communications biology

Article

https://doi.org/10.1038/s42003-024-06835-w

# Multimodal Hox5 activity generates motor neuron diversity

Check for updates

Ritesh KC[1], Raquel López de Boer[1], Minshan Lin[1], Alicia N. Vagnozzi[1], Lucie Jeannotte[2] & Polyxeni Philippidou[1]

Motor neurons (MNs) are the final output of circuits driving fundamental behaviors, such as respiration and locomotion. Hox proteins are essential in generating the MN diversity required for accomplishing these functions, but the transcriptional mechanisms that enable Hox paralogs to assign distinct MN subtype identities despite their promiscuous DNA binding motif are not well understood. Here we show that Hoxa5 modifies chromatin accessibility in all mouse spinal cervical MN subtypes and engages TALE co-factors to directly bind and regulate subtype-specific genes. We identify a paralog-specific interaction of Hoxa5 with the phrenic MN-specific transcription factor Scip and show that heterologous expression of Hoxa5 and Scip is sufficient to suppress limb-innervating MN identity. We also demonstrate that phrenic MN identity is stable after Hoxa5 downregulation and identify Klf proteins as potential regulators of phrenic MN maintenance. Our data identify multiple modes of Hoxa5 action that converge to induce and maintain MN identity.

The motor programs that mediate essential behaviors such as respiration and locomotion rely on the establishment of distinct subtypes of motor neurons (MNs) during development. MN diversity arises from the intersection of dorsoventral and rostrocaudal signaling pathways that drive the combinatorial expression of unique sets of transcription factors (TFs) that specify MN subtype identities along the spinal cord[1,2]. Along the rostrocaudal axis, members of the chromosomally-clustered *Hox* gene family are critical in specifying the identity of segmentally-restricted MN subtypes[3]. Despite the well-described functions of Hox proteins in MN specification, several questions remain regarding the mechanisms that different Hox paralogs employ to induce distinct subtype identities at the transcriptional level and how Hox protein divergent and convergent functions are mediated[4,5]. For example, while several Hox proteins have been shown to converge on common transcriptional targets to redundantly promote limb-innervating Lateral Motor Column (LMC) identity[6], it is less clear how a single Hox paralog may promote multiple MN subtype identities.

Hox proteins bind DNA through their homeodomain, a 60 amino acid domain that recognizes a short TA-rich DNA motif. Homeobox domains are highly similar among different Hox proteins and do not appear to confer DNA-binding selectivity to individual paralogs[7–10]. This contrasts with the unique functions of Hox proteins in vivo, which implies a stringent selectivity of gene targets, giving rise to the Hox-specificity paradox[11,12]. How do Hox proteins achieve their unique functions given the apparent overlap in their DNA-binding motifs? One partial solution to this paradox arises from the cooperative binding of Hox proteins to DNA with a family of cofactors, known as the three aminoacid loop extension (TALE) homeodomain proteins[13]. Pbx proteins, members of the TALE family of TFs, are essential mediators of Hox function in MNs and mutations in Pbx genes recapitulate Hox mutant phenotypes[14]. While Hox/Pbx interactions increase the specificity of the DNA-binding site, it is unlikely that this interaction alone accounts for all the unique functions of individual Hox paralogs, as multiple Hox proteins are able to interact with Pbx proteins, pointing to the existence of additional mechanisms that further contribute to Hox specificity[15].

At cervical levels of the spinal cord, Hox5 paralogs have the ability to promote both Phrenic Motor Column (PMC) and LMC identity[6,16]. Mice lacking *Hox5* genes in MNs die at birth from respiratory failure, largely due to progressive loss and disorganization of phrenic MNs, and a dramatic loss in axon branching and synaptic contacts at the diaphragm[16]. Effects on limb-innervating MNs are subtler, as *Hox5* mutant mice show grossly normal patterns of limb innervation, with only a subset of motor pools adopting abnormal trajectories and targeting inappropriate muscles[17]. The transcriptional mechanisms that underlie the ability of a single Hox TF to induce two opposing MN identities are not well understood. Hox5 proteins are the only Hox paralogs that induce PMC-specific genes in vivo, while the ability to induce genes expressed in LMC neurons is common with other Hox family members (Hox4-8). How do Hox5 proteins accomplish both unique and shared functions in MNs? One possibility is that this distinction arises through different DNA-binding motifs which are highly Hox5-specific in

