Hox binding specificity is directed by DNA sequence preferences and differential abilities to engage inaccessible chromatin.

Milica Bulajić [1,5,6], Divyanshi Srivastava [2,5], Jeremy S Dasen [3], Hynek Wichterle [4], Shaun Mahony [2], Esteban O Mazzoni [1,6].

1: Department of Biology, New York University, New York, NY, USA.

2: Center for Eukaryotic Gene Regulation, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA, USA.

3: Neuroscience Institute, Department of Neuroscience and Physiology, NYU School of Medicine, New York, NY, USA.

4: Department of Physiology and Cellular Biophysics, Columbia University Medical Center, New York, NY, USA.

5: Equal contribution

6: Corresponding authors: Milica Bulajić (mb5443@nyu.edu), Esteban O Mazzoni (eom@nyu.edu).

ABSTRACT

While *Hox* genes encode for conserved transcription factors (TFs), they are further divided into anterior, central, and posterior groups based on their DNA-binding domain similarity. The posterior group expanded in the deuterostome clade and patterns caudal and distal vertebrate structures such as the spinal neuronal diversity required for motor function. Our data revealed that limb-level patterning central Hoxc6, Hoxc8 and posterior Hoxc10 have a reduced ability to access occluded sites compared to other tested Hox TFs. Thus, their genomic binding relies more on cell-specific chromatin accessibility. Although posterior Hoxc9, Hoxc10 and Hoxc13 induce different fates, they share motif preference. However, Hoxc9 and Hoxc13 have a unique ability to access sites occluded by chromatin, resulting in divergent genomic binding patterns. From these results, we propose that the differential abilities of posterior Hox TFs to bind to previously inaccessible chromatin is the predominant force driving their patterning diversification.

INTRODUCTION

Hox genes encode a highly conserved transcription factor family that endow cells with positional identity during embryonic development (Duboule & Dolle, 1989; Lewis, 1978; McGinnis & Krumlauf, 1992). In mammals, *Hox* genes are organized into four clusters located on different chromosomes (*HoxA*, *HoxB*, *HoxC,* and *HoxD*). Each cluster contains a subset of 13 similar paralogous *Hox* genes, genomically arranged in the same order as their spatial and temporal expression patterns in the developing embryo, a phenomenon known as collinearity (Duboule & Morata, 1994; Kmita & Duboule, 2003). Changes in *Hox* gene expression patterns induce gross morphological changes, such as organs transforming into one another (homeotic transformations). However, how similar Hox transcription factors (TFs) assign positional identities during cell differentiation is not entirely understood.

Hox genes encode for homeodomain-containing TFs (Akam, 1989; Regulski et al., 1985). Homeodomains are highly conserved helix turn helix DNA-binding domains that recognize similar consensus DNA sequences (Affolter, Slattery, & Mann, 2008; Berger et al., 2008; Gehring et al., 1994; Noyes et al., 2008). Vertebrate Hox TFs can be divided into anterior (Hox1-5), central (Hox6-8), and posterior (Hox9-13) paralog groups. The posterior Hox group is particularly interesting because it has expanded in the deuterostome clade. While there is one *Drosophila Abd-B*, several *Abd-B*-related *Hox* genes assign different posterior positional identities during vertebrate development (reviewed in (Duboule, 2007; Lanfear, 2010)). Thus, understanding how Hox TFs differentiate their genomic binding activity to specify cell fates is at the core of understanding vertebrate body patterning.

In vitro binding studies have uncovered the intrinsic sequence preferences of Hox TFs alone or in complex with a specific cofactor. These studies revealed slight sequence preference differences between anterior (Hox1-5) and central Hox paralog groups (Hox6-8), which prefer motifs containing the canonical TAAT core sequence, and posterior paralog groups (Hox9-13),

which prefer the TTAT sequence (Berger et al., 2008; Ekker et al., 1994; Mann, Lelli, & Joshi, 2009; Noyes et al., 2008). Moreover, the interaction between Hox TFs and Meis and Pbx cofactors increases the specificity and selectivity of Hox DNA-binding (reviewed in (Mann & Affolter, 1998; Mann & Chan, 1996; Merabet & Mann, 2016)). However, these studies do not address Hox binding in a cell-specific chromatin context, which can be explored by investigating genome-wide binding in differentiating cells. Moreover, the variability among *Abd-B*-derived vertebrate posterior Hox TFs cannot be explored with *Drosophila* genes. Virally-expressed Hoxa9-13 and Hoxd9-13 in primary chicken mesenchymal limb progenitors exhibit some binding specificity differences between the posterior group Hox TFs, with Hoxa/d13 paralogs being the most different (Jerkovic et al., 2017). *Drosophila* Hox proteins have distinct abilities to bind accessible and inaccessible chromatin, and these differential abilities are associated with binding selectivity (Porcelli, Fischer, Russell, & White, 2019). However, how Hox TF abilities to bind to inaccessible chromatin shape their different patterning activities is not well understood. This aspect is of particular interest for the *Abd-B* related posterior Hox group, which have similar sequence preferences but different patterning abilities.

Among the many structures patterned by *Hox* genes, vertebrate spinal cord neuronal diversity requires *Hox* gene patterning activity along its rostro-caudal axis to match the changing musculature along the main body axis (Dasen & Jessell, 2009; Dasen, Liu, & Jessell, 2003; Sweeney et al., 2018). Motor neurons (MNs) maintain a common generic fate along the rostro-caudal axis, but Hox TFs specify their subtype. For example, the limb-innervating lateral motor column (LMC) MN fate is induced by Hoxc6/Hoxc8 at brachial (forelimb) levels, while posterior Hoxc10 induces the lumbar (hindlimb) LMC (Dasen, De Camilli, Wang, Tucker, & Jessell, 2008; Jung et al., 2018; Lacombe et al., 2013; Rousso, Gaber, Wellik, Morrisey, & Novitch, 2008). At both limb levels, neurons express similar canonical LMC genes such as Raldh2 and Foxp1, with dorsally and ventrally projecting neurons expressing Lhx1 and Isl1, respectively (Dasen et al.,

2008; Dasen & Jessell, 2009; Tsuchida et al., 1994). On the other hand, Hoxc9 carves out thoracic fate by repressing brachial-inducing *Hox* genes and the LMC fate (Jung et al., 2010). Similarly, V1 interneurons are also patterned by *Hox* genes, resulting in expression of specific combinations of transcription factors in interneurons formed at forelimb and hindlimb levels, which differ compared to combinations expressed at the thoracic level (Sweeney et al., 2018). Thus, understanding how vertebrate Hox TFs pattern MN and interneuron identity into different subtypes will illuminate mechanisms that establish neuronal diversity.

The role of *Hox* genes in organizing the spinal cord presents an interesting conundrum. The shared limb expression program is controlled by central (Hox6 & Hox8) and posterior (Hox10) Hox TFs, which are expected to have distinct intrinsic DNA binding preferences. Meanwhile, the posterior Hoxc9 should bind to similar DNA sequences as Hox10 TFs, but it instead instructs thoracic fate. Moreover, while all Hox9 paralogs repress rostral *Hox* gene expression, Hoxc9 distinguishes itself from the rest of its paralogs with its ability to induce thoracic fate (Jung et al., 2014). In agreement with their genomic cluster position, *Hox13* paralogs are expressed late during development, distally, and in posterior regions. Moreover, they are associated with the unique ability to serve as “repressors of elongation” or “patterning terminators” by inhibiting cell growth or inducing apoptosis (Denans, Iimura, & Pourquie, 2015; Economides, Zeltser, & Capecchi, 2003; Godwin & Capecchi, 1998). As a model to understand Hox-dependent patterning, we sought to understand how central and posterior Hox TFs bind the genome to induce different spinal cord identities.

To understand how Hox TFs induce different fates, we performed a comparison of global binding preferences and transcriptional targets of different Hox proteins during embryonic stem cell to MN/interneuron differentiation. We focused on a subset of HoxC TFs (Hoxc6, 8, 9, 10, and 13) because of their importance for inducing spinal cord identities. Furthermore, we complemented these studies with additional Hox9 paralog groups. Our results suggest that limb-

innervating fate is not the product of identical central and posterior Hox binding patterns since Hoxc10 does not mimic Hoxc6 and Hoxc8 binding profiles. Although posterior group Hox TF have similar DNA motif preferences, they do not bind to the genome with identical patterns. This difference is mainly due to their abilities to bind to motifs occluded in inaccessible chromatin. Hoxc13 and Hoxc9 bind larger fractions of sites in previously inaccessible chromatin when compared to Hoxc10 and other Hox9 paralogs. In this regard, Hoxc10 behaves more like central Hoxc6 and Hoxc8 that induce a similar limb-innervating fate. From these results, we propose that the differential abilities of posterior Hox TFs to bind to previously inaccessible chromatin is the predominant force driving their patterning diversification.

RESULTS

Hox TFs induce distinct patterns of gene expression during in vitro spinal cord differentiation

The study of Hox TF genomic binding has been made difficult by the lack of availability of homogenous relevant cell populations at scales compatible with chromatin immunoprecipitation studies. To mitigate this technical limitation, we opted for an embryonic stem cell (ESC) differentiation system that recapitulates ventral spinal cord early development (Davis-Dusenbery, Williams, Klim, & Eggan, 2014; Wichterle, Lieberam, Porter, & Jessell, 2002; Wichterle & Peljto, 2008). In response to the dorso-ventral Hedgehog and rostro-caudal retinoic acid (RA) patterning signals, ESCs differentiate into MNs and interneurons transitioning through embryonic progenitor states (Wichterle et al., 2002). The culture acquires a rostral spinal cord identity, with 90% of cells expressing *Hoxa5* (Peljto, Dasen, Mazzoni, Jessell, & Wichterle, 2010; Peljto & Wichterle, 2011). Moreover, differentiating ESCs respond to *Hox* gene overexpression similarly to those in the developing spinal cord (Jung et al., 2010; Machado et

al., 2014; Narendra, Bulajic, Dekker, Mazzoni, & Reinberg, 2016; Tan, Mazzoni, & Wichterle, 2016). The stem cell differentiation system allows for the study of homogenous differentiating cells at an exquisite temporal resolution, producing a large number of cells required for chromatin studies. Therefore, *in vitro* MN/interneuron differentiation is an advantageous platform to study Hox TF binding to the genome during neuronal diversification in a developmentally relevant cellular context.

We initially focused our attention on a subset of HoxC genes because of their cardinal spinal cord patterning activities and generated isogenic ESC cell lines that express Hoxc6 (brachial), Hoxc8 (brachial), Hoxc9 (thoracic) or Hoxc10 (lumbar) upon Doxycycline (Dox) treatment (Figure 1A, B) (Dasen et al., 2008; Dasen et al., 2003; Dasen, Tice, Brenner-Morton, & Jessell, 2005; Iacovino et al., 2011; Jung et al., 2010; Jung et al., 2014; Lacombe et al., 2013; Mazzoni et al., 2011). We refer to these as iHoxc6, iHoxc8, iHoxc9, and iHoxc10 neurons, respectively. To recapitulate *Hox* expression dynamics, Dox was added at a time point when the culture consists of neural progenitors (Figure 1A, Supp Figure 1A). As time progresses, cells become neurons, peaking at 48h after Dox addition when the culture mostly consists of postmitotic MNs as well as interneurons.

To investigate the transcriptional consequences of *Hox* induction on gene expression, we performed RNA-seq in Hoxc6, Hoxc8, Hoxc9, Hoxc10 induced postmitotic neurons as well as control neurons not treated with Dox (no Dox control) (Figure 1A-B, Supp Figure 1A). To examine global similarity in gene expression profiles amongst the four iHox cell lines, we used principal component analysis (PCA). We filtered out genes that were not expressed in any iHox neuron, retaining 19,019 genes (see Methods). The first two principal components explained 80% of the variance in RNA-seq tag counts at these genes, with variance along PC1 reflecting a combination of paralog group and the subtype identities specified by the Hox proteins (Figure 1C). iHoxc6 and iHoxc8 were most similar to each other and the control cells, which mostly

express *Hox5* but harbor some *Hox6* paralog expression at this timepoint (Mazzoni et al., 2013; Narendra et al., 2016; Wichterle et al., 2002). iHoxc9 grouped the furthest away, while iHoxc10 grouped between cells expressing *Hox5-8* genes (control cells, iHoxc6 and iHoxc8) and iHoxc9. Multidimensional scaling (MDS), a non-linear dimensionality reduction technique, produced a similar lower-dimensional embedding (Supp Figure 1B).

Closer analysis of the RNA-seq data revealed that while Hoxc9 and Hoxc10 are both posterior paralogs, their transcriptional impact and the effects on cell fate are distinct, illustrated by the expression of known marker genes. For example, Hoxc6, Hoxc8, and Hoxc10 overexpression induce canonical LMC markers Raldh2 and Foxp1 (Figure 1D, Supp Figure 1C) (Narendra et al., 2016; Tan et al., 2016). A similar recapitulation of the known *in vivo* phenotype can be seen upon *Hoxc9* overexpression, which leads to the repression of anterior *Hox1-5* paralogs (Figure 1D, Supp Figure 1C) (Jung et al., 2010). Importantly, Hox induction during *in vitro* spinal cord differentiation did not affect the ability of cells to acquire an MN identity, as illustrated by the shared downstream expression of Mnx1, Isl1/2, and Lhx3 (Figure 1D, Supp Figure 1C). Thus, Hox TF activity during ESC differentiation induces distinct fates without affecting the generic MN fate, as it does during embryonic development, making it an appropriate strategy to study how Hox TFs bind the genome to activate target gene expression and induce proper positional identities.

Central and posterior Hox TFs have distinct sequence motif preferences during MN differentiation

Hox proteins have similar DNA binding domains, yet they modify the transcriptomes in positional-specific patterns along the spinal cord rostro-caudal axis. To understand how Hox TFs assign positional identity, we assayed genomic binding of Hoxc6, Hoxc8, Hoxc9, and

Hoxc10 by performing ChIP-seq experiments 24h and 48h after Dox treatment (Figure 1A, Supp Figure 1A). These time points were chosen because they align with newborn and young postmitotic MNs, critical times for Hox positional identity patterning. We compared the binding of Hoxc6, Hoxc9, and Hoxc10 Hox TFs at 24h and 48h after Dox induction and concluded that the binding of each Hox protein is mostly the same in young and postmitotic MNs (Supp Figure 2A-C). Importantly, all Hox proteins are N-terminally Flag-tagged, which allows for immunoprecipitation with the same antibody, eliminating any bias that could occur from different antibody affinities and cross-reactivity.

De novo motif discovery, using the ensemble method MEME-ChIP, revealed distinct motifs when comparing central versus posterior Hox ChIP-seq datasets (Machanick & Bailey, 2011). The representative enriched motifs detected for the Hoxc6 and Hoxc8 datasets contained the TAAT sequence, while the identified sequences for the Hoxc9 and Hoxc10 datasets matched the posterior motif TTAT and the bipartite Pbx (TALE cofactor) and posterior Hox motif TGATTTAT (Figure 2A-D). Thus, central and posterior Hox proteins bind to sites in differentiating cells that are in agreement with previous *in vitro* binding preference studies: anterior and central Hox paralogs prefer the TAAT motif while the posterior paralogs prefer the TTAT motif (Mann et al., 2009; Noyes et al., 2008). However, this motif preference did not discriminate within central and posterior group Hox TFs and thus cannot fully explain vertebrate Hox patterning abilities.

Hoxc6, Hoxc9, and Hoxc10 TFs have different genome-wide binding profiles

We next asked how the differing sequence preferences are translated into distinct central versus posterior Hox TF genomic binding distributions. Moreover, posterior Hox TFs share sequence specificity but have different fate-inducing abilities, thus it is not clear how their

genomic binding would be in differentiating cells. We began by comparing Hoxc6 and Hoxc8 because they are central group *Hox* genes and both induce brachial identity. We restricted the analysis to the top 10,000 ChIP-seq sites in each dataset for all downstream analyses to ensure the least amount of bias from comparing different experiments. During cell differentiation, Hoxc6 and Hoxc8 proteins bind the genome in a remarkably similar pattern; less than 2% of tested Hoxc6 and Hoxc8 sites are significantly differentially bound by the two TFs (Figure 3A). Thus, the two main brachial Hox proteins bind similar target sites in differentiating neurons. To facilitate the interpretation of the subsequent comparisons, we picked Hoxc6 to represent the brachial-inducing Hox binding profile.

