryonal carcinoma cells." *Mech Dev* 63.1, pp. 5–14.

KRIWACKI, R. W., SCHULTZ, S. C., STEITZ, T. A., and CARADONNA, J. P. (1992). "Sequence-specific recognition of DNA by zinc-finger peptides derived from the transcription factor Sp1." *Proc Natl Acad Sci U S A* 89.20, pp. 9759–9763.

LAHLIL, R., LÉCUYER, E., HERBLOT, S., and HOANG, T. (2004). "SCL assembles a multifactorial complex that determines glycophorin A expression." *Mol Cell Biol* 24.4, pp. 1439–1452.

LAMPE, X., SAMAD, O. A., GUIGUEN, A., MATIS, C., REMACLE, S., PICARD, J. J., RIJLI, F. M., and REZSOHAZY, R. (2008). "An ultraconserved Hox-Pbx respon-

sive element resides in the coding sequence of Hoxa2 and is active in rhombomere 4." *Nucleic Acids Res* 36.10, pp. 3214–3225.

LANGSTON, A. W., THOMPSON, J. R., and GUDAS, L. J. (1997). "Retinoic acid-responsive enhancers located 3′ of the Hox A and Hox B homeobox gene clusters. Functional analysis." *J Biol Chem* 272.4, pp. 2167–2175.

LARONDE-LEBLANC, N. A. and WOLBERGER, C. (2003). "Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior." *Genes Dev* 17.16, pp. 2060–2072.

LEVINE, M. and HOEY, T. (1988). "Homeobox proteins as sequence-specific transcription factors." *Cell* 55.4, pp. 537–540.

LEWIS, E. B. (1978). "A gene complex controlling segmentation in Drosophila". *Nature* 276.5688, pp. 565–70.

LI, B. and DEWEY, C. N. (2011). "RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome." *BMC Bioinformatics* 12, p. 323.

LI, H. and DURBIN, R. (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." *Bioinformatics* 25.14, pp. 1754–1760.

LI, L. (2009). "GADEM: a genetic algorithm guided formation of spaced dyads coupled with an EM algorithm for motif discovery." *J Comput Biol* 16.2, pp. 317–329.

LIU, T. et al. (2011). "Cistrome: an integrative platform for transcriptional regulation studies." *Genome Biol* 12.8, R83.

LIU, Y., MACDONALD, R. J., and SWIFT, G. H. (2001). "DNA binding and transcriptional activation by a PDX1.PBX1b.MEIS2b trimer and cooperation with a pancreas-specific basic helix-loop-helix complex." *J Biol Chem* 276.21, pp. 17985–17993.

LOHNES, D. (2003). "The Cdx1 homeodomain protein: an integrator of posterior signaling in the mouse." *Bioessays* 25.10, pp. 971–980.

LONGOBARDI, E., IOTTI, G., DI ROSA, P., MEJETTA, S., BIANCHI, F., FERNANDEZ-DIAZ, L. C., MICALI, N., NUCIFORO, P., LENTI, E., PONZONI, M., DOGLIONI, C., CANIATTI, M., DI FIORE, P. P., and BLASI, F. (2010). "Prep1 (pKnox1)-deficiency leads to spontaneous tumor development in mice and accelerates EmuMyc lymphomagenesis: a tumor suppressor role for Prep1." *Mol Oncol* 4.2, pp. 126–134.

LONGOBARDI, E. and BLASI, F. (2003). "Overexpression of PREP-1 in F9 teratocarcinoma cells leads to a functionally relevant increase of PBX-2 by preventing its degradation." *J Biol Chem* 278.40, pp. 39235–39241.

LUFKIN, T., DIERICH, A., LEMEUR, M., MARK, M., and CHAMBON, P. (1991). "Disruption of the Hox-1.6 homeobox gene results in defects in a region corresponding to its rostral domain of expression." *Cell* 66.6, pp. 1105–1119.

MACONOCHIE, M., KRISHNAMURTHY, R., NONCHEV, S., MEIER, P., MANZANARES, M., MITCHELL, P. J., and KRUMLAUF, R. (1999). "Regulation of Hoxa2 in cranial neural crest cells involves members of the AP-2 family." *Development* 126.7, pp. 1483–1494.

MAGNANI, L., BALLANTYNE, E. B., ZHANG, X., and LUPIEN, M. (2011). "PBX1 genomic pioneer function drives ER α signaling underlying progression in breast cancer." *PLoS Genet* 7.11, e1002368.

MAHMOUD, A. I., KOCABAS, F., MURALIDHAR, S. A., KIMURA, W., KOURA, A. S., THET, S., PORRELLO, E. R., and SADEK, H. A. (2013). "Meis1 regulates postnatal cardiomyocyte cell cycle arrest." *Nature* 497.7448, pp. 249–253.

MAHONY, S., MAZZONI, E. O., MCCUINE, S., YOUNG, R. A., WICHTERLE, H., and GIFFORD, D. K. (2011). "Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis." *Genome Biol* 12.1, R2.

MANN, R. S. and AFFOLTER, M. (1998). "Hox proteins meet more partners." *Curr Opin Genet Dev* 8.4, pp. 423–429.

MANN, R. S. and CHAN, S. K. (1996). "Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins." *Trends Genet* 12.7, pp. 258–262.

MANZANARES, M., BEL-VIALAR, S., ARIZA-MCNAUGHTON, L., FERRETTI, E., MARSHALL, H., MACONOCHIE, M. M., BLASI, F., and KRUMLAUF, R. (2001). "Independent regulation of initiation and maintenance phases of Hoxa3 expression in the vertebrate hindbrain involve auto- and cross-regulatory mechanisms." *Development* 128.18, pp. 3595–3607.

MANZANARES, M., CORDES, S., ARIZA-MCNAUGHTON, L., SADL, V., MARUTHAINAR, K., BARSH, G., and KRUMLAUF, R. (1999). "Conserved and distinct roles of kreisler in regulation of the paralogous Hoxa3 and Hoxb3 genes." *Development* 126.4, pp. 759–769.

MARDIS, E. R. (2007). "ChIP-seq: welcome to the new frontier". *Nature methods* 4.8, pp. 613–4.

MAVES, L., TYLER, A., MOENS, C. B., and TAPSCOTT, S. J. (2009). "Pbx acts with Hand2 in early myocardial differentiation." *Dev Biol* 333.2, pp. 409–418.

MAVES, L., WASKIEWICZ, A. J., PAUL, B., CAO, Y., TYLER, A., MOENS, C. B., and TAPSCOTT, S. J. (2007). "Pbx homeodomain proteins direct Myod activity to promote fast-muscle differentiation." *Development* 134.18, pp. 3371–3382.

MCGINNIS, W., GARBER, R. L., WIRZ, J., KUROIWA, A., and GEHRING, W. J. (1984). "A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans." *Cell* 37.2, pp. 403–408.

MCGINNIS, W. and KRUMLAUF, R. (1992). "Homeobox genes and axial patterning." *Cell* 68.2, pp. 283–302.

MCGINNIS, W., LEVINE, M. S., HAFEN, E., KUROIWA, A., and GEHRING, W. J. (1984). "A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes." *Nature* 308.5958, pp. 428–433.

MERCADER, N., LEONARDO, E., AZPIAZU, N., SERRANO, A., MORATA, G., MARTÍNEZ, C., and TORRES, M. (1999). "Conserved regulation of proximodistal limb axis development by Meis1/Hth." *Nature* 402.6760, pp. 425–429.

MERCADER, N., LEONARDO, E., PIEDRA, M. E., MARTÍNEZ-A, C., ROS, M. A., and TORRES, M. (2000). "Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes." *Development* 127.18, pp. 3961–3970.

MIKKELSEN, T. S. et al. (2007). "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells." *Nature* 448.7153, pp. 553–560.

MIKKOLA, H. K. A., KLINTMAN, J., YANG, H., HOCK, H., SCHLAEGER, T. M., FUJIWARA, Y., and ORKIN, S. H. (2003). "Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene." *Nature* 421.6922, pp. 547–551.

MIRKIN, S. M., LYAMICHEV, V. I., DRUSHLYAK, K. N., DOBRYNIN, V. N., FILIPPOV, S. A., and FRANK-KAMENETSKII, M. D. (1987). "DNA H form requires a homopurine-homopyrimidine mirror repeat." *Nature* 330.6147, pp. 495–497.

MOENS, C. B. and SELLERI, L. (2006). "Hox cofactors in vertebrate development". *Developmental biology* 291.2, pp. 193–206.

MONICA, K., GALILI, N., NOURSE, J., SALTMAN, D., and CLEARY, M. L. (1991). "PBX2 and PBX3, new homeobox genes with extensive homology to the human proto-oncogene PBX1." *Mol Cell Biol* 11.12, pp. 6149–6157.

MONTAVON, T., SOSHNIKOVA, N., MASCREZ, B., JOYE, E., THEVENET, L., SPLINTER, E., DE LAAT, W., SPITZ, F., and DUBOULE, D. (2011). "A regulatory archipelago controls Hox genes transcription in digits." *Cell* 147.5, pp. 1132–1145.

MOOREN, M. M., PULLEYBLANK, D. E., WIJMENGA, S. S., BLOMMERS, M. J., and HILBERS, C. W. (1990). "Polypurine/polypyrimidine hairpins form a triple helix structure at low pH." *Nucleic Acids Res* 18.22, pp. 6523–6529.

MORATA, G. and KERRIDGE, S. (1981). "Sequential functions of the bithorax complex of Drosophila." *Nature* 290.5809, pp. 778–781.

MOSKOW, J. J., BULLRICH, F., HUEBNER, K., DAAR, I. O., and BUCHBERG, A. M. (1995). "Meis1, a PBX1-related homeobox gene involved in myeloid leukemia in BXH-2 mice". *Molecular and cellular biology* 15.10, pp. 5434–43.

MUKHERJEE, K. and BURGLIN, T. R. (2007). "Comprehensive analysis of animal TALE homeobox genes: new conserved motifs and cases of accelerated evolution". *Journal of molecular evolution* 65.2, pp. 137–53.

MUROYAMA, Y., FUJIWARA, Y., ORKIN, S. H., and ROWITCH, D. H. (2005). "Specification of astrocytes by bHLH protein SCL in a restricted region of the neural tube." *Nature* 438.7066, pp. 360–363.

NAKAMURA, T., LARGAESPADA, D. A., SHAUGHNESSY JR, J., JENKINS, N. A., and COPELAND, N. G. (1996). "Cooperative activation of Hoxa and Pbx1-related genes in murine myeloid leukaemias." *Nat Genet* 12.2, pp. 149–153.

NOORDERMEER, D., LELEU, M., SPLINTER, E., ROUGEMONT, J., DE LAAT, W., and DUBOULE, D. (2011). "The dynamic architecture of Hox gene clusters." *Science* 334.6053, pp. 222–225.

NOYES, M. B., CHRISTENSEN, R. G., WAKABAYASHI, A., STORMO, G. D., BRODSKY, M. H., and WOLFE, S. A. (2008). "Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites." *Cell* 133.7, pp. 1277–1289.

NÜSSLEIN-VOLHARD, C. and WIESCHAUS, E. (1980). "Mutations affecting segment number and polarity in Drosophila." *Nature* 287.5785, pp. 795–801.

OKADA, Y. et al. (2011). "Multiple ETS family proteins regulate PF4 gene expression by binding to the same ETS binding site." *PLoS One* 6.9, e24837.

OULAD-ABDELGHANI, M., CHAZAUD, C., BOUILLET, P., SAPIN, V., CHAMBON, P., and DOLLÉ, P. (1997). "Meis2, a novel mouse Pbx-related homeobox gene induced by retinoic acid during differentiation of P19 embryonal carcinoma cells." *Dev Dyn* 210.2, pp. 173–183.

PAIGE, S. L., THOMAS, S., STOICK-COOPER, C. L., WANG, H., MAVES, L., SANDSTROM, R., PABON, L., REINECKE, H., PRATT, G., KELLER, G., MOON, R. T., STAMATOYANNOPOULOS, J., and MURRY, C. E. (2012). "A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development." *Cell* 151.1, pp. 221–232.

PASSNER, J. M., RYOO, H. D., SHEN, L., MANN, R. S., and AGGARWAL, A. K. (1999). "Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex". *Nature* 397.6721, pp. 714–9.

PEERS, B., SHARMA, S., JOHNSON, T., KAMPS, M., and MONTMINY, M. (1995). "The pancreatic islet factor STF-1 binds cooperatively with Pbx to a regulatory element in the somatostatin promoter: importance of the FPWMK motif and of the homeodomain." *Mol Cell Biol* 15.12, pp. 7091–7097.

PEIFER, M. and WIESCHAUS, E. (1990). "Mutations in the Drosophila gene extradenticle affect the way specific homeo domain proteins regulate segmental identity". *Genes & development* 4.7, pp. 1209–23.

PELLERIN, I., SCHNABEL, C., CATRON, K. M., and ABATE, C. (1994). "Hox proteins have different affinities for a consensus DNA site that correlate with the positions of their genes on the hox cluster." *Mol Cell Biol* 14.7, pp. 4532–4545.

PELTENBURG, L. T. and MURRE, C. (1996). "Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx." *EMBO J* 15.13, pp. 3385–3393.

PENKOV, D., DI ROSA, P., FERNANDEZ DIAZ, L., BASSO, V., FERRETTI, E., GRASSI, F., MONDINO, A., and BLASI, F. (2005). "Involvement of Prep1 in the αbeta T-cell receptor T-lymphocytic potential of hematopoietic precursors." *Mol Cell Biol* 25.24, pp. 10768–10781.