[1]Department of Neurosciences, Case Western Reserve University, Cleveland, OH, USA. [2]Department of Molecular Biology, Medical Biochemistry & Pathology, Université Laval, Centre Recherche sur le Cancer de l'Université Laval, Centre de recherche du CHU de Québec-Université Laval (Oncology), Québec, Canada. e-mail: pxp282@case.edu

PMC genes but common for multiple Hox proteins in LMC genes. An example of this can be seen in *Drosophila*, where the Hox5 homolog Sex combs reduced (Scr), the only Hox protein that can initiate salivary gland development, can bind cooperatively with the Pbx homolog Extradenticle (Exd) to a unique sequence that other Hox/Pbx complexes are unable to bind[18]. Do Hox5 proteins act in a similar manner in phrenic MNs to bind Hox5/Pbx specific sites? While this mechanism of action might account for the unique ability of Hox5 proteins to induce PMC-specific genes, it would fail to explain how these genes are restricted specifically to PMC neurons given the co-expression of Hox5 and Pbx proteins in other MN populations in the cervical spinal cord. An alternative hypothesis is that additional DNA-binding proteins contribute to the selection of specific targets, either by forming a complex with Hox5/Pbx and altering the preference for a binding site, or by differentially recruiting activators or repressors to the transcriptional complex. In addition to Hox5 and Pbx proteins, PMC neurons also express the POU-domain TF Scip (Pou3f1, Oct6)[16,19,20] while LMC neurons express the TF FoxP1, which is required for the induction of Hox-dependent LMC-specific genes[21,22]. Therefore, one possibility is that, depending on the presence of either Scip or FoxP1, Hox5/Pbx/Scip and Hox5/Pbx/FoxP1 complexes activate two non-overlapping sets of targets, required for PMC and LMC specification respectively.

In addition to their canonical functions as TFs, Hox paralog activities can also diverge based on their differential ability to open chromatin, a characteristic property of pioneer factors[23–25]. For example, Hox13 pioneer activity is essential for initiating developmental programs required for the generation of limb digits and external genitalia in mammals[26,27]. During in vitro MN specification, Hox TFs exhibit differential abilities to bind and open inaccessible chromatin[28]. Hox5 proteins may partly act by promoting the opening of chromatin that is actively-transcribed in specific MN columns. The ability of Hox proteins to alter chromatin state might also contribute to the stable maintenance of subtype-specific MN identity after the downregulation of Hox proteins at postnatal stages.

Here, we utilize RNA-sequencing (RNA-seq) and Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) from isolated mouse embryonic MNs to show that Hox5 TFs can modify chromatin associated with all three major columns of MNs in the cervical spinal cord and engage TALE co-factors to directly bind and regulate subtype-specific genes. We identify a paralog-specific interaction of Hoxa5 with Scip and show that heterologous expression of Hoxa5 and Scip is sufficient to suppress alternative MN identities. We also demonstrate that phrenic MN identity is stable after Hox5 downregulation and identify Klf TFs as potential downstream regulators of phrenic MN maintenance. Our data identify multiple modes of Hox5 action that converge to induce and maintain MN identity.

## Results

### Gene expression and chromatin accessibility profiles in cervical MN subtypes

Spinal MNs are generated from a highly restricted common progenitor domain in the ventral neural tube. As MNs begin to differentiate and exit the cell cycle, they are topographically organized in a stereotypical fashion as discrete motor columns which exhibit distinct transcriptional profiles and subtype-specific molecular markers by embryonic day (e)12.5. MN subtypes are differentially distributed along the rostrocaudal axis of the spinal cord. Cervical levels of the spinal cord contain MNs that can be divided into three major subtypes: Phrenic Motor Column (PMC) neurons which innervate the diaphragm to regulate breathing, Lateral Motor Column (LMC) neurons that project to the upper limbs, and Medial Motor Column (MMC) neurons that project to dorsal axial muscles to control posture, while thoracic levels contain preganglionic (PGC) neurons innervating sympathetic chain ganglia, hypaxial (HMC) neurons projecting to body wall muscles and MMC neurons (Fig. 1a).