The brachial to thoracic transition is marked by phenotypic differences and also by the transition from central Hox TFs to the posterior Hoxc9. Hoxc6 and Hoxc9 exhibit 68% sequence similarity in their homeodomains (Supp Figure 3A). ChIP-seq comparisons with EdgeR revealed that 54% of Hoxc9's top 10,000 binding sites show similar enrichment levels for both Hoxc9 and Hoxc6 (Figure 3B, Supp Figure 3B). Additionally, there are sites differentially bound by Hoxc6 or Hoxc9 (EdgeR q-value <0.01, see Methods). Thus, the different patterning abilities of a brachial and a thoracic Hox protein could in part be explained by their DNA sequence preferences.

The fourth Hox TF important for spinal cord patterning, Hoxc10, has a more similar protein sequence to Hoxc9, yet it induces the limb-innervating fate like central Hoxc6 and Hoxc8 TFs (Figure 1C-D, Supp Figure 1B-C, Supp Figure 3A). Although both induce limb-innervating spinal cord fate, Hoxc6 and Hoxc10 do not have identical binding patterns, as seen in the Hoxc6-Hoxc8 comparison. While there are sites that Hoxc6 and Hoxc10 bind at similar levels, there are also unique Hoxc6 and Hoxc10 sites in differentiating cells (Figure 3C, Supp Figure 3C). However, unlike the comparison between central Hoxc6 and Hoxc8, the posterior Hoxc9-Hoxc10 comparison yielded a new category, sites that are unique for Hoxc9 (Figure 3D, Supp Figure 3D). On the other hand, 90% of Hoxc10 sites are bound at similar levels by Hoxc9. Thus,

although Hoxc9 and Hoxc10 have similar sequence preferences, Hoxc9 binds to additional genomic locations.

To better characterize central versus posterior and limb-innervating versus thoracic binding patterns, we compared binding sites as shared versus differential between Hoxc6, Hoxc9, and Hoxc10 (see Methods) (Figure 3E, Supp Figure 3E). This analysis revealed that 4,892 sites were bound similarly by Hoxc6, Hoxc9, and Hoxc10 ("**c6=c9=c10**") (Figure 3E). Hoxc6 and Hoxc9 differentially bind large sets of sites, with 3,031 sites bound by Hoxc6 alone ("**c6**>c9,c10"), and 3,310 sites bound by Hoxc9 alone ("**c9**>c6,c10"). Consistent with Hoxc10 binding to a subset of Hoxc9 sites, less than 200 sites were categorized as Hoxc10>Hoxc6,Hoxc9. Additionally, there are 1,106 sites differentially bound by posterior Hox TFs ("**c9,c10**>c6"). Finally, there are 1,055 sites bound by Hoxc6 and Hoxc9 ("**c6,c9**>c10"). Mirroring whole ChIP-seq analysis for each factor, MEME-ChIP motif analysis revealed distinct primary motifs when comparing differentially bound sites by a central versus posterior Hox TF (Figure 3F). Altogether, Hox TFs bind to a set of similar sites regardless of their paralog group and fate inducing activity. The brachial and thoracic inducing Hox TFs are different in their binding specificity and bind to unique sites. Although Hoxc10 and Hoxc9 share sequence specificity, Hoxc10 lacks the ability to bind to a large fraction of Hoxc9 sites.

We then asked whether differentially expressed genes in the iHox neurons correlated with specific Hox binding categories. Specifically, we used the logistic regression-based ChIP-Enrich method to identify significant associations (adjusted p-value < 0.01) between RNA-seq derived gene sets and Hox TF binding categories. Genes that are equally upregulated in iHoxc6, iHoxc9 and iHoxc10 neurons associated strongly with **c6=c9=c10** sites, with some association with **c6**>c9,c10 sites and **c9,c10**>c6 sites as well (Figure 4A). Thus, genes upregulated in all Hox inductions correlate with sites bound by all Hox TFs, suggesting a universal "Hox input" in their regulation. Genes differentially expressed in iHoxc6, compared to iHoxc9 or iHoxc10 neurons,

showed a correlation with **c6**>c9,c10 sites (Figure 4A). Similarly, genes differentially expressed in iHoxc9 neurons correlate with **c9**>c6,c10 sites (Figure 4A). Thus, Hox TF binding correlates with transcriptional activity. Moreover, these results suggest that while Hoxc9 expression represses rostral *Hox* gene expression, Hoxc9 is not a general repressing Hox TF.

Hoxc9 has a higher preference for inaccessible chromatin than Hoxc6 and Hoxc10

The central Hoxc6 and posterior TFs (Hoxc9, Hoxc10) bind distinct sequences ("TAAT" vs. "TTAT") explaining Hoxc6 divergent binding. However, Hoxc9 and Hoxc10 binding does not entirely overlap, yet they share sequence preferences ("TTAT", "TGATTTAT"). Since chromatin accessibility is an important determinant of genome-wide TF binding specificity, we decided to explore whether the prior accessibility landscape affects genomic binding patterns of Hox TFs even in instances where they share a primary motif preference.

To investigate whether the chromatin accessibility landscape prior to Hox induction shapes Hox TF binding patterns, we characterized genome-wide chromatin accessibility states by ATAC-seq at the progenitor stage (Figure 5A, Supp Figure 1A). The distribution of progenitor ATAC-seq read density at Hoxc6, Hoxc9, and Hoxc10 sites revealed that Hoxc9 binding sites had the lowest median accessibility before each factor is induced (Figure 5B). Hox factors share some genomic binding locations, yet they also bind to unique sites. Thus, we hypothesized that prior chromatin accessibility might not be equally distributed across all binding sites for each Hox TF. **c6=c9=c10**, **c6**>c9,c10, and **c6,c9**>c10 binding categories harbor similar prior accessibility landscapes (55%, 54%, and 45% of sites overlap accessible domains in neuronal progenitors, respectively). On the other hand, **c9**>c6,c10 and **c9,c10**>c6 binding occurs in genomic locations with much lower prior accessibility (16% and 18% of sites overlap accessible domains in MN progenitors, respectively) (Figure 5C, D). Dissecting Hox binding into different categories

underscored a possible mechanism to complement sequence specificity in generating Hox TF binding profiles. Hoxc6 and Hoxc10 overall binding has a higher correlation with preexisting accessibility than Hoxc9. These results suggest that sites bound by all studied Hox TFs have higher prior accessibility to allow for binding of those factors that cannot bind to inaccessible sites. Hoxc9 has a significant fraction of its binding in previously inaccessible regions, so although it shares motif preference with Hoxc10, it can bind to more sites.

To test if Hox differential preferences for accessible regions were TF intrinsic or due to a MN/interneuron progenitor specific chromatin state or cofactor, we investigated their binding in a different cell type. Undifferentiated cells found in embryoid bodies (EBs) at an early differentiation state before the application of patterning signals have a dissimilar accessibility landscape than MN progenitor state cells and were chosen as the system conducive for experiments utilizing the same inducible system (Figure 5E). Even in this different cell type, Hoxc9 maintains a higher preference for inaccessible chromatin than Hoxc6 (Figure 5F). Thus, the intrinsic ability of Hox TFs to bind to inaccessible chromatin seems to be independent of the particular cellular environment in which they are expressed. Altogether, comparing Hox TF binding with prior accessibility revealed that limb-innervating Hoxc10 and Hoxc6 rely more on chromatin accessibility established at progenitor stages to find their target sites. Moreover, these results divide the posterior paralog group by their ability to bind previously inaccessible chromatin, placing Hoxc9 with a higher ability to bind inaccessible chromatin compared with Hoxc10.

Hox TF binding increases chromatin accessibility

The difference in the abilities of Hox TFs to bind to inaccessible chromatin prompted us to explore whether Hox TFs change the accessibility landscape after binding. To characterize the

accessibility state after Hox binding, we performed ATAC-seq in postmitotic neurons expressing Hoxc6, Hoxc9, or Hoxc10 (Figure 6A, Supp Figure 1A). We then compared accessibility changes from progenitors to post-mitotic neurons at Hoxc6, Hoxc9, and Hoxc10 binding events.

Consistent with Hox TF increasing or maintaining the chromatin status, **c6**=**c9**=**c10** sites gained accessibility after Hoxc6, Hoxc9, and Hoxc10 expression (Figure 6B, C). **c6**>c9,c10 sites gained some accessibility in iHoxc6 neurons, while they lost accessibility in iHoxc9 and iHoxc10 expressing cells. Thus, although with a relatively weaker ability to bind inaccessible sites, Hoxc6 can maintain accessibility. **c9**,**c10**>c6 sites gained accessibility in both iHoxc9 and iHoxc10 neurons, while not significantly changing in iHoxc6 cells. These posterior Hox TFs bind to sites with low accessibility and increase accessibility upon binding. Hoxc9 exclusive sites had the lowest prior accessibility and gained the most accessibility in iHoxc9 neurons. Thus, the large category of sites differentially bound by Hoxc9 went through the biggest change in accessibility status only in Hoxc9-expressing neurons (Figure 6B, C). These results show that although Hox TF binding is generally associated with maintained or increased chromatin accessibility, Hoxc9 stands out in its ability to bind to a large set of sites in inaccessible chromatin and increase the accessibility status after binding.

Posterior group Hox TFs display a range of abilities to bind inaccessible chromatin

Genome-wide chromatin preferences of Hoxc6, Hoxc8, Hoxc9 and Hoxc10 revealed that Hoxc9 binds a larger fraction of previously inaccessible sites. We wondered whether Hoxc9 or Hoxc10 is an outlier compared to other posterior Hox TFs. Hoxc9 thoracic inducing ability is unique among Hox9 paralogs (Jung et al., 2014). Thus, the ability of Hoxc9 to bind to inaccessible chromatin could also be unique among Hox9 paralogs. We captured the genomic binding of Hoxa9 and Hoxd9 by performing ChIP-seq experiments 24h after Dox treatment (Figure 7A-B).

The detected motifs for Hoxa9 and Hoxd9 resemble the posterior Hox TTAT and bipartite Pbx and posterior TGATTTAT motifs (Figure 7A-B). A comparison of Hoxa9 and Hoxd9 ChIP-seq using EdgeR revealed that they share most of their binding sites (Figure 7C), while comparing Hoxc9 and Hoxa9 binding patterns showed that Hoxc9 binds a large category of sites alone (Figure 7D). Among the Hox9 paralogs, Hoxc9 binds to sites with the lowest median prior accessibility (Figure 7E). Accordingly, the sites uniquely bound by Hoxc9 showed a lower prior accessibility than other sets of sites (Figure 7F, G). Hence, along with a strong ability to induce thoracic fate, the ability of Hoxc9 to bind to a large set of inaccessible sites is not shared by other Hox9 paralogs.

The posterior Hox binding profiles we have analyzed revealed that in spite of sharing motif preferences, they do not entirely overlap in their genomic binding. Thus, we sought to expand these analyses to another relevant posterior *Hox* gene. *Hox13* paralogs are also posterior group genes but have the unique ability to terminate axial elongation. Analysis of Hoxc13 ChIP-seq data revealed a posterior Hox TTAT motif (Figure 8A). To assess whether Hoxc13 would bind like Hoxc9 or other posterior Hox TFs, we compared Hoxc9, Hoxc10, and Hoxc13 genomic binding overlaps (Figure 8B). A small fraction of all sites (1,024) were shared by Hoxc9, Hoxc10, and Hoxc13 ("**c9=c10=c13**"). Hoxc9 and Hoxc10 ("**c9,c10**>c13") bind similarly to 6,257 sites, while only 1,208 sites were bound similarly by Hoxc9 and Hoxc13 ("**c9,c13**>c10"). As in all previous comparisons, Hoxc9 retained a subset of private sites 2,543 ("**c9**>c10,c13"). However, in this comparison, a large category (5,652) of sites was bound by Hoxc13 alone ("**c13**>c9,c10"). A direct comparison of Hoxc9 and Hoxc13 binding profiles supported the finding that they largely bind distinct sets of sites (Supp Figure 4A). Thus, while Hoxc13 shares a posterior motif preference, it binds to unique sites increasing the posterior Hox TF binding diversity.

We then asked if the posterior Hox group TF binding diversity correlates with their differential ability to bind to previously inaccessible chromatin. The distribution of progenitor ATAC-seq read density at Hoxc9, Hoxc10, and Hoxc13 sites revealed that Hoxc13 binding sites had the lowest median accessibility in MN progenitors, even lower than Hoxc9 (Figure 8C). Dissecting the accessibility at the different Hox binding categories underscored the ability of Hoxc13 to bind to sites with the lowest prior accessibility (Figure 8D, E).

To visualize the overall variation in Hox TF genome-wide binding profiles, we performed Principal Components Analysis (PCA) on Hox TF ChIP-seq read counts associated with the set of genomic sites that belong to the top 10,000 binding sites for at least one Hox TF (38,874 unique genomic locations; Figure 9A). PC1, which explains 43% of the variance between the TFs, separates Hoxc13 from the anterior Hox TFs, Hox9 paralogs, and Hoxc10. On the other hand, PC2 and PC3, which cumulatively explain 32% of the variance, separate the TFs into three clusters. The limb-level-specifying Hoxc6 and Hoxc8 cluster together, Hoxc9 clusters by itself, and finally Hoxc10 and Hox9 paralogs Hoxa9 and Hoxd9 cluster together. Of note, the binding pattern of HoxC TFs to the *HoxC* gene cluster demonstrates their differential ability to engage inaccessible chromatin (Figure 9B). While Hoxc6 only binds *Hoxc4-5* genes, which are transcriptionally active, Hoxc10, Hoxc9 and Hoxc13 bind progressively deeper into the cluster at repressed *HoxC* genes which are covered in the catalytic product of PRC2 (Mazzoni et al., 2013; Narendra et al., 2015). Thus, differences in genome-wide binding profiles of the Hox TFs reflect sequence preferences, as well as the differential abilities of the Hox TFs to bind previously inaccessible chromatin.

DISCUSSION

Hox TFs have crucial roles in body patterning during animal development. However, surprisingly little is known about how the genomic binding patterns of Hox TFs compare with one another in vertebrate cells. In this study, we performed a multilevel comparison of global binding site patterns, chromatin accessibility preferences, and transcriptional target genes of seven Hox proteins expressed under the same developmentally relevant conditions. These studies highlight two main aspects of Hox TF genomic binding properties. First, intrinsic sequence preferences differ between central and posterior Hox TFs. Second, Hox TFs within the expanded posterior group have a range of abilities to bind to inaccessible chromatin. These two variables allow for Hox TFs to bind to similar regulatory regions based on cell-specific accessibility and to Hox-unique sites based on their differential ability to engage inaccessible chromatin.

The central group Hoxc6 and Hoxc8 TFs induce limb-innervating fate at brachial spinal cord levels. They do so by binding to similar sites in the genome. On the other hand, Hoxc10 induces a similar limb-innervating fate, but at lumbar levels. Although these anatomically distinct fates express the same canonical fate markers, Hoxc10 has a more similar binding profile to the thoracic Hoxc9 (repressor of limb-level fates) than to Hoxc6. Therefore, alike cell fates such as limb-innervating LMC neurons, are not always the product of identical Hox TF binding patterns. During embryonic development, LMC neurons are further patterned by Hox proteins into specific pools (Dasen et al., 2005). During this second stage, Hoxc6 and Hoxc8 TFs differ in their fine fate inducing identity. The *in vitro* differentiation strategy and the forced expression of a single *Hox* gene does not recapitulate this late differentiation event. Thus, we speculate that pool patterning by *Hox* genes does require a more complex developmental context expressing more *Hox* genes and signaling molecules important for pool identity acquisition (Arber, Ladle, Lin, Frank, & Jessell, 2000; Dasen et al., 2005; Haase et al., 2002).

In agreement with *in vitro* binding preference studies, central and posterior Hox proteins bind to different motifs in neurons (Mann et al., 2009; Noyes et al., 2008). While central Hox TFs bind to sites with the central TAAT motif, posterior Hox TFs bind to TTAT central motifs. MEME-ChIP also discovered bipartite cofactor and Hox motifs with high enrichment in some sets of sites. An interaction with TALE cofactors, Meis and Pbx, can change the affinity and selectivity of Hox DNA-binding (Slattery, Riley, et al., 2011) (reviewed in (Mann & Affolter, 1998; Mann & Chan, 1996; Merabet & Mann, 2016)). In addition, partnering with TALE cofactors modifies Hox binding specificity through the recognition of DNA shape (Abe et al., 2015; Joshi et al., 2007; Zeiske et al., 2018). Although *Hox* expression changes the fate of *in vitro* differentiating neurons, we have not seen differential binding of the canonical Meis Hox co-factor that could explain differential Hox TF binding patterns (Supp Figure 5). Furthermore, Pbx1-4 and Meis1-3 expression levels are similar across the different Hox TF inductions and thus cannot explain binding differences (Figure 1D, Supp Figure 1C). In search for additional factors, we failed to detect a CTCF motif at Hox binding sites, which was reported to co-bind with some HoxA/D proteins (Jerkovic et al., 2017). Thus, CTCF as a Hox-cofactor could be cell fate dependent.