PHELAN, M. L. and FEATHERSTONE, M. S. (1997). "Distinct HOX N-terminal arm residues are responsible for specificity of DNA recognition by HOX monomers and HOX.PBX heterodimers." *J Biol Chem* 272.13, pp. 8635–8643.

PIPER, D. E., BATCHELOR, A. H., CHANG, C. P., CLEARY, M. L., and WOLBERGER, C. (1999). "Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation." *Cell* 96.4, pp. 587–597.

PÖPPERL, H., BIENZ, M., STUDER, M., CHAN, S. K., APARICIO, S., BRENNER, S., MANN, R. S., and KRUMLAUF, R. (1995). "Segmental expression of Hoxb-1 is controlled by a highly conserved autoregulatory loop dependent upon exd/pbx." *Cell* 81.7, pp. 1031–1042.

PÖPPERL, H. and FEATHERSTONE, M. S. (1993). "Identification of a retinoic acid response element upstream of the murine Hox-4.2 gene." *Mol Cell Biol* 13.1, pp. 257–265.

PORCHER, C., LIAO, E. C., FUJIWARA, Y., ZON, L. I., and ORKIN, S. H. (1999). "Specification of hematopoietic and vascular development by the bHLH transcription factor SCL without direct DNA binding." *Development* 126.20, pp. 4603–4615.

RAUSCH, J. W. and LE GRICE, S. F. J. (2004). "'Binding, bending and bonding': polypurine tract-primed initiation of plus-strand DNA synthesis in human immunodeficiency virus." *Int J Biochem Cell Biol* 36.9, pp. 1752–1766.

RAUSKOLB, C., PEIFER, M., and WIESCHAUS, E. (1993). "extradenticle, a regulator of homeotic gene activity, is a homolog of the homeobox-containing human proto-oncogene pbx1." *Cell* 74.6, pp. 1101–1112.

RAUSKOLB, C., SMITH, K. M., PEIFER, M., and WIESCHAUS, E. (1995). "extradenticle determines segmental identities throughout Drosophila development". *Development* 121.11, pp. 3663–73.

REEMS, J.-A., WANG, W., TSUBATA, K., ABDURRAHMAN, N., SUNDELL, B., TIJSSEN, M. R., VAN DER SCHOOT, E., DI SUMMA, F., PATEL-HETT, S., ITALIANO JR, J., and GILLIGAN, D. M. (2008). "Dynamin 3 participates in the growth and development of megakaryocytes." *Exp Hematol* 36.12, pp. 1714–1727.

REN, B., ROBERT, F., WYRICK, J. J., APARICIO, O., JENNINGS, E. G., SIMON, I., ZEITLINGER, J., SCHREIBER, J., HANNETT, N., KANIN, E., VOLKERT, T. L., WILSON, C. J., BELL, S. P., and YOUNG, R. A. (2000). "Genome-wide location and function of DNA binding proteins". *Science* 290.5500, pp. 2306–9.

REYMOND, A., MARIGO, V., YAYLAOGLU, M. B., LEONI, A., UCLA, C., SCAMUFFA, N., CACCIOPPOLI, C., DERMITZAKIS, E. T., LYLE, R., BANFI, S., EICHELE, G., ANTONARAKIS, S. E., and BALLABIO, A. (2002). "Human chromosome 21 gene expression atlas in the mouse." *Nature* 420.6915, pp. 582–586.

RHEE, H. S. and PUGH, B. F. (2011). "Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution." *Cell* 147.6, pp. 1408–1419.

RIECKHOF, G. E., CASARES, F., RYOO, H. D., ABU-SHAAR, M., and MANN, R. S. (1997). "Nuclear translocation of extradenticle requires homothorax, which encodes an extradenticle-related homeodomain protein". *Cell* 91.2, pp. 171–83.

ROBERTSON, G. et al. (2007). "Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing". *Nature methods* 4.8, pp. 651–7.

ROSELLÓ-DÍEZ, A., ROS, M. A., and TORRES, M. (2011). "Diffusible signals, not autonomous mechanisms, determine the main proximodistal limb subdivision." *Science* 332.6033, pp. 1086–1088.

SALEH, M., HUANG, H., GREEN, N. C., and FEATHERSTONE, M. S. (2000). "A conformational change in PBX1A is necessary for its nuclear localization." *Exp Cell Res* 260.1, pp. 105–115.

SALEH, M., RAMBALDI, I., YANG, X. J., and FEATHERSTONE, M. S. (2000). "Cell signaling switches HOX-PBX complexes from repressors to activators of transcription mediated by histone deacetylases and histone acetyltransferases." *Mol Cell Biol* 20.22, pp. 8623–8633.

SCHMIDT, D., WILSON, M. D., BALLESTER, B., SCHWALIE, P. C., BROWN, G. D., MARSHALL, A., KUTTER, C., WATT, S., MARTINEZ-JIMENEZ, C. P., MACKAY, S., TALIANIDIS, I., FLICEK, P., and ODOM, D. T. (2010). "Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding." *Science* 328.5981, pp. 1036–1040.

SCHNABEL, C. A., JACOBS, Y., and CLEARY, M. L. (2000). "HoxA9-mediated immortalization of myeloid progenitors requires functional interactions with TALE cofactors Pbx and Meis." *Oncogene* 19.5, pp. 608–616.

SCHNEIDER, T. D. and STEPHENS, R. M. (1990). "Sequence logos: a new way to display consensus sequences." *Nucleic Acids Res* 18.20, pp. 6097–6100.

SCOTT, M. P., WEINER, A. J., HAZELRIGG, T. I., POLISKY, B. A., PIRROTTA, V., SCALENGHE, F., and KAUFMAN, T. C. (1983). "The molecular organization of the Antennapedia locus of Drosophila." *Cell* 35.3 Pt 2, pp. 763–776.

SELLERI, L., DEPEW, M. J., JACOBS, Y., CHANDA, S. K., TSANG, K. Y., CHEAH, K. S., RUBENSTEIN, J. L., O'GORMAN, S., and CLEARY, M. L. (2001). "Requirement for Pbx1 in skeletal patterning and programming chondrocyte proliferation and differentiation". *Development* 128.18, pp. 3543–57.

SHEN, W. F., MONTGOMERY, J. C., ROZENFELD, S., MOSKOW, J. J., LAWRENCE, H. J., BUCHBERG, A. M., and LARGMAN, C. (1997). "AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins." *Mol Cell Biol* 17.11, pp. 6448–6458.

SHEN, W. F., ROZENFELD, S., KWONG, A., KÖM VES, L. G., LAWRENCE, H. J., and LARGMAN, C. (1999). "HOXA9 forms triple complexes with PBX2 and MEIS1 in myeloid cells." *Mol Cell Biol* 19.4, pp. 3051–3061.

SHEN, W. F., ROZENFELD, S., LAWRENCE, H. J., and LARGMAN, C. (1997). "The Abd-B-like Hox homeodomain proteins can be subdivided by the ability to form complexes with Pbx1a on a novel DNA target." *J Biol Chem* 272.13, pp. 8198–8206.

SHEN, Y., YUE, F., MCCLEARY, D. F., YE, Z., EDSALL, L., KUAN, S., WAGNER, U., DIXON, J., LEE, L., LOBANENKOV, V. V., and REN, B. (2012). "A map of the cis-regulatory sequences in the mouse genome." *Nature* 488.7409, pp. 116–120.

SHIN, H., LIU, T., MANRAI, A. K., and LIU, X. S. (2009). "CEAS: cis-regulatory element annotation system." *Bioinformatics* 25.19, pp. 2605–2606.

SIEPEL, A. et al. (2005). "Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes." *Genome Res* 15.8, pp. 1034–1050.

SIMEONE, A., ACAMPORA, D., ARCIONI, L., ANDREWS, P. W., BONCINELLI, E., and MAVILIO, F. (1990). "Sequential activation of HOX2 homeobox genes by retinoic acid in human embryonal carcinoma cells." *Nature* 346.6286, pp. 763–766.

SLATTERY, M., RILEY, T., LIU, P., ABE, N., GOMEZ-ALCALA, P., DROR, I., ZHOU, T., ROHS, R., HONIG, B., BUSSEMAKER, H. J., and MANN, R. S. (2011). "Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins." *Cell* 147.6, pp. 1270–1282.

SOLOMON, M. J., LARSEN, P. L., and VARSHAVSKY, A. (1988). "Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene". *Cell* 53.6, pp. 937–47.

STANKUNAS, K., SHANG, C., TWU, K. Y., KAO, S.-C., JENKINS, N. A., COPELAND, N. G., SANYAL, M., SELLERI, L., CLEARY, M. L., and CHANG, C.-P. (2008). "Pbx/Meis deficiencies demonstrate multigenetic origins of congenital heart disease." *Circ Res* 103.7, pp. 702–709.

SUSKE, G., BRUFORD, E., and PHILIPSEN, S. (2005). "Mammalian SP/KLF transcription factors: bring in the family." *Genomics* 85.5, pp. 551–556.

SWIATEK, P. J. and GRIDLEY, T. (1993). "Perinatal lethality and defects in hindbrain development in mice homozygous for a targeted mutation of the zinc finger gene Krox20." *Genes Dev* 7.11, pp. 2071–2084.

SWIFT, G. H., LIU, Y., ROSE, S. D., BISCHOF, L. J., STEELMAN, S., BUCHBERG, A. M., WRIGHT, C. V., and MACDONALD, R. J. (1998). "An endocrine-exocrine switch in the activity of the pancreatic homeodomain protein PDX1 through formation of a trimeric complex with PBX1b and MRG1 (MEIS2)." *Mol Cell Biol* 18.9, pp. 5109–5120.

TAMPLIN, O. J., KINZEL, D., COX, B. J., BELL, C. E., ROSSANT, J., and LICKERT, H. (2008). "Microarray analysis of Foxa2 mutant mouse embryos reveals novel gene expression and inductive roles for the gastrula organizer and its derivatives." *BMC Genomics* 9, p. 511.

THORSTEINSDOTTIR, U., KROON, E., JEROME, L., BLASI, F., and SAUVAGEAU, G. (2001). "Defining roles for HOX and MEIS1 genes in induction of acute myeloid leukemia." *Mol Cell Biol* 21.1, pp. 224–234.

TIJSSEN, M. R. et al. (2011). "Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators." *Dev Cell* 20.5, pp. 597–609.

TREISMAN, J., HARRIS, E., WILSON, D., and DESPLAN, C. (1992). "The homeodomain: a new face for the helix-turn-helix?" *Bioessays* 14.3, pp. 145–150.

TUERK, C. and GOLD, L. (1990). "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase." *Science* 249.4968, pp. 505–510.

TÜMPEL, S., CAMBRONERO, F., FERRETTI, E., BLASI, F., WIEDEMANN, L. M., and KRUMLAUF, R. (2007). "Expression of Hoxa2 in rhombomere 4 is regulated by a conserved cross-regulatory mechanism dependent upon Hoxb1." *Dev Biol* 302.2, pp. 646–660.

VAN DIJK, M. A., VOORHOEVE, P. M., and MURRE, C. (1993). "Pbx1 is converted into a transcriptional activator upon acquiring the N-terminal region of E2A

in pre-B-cell acute lymphoblastoid leukemia." *Proc Natl Acad Sci U S A* 90.13, pp. 6061–6065.

VIEILLE-GROSJEAN, I. and HUBER, P. (1995). "Transcription factor GATA-1 regulates human HOXB2 gene expression in erythroid cells." *J Biol Chem* 270.9, pp. 4544–4550.

WAGNER, K., MINCHEVA, A., KORN, B., LICHTER, P., and PÖPPERL, H. (2001). "Pbx4, a new Pbx family member on mouse chromosome 8, is expressed during spermatogenesis." *Mech Dev* 103.1-2, pp. 127–131.

WAMSTAD, J. A. et al. (2012). "Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage." *Cell* 151.1, pp. 206–220.

WANG, G. and VASQUEZ, K. M. (2004). "Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells." *Proc Natl Acad Sci U S A* 101.37, pp. 13448–13453.

WANG, W., GILLIGAN, D. M., SUN, S., WU, X., and REEMS, J.-A. (2011). "Distinct functional effects for dynamin 3 during megakaryocytopoiesis." *Stem Cells Dev* 20.12, pp. 2139–2151.

WHITE, M. A., MYERS, C. A., CORBO, J. C., and COHEN, B. A. (2013). "Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks." *Proc Natl Acad Sci U S A* 110.29, pp. 11952–11957.

WILLIAMS, T. M., WILLIAMS, M. E., and INNIS, J. W. (2005). "Range of HOX/TALE superclass associations and protein domain requirements for HOXA13:MEIS interaction." *Dev Biol* 277.2, pp. 457–471.

WILSON, N. K., FOSTER, S. D., WANG, X., KNEZEVIC, K., SCHUTTE, J., KAIMAKIS, P., CHILARSKA, P. M., KINSTON, S., OUWEHAND, W. H., DZIERZAK, E., PIMANDA, J. E., BRUIJN, M. F. DE, and GOTTGENS, B. (2010). "Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators". *Cell stem cell* 7.4, pp. 532–44.

WONG, P., IWASAKI, M., SOMERVAILLE, T. C. P., SO, C. W. E., SO, C. W. E., and CLEARY, M. L. (2007). "Meis1 is an essential and rate-limiting regulator of MLL leukemia stem cell potential." *Genes Dev* 21.21, pp. 2762–2774.