This study revealed a gradation of Hox TF abilities to bind to inaccessible chromatin. This idea is supported further by previous studies reported that anterior Hox TF Hoxa1 does not act as a pioneer factor and instead binds to sites that were previously accessible (De Kumar et al., 2017). Moreover, a study of the binding profiles of Ultrabithorax (Ubx), Abdominal A (Abd-A) and Abdominal B (Abd-B) in *Drosophila* cells revealed that Abd-B harbored a unique ability to bind to inaccessible chromatin alone (Beh et al., 2016). Our results suggest that posterior group Hox TFs, which are related to Abd-B, have different abilities to engage inaccessible chromatin. Hoxc9 has a higher preference for inaccessible chromatin than any other spinal cord Hox TF. Furthermore, Hoxc13 has a high preference for inaccessible chromatin, unlike any other posterior Hox TF we investigated. Hox13 factors have a unique role in the termination of

posterior structures during development (Denans et al., 2015; Economides et al., 2003; Godwin & Capecchi, 1998; Young et al., 2009). Thus, we propose a model of posterior diversification based on their ability to bind inaccessible chromatin.

Limb innervating Hox TFs from the central and posterior groups do not share binding motif preference but share a reduced ability to bind to inaccessible sites. They rely more on cell-specific accessibility landscape generated during differentiation. Indeed, the binding patterns of the Hoxc8 orthologue Ultrabithorax (Ubx) in T3 leg and haltere imaginal discs are not identical (Slattery, Ma, Negre, White, & Mann, 2011). Thus, a Hox TF binding model should include motif preference, co-factor binding and Hox TF-specific ability to bind DNA motifs occluded by chromatin.

DATA AVAILABILITY

Sequencing data have been submitted to the GEO under accession code GSE142379.

ACKNOWLEDGEMENTS

This work was supported by the NICHD (R01HD079682) grant to E.O.M, the NYSTEM pre-doctoral training grant (C322560GG) to M.B, the Penn State Academic Computing Fellowship to D.S, the NINDS (R01NS062822) grant to J.S.D. The authors would like to thank Ayana Sawai-Frantz for sharing the pCAGGS mHoxA9 (JD-114) and pCAGGS mHoxD9 (JD-237) plasmids; the Mazzoni and Mahony lab members for feedback and constructive comments; the NYU Genomics Core facility for technical support and data processing.

AUTHOR CONTRIBUTIONS

M.B performed *in vitro* differentiation, ChIP-seq, RNA-seq, and ATAC-seq experiments and established inducible cell lines. D.S performed the analysis of sequencing data with help from M.B and S.M. J.S.D and H.W provided plasmids, cell lines, and guidance. M.B, D.S, S.M, and E.O.M conceived the experiments and set up the computational analysis framework. M.B and E.O.M co-wrote the original draft of the manuscript. All authors read and approved the final manuscript.

DECLARATION OF INTEREST

Authors declare no competing interests.

METHODS

Cell line generation

The inducible Hoxc6, Hoxc8, Hoxc9, and Hoxc10 cell lines were generated as previously described (Mazzoni et al., 2011). The inducible cassette exchange (ICE) system was used to generate cell lines (Iacovino et al., 2011). The resulting inducible cell lines harbor a single copy of the transgene inserted at the expression-competent HPRT locus. Inducible Hoxa9, Hoxd9 and Hoxc13 cell lines were generated for this study. Hoxa9, Hoxd9, and Hoxc13 cDNA was amplified using Phusion polymerase (Thermo Scientific) from pCAGGS mHoxA9 (JD-114), pCAGGS mHoxD9 (JD-237) and hHoxc13 cDNA (Dharmacon, Accession: BC090850), respectively. Flag and HA tags were introduced during the amplification step at the amino- or carboxyl-terminus, respectively. p2lox Hoxa9, p2lox Hoxd9, and p2lox Hoxc13 plasmids were generated by In-Fusion cloning (Takara) the respective cDNAs into the p2lox plasmid backbone. The recipient mESCs were treated with 1 µg/mL Doxycycline (Sigma, D9891) for 16h to induce

Cre recombinase expression prior to electroporation of the respective plasmids. After selection with G418 (250ng/mL, Cellgro), cell lines were characterized by performing antibody staining for Flag (mouse anti-FLAG; Sigma, F1804) and HA (rabbit anti-HA; Abcam, ab9110) and expanded.

Cell culture

mESC lines were cultured in 2-inhibitors based medium (Advanced DMEM/F12:Neurobasal (1:1) medium (Gibco), supplemented with 2.5% ESC-grade fetal bovine serum (vol/vol, Corning), N2 (Gibco), B27 (Gibco), 2mM L-glutamine (Gibco), 0.1 mM β-mercaptoethanol (Gibco), 1000 U/mL leukemia inhibitory factor (Millipore), 3μM CHIR (BioVision) and 1 μM PD0325901 (Sigma)) on 0.1% gelatin (Millipore) coated plates at 37 °C, 8% CO2.

In vitro differentiation of mESCs to motor neurons was described previously (Tan et al., 2016; Wichterle et al., 2002; Wichterle & Peljto, 2008). Briefly, embryoid bodies (EBs) were obtained by plating trypsinized (Gibco) mESCs in AK medium (Advanced DMEM/F12:Neurobasal (1:1) medium (Gibco), 7% KnockOut SR (vol/vol) (Gibco), 2 mM L-glutamine, 0.1 mM ß-mercaptoethanol and penicillin–streptomycin (Gibco),) at 37 °C, 5% CO2 (day -2). On day 0, EBs were split 1:2 and AK medium was replenished and supplemented with 1 μM all-trans retinoic acid and 0.5 μM smoothened agonist (SAG) (Millipore, 566660). TF induction was performed by adding 3 μg/mL of Doxycycline (Sigma, D9891) on day 2. For RNA-seq and ATAC-seq experiments, $3.5x10^5$ mESCs were plated in 100mm suspension dishes (Corning). For ChIP-seq experiments, $3\text{-}3.5x10^6$ mESCs were plated in 245mm x 245mm square dishes (Corning).

ChIP-seq

Cells were collected 24h and 48h after Doxycycline (Dox) treatment. Crosslinking was performed at room temperature in 1mM DSG (ProteoChem) for 15min, followed by the addition

of 1% FA (vol/vol) for an additional 15min. After quenching with Glycine, cells were divided into ~25-30x10^6 aliquots, pelleted by centrifugation and frozen at -80°C. After thawing cells on ice, lysis was performed in 5mL of 50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8.0, 10% glycerol (vol/vol), 0.5% Igepal (vol/vol), 0.25% Triton X-100 (vol/vol) with 1× protease inhibitors (Roche, 11697498001) for 10 min at 4 °C. Cells were centrifuged for 5min, resuspended in 5mL of 10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0 with 1× protease inhibitors, and incubated for 10 min at 4 °C on a rotating platform. Cells were centrifuged for 5min and resuspended in 2mL of Sonication Buffer (50 mM Hepes pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8.0, 1 mM EGTA pH 8.0, 1% Triton X-100 (vol/vol), 0.1% sodium deoxycholate (wt/vol), 0.1% SDS (vol/vol) with 1× protease inhibitors). Sonication was performed by splitting each sample to 2 Bioruptor tubes with sonication beads and using the Bioruptor Pico (Diagenode) for 18 cycles of 30 sec on and 30 sec off into an average size of approximately 200 bp. Immunoprecipitation was performed for 16h at 4 °C on a rotating platform by incubating with Dynabeads protein-G (Thermo Fisher Scientific) conjugated with 5 µg of one of the following antibodies: mouse monoclonal antibody to Flag (Sigma, F1804) or rabbit polyclonal antibody to HA (Abcam, ab9110). Following the immunoprecipitation, washes were performed with the following buffers (cold): sonication buffer, sonication buffer with 500 mM NaCl, LiCl wash buffer (20 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 250 mM LiCl, 0.5% Igepal (vol/vol), 0.5% sodium deoxycholate (wt/vol)), and TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8). Elution was performed by incubating in elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0, 1% SDS (vol/vol)) for 45 min at 65 °C. Reversal of crosslinks was performed by incubating for 16h at 65 °C. RNA digestion was performed by the addition of 200 µL of TE and RNAse A (Sigma) at a final concentration of 0.2 mg/mL and incubation for 2 h at 37°C. Proteinase K (Invitrogen) was added at a final concentration of 0.2 mg/mL, supplemented with CaCl2, to digest protein at 55°C for 30 min. DNA was purified with phenol:chloroform:isoamyl alcohol (25:24:1; vol/vol) (Invitrogen) and by performing ethanol

precipitation. DNA pellets were resuspended in water. Illumina DNA sequencing libraries were prepared with one third of the ChIP sample or a 1:100 dilution of the input sample in water. Library preparation was performed by end repair, A-tailing and ligating Illumina-compatible Bioo Scientific multiplexed adapters. Unligated adapters were removed using Agencourt AmpureXP beads (Beckman Coulter). Amplification was performed by PCR with Phusion polymerase (New England Biolabs) and TruSeq primers (Sigma). Libraries were gel purified (Qiagen) between 250 and 550bp in size. Final quantification of the library was performed using the KAPA library amplification kit on the Roche Lightcycler 480 before pooling. The libraries were sequenced on Illumina NextSeq 500 using V2 and V2.5 chemistry (75 cycles, single-end 75bp) at the Genomics Core Facility at NYU.

RNA-seq

Cells were collected prior to TF induction (day 2 of RA/SAG differentiation) and 48h after Doxycycline (Dox) treatment. RNA was extracted by using TRIzol LS Reagent (Life Technologies) and purified using the RNAeasy mini kit (Qiagen). Agilent High Sensitivity RNA Screentape (Agilent, 5067-5579) was used to check RNA integrity. 500ng of RNA was used to prepare RNA-seq libraries and spiked-in with ERCC Exfold Spike-in mixes (Thermo Fisher, 4456739). RNA-seq libraries were prepared using TruSeq Stranded mRNA Library Preparation kit (Illumina, 20020594). Library size was verified using High Sensitivity DNA ScreenTape (Agilent, 5067-5584). The KAPA library amplification kit was used on Roche Lightcycler 480 for library quantification before pooling. The libraries were sequenced on Illumina NextSeq 500 using V2.5 chemistry (75 cycles, single-end 75bp) at the Genomics Core Facility at NYU.

ATAC-seq

Cells were collected prior to TF induction (day 2 of RA/SAG differentiation) and 48h after Doxycycline (Dox) treatment. 50,000 cells were aliquoted and washed twice in 1 × PBS (cold).

Cell pellets were resuspended in 10 mM Tris pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, and freshly added 0.1% NP-40 (vol/vol), and centrifuged at 4 °C. Pellets were resuspended in 25 µL of 2 × TD buffer, 2.5 µL TDE1 (Nextera DNA sample preparation kit, FC-121–1030) and 22.5 µL of water and incubated at 37 °C for 30 min. The sample was purified using the MinElute PCR purification kit (Qiagen, 28004). A quantitative PCR reaction with 1 × SYBR Green (Invitrogen), custom-designed primers and 2 × NEB MasterMix (New England Labs, M0541) was performed to determine the optimal number of PCR cycles (one-third of the maximum measured fluorescence) (Buenrostro, Giresi, Zaba, Chang, & Greenleaf, 2013). PCR enrichment of the library was performed with custom-designed primers and 2 × NEB MasterMix (New England Labs, M0541). The libraries were purified using the MinElute PCR purification kit. High Sensitivity DNA ScreenTape (Agilent, 5067-5584) was used to verify the fragment length distribution of the library. Library quantification was performed using the KAPA library amplification kit on the Roche Lightcycler 480. The libraries were sequenced on Illumina NextSeq 500 using V2 and V2.5 chemistry (150 cycles, paired-end 75bp) at the Genomics Core Facility at NYU.

ChIP-seq Data Processing

ChIP-seq reads were aligned to the mm10 genome using Bowtie (v1.0.1) (Langmead, Trapnell, Pop, & Salzberg, 2009), using options “-q --best --strata -m 1 --chunkmbs 1024”. Genome-wide TF binding events were called in each condition using MultiGPS (v0.74) (Mahony et al., 2014). EdgeR (v3.22.5) was used within the MultiGPS framework to test whether genomic sites were differentially bound by TFs (Robinson, McCarthy, & Smyth, 2010). Specifically, EdgeR uses a negative binomial generalized linear model (GLM) to test whether ChIP-seq reads in one condition are significantly greater than in alternative condition. A genomic site was defined as shared by TFs if significant binding events were called in both TF ChIP-seq experiments (q-value < 0.001), and the TFs did not display differential read enrichment at that site as estimated

by EdgeR (q-value > 0.01). A binding event was defined as “TF1>TF2” if a MultiGPS peak was called in TF1 ChIP-seq (q-value < 0.001), TF1 exhibited a greater log fold-change with respect to the input ChIP-seq than TF2, and TF1 and TF2 were significantly differentially bound as defined by EdgeR (q-value < 0.01). A similar strategy was applied to perform multi-way ChIP-seq comparisons. For example, when comparing Hoxc6, Hoxc9, and Hoxc10 ChIP-seq experiments, “shared” binding sites were defined if significant binding events were detected in all three ChIP-seq datasets, and no two TFs were differentially bound with respect to each other. Sites were defined as “TF1> TF2, TF3” if a significant binding event was called for TF1, and a significantly greater number of reads were detected by EdgeR in TF1 ChIP-seq when compared to both TF2 and TF3. Finally, “TF1, TF2 > TF3” events were defined if significant binding events were called in both TF1 and TF2, no differential binding was detected between TF1 and TF2, and both TF1 and TF2 has a significantly greater number of reads than TF2. Only binding categories containing at least 500 binding events were retained

RNA-seq Data Processing

Fastq files obtained from RNA-sequencing were aligned to the genome using the splice-aware STAR (Spliced Transcripts Alignment to a Reference) aligner (v2.7.0c) (Dobin & Gingeras, 2016). Mapped reads were assigned to NCBI RefSeq annotated mm10 genes using the featureCount function in Rsubread (v1.30.9) (Liao, Smyth, & Shi, 2019). RefSeq genes with matching Entrez IDs were merged into a single gene by Rsubread. Following read summarization, read counts were normalized using the ‘rlog’ or regularized log transformation in DESeq2 (v1.20.0) (Love, Huber, & Anders, 2014). Transformed read counts were used as input features into the dimensionality reduction techniques PCA and MDS. The $\log_2$ fold change (LFC) in gene expression levels between iHox motor neuron vs. progenitor motor neurons was estimated using DESeq2. A q-value < 0.01 and LFC > 2 was used to define differentially

expressed genes between progenitor motor neurons, noDox controls, iHoxc6 MNs, iHoxc8 MNs, iHoxc9 MNs, and iHoxc10 MNs.

ATAC-seq Data Processing

Paired-end ATAC-seq reads were mapped to the mouse mm10 genome using Bowtie2 (v2.2.2) (Langmead et al., 2009). Genome-wide ATAC-seq derived accessible domains were defined using DomainFinder in the SeqCode project

(https://github.com/seqcode/seqcode-core/blob/master/src/org/seqcode/projects/seed/DomainFinder.java).

ChIP-seq and ATAC-seq Data Visualization

Heatmaps were used to plot the ChIP-seq and ATAC-seq reads at multiple categories of genomic sites in iHox neurons. Heatmaps were made using the MetaMaker program in SeqCode. (https://github.com/seqcode/seqcode-core/blob/master/src/org/seqcode/viz/metaprofile/MetaMaker.java). Raw reads from the ChIP-seq data were extended by 100 bps and read counts were binned into 100 bp bins. Binned reads were plotted over 1000 base pair windows centered on MultiGPS binding events. Color thresholds used to produce heatmaps were determined from binding data using the MetaPlotLineMaxEstimator program in SeqCode. Specifically, all 100 bp bins were ordered by the number of reads overlapping each bin. The maximum value for the heatmap color scale was set to the number of read counts at the 85th percentile bin. The minimum value for all heatmaps was set to 5 reads. Finally, all binding events represented were ordered based on their genomic co-ordinates.

ATAC-seq Composite Plots

For each ATAC-seq experiment, reads mapping to mitochondrial DNA or unannotated regions were filtered out. All experiments were performed in replicates; composite read distributions were initially calculated independently for each replicate. Filtered bedfiles were used to calculate the total number of reads overlapping each position in a 2000 base pair window surrounding a binding event of interest. The number of reads at each position was summed over the set of genomic sites being analyzed. The reads were total tag normalized, and finally the resulting read counts were averaged over replicates. Read density (reads per million per site) was plotted using seaborn in Python.

Association between gene sets and binding events

In order to construct gene sets, pairwise comparisons of gene expression levels between iHoxc6, iHoxc9 and iHoxc10 vs. progenitor motor neurons were performed using DESeq2. Additionally, pairwise comparisons were also performed between the iHox motor neurons: iHoxc6 vs. iHoxc9, iHoxc6 vs. iHoxc10 and iHoxc9 vs. iHoxc10. Genes that were upregulated in all iHox vs. pMN comparisons ($\log_2$ fold change (LFC) > 2 and adjusted p-value < 0.01), as well as not differentially expressed between the iHox MNs (|LFC| > 2) were assigned to the gene set “shared-upregulated”. A similar logic was used to identify the “shared-downregulated” genes. Pairwise comparisons between iHox motor neurons were used to identify genes significantly upregulated or downregulated (|LFC| > 2 and adjusted p-value < 0.01) in individual iHox MNs. Genes that were significantly upregulated or downregulated in two out of three iHox motor neurons were assigned to gene sets using the same thresholds.

In order to test whether specific categories of ChIP-seq binding events were associated with differentially regulated gene sets, we used ChIP-Enrich (v.1.10.0). Specifically, each peak was assigned to the gene with the nearest TSS. ChIP-Enrich uses a logistic regression model to test whether gene set membership (controlled by mappable gene length) predicts whether a gene is associated with a ChIP-seq peak (Welch et al., 2014). The significance of the weight associated

with gene set membership is estimated using the 'Wald' statistic. P-values obtained from the Wald statistic adjusted using the Benjamini-Hochberg multiple testing correction (Welch et al., 2014). For each category of binding events, adjusted p-values were plotted for each pre-defined gene set.

REFERENCES

Abe, N., Dror, I., Yang, L., Slattery, M., Zhou, T., Bussemaker, H. J., . . . Mann, R. S. (2015). Deconvolving the recognition of DNA shape from sequence. *Cell, 161*(2), 307-318. doi:10.1016/j.cell.2015.02.008

Affolter, M., Slattery, M., & Mann, R. S. (2008). A lexicon for homeodomain-DNA recognition. *Cell, 133*(7), 1133-1135. doi:10.1016/j.cell.2008.06.008

Akam, M. (1989). Hox and HOM: homologous gene clusters in insects and vertebrates. *Cell, 57*(3), 347-349.

Arber, S., Ladle, D. R., Lin, J. H., Frank, E., & Jessell, T. M. (2000). ETS gene Er81 controls the formation of functional connections between group Ia sensory afferents and motor neurons. *Cell, 101*(5), 485-498.

Beh, C. Y., El-Sharnouby, S., Chatzipli, A., Russell, S., Choo, S. W., & White, R. (2016). Roles of cofactors and chromatin accessibility in Hox protein target specificity. *Epigenetics Chromatin, 9*, 1. doi:10.1186/s13072-015-0049-x

Berger, M. F., Badis, G., Gehrke, A. R., Talukder, S., Philippakis, A. A., Pena-Castillo, L., . . . Hughes, T. R. (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell, 133*(7), 1266-1276. doi:10.1016/j.cell.2008.05.024

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods, 10*(12), 1213-1218. doi:10.1038/nmeth.2688

Dasen, J. S., De Camilli, A., Wang, B., Tucker, P. W., & Jessell, T. M. (2008). Hox repertoires for motor neuron diversity and connectivity gated by a single accessory factor, FoxP1. *Cell, 134*(2), 304-316. doi:10.1016/j.cell.2008.06.019

Dasen, J. S., & Jessell, T. M. (2009). Hox networks and the origins of motor neuron diversity. *Curr Top Dev Biol, 88*, 169-200. doi:10.1016/S0070-2153(09)88006-X

Dasen, J. S., Liu, J. P., & Jessell, T. M. (2003). Motor neuron columnar fate imposed by sequential phases of Hox-c activity. *Nature, 425*(6961), 926-933. doi:10.1038/nature02051

Dasen, J. S., Tice, B. C., Brenner-Morton, S., & Jessell, T. M. (2005). A Hox regulatory network establishes motor neuron pool identity and target-muscle connectivity. *Cell, 123*(3), 477-491. doi:10.1016/j.cell.2005.09.009

Davis-Dusenbery, B. N., Williams, L. A., Klim, J. R., & Eggan, K. (2014). How to make spinal motor neurons. *Development, 141*(3), 491-501. doi:10.1242/dev.097410

De Kumar, B., Parker, H. J., Parrish, M. E., Lange, J. J., Slaughter, B. D., Unruh, J. R., . . . Krumlauf, R. (2017). Dynamic regulation of Nanog and stem cell-signaling pathways by Hoxa1 during early neuro-ectodermal differentiation of ES cells. *Proc Natl Acad Sci U S A, 114*(23), 5838-5845. doi:10.1073/pnas.1610612114

Denans, N., Iimura, T., & Pourquie, O. (2015). Hox genes control vertebrate body elongation by collinear Wnt repression. *Elife, 4*. doi:10.7554/eLife.04379

Dobin, A., & Gingeras, T. R. (2016). Optimizing RNA-Seq Mapping with STAR. *Methods Mol Biol, 1415*, 245-262. doi:10.1007/978-1-4939-3572-7_13

Duboule, D. (2007). The rise and fall of Hox gene clusters. *Development, 134*(14), 2549-2560. doi:10.1242/dev.001065

Duboule, D., & Dolle, P. (1989). The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes. *EMBO J, 8*(5), 1497-1505.

Duboule, D., & Morata, G. (1994). Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet, 10*(10), 358-364.

Economides, K. D., Zeltser, L., & Capecchi, M. R. (2003). Hoxb13 mutations cause overgrowth of caudal spinal cord and tail vertebrae. *Dev Biol, 256*(2), 317-330. doi:10.1016/s0012-1606(02)00137-9

Ekker, S. C., Jackson, D. G., von Kessler, D. P., Sun, B. I., Young, K. E., & Beachy, P. A. (1994). The degree of variation in DNA sequence recognition among four Drosophila homeotic proteins. *EMBO J, 13*(15), 3551-3560.

Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., . . . Wuthrich, K. (1994). Homeodomain-DNA recognition. *Cell, 78*(2), 211-223.

Godwin, A. R., & Capecchi, M. R. (1998). Hoxc13 mutant mice lack external hair. *Genes Dev, 12*(1), 11-20. doi:10.1101/gad.12.1.11

Haase, G., Dessaud, E., Garces, A., de Bovis, B., Birling, M., Filippi, P., . . . deLapeyriere, O. (2002). GDNF acts through PEA3 to regulate cell body positioning and muscle innervation of specific motor neuron pools. *Neuron, 35*(5), 893-905.

Iacovino, M., Bosnakovski, D., Fey, H., Rux, D., Bajwa, G., Mahen, E., . . . Kyba, M. (2011). Inducible cassette exchange: a rapid and efficient system enabling conditional gene expression in embryonic stem and primary cells. *Stem Cells, 29*(10), 1580-1588.

Jerkovic, I., Ibrahim, D. M., Andrey, G., Haas, S., Hansen, P., Janetzki, C., . . . Mundlos, S. (2017). Genome-Wide Binding of Posterior HOXA/D Transcription Factors Reveals Subgrouping and Association with CTCF. *PLoS Genet, 13*(1), e1006567. doi:10.1371/journal.pgen.1006567

Joshi, R., Passner, J. M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M. A., . . . Mann, R. S. (2007). Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell, 131*(3), 530-543. doi:10.1016/j.cell.2007.09.024

Jung, H., Baek, M., D'Elia, K. P., Boisvert, C., Currie, P. D., Tay, B. H., . . . Dasen, J. S. (2018). The Ancient Origins of Neural Substrates for Land Walking. *Cell, 172*(4), 667-682 e615. doi:10.1016/j.cell.2018.01.013

Jung, H., Lacombe, J., Mazzoni, E. O., Liem, K. F., Jr., Grinstein, J., Mahony, S., . . . Dasen, J. S. (2010). Global control of motor neuron topography mediated by the repressive actions of a single hox gene. *Neuron, 67*(5), 781-796. doi:10.1016/j.neuron.2010.08.008

Jung, H., Mazzoni, E. O., Soshnikova, N., Hanley, O., Venkatesh, B., Duboule, D., & Dasen, J. S. (2014). Evolving Hox activity profiles govern diversity in locomotor systems. *Dev Cell, 29*(2), 171-187. doi:10.1016/j.devcel.2014.03.008

Kmita, M., & Duboule, D. (2003). Organizing axes in time and space; 25 years of colinear tinkering. *Science, 301*(5631), 331-333.

Lacombe, J., Hanley, O., Jung, H., Philippidou, P., Surmeli, G., Grinstein, J., & Dasen, J. S. (2013). Genetic and functional modularity of Hox activities in the specification of limb-innervating motor neurons. *PLoS Genet, 9*(1), e1003184. doi:10.1371/journal.pgen.1003184

Lanfear, R. (2010). Are the deuterostome posterior Hox genes a fast-evolving class? *Adv Exp Med Biol, 689*, 111-122. doi:10.1007/978-1-4419-6673-5_8

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol, 10*(3), R25. doi:10.1186/gb-2009-10-3-r25

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature, 276*(5688), 565-570.

Liao, Y., Smyth, G. K., & Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res, 47*(8), e47. doi:10.1093/nar/gkz114

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol, 15*(12), 550. doi:10.1186/s13059-014-0550-8

Machado, C. B., Kanning, K. C., Kreis, P., Stevenson, D., Crossley, M., Nowak, M., . . . Lieberam, I. (2014). Reconstruction of phrenic neuron identity in embryonic stem cell-derived motor neurons. *Development, 141*(4), 784-794. doi:10.1242/dev.097188

Machanick, P., & Bailey, T. L. (2011). MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics, 27*(12), 1696-1697. doi:10.1093/bioinformatics/btr189

Mahony, S., Edwards, M. D., Mazzoni, E. O., Sherwood, R. I., Kakumanu, A., Morrison, C. A., . . . Gifford, D. K. (2014). An integrated model of multiple-condition ChIP-Seq data reveals predeterminants of Cdx2 binding. *PLoS Comput Biol, 10*(3), e1003501. doi:10.1371/journal.pcbi.1003501

Mann, R. S., & Affolter, M. (1998). Hox proteins meet more partners. *Curr Opin Genet Dev, 8*(4), 423-429.

Mann, R. S., & Chan, S. K. (1996). Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. *Trends Genet, 12*(7), 258-262.

Mann, R. S., Lelli, K. M., & Joshi, R. (2009). Hox specificity unique roles for cofactors and collaborators. *Curr Top Dev Biol, 88*, 63-101. doi:10.1016/S0070-2153(09)88003-4

Mazzoni, E. O., Mahony, S., Iacovino, M., Morrison, C. A., Mountoufaris, G., Closser, M., . . . Wichterle, H. (2011). Embryonic stem cell-based mapping of developmental transcriptional programs. *Nat Methods, 8*(12), 1056-1058. doi:10.1038/nmeth.1775

Mazzoni, E. O., Mahony, S., Peljto, M., Patel, T., Thornton, S. R., McCuine, S., . . . Wichterle, H. (2013). Saltatory remodeling of Hox chromatin in response to rostrocaudal patterning signals. *Nat Neurosci, 16*(9), 1191-1198. doi:10.1038/nn.3490

McGinnis, W., & Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell, 68*(2), 283-302.

Merabet, S., & Mann, R. S. (2016). To Be Specific or Not: The Critical Relationship Between Hox And TALE Proteins. *Trends Genet, 32*(6), 334-347. doi:10.1016/j.tig.2016.03.004

Narendra, V., Bulajic, M., Dekker, J., Mazzoni, E. O., & Reinberg, D. (2016). CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Genes Dev, 30*(24), 2657-2662. doi:10.1101/gad.288324.116

Narendra, V., Rocha, P. P., An, D., Raviram, R., Skok, J. A., Mazzoni, E. O., & Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science, 347*(6225), 1017-1021. doi:10.1126/science.1262088

Noyes, M. B., Christensen, R. G., Wakabayashi, A., Stormo, G. D., Brodsky, M. H., & Wolfe, S. A. (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell, 133*(7), 1277-1289. doi:10.1016/j.cell.2008.05.023

Peljto, M., Dasen, J. S., Mazzoni, E. O., Jessell, T. M., & Wichterle, H. (2010). Functional diversity of ESC-derived motor neuron subtypes revealed through intraspinal transplantation. *Cell Stem Cell, 7*(3), 355-366. doi:10.1016/j.stem.2010.07.013

Peljto, M., & Wichterle, H. (2011). Programming embryonic stem cells to neuronal subtypes. *Curr Opin Neurobiol, 21*(1), 43-51. doi:10.1016/j.conb.2010.09.012

Porcelli, D., Fischer, B., Russell, S., & White, R. (2019). Chromatin accessibility plays a key role in selective targeting of Hox proteins. *Genome Biol, 20*(1), 115. doi:10.1186/s13059-019-1721-4

Regulski, M., Harding, K., Kostriken, R., Karch, F., Levine, M., & McGinnis, W. (1985). Homeo box genes of the Antennapedia and bithorax complexes of Drosophila. *Cell, 43*(1), 71-80.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140. doi:10.1093/bioinformatics/btp616

Rousso, D. L., Gaber, Z. B., Wellik, D., Morrisey, E. E., & Novitch, B. G. (2008). Coordinated actions of the forkhead protein Foxp1 and Hox proteins in the columnar organization of spinal motor neurons. *Neuron, 59*(2), 226-240. doi:10.1016/j.neuron.2008.06.025

Slattery, M., Ma, L., Negre, N., White, K. P., & Mann, R. S. (2011). Genome-wide tissue-specific occupancy of the Hox protein Ultrabithorax and Hox cofactor Homothorax in Drosophila. *PLoS One, 6*(4), e14686. doi:10.1371/journal.pone.0014686

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., . . . Mann, R. S. (2011). Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell, 147*(6), 1270-1282. doi:10.1016/j.cell.2011.10.053

Sweeney, L. B., Bikoff, J. B., Gabitto, M. I., Brenner-Morton, S., Baek, M., Yang, J. H., . . . Jessell, T. M. (2018). Origin and Segmental Diversity of Spinal Inhibitory Interneurons. *Neuron, 97*(2), 341-355 e343. doi:10.1016/j.neuron.2017.12.029

Tan, G. C., Mazzoni, E. O., & Wichterle, H. (2016). Iterative Role of Notch Signaling in Spinal Motor Neuron Diversification. *Cell Rep, 16*(4), 907-916. doi:10.1016/j.celrep.2016.06.067

Tsuchida, T., Ensini, M., Morton, S. B., Baldassare, M., Edlund, T., Jessell, T. M., & Pfaff, S. L. (1994). Topographic organization of embryonic motor neurons defined by expression of LIM homeobox genes. *Cell, 79*(6), 957-970.

Welch, R. P., Lee, C., Imbriano, P. M., Patil, S., Weymouth, T. E., Smith, R. A., . . . Sartor, M. A. (2014). ChIP-Enrich: gene set enrichment testing for ChIP-seq data. *Nucleic Acids Res, 42*(13), e105. doi:10.1093/nar/gku463

Wichterle, H., Lieberam, I., Porter, J. A., & Jessell, T. M. (2002). Directed differentiation of embryonic stem cells into motor neurons. *Cell, 110*(3), 385-397.

Wichterle, H., & Peljto, M. (2008). Differentiation of mouse embryonic stem cells to spinal motor neurons. *Curr Protoc Stem Cell Biol, Chapter 1*, Unit 1H 1 1-1H 1 9. doi:10.1002/9780470151808.sc01h01s5

Young, T., Rowland, J. E., van de Ven, C., Bialecka, M., Novoa, A., Carapuco, M., . . . Deschamps, J. (2009). Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos. *Dev Cell, 17*(4), 516-526. doi:10.1016/j.devcel.2009.08.010

Zeiske, T., Baburajendran, N., Kaczynska, A., Brasch, J., Palmer, A. G., 3rd, Shapiro, L., . . . Mann, R. S. (2018). Intrinsic DNA Shape Accounts for Affinity Differences between Hox-Cofactor Binding Sites. *Cell Rep, 24*(9), 2221-2230. doi:10.1016/j.celrep.2018.07.100

FIGURE LEGENDS

Figure 1: Hox TFs control cell fates during in vitro spinal cord differentiation.