YANG, Y., HWANG, C. K., D'SOUZA, U. M., LEE, S. H., JUNN, E., and MOURADIAN, M. M. (2000). "Three-amino acid extension loop homeodomain proteins Meis2 and TGIF differentially regulate transcription." *J Biol Chem* 275.27, pp. 20734–20741.

YOKOYAMA, S. et al. (2009). "A systems approach reveals that the myogenesis genome network is regulated by the transcriptional repressor RP58." *Dev Cell* 17.6, pp. 836–848.

ZHANG, X., ROBERTSON, G., KRZYWINSKI, M., NING, K., DROIT, A., JONES, S., and GOTTARDO, R. (2011). "PICS: probabilistic inference for ChIP-seq." *Biometrics* 67.1, pp. 151–163.

ZHANG, Y., LIU, T., MEYER, C. A., EECKHOUTE, J., JOHNSON, D. S., BERNSTEIN, B. E., NUSBAUM, C., MYERS, R. M., BROWN, M., LI, W., and LIU, X. S. (2008). "Model-based analysis of ChIP-Seq (MACS)." *Genome Biol* 9.9, R137.

Publications

Cell Reports
Resource

Analysis of the DNA-Binding Profile and Function of TALE Homeoproteins Reveals Their Specialization and Specific Interactions with Hox Genes/Proteins

Dmitry Penkov,[1,2,7] Daniel Mateos San Martín,[3,7] Luis C. Fernandez-Díaz,[1] Catalina A. Rosselló,[3] Carlos Torroja,[4] Fátima Sánchez-Cabo,[4] H.J. Warnatz,[5] Marc Sultan,[5] Marie L. Yaspo,[5] Arianna Gabrieli,[1] Vsevolod Tkachuk,[2] Andrea Brendolan,[6] Francesco Blasi,[1,*] and Miguel Torres[3,*]

[1]IFOM (Foundation FIRC Institute of Molecular Oncology) at the IFOM-IEO Campus, via Adamello 16, 20139 Milan, Italy
[2]Department of Basic Medicine, Lomonosov Moscow State University, Lomonosov Prospect, 31/5, 119192, Moscow, Russia
[3]Cardiovascular Development and Repair Department
[4]Bioinformatics Unit
Centro Nacional de Investigaciones Cardiovasculares (CNIC), Melchor Fernández Almagro 3, 28029 Madrid, Spain
[5]Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestrasse 63-73, 14195 Berlin, Germany
[6]San Raffaele Scientific Institute, Division of Molecular Oncology, via Olgettina 60, 20123, Milan, Italy
[7]These authors contributed equally to this work
*Correspondence: francesco.blasi@ifom.eu (F.B.), mtorres@cnic.es (M.T.)
http://dx.doi.org/10.1016/j.celrep.2013.03.029

SUMMARY

The interactions of Meis, Prep, and Pbx1 TALE homeoproteins with Hox proteins are essential for development and disease. Although Meis and Prep behave similarly in vitro, their in vivo activities remain largely unexplored. We show that Prep and Meis interact with largely independent sets of genomic sites and select different DNA-binding sequences, Prep associating mostly with promoters and housekeeping genes and Meis with promoter-remote regions and developmental genes. Hox target sequences associate strongly with Meis but not with Prep binding sites, while Pbx1 cooperates with both Prep and Meis. Accordingly, *Meis1* shows strong genetic interaction with *Pbx1* but not with *Prep1*. *Meis1* and *Prep1* nonetheless coregulate a subset of genes, predominantly through opposing effects. Notably, the TALE homeoprotein binding profile subdivides Hox clusters into two domains differentially regulated by *Meis1* and *Prep1*. During evolution, Meis and Prep thus specialized their interactions but maintained significant regulatory coordination.

INTRODUCTION

The specificity of transcription in a crowded eukaryotic chromatin is something of a mystery. Different members of closely related transcription factor families bind near-identical DNA sequences in vitro, but their individual function in vivo is rarely known. Transcription factors may also bind different cofactors, resulting in differing patterns of DNA recognition and binding. An example is provided by the Hox and TALE (three amino acid loop extension) families (Moens and Selleri, 2006), which have similar DNA-binding domains. Interaction between the *Drosophila* TALE proteins Extradenticle (Exd) and Homothorax (Hth) targets the two proteins to the nucleus (Chan et al., 1994; Rieckhof et al., 1997) where the complex interacts with Hox proteins, determining their DNA-binding specificity and thereby anteroposterior segmental identity (reviewed in Mann and Affolter, 1998).

The genomes of mammals contain four *Exd*-related genes (*Pbx*) and two *Hth*-related subfamilies, *Meis* and *Prep* (the latter also known as *pKnox*), respectively comprising three and two members. The interaction of Exd with Hth or Hox has been retained in all species, and hence in vertebrates Pbx proteins form complexes with Hox, Meis, and Prep. Pbx proteins interact with Prep or Meis through a conserved amino-terminal domain (Berthelsen et al., 1998; Chang et al., 1997; Knoepfler et al., 1997) and with Hox proteins through the homeodomain (Piper et al., 1999). The independent interaction surfaces allow Pbx to form trimers with Prep or Meis and Hox, and this interaction alters the DNA-binding selectivity of the individual Hox proteins (Ferretti et al., 2000; Jacobs et al., 1999; Ryoo et al., 1999). Meis, but not Prep, can also interact directly with posterior Hox proteins (Williams et al., 2005).

The full complexity of the TALE transcriptional regulatory network in vivo has not even been estimated. Our knowledge of these factors' DNA sequence specificity is based on in vitro selection of target sequences by purified or in-vitro-translated protein complexes and on the analysis of a limited number of endogenous target sequences. A general observation is that affinity for DNA is low for monomers and increases with heterologous complex formation. Prep and Meis alone preferentially bind the TGACAG hexameric sequence (PM sites) (Berthelsen et al., 1998; Chang et al., 1997; Ferretti et al., 2000; Shen et al., 1997) and Pbx to the TGATTGAT sequence (LeBrun and Cleary, 1994). Prep-Pbx and Meis-Pbx dimers both preferentially bind the decameric sequence TGATTGACAG (Chang et al., 1997; Knoepfler et al., 1997). Pbx-Hox dimers bind octameric motifs

of the type TGATNNAT, in which the variable core determines the Hox paralog group binding (Shen et al., 1997). Studies combining oligonucleotide selection (SELEX) with deep sequencing (SELEX-seq) in *Drosophila* show that the site variants at the variable core can be grouped into three main classes of specificity that obey the colinearity rules and underline the preference of Hox for distinct DNA minor groove topographies (Slattery et al., 2011). X-ray studies showed that in Pbx-Hox binding to the octameric sites, each monomer binds one half-site (LaRonde-LeBlanc and Wolberger, 2003; Piper et al., 1999). Ternary complexes take place through the interaction between Meis/Prep bound to hexameric sites and nearby Pbx-Hox bound to octameric sites through direct Prep/Meis-Pbx interaction (Berthelsen et al., 1998; Ferretti et al., 2000, 2005; Ryoo et al., 1999). Ternary complexes can also form by Meis1 interaction with DNA-bound Pbx-Hox dimers without Meis1 binding DNA (Shanmugam et al., 1999).

Meis and Prep proteins contain two homologous functional domains: the Pbx-interacting domain and the homeodomain. The homeodomains and Pbx-interacting regions of Meis1 and Prep1 are 84% and 63% identical. However, other regions of the two proteins are not conserved, including the C-terminal domain, which is essential for Meis1 oncogenic activity (Bisaillon et al., 2011; Wong et al., 2007). Both Prep1 and Meis1 dimerize with Pbx and recognize similar DNA sequences in vitro. Although some specific functions have been identified for Prep and Meis, there is no information about whether their activities are coordinated in vivo.

Prep1 is ubiquitously expressed from the oocyte to the embryo and the adult (Fernandez-Diaz et al., 2010; Ferretti et al., 1999, 2006). *Meis1* and *Meis2* encode very similar proteins (Moskow et al., 1995; Nakamura et al., 1996), and their expression starts around gastrulation and is regionalized (Cecconi et al., 1997; Oulad-Abdelghani et al., 1997). *Pbx1*, *Prep1*, and *Meis1* are developmentally essential genes. *Prep1* null embryos die shortly after implantation, with massive apoptosis and proliferation defects (Fernandez-Diaz et al., 2010). *Pbx1* deletion is embryonically lethal at embryonic day (E) 15.5, and embryos display major homeotic anomalies, organ absence or hypoplasia, hematopoietic defects, and other features (DiMartino et al., 2001; Selleri et al., 2001). *Meis1*-deficient mice die at E14.5 with definitive hematopoietic stem cell failure, megakaryocyte lineage aplasia, lymphatic vasculature defects, heart defects, and eye hypoplasia (Azcoitia et al., 2005; Hisa et al., 2004). While *Prep1* null embryos die early, hypomorphic mutants (*Prep1*$^{i/i}$), in which only 3%–7% of the wild-type protein is produced, show variable viability during gestation. *Prep1* hypomorphs show defects in hematopoiesis, including hematopoietic stem cells, eye development, and angiogenesis (Di Rosa et al., 2007; Ferretti et al., 2006). Although impairment of eye development, hematopoiesis, and angiogenesis is common in *Meis1* and *Prep1*^{i} mutants, the specific aspects affected are different. In addition, the involvement of these factors in disease is clearly divergent, since *Prep1* acts as a tumor suppressor (Iotti et al., 2011; Longobardi et al., 2010) while *Meis1* is leukemogenic (Moskow et al., 1995; Wong et al., 2007) and *Meis1* leukemogenic activity cannot be replaced by *Prep1* (Thorsteinsdottir et al., 2001).

We have undertaken a comprehensive comparative analysis of the genomic interaction and function of Meis, Prep, and Pbx1 in mouse embryos in vivo. We show that Meis and Prep mostly select distinct genomic sites and DNA motifs and show differential interactions with Hox genes and proteins. Our analysis establishes a framework for understanding the mechanisms of action of TALE proteins in development and disease.

RESULTS

Prep and Meis Select, and Drive Pbx1 to, Different Genomic Sites

Chromatin immunoprecipitation sequencing (ChIP-seq) on E11.5 embryos with antibodies to Prep1/2, Pbx1, or Meis1/2 (see Experimental Procedures) detected 3,331 peaks for Prep, 5,686 for Meis, and 3,504 for Pbx1 (Table S1) (Gene Expression Ominbus [GEO] accession number GSE39609). The nonredundant peak list contains 10,326 genomic regions, of which 82% correspond to single-factor-bound regions, 16% to two-factor-bound regions, and 2% to regions bound by all three (Figure 1A). About half of the Pbx1 and Prep peaks were exclusively bound by these factors (Pbx1^{exc}, Prep^{exc}), while 85% of Meis peaks were exclusive (Meis^{exc}), suggesting more independent activities for Meis compared with Prep and Pbx1 (Figure 1A). Analysis of peak overlaps revealed a lower coincidence of Prep with Meis (Prep-Meis^{com}) than with Pbx1 (5.6% versus 30% of Prep peaks; Figure 1A). Almost 30% of the Pbx1 peaks were also bound by Prep, a larger proportion than by Meis (12.6%; Figure 1A). An additional 6% of Pbx1 peaks were simultaneously bound by Prep and Meis (triple peaks).

ChIP-re-ChIP assays of double and triple peaks confirmed that Pbx1 binds simultaneously with either Prep or Meis in a majority of sites (10/17 for Prep and 17/21 for Meis) (Figure S1A). In contrast, in 15/17 Meis-Prep common peaks, these factors do not show simultaneous binding (Figures 1B and S1A). In triple peaks, the most frequent situation was thus alternative binding by either Prep+Pbx1 or Meis+Pbx. The mapping of the relative positions between pairs of the three factors in triple peaks indicates that their binding preferences are cocentered (Figure S1B), indicating that in most cases they bind to the same sequences. Given that in most sites there is no simultaneous binding of Meis and Prep, these factors may compete for binding to the same sequences. The infrequent cases of the simultaneous cobinding of Meis and Prep may thus correspond to the independent binding of the two factors to neighboring target sequences.

In relative terms, Prep-Pbx1 cobinding is therefore predominant with respect to Meis-Pbx1 cobinding. In addition, the analysis of Prep-Pbx2 site occupancy in the thymus, where only Pbx2 is expressed, indicates that the embryonic Pbx1-Prep^{com} peaks can be bound by Prep-Pbx2 when Pbx1 is not available (Tables S2 and S3). These data show that Prep-Pbx interactions are predominant in Prep targets and can occur with different Pbx partners.