(A) Experimental scheme. (B) Diagram of neuronal cell fate control by *Hox* gene overexpression. (C) Principal Component Analysis (PCA) of the RNA-seq datasets reveals similarities in the gene expression profiles induced by Hox TFs. (D) RNA–seq heatmap showing the expression of representative marker genes in No Dox control, iHoxc6, iHoxc8, iHoxc9 and iHoxc10 relative to Day2 progenitors.

Figure 2: Central and posterior Hox TFs have distinct sequence motif preferences during MN differentiation.

(A-D) ChIP-seq heatmap of the top 10,000 sites of the indicated Hox TF plotted within a 1kb window around the peak center. Selected enriched motifs discovered via MEME-ChIP indicated on the right.

Figure 3: Hoxc6, Hoxc9 and Hoxc10 TFs have different genome-wide binding profiles.

(A-D) ChIP-seq heatmap showing binding comparisons of Hox TFs in differentiating MNs, at Day 3. Sites bound by both indicated Hox TFs noted as “=” sites. Preferentially bound sites by Hoxc6, Hoxc8, Hoxc9 or Hoxc10 noted as “c6 >”, “c8 >”, “c9 >” and “c10 >”, respectively. (E) ChIP-seq heatmap showing binding comparisons of Hoxc6, Hoxc9 and Hoxc10 in differentiating MNs, at Day 3. Sites bound by all three Hox TFs noted as “c6 = c9 = c10” sites. Preferentially bound sites by Hoxc6, Hoxc9, Hoxc6 and Hoxc9 or Hoxc9 and Hoxc10 noted as “c6 > c9, c10”, “c9 > c6, c10”, “c6, c9 > c10” and “c9, c10 > c6”, respectively. (F) Selected enriched motifs discovered via MEME-ChIP at the indicated Hox binding categories.

Figure 4: Differentially expressed gene sets correlate with differential Hox binding events.

(A) Heatmap representing associations between Hox binding categories and the indicated gene sets. Upregulated/Downregulated in iHoxc6 MNs, iHoxc9 MNs and iHoxc10 is relative to Day 2 progenitors (prior to Hox induction).

Figure 5: Hoxc9 has a higher preference for inaccessible chromatin than Hoxc6 and Hoxc10.

(A) Experimental scheme. (B) The distribution of MN progenitor ATAC-seq read density at the top 10,000 Hoxc6, Hoxc9 and Hoxc10 sites at Day 3. Data is ordered based on normalized read density (tags per million per site) and divided into quartiles. The box displays the central 50% (quartile 2 and quartile 3) of the data, while the top and bottom whiskers represent the top 25% and bottom 25% (the top and the bottom quartiles) of the data, respectively. (C) ATAC-seq heatmap displaying the accessibility in MN progenitors (before Hox induction) at the indicated Hox binding categories. (D) Metagene plots of accessibility (ATAC-seq reads) in MN progenitors displaying the prior accessibility (before Hox induction) at the indicated binding category. Normalized read density represents tags per million per 1000 sites. (E) Experimental scheme for F. (F) The distribution of EB ATAC-seq read density at the top 10,000 Hoxc6 and Hoxc9 sites in EBs (12h Dox induction). Data is ordered based on normalized read density (tags per million per site) and divided into quartiles. The box displays the central 50% (quartile 2 and quartile 3) of the data, while the top and bottom whiskers represent the top 25% and bottom 25% (the top and the bottom quartiles) of the data, respectively.

Figure 6: Hox TF binding increases the local chromatin accessibility.

(A) Experimental scheme. (B) ATAC-seq heatmaps displaying the gain of accessibility in iHoxc6 MNs versus iHoxc9 MNs versus iHoxc10 MNs at the indicated binding categories. (C) Metagene plots of accessibility (ATAC-seq reads) showing the accessibility gain in iHoxc6 MNs versus

iHoxc9 MNs versus iHoxc10 MNs at the indicated binding category. Normalized read density represents tags per million per 1000 sites.

Figure 7: Hoxc9 binds to a larger fraction of sites in inaccessible chromatin than Hoxa9 and Hoxd9.

(A-B) ChIP-seq heatmap of the top 10,000 sites of the indicated Hox TF plotted within a 1kb window around the peak center. Selected enriched motifs discovered via MEME-ChIP indicated on the right. (C-D) ChIP-seq heatmap showing binding comparisons of the indicated Hox TFs in differentiating MNs, at Day 3. Sites bound by both indicated Hox TFs noted as "=" sites. Preferentially bound sites by Hoxa9, Hoxd9 or Hoxc9 noted as "a9 >", "d9 >" or "c9 >". (E) The distribution of MN progenitor ATAC-seq read density at the top 10,000 Hoxc9, Hoxa9 and Hoxd9 sites at Day 3. Data is ordered based on normalized read density (tags per million per site) and divided into quartiles. The box displays the central 50% (quartile 2 and quartile 3) of the data, while the top and bottom whiskers represent the top 25% and bottom 25% (the top and the bottom quartiles) of the data, respectively. (F) ATAC-seq heatmap displaying the accessibility in MN progenitors (before Hox induction) at the indicated Hox binding categories. (G) Metagene plots of accessibility (ATAC-seq reads) in MN progenitors displaying the prior accessibility (before Hox induction) at the indicated binding categories. Normalized read density represents tags per million per 1000 sites.

Figure 8: Hoxc9 and Hoxc13 have a higher preference for inaccessible chromatin than Hoxc10.

(A) ChIP-seq heatmap of the top 10,000 Hoxc13 sites plotted within a 1kb window around the peak center. Selected enriched motifs discovered via MEME-ChIP indicated on the right. (B) ChIP-seq heatmap showing binding comparisons of Hoxc9, Hoxc10 and Hoxc13 in differentiating MNs, at Day 3. Sites bound by all three Hox TFs noted as "c9 = c10 = c13" sites.

Preferentially bound sites by Hoxc9, Hoxc13, Hoxc9 and Hoxc10 or Hoxc9 and Hoxc13 noted as “c9 > c10, c13”, “c13 > c9, c10”, “c9, c10 > c13” and “c9, c13 > c10”, respectively. (C) The distribution of MN progenitor ATAC-seq read density at the top 10,000 Hoxc9, Hoxc10 and Hoxc13 sites at Day 3. Data is ordered based on normalized read density (tags per million per site) and divided into quartiles. The box displays the central 50% (quartile 2 and quartile 3) of the data, while the top and bottom whiskers represent the top 25% and bottom 25% (the top and the bottom quartiles) of the data, respectively. (D) ATAC-seq heatmap displaying the accessibility in MN progenitors (before Hox induction) at the indicated Hox binding categories. (E) Metagene plots of accessibility (ATAC-seq reads) in MN progenitors displaying the prior accessibility (before Hox induction) at the indicated binding categories. Normalized read density represents tags per million per 1000 sites.

Figure 9: Posterior Hox TFs have distinct genomic binding patterns.

(A) Principal Component Analysis (PCA) of the ChIP-seq datasets reveals similarities in the binding patterns of Hox TFs. (B) Browser screenshots of the indicated Hox ChIP-seqs, Day 2 ATAC-seq and H3K27me3 ChIP-seq at the *HoxC* gene cluster.

Supp Figure 1: Hox TFs control cell fates during in vitro spinal cord differentiation.

(A) Experimental scheme. (mESCs: mouse embryonic stem cells. EBs: embryoid bodies) (B) Multidimensional scaling (MDS) of the RNA-seq datasets reveals similarities in the gene expression profiles induced by Hox TFs. (C) RNA–seq heatmap showing the expression of representative genes in iHoxc6, iHoxc8, iHoxc9 and iHoxc10 relative to No Dox control (Day 4).

Supp Figure 2: Hoxc6, Hoxc9 and Hoxc10 TFs have largely the same binding patterns at Day 3 vs Day 4.

(A-C) ChIP-seq heatmaps showing binding comparisons of the indicated Hox TFs at Day 3 versus Day 4. Sites equally bound at both time points noted as “Day 3 = Day 4” sites. Differentially bound sites noted as “Day 3 > Day 4” and “Day 4 > Day 3”.

Supp Figure 3: Hoxc6, Hoxc9 and Hoxc10 TFs have different genome-wide binding profiles at Day 4.

(A) Schematic indicating the percentage of conserved amino acids in the homeodomain (HD) and total protein between the indicated Hox TFs. (B-D) ChIP-seq heatmap showing binding comparisons in differentiating MNs, at Day 4. Sites bound by both indicated Hox TFs noted as “=” sites. Preferentially bound sites by Hoxc6, Hoxc9 or Hoxc10 noted as “c6 >”, “c9 >” and “c10 >”, respectively. (E) ChIP-seq heatmap showing binding comparisons of Hoxc6, Hoxc9 and Hoxc10 in differentiating MNs, at Day 4. Sites bound by all three Hox TFs noted as “c6 = c9 = c10” sites. Preferentially bound sites by Hoxc6, Hoxc9, Hoxc6 and Hoxc9 or Hoxc9 and Hoxc10 noted as “c6 > c9, c10”, “c9 > c6, c10”, “c6, c9 > c10” and “c9, c10 > c6”, respectively.

Supp Figure 4: Hoxc9 and Hoxc13 genome-wide binding profiles differ from other posterior Hox TFs.

(A) ChIP-seq heatmap showing binding comparisons of Hoxc9 and Hoxc13 in differentiating MNs, at Day 3. Sites bound by both indicated Hox TFs noted as “c9 = c13” sites. Preferentially bound sites by Hoxc9 or Hoxc13 noted as “c9 > c13” or “c13 > c9“. (B-C) Principal Component Analysis (PCA) of the ChIP-seq datasets reveals similarities in the binding patterns of Hox TFs.

Supp Figure 5: Meis TFs co-bind with Hoxc6, Hoxc9 and Hoxc10

(A) ChIP-seq of Meis performed in the indicated cell line reveals co-binding with the respective Hox TF at the Day 3 binding categories.

Figure 1: Hox TFs control cell fates during in vitro spinal cord differentiation.

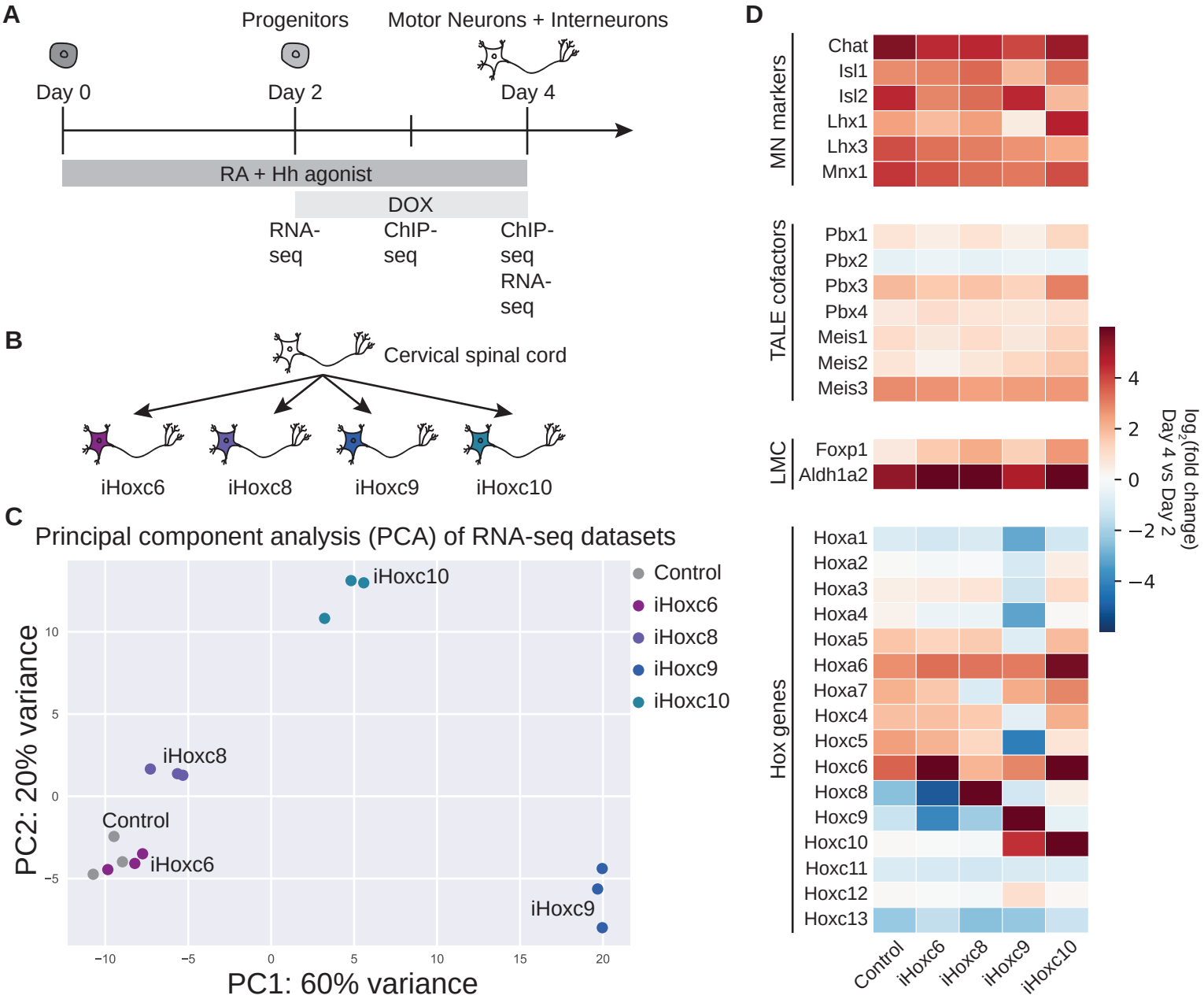


Figure 2: Central and posterior Hox TFs have distinct sequence motif preferences during MN differentiation.

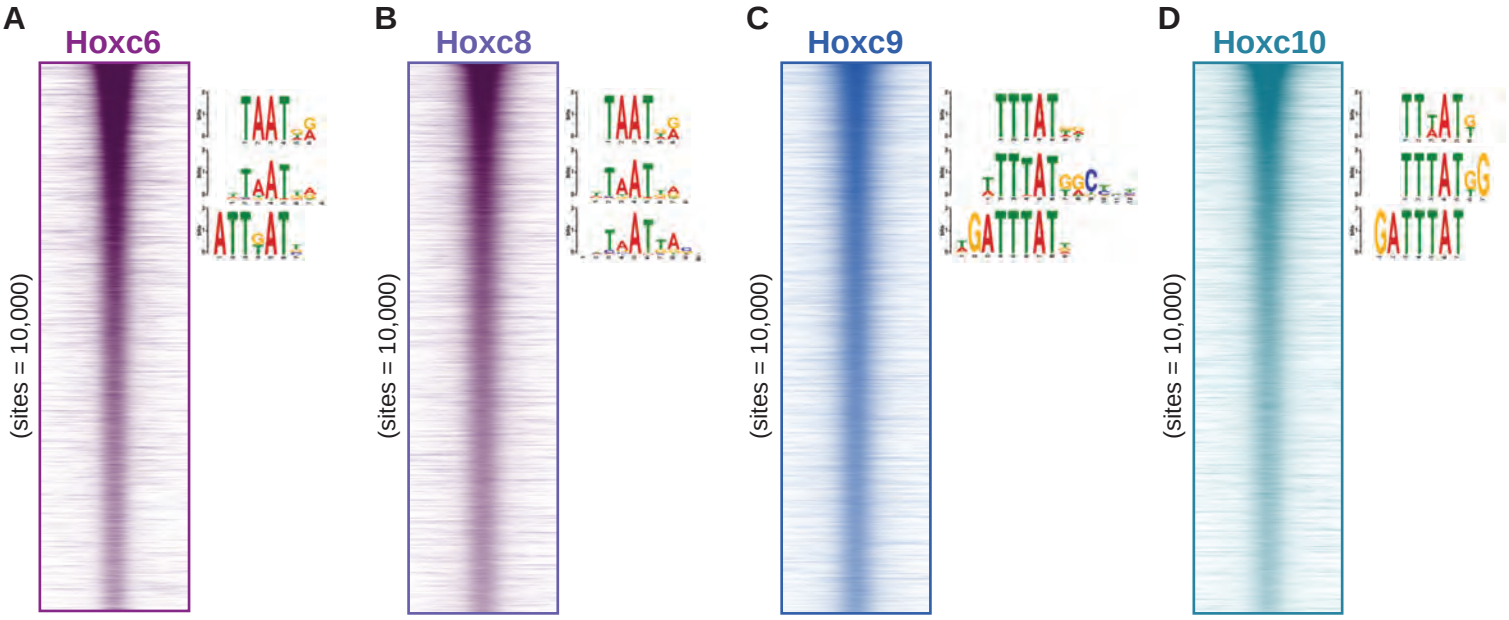


Figure 3: Hoxc6, Hoxc9 and Hoxc10 TFs have different genome-wide binding profiles.