Prep Binding Sites Correlate with Transcription Start Sites, while Meis Binding Sites Concentrate in Transcription-Start-Site-Remote Regions

We next analyzed the distribution of the peaks according to their position with respect to RefSeq genes. We classified peaks as transcription start site associated (TSSA) when they appeared

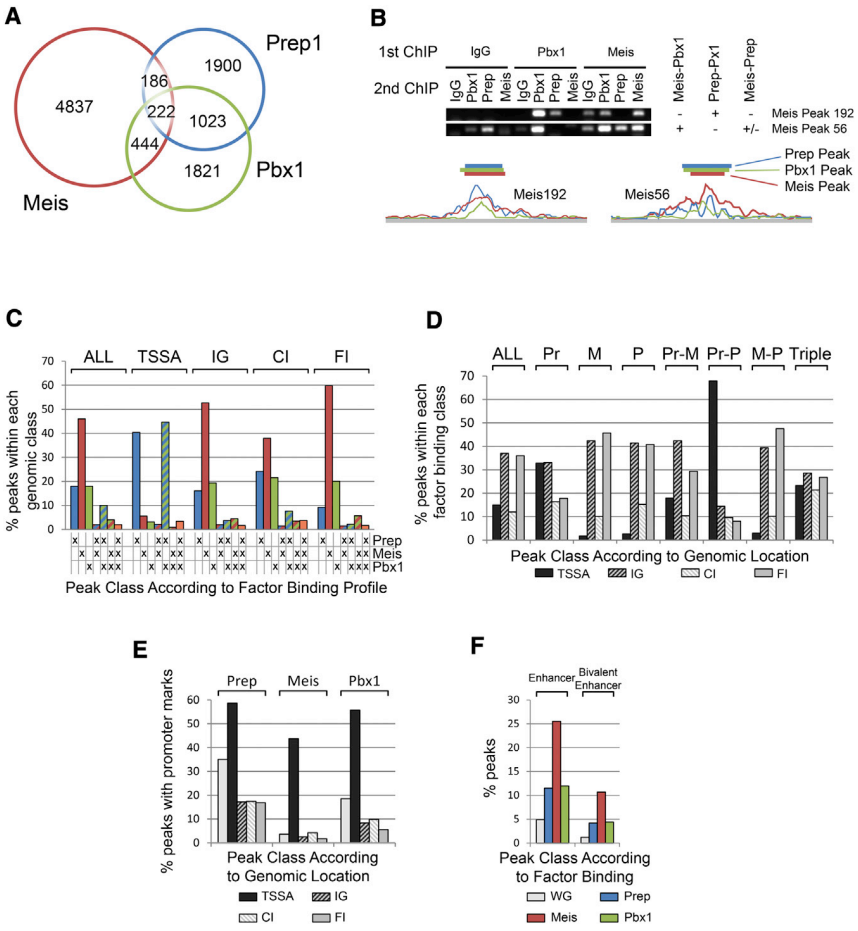

Figure 1. Meis and Prep Select Different Binding Sites and Gene-Regulatory Regions in Cooperation with Pbx1

(A) Venn diagram of peak classes containing single, double, and triple binding by Meis, Prep, and Pbx1. Prep-Meis overlap versus Prep-Pbx overlap and Pbx1-Prep versus Pbx1-Meis adjusted p values (adjp) < 0.0001.

(B) ChIP-re-ChIP experiment. Top: PCR amplification of consecutive immunoprecipitations with anti-immunoglobulin G (IgG), anti-Meis, anti-Prep or anti-Pbx1 antibodies. Bottom: Read profile of peaks tested above. Color bands represent peaks as called by PICS, and color lines represent read density. The interpretation of the cobinding is shown to the right of the gel and is based on the comparison of the specific band intensity with that of the control immunoprecipitations (IgG).

(C) Percentage of peaks located in transcription-start site-associated (TSSA), intragenic (IG), close-intergenic (CI), or far-intergenic (FI) regions that belong to each factor binding class.

(D) Percentage distribution of genomic location classes within each factor binding profile category (adjp = 1 for Meis^{exc} versus Meis-Pbx1^{com} and p = 0.26 for Pbx1^{exc} versus Meis-Pbx1^{com}; adjp < 0.0001 for Prep enrichment in TSSA and deployment in FI classes and for Prep-Pbx1^{com} profile versus that of either Prep or Pbx1 alone).

(E) Percentage of Meis, Prep, or Pbx1 peaks containing promoter marks within each genomic location (adjp < 0.0001 for Prep promoter marks preference in any class and for Meis, only in the TSSA class).

(F) Percentage of Meis, Prep, and Pbx1 peaks containing enhancer marks (adjp < 0.0001 for Meis association with enhancer marks and adjp < 0.01 for Prep and Pbx1). WG, whole genome.

See also Figures S1, S2, and Tables S1, S2, and S3.

within −500 to +100 bp from a transcription start site (TSS), intragenic (IG) when they overlapped a transcription unit, close intergenic (CI) when they appeared <20 kb from a TSS, and far intergenic (FI) when they were located >20 kb from the closest TSS. We first studied the abundance of the different peak classes defined by factor binding profile within each of these genomic regions. Within the TSSA class, the most abundant peaks were Prep^{exc} and Prep-Pbx1^{com} sites, which together represented 85% of all TSSA peaks (Figure 1C). In contrast, in all other genomic regions, single-factor-bound sites were the most abundant peak classes, with Meis^{exc} peaks predominating in all classes, but especially in the IG and FI classes. Peaks in the Prep^{exc} and Pbx^{exc} lists were moderately represented in non-TSSA classes, where peaks bound by more than one factor were generally of low abundance. Within the IG class, peaks for Prep and Pbx1 show a neutral distribution between exonic and intronic regions; however, Meis peaks show a 4-fold reduction in the expected occurrence in exons (p < 0.0001 for Meis, p = 0.1 for Prep, and p = 1 for Pbx1).

To determine how cobinding modifies the binding preferences of each factor, we studied the distribution of different genomic regions across the peak classes defined by factor binding profile (Figure 1D). Binding distributions for Pbx1^{exc}, Meis^{exc}, and Meis-Pbx1^{com} were very similar, with low preference for TSSA regions and CI and high preference for IG and FI compared with the distribution shown by all peaks (Figure 1D). These data indicate that Pbx1 and Meis have similar preferences individually and that their cobinding does not change these preferences. Prep alone, in contrast, showed a strong preference for TSSA regions (41-fold enrichment compared to genomic TSSA region content) and a low preference for FI regions. Unlike Meis-Pbx1^{com} peaks, the Prep-Pbx1^{com} profile diverged sharply from that observed for each factor in isolation, with a marked prevalence of binding to TSSA regions (71.5% for Prep-Pbx1^{com} versus 2.6% and 32.8% for Pbx1^{exc} and Prep^{exc}, respectively) and underrepresentation of all other regions with respect to the Prep^{exc} and Pbx1^{exc} profiles. These data indicate a strong preference of Prep-Pbx1 dimers for TSSA regions, which is led mainly by Prep since Pbx1 alone does not show any such preference. Common binding of Prep and Meis mostly affected the TSSA and FI classes, appearing at frequencies between those observed for the single factors. Prep-Meis cobinding thus displays mixed properties of the two independent factors and does not generate new binding preferences. The peaks bound by all three factors are predominantly enriched in the TSSA and CI classes in comparison with the whole genome. The

binding preferences of TALE factors in the genome correlate with the global distribution of their occupancy levels (Figure S2).

We next studied the correlations between the identified peaks and known epigenetic marks. Peaks located close to a TSS could coincide with promoters, which are associated with H3K4Me3 and RNAPolII marks (Mikkelsen et al., 2007). Prep peaks are strongly enriched in promoter marks (35% of Prep peaks versus 0.4% in the whole genome), not only in the TSSA category but also within the CI and IG categories and, to a lesser extent, the FI category (Figure 1E). Prep thus appears to have a strong binding preference for promoters and sequences with promoter-like epigenetic marks. This tendency was weaker for Meis peaks, which only correlated significantly with promoter marks in the TSSA and CI peaks, and at a lower proportion than Prep. An intermediate situation was found for Pbx1, which showed a very strong association with promoter marks for the TSSA peaks and a significant association, but weaker than that observed for Prep peaks, in other genomic regions (Figure 1E). Similar analyses of the coincidence of peaks with murine embryonic fibroblast enhancer ($H3K4Me1^+$, $H3K4Me3^-$) and bivalent enhancer ($H3K4Me1^+$, $H3K27Ac^+$) marks (Shen et al., 2012) revealed more than a 5-fold enrichment of Meis peaks with both enhancer and bivalent enhancer marks with respect to the whole genome and 2-fold with respect to Prep and Pbx1 peaks (Figure 1F).

Thus, while many Prep and Pbx1 sites are located in promoters, Meis peaks show a preference for enhancers. Interestingly, however, Meis peak sequences are more conserved than those of Prep, Pbx1, and several other developmental transcription factors (Figure 2A). A notable exception is the conservation of HoxC9 binding sites in the embryonic spinal cord (Jung et al., 2010), whose conservation profile is very similar to that of the Meis peaks. The degree of conservation of Meis and HoxC9 peaks is only surpassed by that of p300 peaks in forebrain (Blow et al., 2010).

Prep and Meis Select Different DNA-Binding Sequences in the Genome, Alone or in Combination with Pbx1

To identify consensus DNA sequences in the identified peaks, we performed an unbiased search using rGADEM software (comparable results were obtained with MEME; data not shown) (Figure S3). For each peak, we searched 300 bp centered on the peak maximum. We obtained two types of motifs: those mapping at a single maximum coinciding with the peak center, which we call core motifs, and those showing a bimodal distribution with maxima symmetrically flanking the peak center or showing a spread distribution, which we call accessory motifs. Within the core motifs, we identified the following known motifs: hexameric sequences resembling or identical to the previously in-vitro-described Meis/Prep consensus (HEXA), octameric sequences similar to Pbx/Hox sites (OCTA), a decameric sequence containing a 5′ Pbx1 half-site followed by a Meis/Prep site (DECA), and an extended version of the DECA sequence containing a CCAAT sequence at a fixed distance ($DECA^{ext}$) (Figure S3).

Within the $Prep^{exc}$ sites, an unbiased motif search only identified the core motif DECA (Figure 2B). In addition, the DECA motif always appeared in the binding classes in which Prep was present in combination with any other factor/s. In contrast, the $DECA^{ext}$ domain only appeared in the Prep-$Pbx1^{com}$ class. In contrast, within the $Meis^{exc}$ class, both the HEXA and OCTA motifs were identified but not the DECA motif. HEXA and OCTA motifs also appeared in Meis-$Pbx1^{com}$, while the OCTA motif appeared in all categories in which Meis was present. In the $Pbx1^{exc}$ class, a previously undescribed and poorly defined consensus motif was detected. Given the poor definition of this motif, we excluded it from further analysis. The sites identified in the Pbx1 combinations with either Meis or Prep represent, respectively, the binding preferences of Meis or Prep alone, with the previously mentioned exception of $DECA^{ext}$ in Prep-$Pbx1^{com}$ peaks. The accessory motifs mostly consisted of sequences of low complexity, which nonetheless occurred preferentially in association with specific factors and may enhance binding or allow the binding of cofactors (Figure S3).

We then performed directed searches to determine the abundance of the identified core motifs in the peak sets for each factor and their combinations (Figures 2C and 2D). Overall, 62% of all peaks contained at least one core motif; the HEXA motif was present in about 20% of all peaks, the OCTA in 27%, and the DECA in 29% (Figure 2C). Core sequences were present in 82% of $Prep^{exc}$ peaks and 68% of $Meis^{exc}$ peaks (Figure 2D).

In $Prep^{exc}$ peaks, DECA or $DECA^{ext}$ motifs were predominant (48% and 37%, respectively; Figure 2D), while OCTA motifs were not represented over random expectation (4.7% versus 6.4%; adjp = 1). In contrast, DECA and especially $DECA^{ext}$ motifs were not overrepresented in $Meis^{exc}$ peaks (adjp = 1), while the HEXA (25%) and the OCTA motifs (42%) were predominant and represented over random expectation (adjp < 0.0001; Figure 2D). $Pbx1^{exc}$ peaks did not show a strong preference for any of the core motifs, but participation of Pbx1 in the binding increased the presence of the $DECA^{ext}$ motif in the Prep profile (74.8% versus 37%; adjp < 0.0001) and of the OCTA motif in the Meis profile (55% versus 42%; adjp < 0.0001) (Figure 2D).

Regarding Meis-$Prep^{com}$ peaks, all motifs show an abundance intermediate between that found for each factor independently, with the exception of the HEXA motif, which is more abundant in the common peaks than in the single-factor peaks. The triple-factor peaks have a profile similar to that of the Meis-Prep peaks (chi-square p value = 0.37), except for a clear increase in the OCTA sequence (36% versus 50%; p = 0.005), again indicating correlation between Pbx1 and the OCTA sequence, provided that Meis is also involved in the binding.

Electrophoretic mobility shift assays (EMSA) of peak sequences from the identified binding motifs showed that while Prep and Pbx can bind any of the core motifs identified, Meis can bind the HEXA and OCTA sequences but can only weakly bind to the DECA sequence (Figure 2E; Extended Results; Figure S4A).

Meis Binding to the OCTA Motif Corresponds to Pbx-Hox Binding Sites

The abundance of OCTA sites in Meis targets could correspond to a strong association between Meis and Pbx-Hox target sites. In contrast, the low representation of the OCTA motif in the $Prep^{exc}$ peaks would then indicate that Prep-Pbx1 mainly selects non-Hox binding sites. Within the OCTA motif, not all base combinations at the variable core of the OCTA motif stimulate

Cell PRESS
Open ACCESS

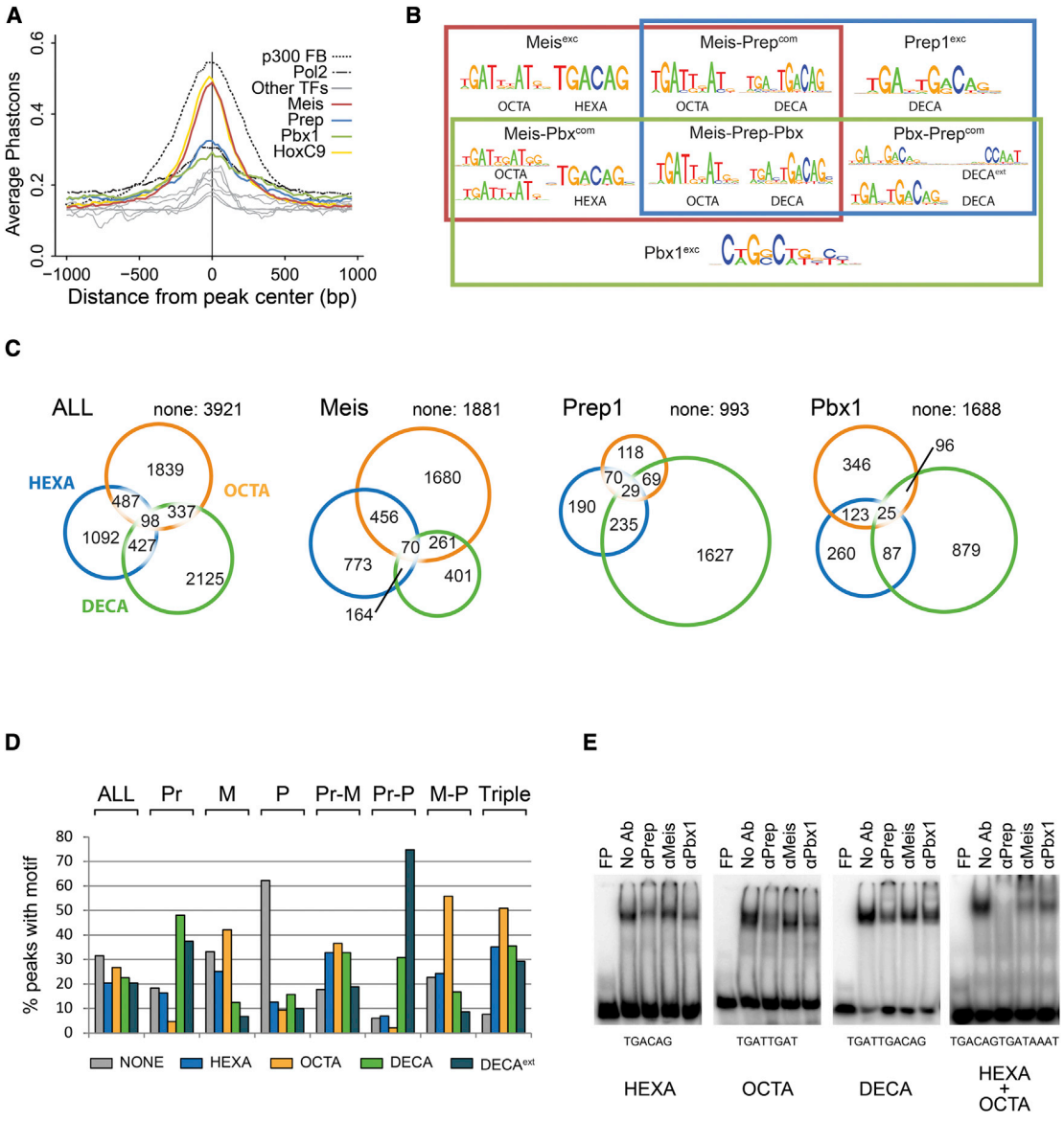

Figure 2. Meis and Prep Select Different DNA Target Sequences in the Genome

(A) DNA sequence conservation (vertebrate PhastCons) profile of Meis, Prep, and Pbx1 peaks. For comparison, the plot shows binding sites for HoxC9, HoxA2, p300 forebrain, and other transcription factors (Mahony et al., 2011; Schmidt et al., 2010).

(B) Core sequence motifs identified in exclusive, double, and triple peaks.

(C) co-occurrence of core sequence motifs in each binding class. Boxplots show Pbx1 enrichment factors for peaks cobound by dimers and trimers (Pbx1-Prep, Pbx1-Meis, and Pbx1-Prep-Meis).

(D) Abundance of core sequence motifs in each factor binding class.

(E) EMSA testing of the in vitro binding ability of the TALE factors. FP, free probe.

See also Figures S3 and S4.

Pbx-Hox dimer binding (Berger et al., 2008; Chan et al., 1994; Chang et al., 1996; Lu and Kamps, 1997; Mann and Chan, 1996; Noyes et al., 2008). We therefore examined the enrichment of each dinucleotide combination at the OCTA variable core (bases 5 and 6) in the peak sets for each factor and their combinations (Figure 3A). Eight two-base combinations have been reported to promote Pbx-Hox binding, while the remaining eight have not (Slattery et al., 2011; Tümpel et al., 2007). Of the eight that do, five were strongly and significantly overrepresented in all but one of the peak sets (Figure 3A). In contrast, only one of the combinations (GA in the variable core) not previously found to bind Pbx-Hox was overrepresented in various peak sets (Figure 3A). EMSA analyses of sequences from OCTA-containing peaks confirmed various Hox protein binding to the previously

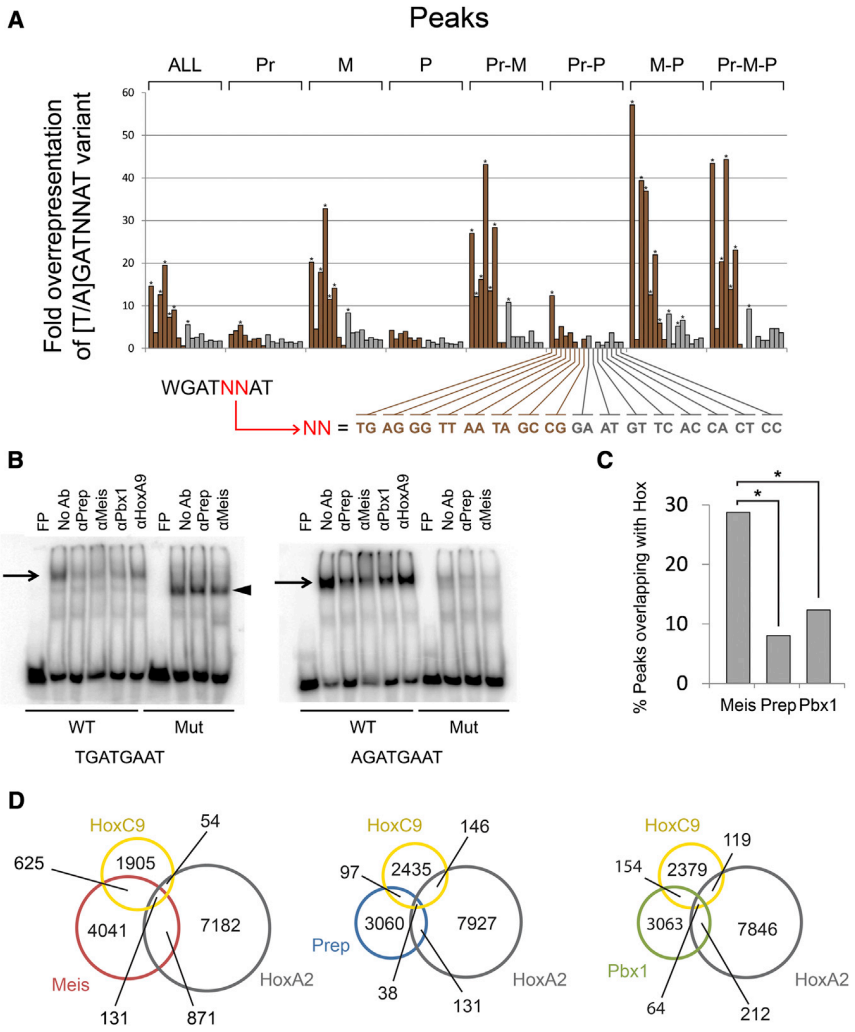

Figure 3. Hox Binding Motifs and Sites Strongly Correlate with Meis Peaks and Not with Prep Peaks

(A) Overrepresentation of OCTA variants for each base combination at positions 5 and 6 of the sequence in the peak sets for each factor and their combinations. The two-base combinations that have been previously described to bind Pbx-Hox are shown in brown, while those that have not are shown in gray. Cases in which the screened sequences were found more than five times overrepresented and deviated from random expectation with $p < 0.001$ are indicated with an asterisk.

(B) EMSA testing of in vitro binding ability of the TALE factors and an example Hox protein to the candidate new Hox binding sequence. Mutant probes contain WGATCCAT instead of WGATGAAT. FP, free probe. Arrows indicate the migration of complexes formed between nuclear proteins and DNA. Arrowhead indicates a nonspecific complex.

(C) Percentage overlap of $Prep^{total}$, $Meis^{total}$, and $Pbx1^{total}$ peaks with either HoxC9 or HoxA2. Asterisks show $p < 0.0001$.

(D) Overlap of Prep, Meis, and Pbx1 peaks with HoxC9 and HoxA2 peaks.

known sequences (Figure S4B) and weak binding of Hoxa9 to the OCTA motif with GA in the variable core (Figure 3B).

All peak sets that showed enrichment were Meis-bound, while peak sets in which Meis was not involved showed marginal or no enrichment for Hox-bound base combinations. An exception was the enrichment for the TGATTGAT sequence in the Prep-$Pbx1^{com}$ peaks; however, this sequence might be a variant of the DECA sequence. Interestingly, the degree of enrichment in Hox-type sequences increased with cobinding of Pbx1 or Prep with Meis, being maximal in peaks bound by all three factors. These data support the idea that the OCTA sequence represents Pbx-Hox targets and that Meis is the factor most associated with Hox binding sites in the genome. In contrast, Prep does not normally select Hox binding sequences unless the peak is also Meis-bound.

These data suggest that Meis peaks containing an OCTA motif could represent Hox targets. In line with this suggestion, ChIP-seq peaks identified for HoxA2 in E11.5 second branchial arch (Donaldson et al., 2012) and for HoxC9 in E11.5 spinal cord (Jung et al., 2010), despite representing the targets of just 2 of the 39 Hox proteins in embryonic tissues different from those analyzed here, show a strong overlap with $Meis^{total}$ (28% versus 0.42% expected by chance; $p < 0.0001$), a much lower overlap with $Prep^{total}$, and moderate overlap with $Pbx1^{total}$ (Figures 3C and 3D). Common Meis-Pbx1 and Meis-Prep peaks show an increased chance to overlap with Hox (51% of Meis-Pbx1 and 44% of Meis-$Prep^{com}$ peaks). Moreover, 68% of Prep-Hox^{com} peaks and 79% of Pbx1-Hox^{com} peaks are also Meis peaks (data not shown), again indicating that Meis binding shows the strongest association with Hox binding in these experiments. It is noteworthy, however, that while our analysis is comprehensive for Meis proteins, it is not so for Pbx proteins, so that the lower overlap of Pbx1 with Hox binding sites may be due to the participation of other Pbx family members instead of Pbx1.

Prep1 and *Meis1* Coordinately Regulate a Subset of Their Target Genes

To investigate functional interactions between Prep1 and Meis1, we compared changes to the transcriptome caused by elimination of either *Meis1* or *Prep1* in mouse embryos. To this end, we performed total RNA sequencing (RNA-seq) in *Meis1*-deficient and $Prep1^{i/i}$ E11.5 embryos. RNA-seq identified 855 upregulated and 631 downregulated transcripts in $Prep1^{i/i}$ embryos and 210 upregulated and 198 downregulated transcripts in *Meis1*-deficient embryos (Figure 4A; Table S4). The affected transcripts in *Meis1* mutants probably represent only a fraction of all *Meis*-regulated genes, since the expression patterns of *Meis1* and *Meis2* overlap considerably in the embryo. To estimate the