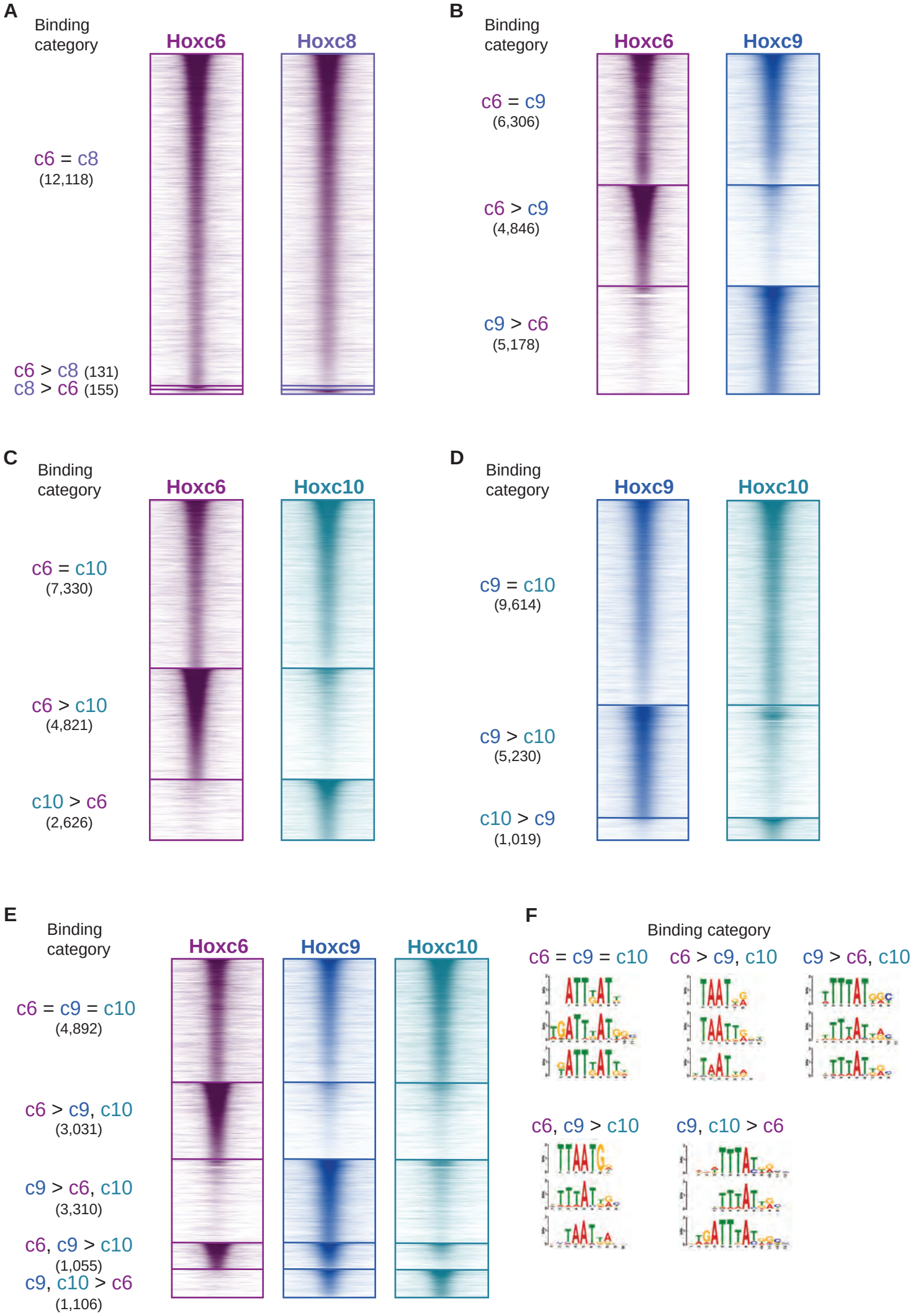


Figure 4: Differentially expressed gene sets correlate with differential Hox binding events.

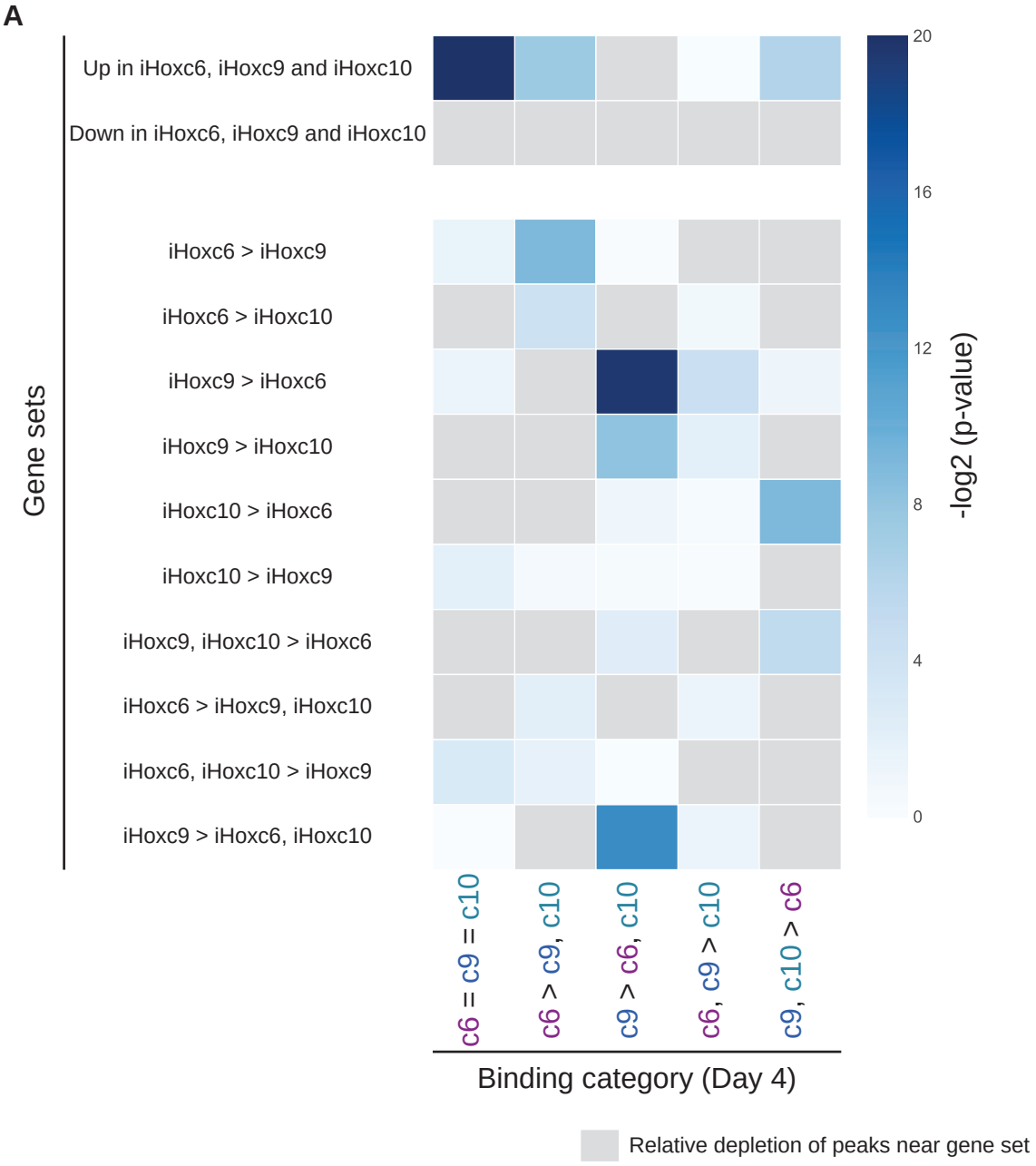


Figure 5: Hoxc9 has a higher preference for inaccessible chromatin than Hoxc6 and Hoxc10.

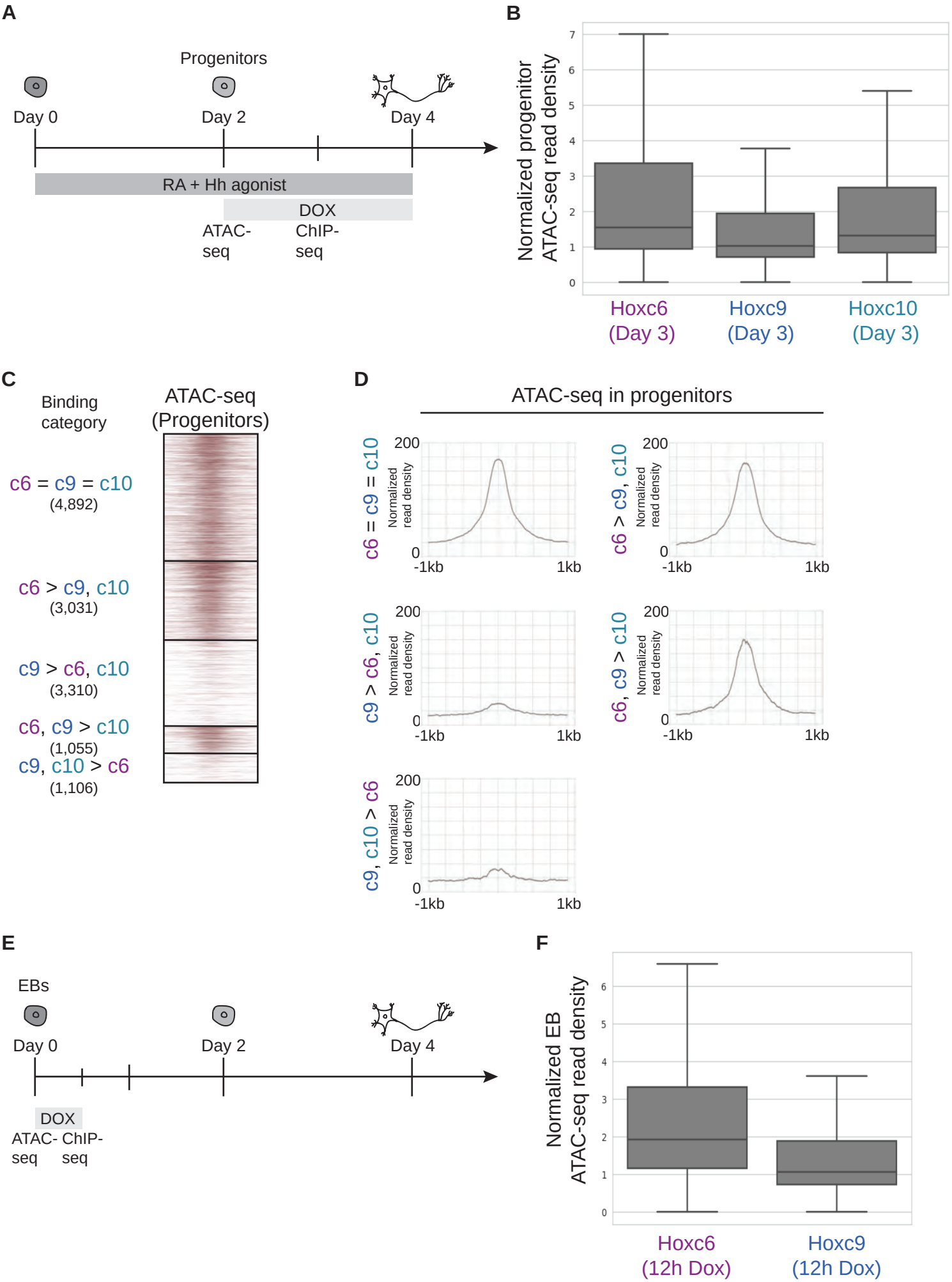


Figure 6: Hox TF binding increases the local chromatin accessibility.

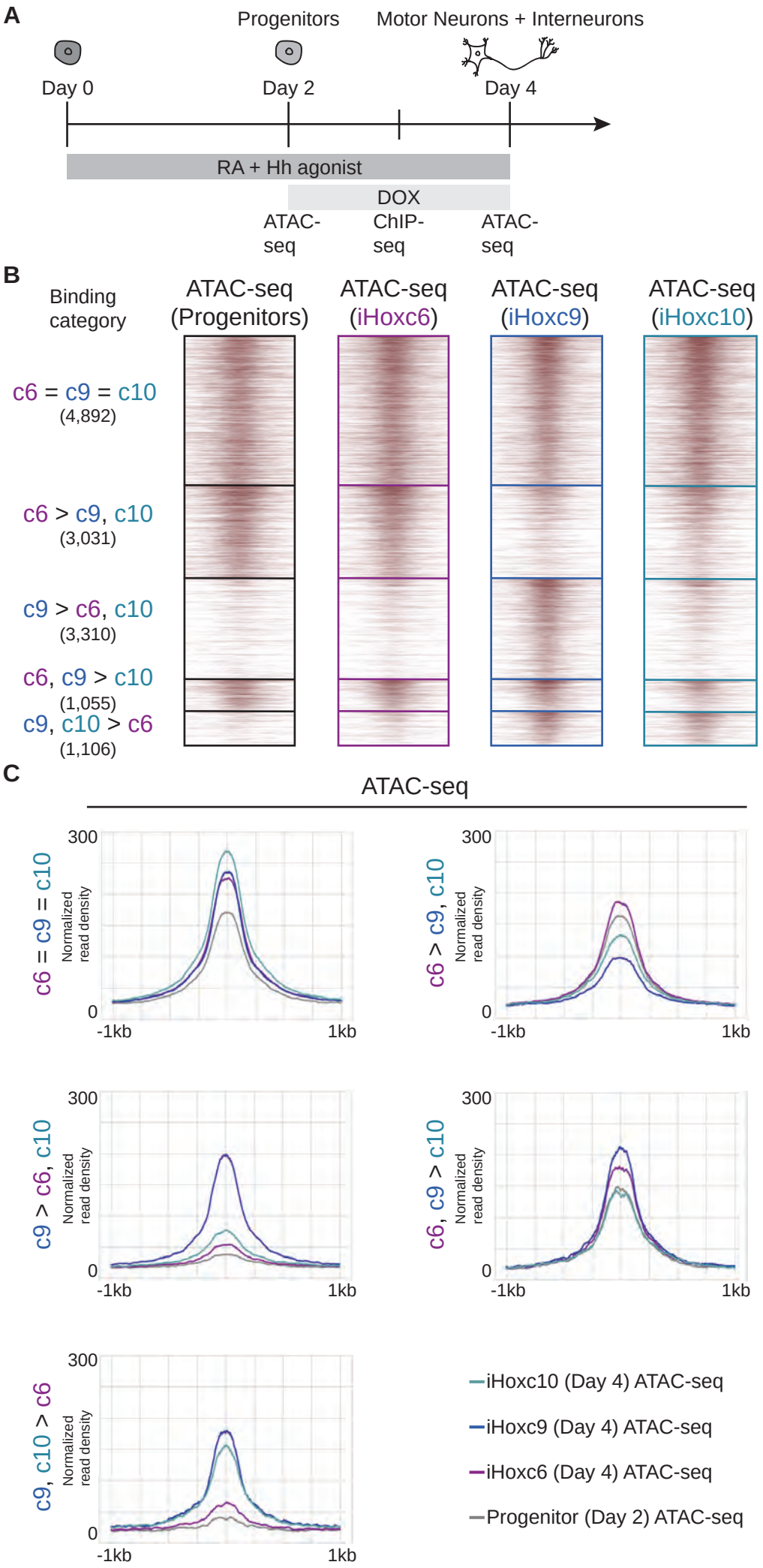


Figure 7: Hoxc9 binds to a larger fraction of sites in inaccessible chromatin than Hoxa9 and Hoxd9.