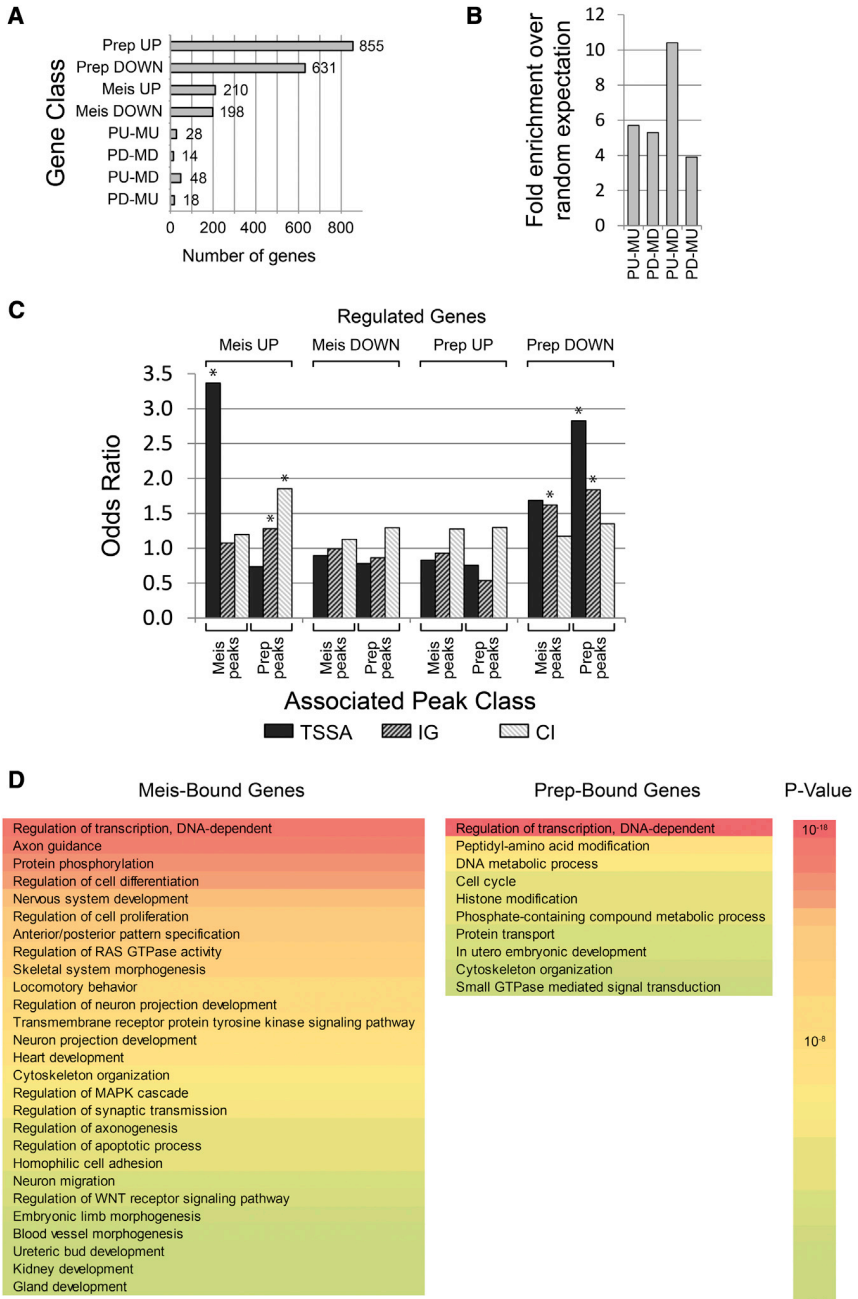

Figure 4. Meis and Prep Target Gene Coregulation and Functional Annotation

(A) Numbers of genes up- or downregulated in *Prep1*$^{i/i}$ mutant and *Meis1*-deficient embryos showing coregulation. Prep UP (PU) are genes upregulated and Prep DOWN (PD) are genes downregulated in *Prep1*$^{i/i}$ embryos; Meis UP (MU) are genes upregulated and Meis DOWN (MD) are genes downregulated in *Meis1* loss-of-function embryos.

(B) Extent of coregulation. The most overrepresented set of genes is that composed of genes upregulated in *Prep1*$^{i/i}$ and downregulated in *Meis1*-deficient mutants. Asterisks show chi-square $p < 0.0001$.

(C) Association of regulated genes with peak classes and their genomic location. Graph shows fold enrichment in peak density over whole-genome average (i.e., odds ratio). Asterisk shows $p < 0.001$.

(D) Selected Gene Ontology terms associated with bound genes.

See also Tables S4 and S9.

extent of gene coregulation by *Meis1* and *Prep1*, we determined the frequency of coregulated genes and the nature of the coregulation. All classes of coregulated genes occurred at frequencies 4- to 10-fold higher than expected under the null hypothesis of independence of gene subsets, indicating coordinated actions of these transcription factors in the regulation of specific sets of genes (Figure 4B). We found 108 genes coregulated by the two factors, corresponding to 26% of *Meis1*-regulated genes and 7% of *Prep1*-regulated genes. Interestingly, the most enriched class was genes downregulated in *Meis1*-deficient and upregulated in the *Prep1*-deficient embryos. This class included 48 genes, or 24% of the genes downregulated in *Meis1*-deficient embryos, and was 10-fold higher than random expectation. These results indicate considerable functional interactions between *Meis1* and *Prep1* in the regulation of gene expression, including cooperative and, more frequently, antagonistic actions.

To identify putative direct transcriptional targets of TALE factors, we examined the correlation between ChIP-seq peaks and the set of genes regulated by *Prep1* and *Meis1* (Figure 4C). Among the genes upregulated in *Meis1*-deficient embryos, we found significant enrichment in TSSA Meis peaks but not in those in other gene regions, suggesting a correlation between TSSA Meis binding and negative regulation of transcription (Figure 4C). Surprisingly, within the genes upregulated in *Meis1*-deficient embryos, there was a significant enrichment in Prep IG and CI peaks and, conversely, we found some enrichment of IG Meis peaks among the genes upregulated in *Prep1*$^{i/i}$ embryos, again suggesting a functional interaction between Meis and Prep in gene regulation (Figure 4C). In contrast, we found no enrichment for any factor peaks among genes downregulated in *Meis1*-deficient embryos or upregulated in *Prep1*$^{i/i}$ embryos. Finally, TSSA and IG Prep peaks were overrepresented among *Prep1*$^{i/i}$-downregulated genes and underrepresented among *Prep1*$^{i/i}$-upregulated genes, indicating a transcriptional activator function for these Prep sites.

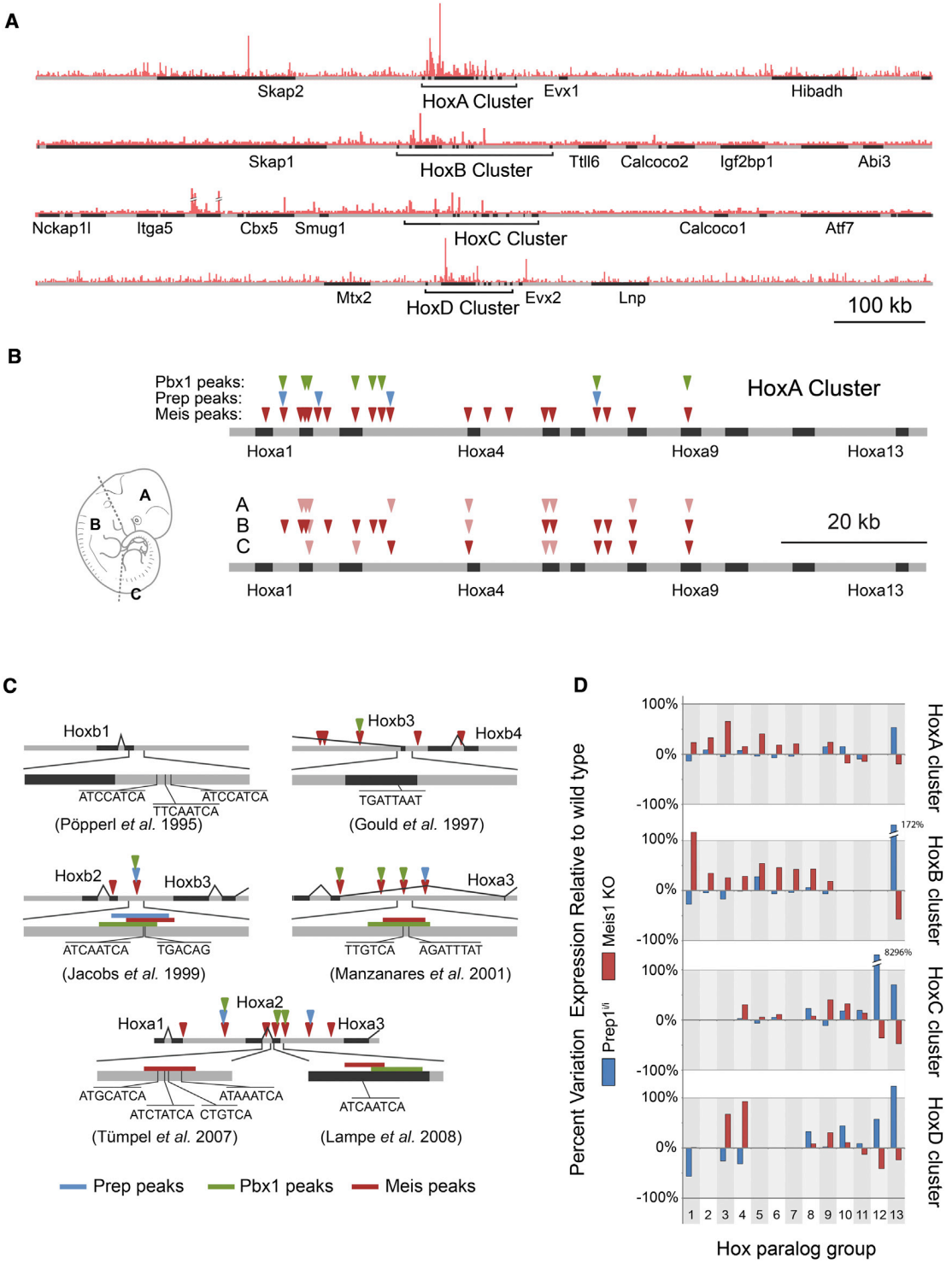

Figure 5. TALE-Factor Binding Subdivides the Hox Clusters in Two Regulatory Domains

(A) Meis ChIP-seq read profiles (in red) in the Hox cluster environment. Black bars show coding regions.

(B) Top: The HoxA cluster is shown, with Meis, Prep, and Pbx1 ChIP-seq peaks represented in different colors as indicated. Bottom: A similar representation of the HoxA cluster shows Meis ChIP-PCR signals obtained from different embryo portions, as indicated on the scheme on the left. An absent triangle indicates no binding detected, a light-colored triangle indicates positive but not predominant binding compared to other embryo regions, and a dark-colored triangle indicates predominant binding.

(legend continued on next page)

These results indicate an association of Meis TSSA binding with the repression of Meis1 target genes and an association of Prep TSSA binding with the activation of Prep1 target genes. The reported context-dependent repressive activity of Meis/Hth proteins is in agreement with these findings (Elkouby et al., 2012; Huang et al., 2005). In addition, the association of Meis peaks with Prep1-regulated genes and of Prep peaks with Meis1-regulated genes suggests coordinated actions of Meis and Prep on some of their targets.

We next profiled the Gene Ontology annotations of potential Meis and Prep targets (Figure 4D), considering the set of genes with Meis or Prep peaks in their promoters or transcriptional units as potential direct targets. For both factors, target genes encoding transcriptional regulators are strongly overrepresented. Meis-bound genes are strongly enriched for functions involved in several aspects of development, such as AP pattern specification, heart development, nervous system development, and blood vessel morphogenesis. Meis also binds to genes involved in cell processes that potentially mediate its leukemogenic properties, such as cell differentiation and proliferation. In contrast, developmentally associated genes are only weakly overrepresented among Prep targets, which are instead annotated to basal cell functions like DNA and histone modification, protein transport, and signal transduction. These data correlate with the fact that Meis proteins are expressed in a developmentally restricted manner, while Prep is a ubiquitously expressed protein that regulates essential cell functions (Fernandez-Diaz et al., 2010; Iotti et al., 2011).

The TALE-Factor Binding Landscape Subdivides Hox Clusters into Two Regions with Differential Transcriptional Responses

Examination of the binding sites of Prep, Meis, and Pbx1 in the *Hox* clusters reveals abundant interaction sites (Figures 5A and 5B; Figure S5A), suggestive of extensive crosstalk and autoregulation within the Hox/TALE network. In the Hox clusters, Meis peaks are the most abundant, occurring mostly in the HoxA and least in the HoxC cluster. Pbx1 binding sites are less abundant, and of the three factors, Prep binding sites are the least abundant. In all cases but one, Pbx1 and Prep sites coincide with Meis peaks. Interestingly, all peaks concentrate in paralog groups 1–9, with no peak present in paralogs 10–13 in any Hox cluster, indicating a subdivision of the Hox clusters into TALE-interactive and TALE-noninteractive regions (Figure 5A; Figure S5A). ChIP-PCR analysis of Meis binding sites in the HoxA cluster indicated that the binding profile was variable in different regions of the embryo and correlated with the expression status of the HoxA cluster in these regions (Figure 5B).

Previous studies had identified six TALE protein binding regions in the Hox clusters that were mostly involved in cooperation with Hox proteins in auto- and cross-regulatory interactions (Gould et al., 1997; Jacobs et al., 1999; Lampe et al., 2008; Manzanares et al., 2001; Pöpperl et al., 1995; Tümpel et al., 2007). Although those interactions were described at a different developmental stage and only affect a subset of the tissues analyzed here, we found interactions at the precise sites previously described in four out of the six regions (Figure 5C).

We then compared *Hox* gene expression in *Meis1*-deficient and *Prep1*$^{i/i}$ mutant E11.5 embryos with that in wild-type littermates by RNA-seq. In *Meis1*-deficient embryos, 22 of the 27 Hox genes from paralog groups 1–9 increased their expression, while expression of the remaining five decreased or was maintained (Figure 5D). In contrast, seven of the eight Hox genes from paralog groups 11–13 decreased their expression and expression of the other was maintained (Figure 5D). Paralog 10 genes showed variable behavior, with *Hoxa10* expression being reduced, *Hoxc10* increased, and *Hoxd10* maintained. Although many of the expression changes are moderate and would not be significant in isolation, the correlation of the expression changes with the position of the genes in the cluster significantly ($p < 0.05$) diverges from the transcriptomic average for all clusters except the 3′ part of cluster C (see Experimental Procedures). These results suggest that Meis function moderates the expression of paralog groups 1–9 while enhancing expression of paralog groups 11–13. Paralog group *Hox10* seems to be placed in a frontier region, with the influence of Meis activity depending on the specific Hox cluster.

While expression of many *Hox* cluster genes does not change in *Prep1*$^{i/i}$ E11.5 mutants, the 5′ genes show changes opposite to those observed in *Meis1*-deficient embryos (Figure 5D). An opposite regulation to that observed in *Meis1*-deficient embryos was also observed in all cases for paralog group 1, extending to paralogs 1–4 in the case of the HoxD cluster (Figure 5D).