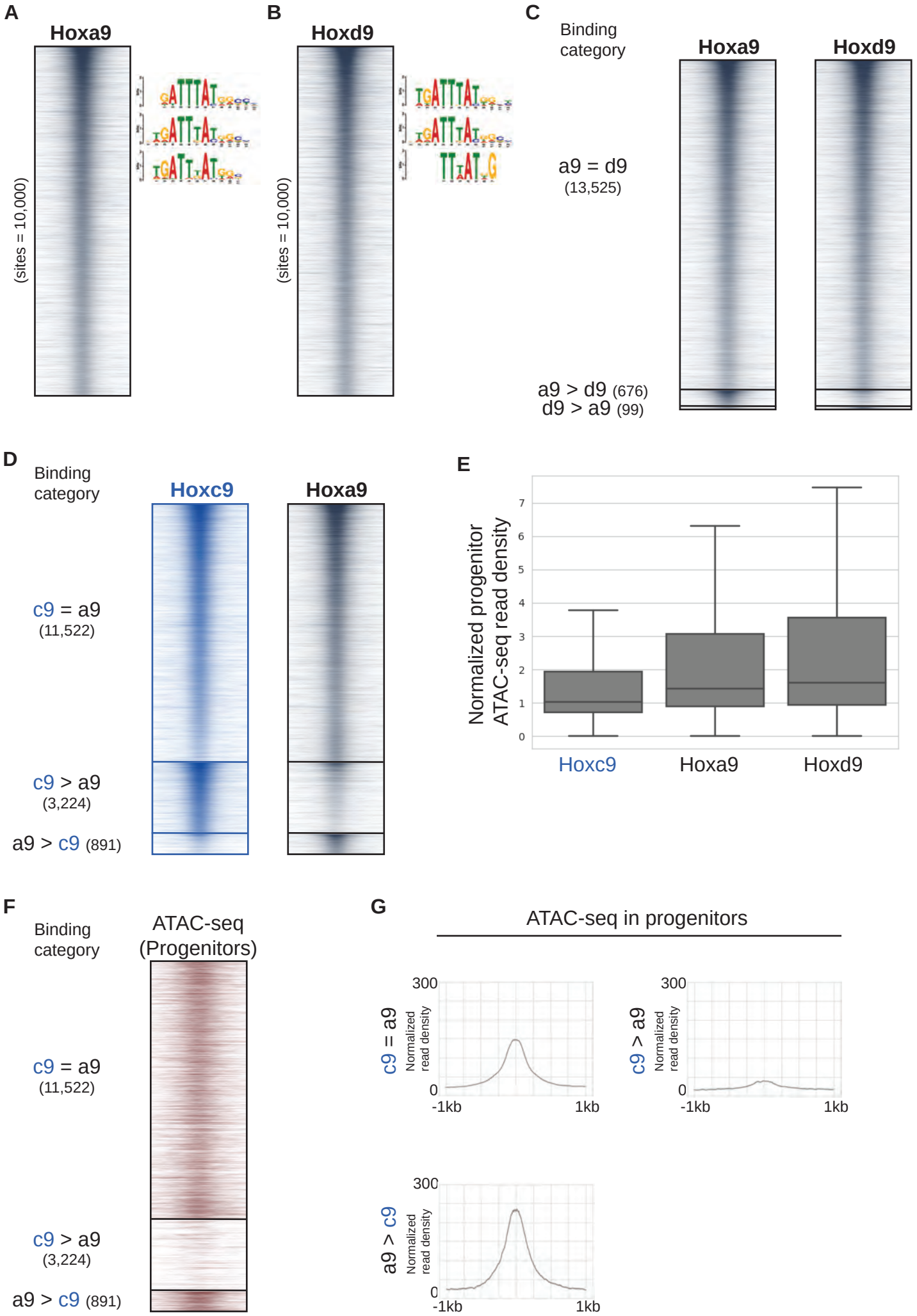


Figure 8: Hoxc9 and Hoxc13 have a higher preference for inaccessible chromatin than Hoxc10.

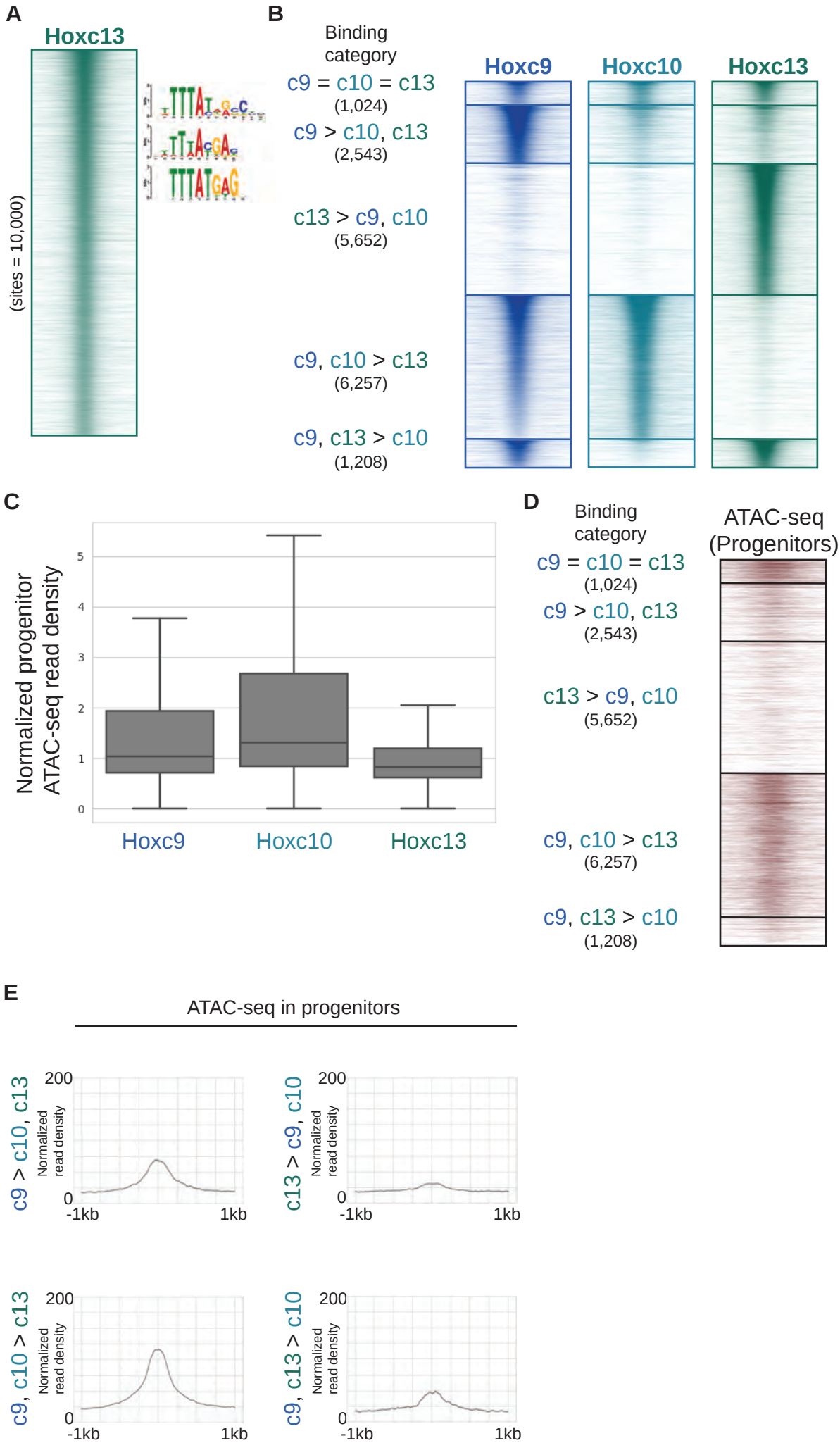
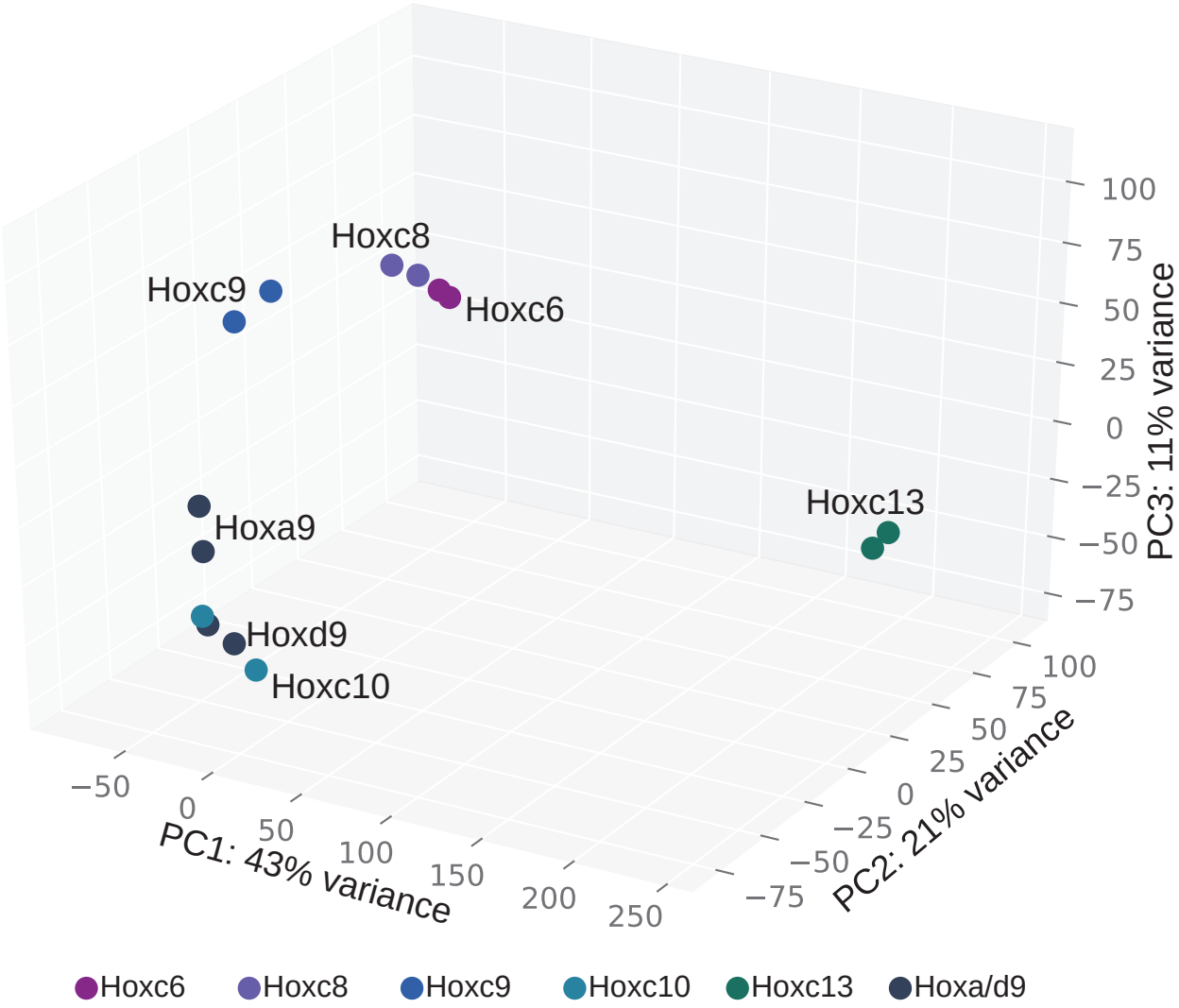


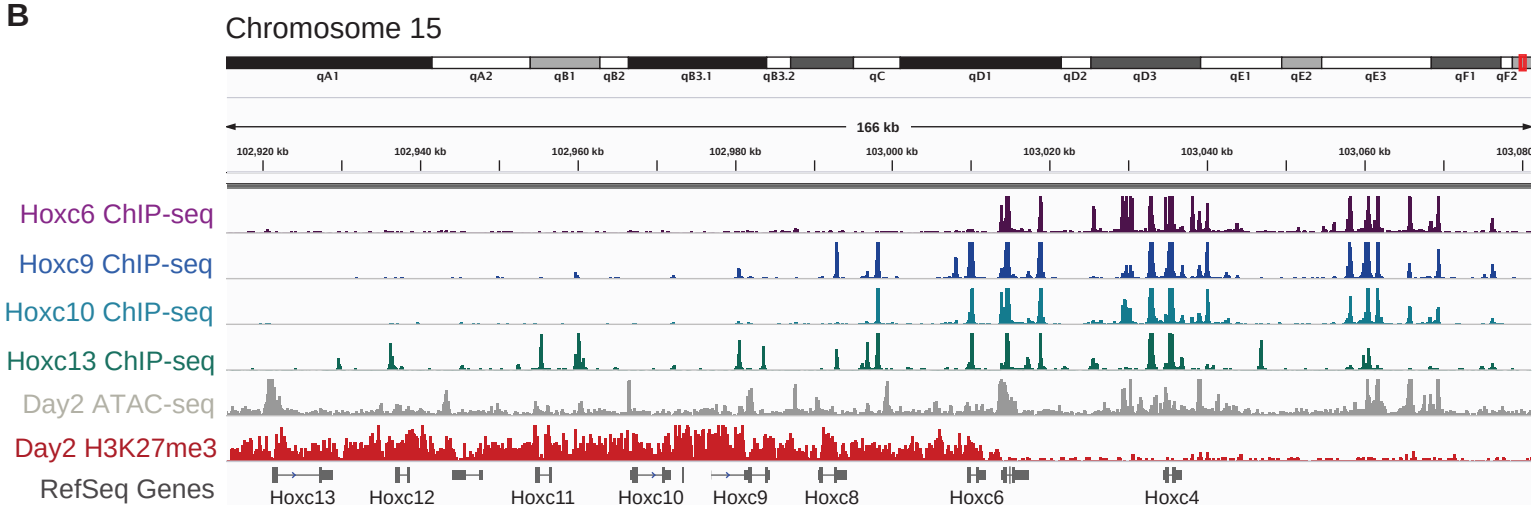
Figure 9: Posterior Hox TFs have distinct genomic binding patterns.

A

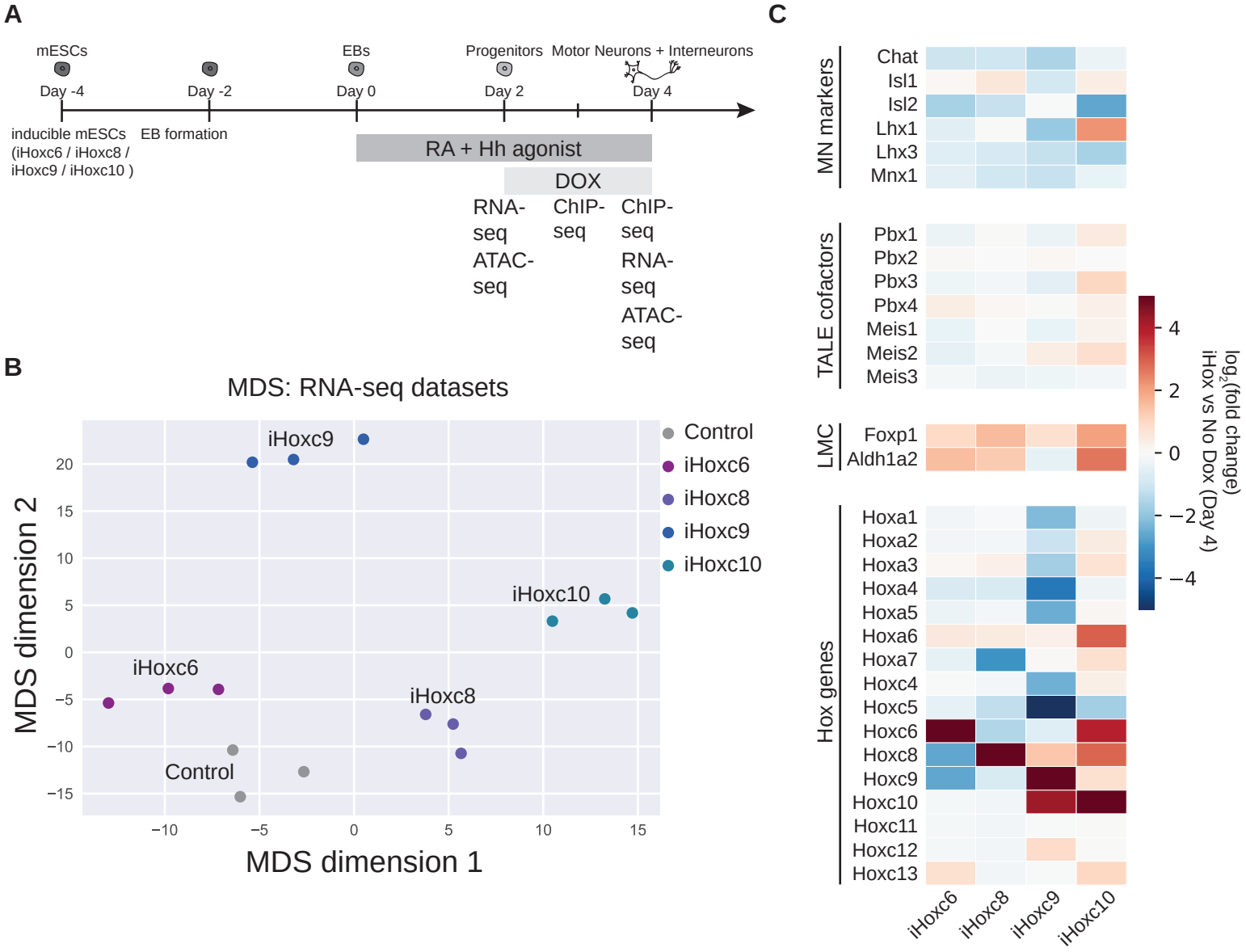
Principal component analysis (PCA) of ChIP-seq datasets