These results show interactions between Prep and Meis in the global modulation of *Hox* gene expression. From the six previously described regulatory interactions, four are detected in our study, suggesting that the observed regulatory effects involve direct interactions linked to the described binding sites. This view is further supported by the correlation between the Meis binding profile and HoxA cluster expression and by the coincidence between the Meis/Prep/Pbx binding profiles and the transcriptional response landscape in the Hox clusters.

Additional genetic interaction studies showed no interaction between *Prep1* and *Meis1* loss-of-function alleles (Extended Results; Table S5; Figure S6), suggesting that the critically affected functions in these mutants are independent. This is consistent with the predominantly independent DNA-binding activities observed for each factor. In contrast, the strong genetic interaction between *Meis1* and *Pbx1* (Table S8) correlates with the predominance of Pbx-Hox binding sites within the Meis ChIP peaks.

(C) Representation of previously described TALE factor binding sites in the Hox clusters. Black bars indicate exons, and gray bars indicate introns or intergenic regions. For each case, a representation of the Hox cluster subregion with the ChIP-seq peaks is shown above and a zoom showing the specific sequences previously described and their position with respect to the peaks here described (color stripes) is shown below.
(D) Differential transcriptional response of the 3′ and 5′ halves of the Hox clusters to *Meis1* and *Prep1* deficiency. Graphs show the change in transcript levels (percentage) between controls and mutant *Meis1 ko* and *Prep1*$^{i/i}$ E11.5 embryos.
See also Figure S5.

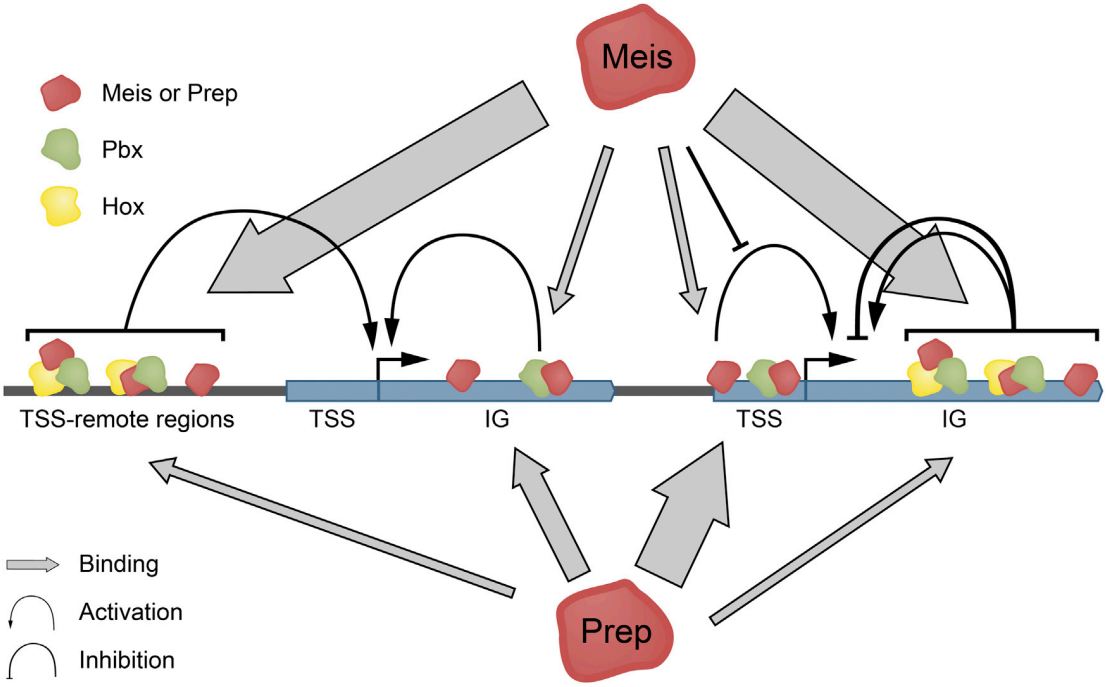

Figure 6. A Representation of PREP and MEIS Homeodomain Factor Activity in Regulating Gene Expression
Meis and Prep proteins cooperate to regulate gene expression. Prep binds mostly to promoters in conjunction with Pbx. Meis binds mainly to non-TSSA regions in cooperation with Hox proteins, often without contacting DNA. They often show opposing activities.

DISCUSSION

In this study, we have analyzed genomic binding sites for Pbx1, Meis1/2, and Prep1/2. While Pbx3/4 and Meis3 were not studied, the similarities of these proteins to the members of the family studied here suggest that their binding repertoires will be comparable. In addition, the expression patterns of the studied proteins cover the majority of tissues in which their counterparts are expressed. This analysis is thus a near-comprehensive picture of the general binding abilities of TALE factors in the mammalian embryo. The results highlight the specialization of Prep and Meis in binding largely independent genomic elements through selection of different DNA sequences. The contrast with the previously reported in vitro activities suggest that Meis and Prep gain additional binding specificity in vivo through interaction with cofactors or chromatin landmarks. While Prep1 interacts preferentially with promoters and nearby regions, Meis shows preference for intergenic and intragenic regions away from TSSA regions. Prep could thus directly control promoter activity, while a substantial part of the Meis sites coincides with enhancers. A number of Meis binding sites remain functionally undefined and, despite their high evolutionary conservation, do not correlate with the described marks of known constitutive chromatin factors such as CTCF and others (not shown). These results suggest that a proportion of Meis sites are evolutionarily conserved protein-DNA interaction regions whose function remains to be explored. In addition, Prep mostly participates in dimer formation with Pbx, while Meis is predominant in Pbx-Hox interactions on targets.

These findings suggest that during evolution, Meis and Prep proteins acquired specialized functions enabling them to interact with specific subsets of regulatory regions and target genes (Figure 6). Functions of the MEINOX and PBC proteins extend beyond the regulation of Hox protein activity. The complete loss of function of *exd* or *hth* in *Drosophila* results in the early failure of embryonic development due to defects in cell division at stages when Hox genes are not required (Rauskolb et al., 1993; Salvany et al., 2009). The ontology analysis of the targets bound by Meis and Prep, together with the functions previously described for *Meis1* and *Prep1* in mice, suggest that Prep has specialized in the basic cellular functions, which would be Hox independent, and Meis has specialized in patterning functions more related to Hox activity. Whether the sum of the functions of Meis and Prep corresponds to those exerted by the single MEINOX in flies (Hth) or, alternatively, involves the acquisition by either factor of new functions related to vertebrate evolution remains to be explored.

Regarding the interaction with Pbx, we found that Prep binds DNA preferentially as a dimer with any of the Pbx proteins. The strong overlap between thymic Pbx2 sites and embryonic Pbx1 sites also shows that the two proteins can substitute each other, and hence provides a molecular basis for the concept of Pbx redundancy (Selleri et al., 2004).

The data presented point to Meis factors as major in vivo partners of Hox proteins, cooperating with them in target selection with little contribution from Prep. Identifying the genomic binding sites of the 39 mammalian Hox proteins is a major challenge that is still far from being achieved. Given the general requirement of

Hox proteins for cooperation with TALE factors, and the fact that TALE proteins can interact promiscuously with Hox proteins, the putative binding sites presented in this study likely represent the most comprehensive set of in vivo Hox genomic targets yet identified.

Despite the extensive divergence in their genomic binding patterns, *Meis1* and *Prep1* do show coregulation of some downstream genes, with opposing effects predominating. Some of these antagonistic interactions might underlie the opposing roles of *Meis1* and *Prep1* in tumor formation, where *Meis1* function promotes tumor formation while *Prep1* behaves as a tumor suppressor (Iotti et al., 2011; Longobardi et al., 2010; Thorsteinsdottir et al., 2001).

A striking case of coordinated regulation was observed in *Hox* gene regulation, where the TALE protein binding profile and transcriptional regulatory activity subdivide *Hox* clusters into two regions: the paralog 1–9 region and the paralog 10–13 region. These results highlight the important role of TALE factors in globally regulating *Hox* gene expression, in addition to serving as cofactors of Hox proteins. The modulation of *Hox* cluster transcriptional activity may be the result of global conformational changes promoted by TALE factors, since paralog groups 10–13 are not directly bound by TALE factors, yet they are sensitive to their levels.

Our work thus identifies TALE and TALE-Hox binding sites, target genes, and in vivo specificities that increase our understanding of the molecular pathways controlled by this regulatory network in development and disease.

EXPERIMENTAL PROCEDURES

ChIP-Seq and ChIP-Re-ChIP

ChIPs were performed using standard methods on E11.5 mice embryo trunks. We used anti-Prep1/2 antibody, anti-Pbx1 antibody, and a mix of anti-Meis antibodies. The same antibodies were used for ChIP-re-ChIP.

ChIP-Chip

ChIPs of mouse thymocyte lysates were performed as described above with anti-Prep1/2 antibody and anti-Pbx2 antibody on thymuses from 6- to 8-week-old C57B6 mice. The resulting DNA was hybridized to a Nimblegen mouse RefSeq promoter array.

ChIP-Seq Data Analysis

ChIP DNA was sequenced using an Illumina GAII analyzer. Single-end 36 bp reads were mapped with BWA software against mm9 version of the mouse genome. The alignments were then used for peak calling, de novo motif discovery, and motif identification and validation with MotIV.

ChIP-PCR

For comparison of AP occupancy of Meis peaks, E11.5 embryos were dissected as shown in the diagram in Figure 5B and ChIP was carried out as described above. DNA was then subjected to PCR for 30 or 35 cycles, depending on primers, to avoid saturation.

RNA-Seq Data Analysis

Total RNA was purified from whole E11.5 mouse embryos, and a library was prepared and sequenced on the Illumina platform according to the manufacturer's instructions. Approximately 7M reads per sample were aligned to mouse mm9, and transcript expression was estimated with mouse Ensembl 63 genebuild as a reference.

Exploratory Data Analysis

Peak overlapping, correlation with RNA-seq data, and conservation data aggregation were performed on the Galaxy platform.

Individual instances of the core motifs within all peaks were searched with a local install of the FIMO (Find Individual Motif Occurrences) program from the MEME suite.

Peak profiling was performed using custom Python scripts and the CEAS (Cis-regulatory Element Annotation System) tool.

EMSA

Nuclear extracts were isolated from cells prepared from E11.5 mouse embryonic body. EMSA reactions were performed following the standard protocol.

Gene Ontology Analysis

GO term overrepresentation was assessed with GOrilla, comparing the lists of genes with Meis or Prep binding sites against the list of all nuclear genes in Ensembl v63, with a p value cutoff of 10^{-5}. We considered those genes that have a Meis or Prep peak in their promoter (−500 to +100) or within the transcriptional unit.

Animal Procedures

All animal procedures have been reviewed and approved by the CNIC Animal Experimentation Ethics Committee, according to the National and European regulations.

For further details, see Extended Experimental Procedures.

ACCESSION NUMBERS

The GEO accession number for the ChIP-seq results reported in this paper is GSE39609.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Results, Extended Experimental Procedures, six figures, and nine tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2013.03.029.

LICENSING INFORMATION

ACKNOWLEDGMENTS

D.P., F.B., and M.T. are grateful to D. Pasini, G. Natoli, A. Roselló, and members of the Blasi and Torres labs for their helpful discussions. We thank L. Modica for Prep1$^{i/i}$ RNA preparation, J. Wilde and M. Linser for assistance in ChIP-seq library preparation, A. Dopazo for sequencing, V. Amstislavskiy for sequencing data processing, L. Selleri for providing the Pbx1 KO mice, and Jeremy Dasen for antibodies. The collaboration among F.B., M.T., and D.P. was made possible by COST Action BM0805. F.B. was supported by grants from the AIRC (Associazione Italiana Ricerche sul Cancro, 8929), Ministero dell'Università e Ricerca (MERIT; EU FP7 Prepobedia, MIUR-FIRB RBNE08NKH7), the Cariplo Foundation, and the Italian Ministry of Health. M.T. was supported by grants RD06/0010/0008 and BFU2009-08331/BMC from the Spanish Ministerio de Economia y Competitividad (MINECO). D.M.S.M. was supported by a fellowship from the Consejería de Educación de la Comunidad de Madrid and the European Social Fund. D.P. and V.T. were supported by grant 02.740.11.0872 from the Russian Ministry of Education and Science. D.P. was supported by grant 12-04-01659-a from the Russian Foundation for Basic Research. A.B. was supported by Associazione Italiana Ricerca sul Cancro start-up grant #4780. The CNIC is supported by the MINECO and the pro-CNIC Foundation. IFOM is supported by the FIRC

(Fondazione Italiana Ricerche sul Cancro). S. Bartlett (CNIC) provided English editing.

Received: July 18, 2012
Revised: February 19, 2013
Accepted: March 20, 2013
Published: April 18, 2013

REFERENCES

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Azcoitia, V., Aracil, M., Martínez-A, C., and Torres, M. (2005). The homeodomain protein Meis1 is essential for definitive hematopoiesis and vascular patterning in the mouse embryo. Dev. Biol. *280*, 307–320.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Peña-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., et al. (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell *133*, 1266–1276.

Berthelsen, J., Zappavigna, V., Mavilio, F., and Blasi, F. (1998). Prep1, a novel functional partner of Pbx proteins. EMBO J. *17*, 1423–1433.