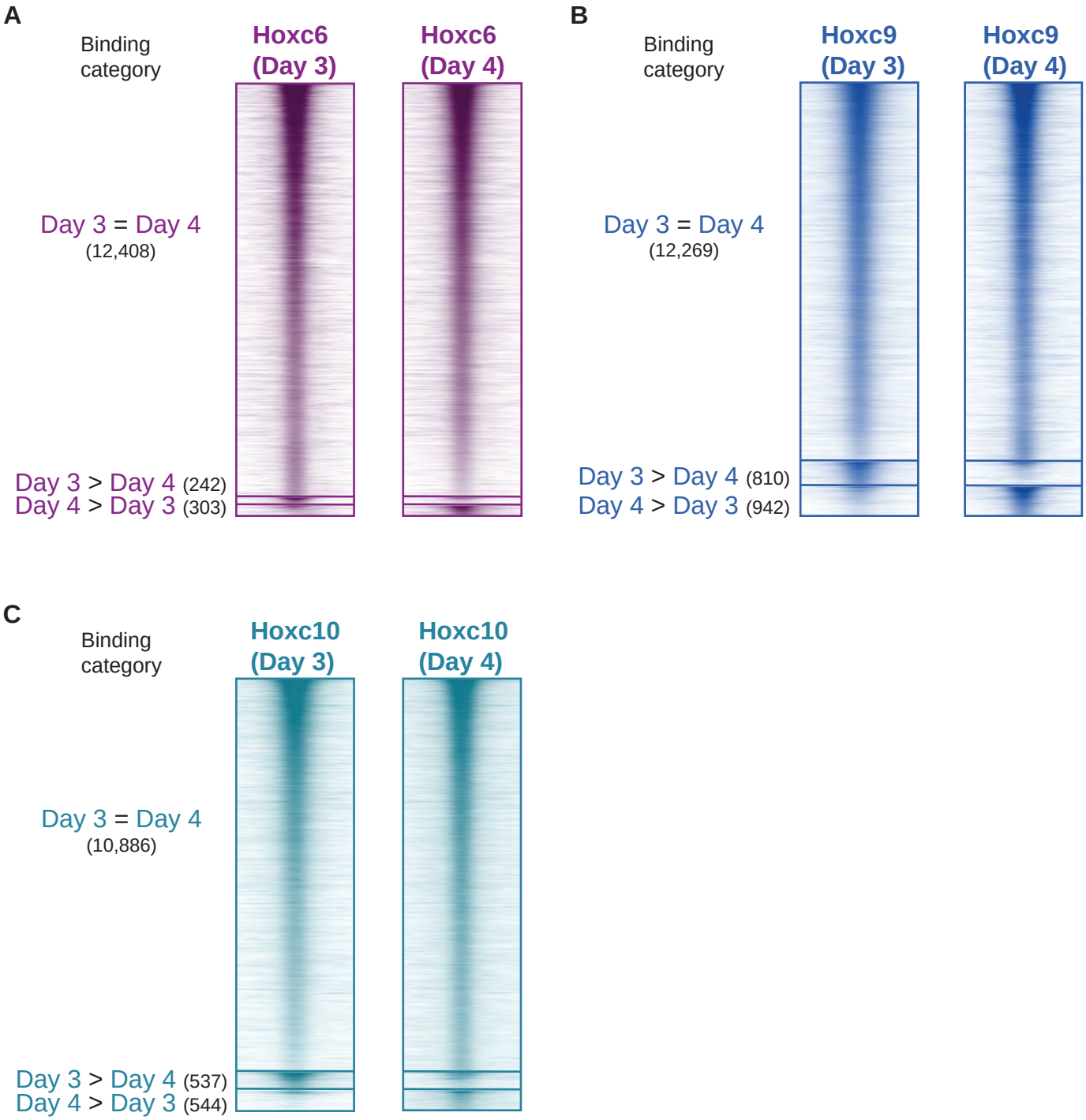
B



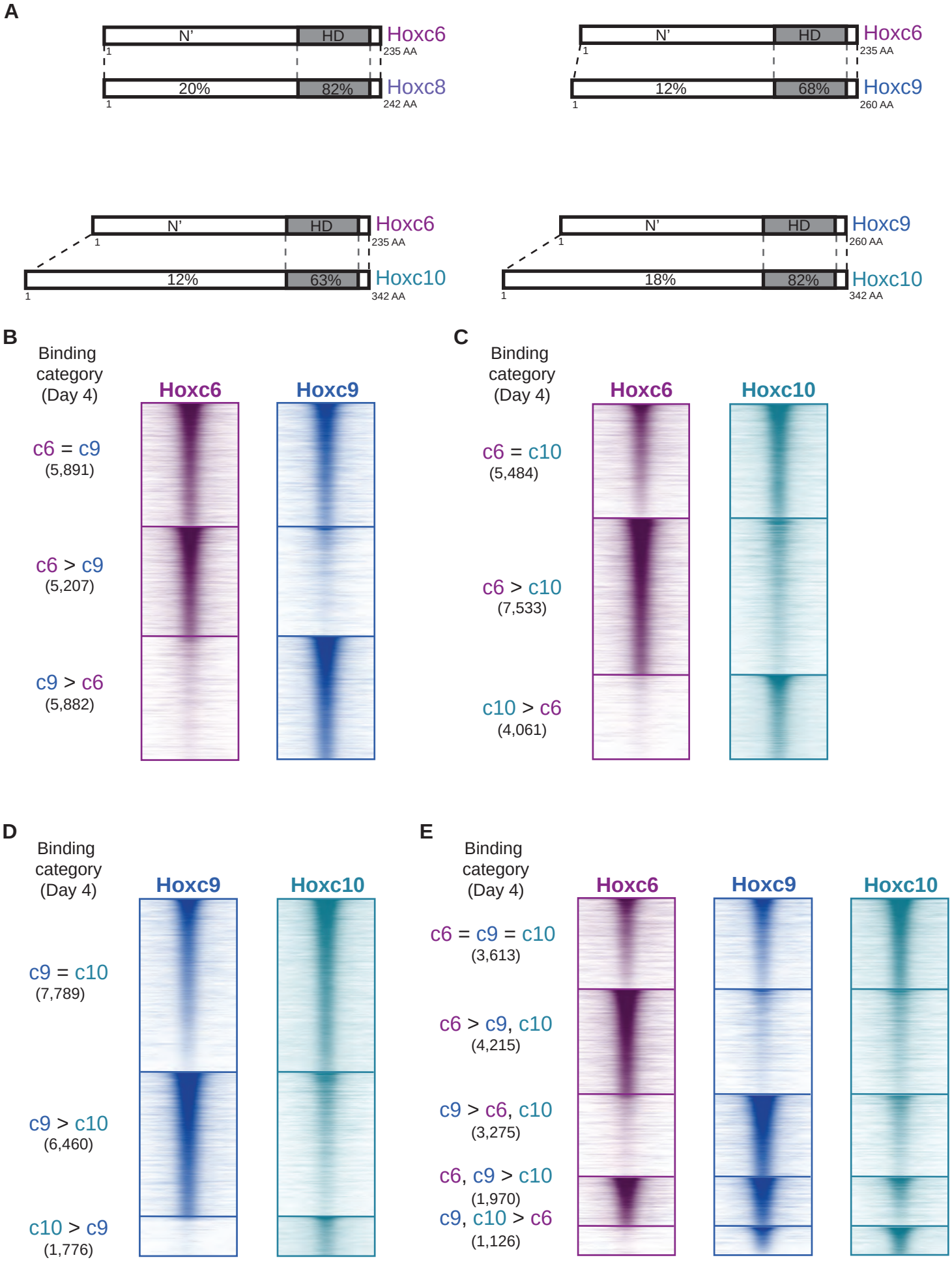
Supp Figure 1: Hox TFs control cell fates during in vitro spinal cord differentiation.



Supp Figure 2: Hoxc6, Hoxc9 and Hoxc10 TFs have largely the same binding patterns at Day 3 vs Day 4.

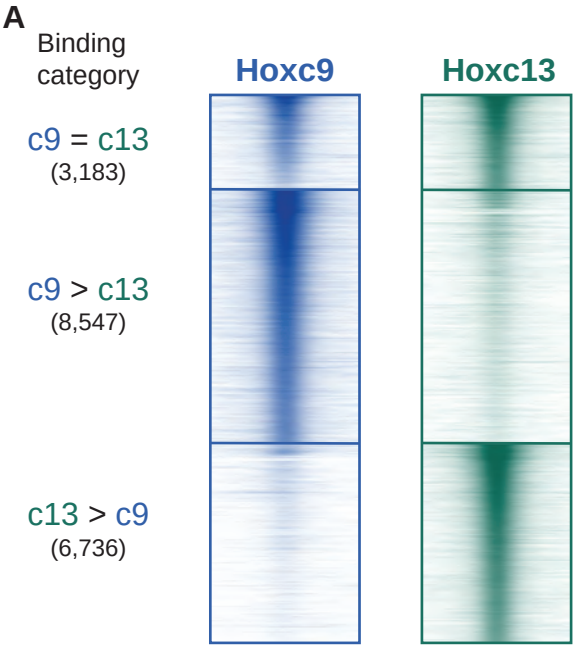


Supp Figure 3: Hoxc6, Hoxc9 and Hoxc10 TFs have different genome-wide binding profiles at Day4.

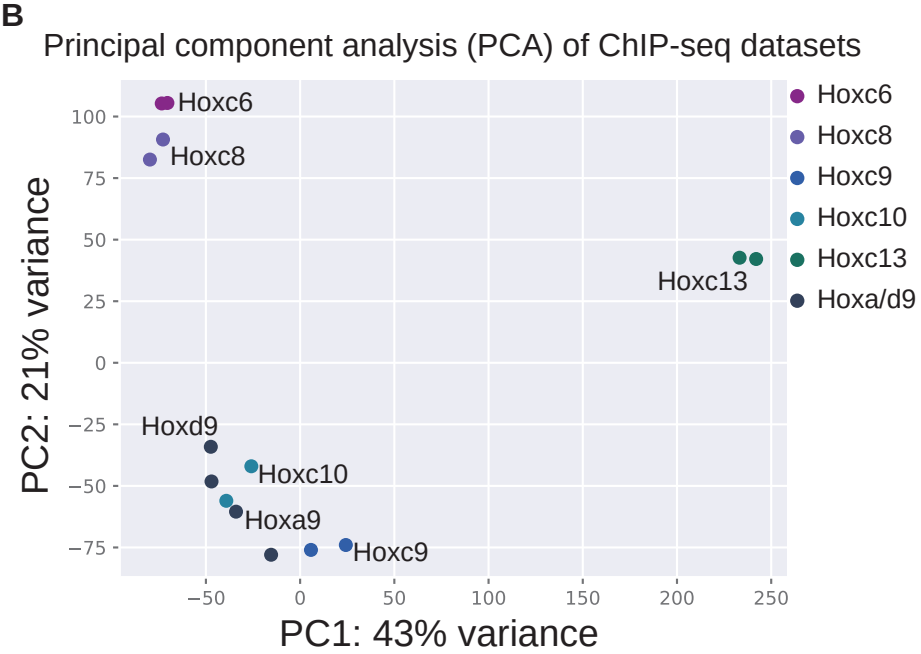


Supp Figure 4: Hoxc9 and Hoxc13 genome-wide binding profiles differ from other posterior Hox TFs.

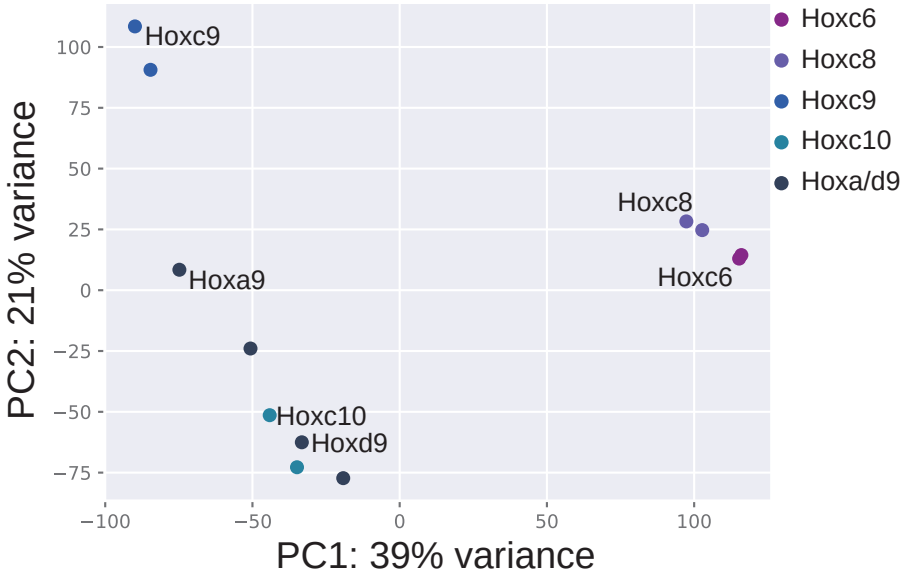
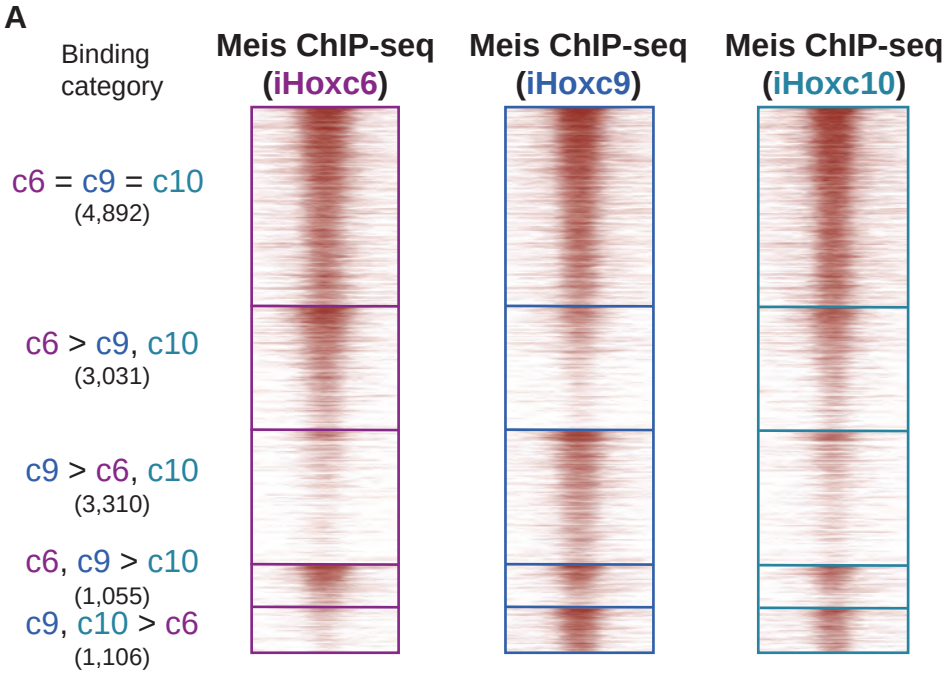
A



B



C



Supp Figure 5: Meis TFs co-bind with Hoxc6, Hoxc9 and Hoxc10.