Bisaillon, R., Wilhelm, B.T., Krosl, J., and Sauvageau, G. (2011). C-terminal domain of MEIS1 converts PKNOX1 (PREP1) into a HOXA9-collaborating oncoprotein. Blood *118*, 4682–4689.

Blow, M.J., McCulley, D.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2010). ChIP-Seq identification of weakly conserved heart enhancers. Nat. Genet. *42*, 806–810.

Cecconi, F., Proetzel, G., Alvarez-Bolado, G., Jay, D., and Gruss, P. (1997). Expression of Meis2, a Knotted-related murine homeobox gene, indicates a role in the differentiation of the forebrain and the somitic mesoderm. Dev. Dyn. *210*, 184–190.

Chan, S.K., Jaffe, L., Capovilla, M., Botas, J., and Mann, R.S. (1994). The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. Cell *78*, 603–615.

Chang, C.P., Brocchieri, L., Shen, W.F., Largman, C., and Cleary, M.L. (1996). Pbx modulation of Hox homeodomain amino-terminal arms establishes different DNA-binding specificities across the Hox locus. Mol. Cell. Biol. *16*, 1734–1745.

Chang, C.P., Jacobs, Y., Nakamura, T., Jenkins, N.A., Copeland, N.G., and Cleary, M.L. (1997). Meis proteins are major in vivo DNA binding partners for wild-type but not chimeric Pbx proteins. Mol. Cell. Biol. *17*, 5679–5687.

Di Rosa, P., Villaescusa, J.C., Longobardi, E., Iotti, G., Ferretti, E., Diaz, V.M., Miccio, A., Ferrari, G., and Blasi, F. (2007). The homeodomain transcription factor Prep1 (pKnox1) is required for hematopoietic stem and progenitor cell activity. Dev. Biol. *311*, 324–334.

DiMartino, J.F., Selleri, L., Traver, D., Firpo, M.T., Rhee, J., Warnke, R., O'Gorman, S., Weissman, I.L., and Cleary, M.L. (2001). The Hox cofactor and proto-oncogene Pbx1 is required for maintenance of definitive hematopoiesis in the fetal liver. Blood *98*, 618–626.

Donaldson, I.J., Amin, S., Hensman, J.J., Kutejova, E., Rattray, M., Lawrence, N., Hayes, A., Ward, C.M., and Bobola, N. (2012). Genome-wide occupancy links Hoxa2 to Wnt-β-catenin signaling in mouse embryonic development. Nucleic Acids Res. *40*, 3990–4001.

Elkouby, Y.M., Polevoy, H., Gutkovich, Y.E., Michaelov, A., and Frank, D. (2012). A hindbrain-repressive Wnt3a/Meis3/Tsh1 circuit promotes neuronal differentiation and coordinates tissue maturation. Development *139*, 1487–1497.

Fernandez-Diaz, L.C., Laurent, A., Girasoli, S., Turco, M., Longobardi, E., Iotti, G., Jenkins, N.A., Fiorenza, M.T., Copeland, N.G., and Blasi, F. (2010). The absence of Prep1 causes p53-dependent apoptosis of mouse pluripotent epiblast cells. Development *137*, 3393–3403.

Ferretti, E., Schulz, H., Talarico, D., Blasi, F., and Berthelsen, J. (1999). The PBX-regulating protein PREP1 is present in different PBX-complexed forms in mouse. Mech. Dev. *83*, 53–64.

Ferretti, E., Marshall, H., Pöpperl, H., Maconochie, M., Krumlauf, R., and Blasi, F. (2000). Segmental expression of Hoxb2 in r4 requires two separate sites that integrate cooperative interactions between Prep1, Pbx and Hox proteins. Development *127*, 155–166.

Ferretti, E., Cambronero, F., Tümpel, S., Longobardi, E., Wiedemann, L.M., Blasi, F., and Krumlauf, R. (2005). Hoxb1 enhancer and control of rhombomere 4 expression: complex interplay between PREP1-PBX1-HOXB1 binding sites. Mol. Cell. Biol. *25*, 8541–8552.

Ferretti, E., Villaescusa, J.C., Di Rosa, P., Fernandez-Diaz, L.C., Longobardi, E., Mazzieri, R., Miccio, A., Micali, N., Selleri, L., Ferrari, G., and Blasi, F. (2006). Hypomorphic mutation of the TALE gene Prep1 (pKnox1) causes a major reduction of Pbx and Meis proteins and a pleiotropic embryonic phenotype. Mol. Cell. Biol. *26*, 5650–5662.

Gould, A., Morrison, A., Sproat, G., White, R.A., and Krumlauf, R. (1997). Positive cross-regulation and enhancer sharing: two mechanisms for specifying overlapping Hox expression patterns. Genes Dev. *11*, 900–913.

Hisa, T., Spence, S.E., Rachel, R.A., Fujita, M., Nakamura, T., Ward, J.M., Devor-Henneman, D.E., Saiki, Y., Kutsuna, H., Tessarollo, L., et al. (2004). Hematopoietic, angiogenic and eye defects in Meis1 mutant animals. EMBO J. *23*, 450–459.

Huang, H., Rastegar, M., Bodner, C., Goh, S.L., Rambaldi, I., and Featherstone, M. (2005). MEIS C termini harbor transcriptional activation domains that respond to cell signaling. J. Biol. Chem. *280*, 10119–10127.

Iotti, G., Longobardi, E., Masella, S., Dardaei, L., De Santis, F., Micali, N., and Blasi, F. (2011). Homeodomain transcription factor and tumor suppressor Prep1 is required to maintain genomic stability. Proc. Natl. Acad. Sci. USA *108*, E314–E322.

Jacobs, Y., Schnabel, C.A., and Cleary, M.L. (1999). Trimeric association of Hox and TALE homeodomain proteins mediates Hoxb2 hindbrain enhancer activity. Mol. Cell. Biol. *19*, 5134–5142.

Jung, H., Lacombe, J., Mazzoni, E.O., Liem, K.F., Jr., Grinstein, J., Mahony, S., Mukhopadhyay, D., Gifford, D.K., Young, R.A., Anderson, K.V., et al. (2010). Global control of motor neuron topography mediated by the repressive actions of a single hox gene. Neuron *67*, 781–796.

Knoepfler, P.S., Calvo, K.R., Chen, H., Antonarakis, S.E., and Kamps, M.P. (1997). Meis1 and pKnox1 bind DNA cooperatively with Pbx1 utilizing an interaction surface disrupted in oncoprotein E2a-Pbx1. Proc. Natl. Acad. Sci. USA *94*, 14553–14558.

Lampe, X., Samad, O.A., Guiguen, A., Matis, C., Remacle, S., Picard, J.J., Rijli, F.M., and Rezsohazy, R. (2008). An ultraconserved Hox-Pbx responsive element resides in the coding sequence of Hoxa2 and is active in rhombomere 4. Nucleic Acids Res. *36*, 3214–3225.

LaRonde-LeBlanc, N.A., and Wolberger, C. (2003). Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. Genes Dev. *17*, 2060–2072.

LeBrun, D.P., and Cleary, M.L. (1994). Fusion with E2A alters the transcriptional properties of the homeodomain protein PBX1 in t(1;19) leukemias. Oncogene *9*, 1641–1647.

Longobardi, E., Iotti, G., Di Rosa, P., Mejetta, S., Bianchi, F., Fernandez-Diaz, L.C., Micali, N., Nuciforo, P., Lenti, E., Ponzoni, M., et al. (2010). Prep1 (pKnox1)-deficiency leads to spontaneous tumor development in mice and accelerates EmuMyc lymphomagenesis: a tumor suppressor role for Prep1. Mol. Oncol. *4*, 126–134.

Lu, Q., and Kamps, M.P. (1997). Heterodimerization of Hox proteins with Pbx1 and oncoprotein E2a-Pbx1 generates unique DNA-binding specifities at nucleotides predicted to contact the N-terminal arm of the Hox homeodomain—demonstration of Hox-dependent targeting of E2a-Pbx1 in vivo. Oncogene *14*, 75–83.

Mahony, S., Mazzoni, E.O., McCuine, S., Young, R.A., Wichterle, H., and Gifford, D.K. (2011). Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis. Genome Biol. *12*, R2.

Mann, R.S., and Chan, S.K. (1996). Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. Trends Genet. *12*, 258–262.

Mann, R.S., and Affolter, M. (1998). Hox proteins meet more partners. Curr. Opin. Genet. Dev. *8*, 423–429.

Manzanares, M., Bel-Vialar, S., Ariza-McNaughton, L., Ferretti, E., Marshall, H., Maconochie, M.M., Blasi, F., and Krumlauf, R. (2001). Independent regulation of initiation and maintenance phases of Hoxa3 expression in the vertebrate hindbrain involve auto- and cross-regulatory mechanisms. Development *128*, 3595–3607.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Moens, C.B., and Selleri, L. (2006). Hox cofactors in vertebrate development. Dev. Biol. *291*, 193–206.

Moskow, J.J., Bullrich, F., Huebner, K., Daar, I.O., and Buchberg, A.M. (1995). Meis1, a PBX1-related homeobox gene involved in myeloid leukemia in BXH-2 mice. Mol. Cell. Biol. *15*, 5434–5443.

Nakamura, T., Jenkins, N.A., and Copeland, N.G. (1996). Identification of a new family of Pbx-related homeobox genes. Oncogene *13*, 2235–2242.

Noyes, M.B., Christensen, R.G., Wakabayashi, A., Stormo, G.D., Brodsky, M.H., and Wolfe, S.A. (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. Cell *133*, 1277–1289.

Oulad-Abdelghani, M., Chazaud, C., Bouillet, P., Sapin, V., Chambon, P., and Dollé, P. (1997). Meis2, a novel mouse Pbx-related homeobox gene induced by retinoic acid during differentiation of P19 embryonal carcinoma cells. Dev. Dyn. *210*, 173–183.

Piper, D.E., Batchelor, A.H., Chang, C.P., Cleary, M.L., and Wolberger, C. (1999). Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. Cell *96*, 587–597.

Pöpperl, H., Bienz, M., Studer, M., Chan, S.K., Aparicio, S., Brenner, S., Mann, R.S., and Krumlauf, R. (1995). Segmental expression of Hoxb-1 is controlled by a highly conserved autoregulatory loop dependent upon exd/pbx. Cell *81*, 1031–1042.

Rauskolb, C., Peifer, M., and Wieschaus, E. (1993). extradenticle, a regulator of homeotic gene activity, is a homolog of the homeobox-containing human proto-oncogene pbx1. Cell *74*, 1101–1112.

Rieckhof, G.E., Casares, F., Ryoo, H.D., Abu-Shaar, M., and Mann, R.S. (1997). Nuclear translocation of extradenticle requires homothorax, which encodes an extradenticle-related homeodomain protein. Cell *91*, 171–183.

Ryoo, H.D., Marty, T., Casares, F., Affolter, M., and Mann, R.S. (1999). Regulation of Hox target genes by a DNA bound Homothorax/Hox/Extradenticle complex. Development *126*, 5137–5148.

Salvany, L., Aldaz, S., Corsetti, E., and Azpiazu, N. (2009). A new role for hth in the early pre-blastodermic divisions in Drosophila. Cell Cycle *8*, 2748–2755.

Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S., et al. (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science *328*, 1036–1040.

Selleri, L., Depew, M.J., Jacobs, Y., Chanda, S.K., Tsang, K.Y., Cheah, K.S., Rubenstein, J.L., O'Gorman, S., and Cleary, M.L. (2001). Requirement for Pbx1 in skeletal patterning and programming chondrocyte proliferation and differentiation. Development *128*, 3543–3557.

Selleri, L., DiMartino, J., van Deursen, J., Brendolan, A., Sanyal, M., Boon, E., Capellini, T., Smith, K.S., Rhee, J., Pöpperl, H., et al. (2004). The TALE homeodomain protein Pbx2 is not essential for development and long-term survival. Mol. Cell. Biol. *24*, 5324–5331.

Shanmugam, K., Green, N.C., Rambaldi, I., Saragovi, H.U., and Featherstone, M.S. (1999). PBX and MEIS as non-DNA-binding partners in trimeric complexes with HOX proteins. Mol. Cell. Biol. *19*, 7577–7588.

Shen, W.F., Montgomery, J.C., Rozenfeld, S., Moskow, J.J., Lawrence, H.J., Buchberg, A.M., and Largman, C. (1997). AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins. Mol. Cell. Biol. *17*, 6448–6458.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., and Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116–120.

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H.J., and Mann, R.S. (2011). Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. Cell *147*, 1270–1282.

Thorsteinsdottir, U., Kroon, E., Jerome, L., Blasi, F., and Sauvageau, G. (2001). Defining roles for HOX and MEIS1 genes in induction of acute myeloid leukemia. Mol. Cell. Biol. *21*, 224–234.

Tümpel, S., Cambronero, F., Ferretti, E., Blasi, F., Wiedemann, L.M., and Krumlauf, R. (2007). Expression of Hoxa2 in rhombomere 4 is regulated by a conserved cross-regulatory mechanism dependent upon Hoxb1. Dev. Biol. *302*, 646–660.

Williams, T.M., Williams, M.E., and Innis, J.W. (2005). Range of HOX/TALE superclass associations and protein domain requirements for HOXA13:MEIS interaction. Dev. Biol. *277*, 457–471.

Wong, P., Iwasaki, M., Somervaille, T.C., So, C.W., and Cleary, M.L. (2007). Meis1 is an essential and rate-limiting regulator of MLL leukemia stem cell potential. Genes Dev. *21*, 2762–2774.