To gain insights into the transcriptional programs that regulate MN specification and diversity, we performed RNA-seq to compare gene expression profiles between cervical and thoracic MN subtypes. We used fluorescence activated cell sorting (FACS) to sort GFP+ MNs from *Hb9::GFP* mice, which selectively express GFP in MNs[29], from the cervical and thoracic spinal cord at e12.5, when motor columns have acquired their distinct identities (Fig. 1a). We identified 417 and 543 genes that are preferentially expressed in cervical and thoracic MNs, respectively (fold change cutoff of 1.2-fold and significance cutoff of FDR < 0.05, Fig. 1b). Consistent with previous studies, we find that Hox TFs show differential expression along the rostrocaudal axis, with Hox5 and Hox9 paralogs enriched at cervical and thoracic levels respectively[3]. We also find enrichment of known PMC- (Ptprt) and LMC- (Lhx1) specific genes at cervical levels of the spinal cord, further validating our approach.

MN diversity relies on the function of genetic programs during development[2,3]. To define additional transcriptional regulators that may contribute to MN subtype identities, we examined differential TF expression between cervical and thoracic MNs and identified several rostrocaudally restricted TFs (Fig. 1c). Notably, about half of cervical MN-enriched TFs were Hox TFs belonging to Hox1-8 paralog groups. Hox paralog diversity at cervical levels of the spinal cord contributes to the specialization of MN subtypes[6,30]; however, it is not well understood how closely related Hox TFs elicit different genetic programs. To further define regulatory mechanisms that contribute to MN diversity, we focused on the cervical spinal cord and performed ATAC-seq to identify regions of actively-transcribed open chromatin in cervical MNs. We utilized *Choline Acetyltransferase (ChAT)::GFP* transgenic reporter mice, which express GFP in ventrally located Isl1/2 + MNs[31], to sort MNs from the cervical spinal cord at e12.5 (Figure S1a), as these reporter mice produced a slightly higher yield of MNs than *Hb9::GFP* mice at this age.

We generated ATAC-seq biological replicates with a mean of 107 M unique paired-end mapped reads per sample and identified 85,886 peaks of transposase accessible chromatin that were distributed across both intronic and exonic regions, with about 19% being located in promoters (1 kb upstream or downstream of the transcriptional start site (TSS)), including peaks at the *ChAT* promoter and the pan-MN TFs, *Isl1* and *Mnx1 (Hb9)* (Fig. 1d, S1b). Next, we used HOMER[32] to perform de novo motif search using ATAC-seq peaks to determine the relative abundance of sequence-specific TF consensus motifs. We found enrichment of CTCF motifs along with known MN TF motifs, such as Isl1 and Ebf1, as well as prominent homeobox recognition motifs such as Evx2 and Hoxc9 (Fig. 1e).

To identify chromatin accessibility regions that correspond to distinct MN columnar subtypes (PMC, LMC and MMC), we compared our ATAC-seq generated peaks from e12.5 MNs to column-enriched genes identified by single-cell (sc) RNA-seq at e13.5[33]. We reasoned that regions of open chromatin at e12.5 would correlate well with MN-specific gene expression at e13.5, as chromatin opening precedes active gene transcription during development[34]. We employed a graph-based clustering approach, Seurat[35], to identify differential gene expression. We assigned columnar identities based on the average expression of key MN marker genes in distinct clusters. For example, MN clusters exhibiting high expression of Foxp1 and Aldh1a2 were combined and assigned as LMC. Similarly, MNs exhibiting high expression of Mecom and Lhx3 were combined and assigned as MMC. With this approach, however, we were unable to confidently identify a phrenic MN cluster, likely due to the fact that PMC neurons are a rare population that may not form a distinct cluster in embryonic scRNA-seq data. Therefore, we instead utilized a list of genes shown to be selectively enriched in phrenic MNs by microarray analysis (Table S1, modified from table S4 in ref. 19. We then assigned ATAC-seq peaks to the gene of their nearest TSS and intersected genes associated with ATAC-seq peaks with column-enriched genes (Table S2). While we limited this analysis to specific MN subtype-enriched genes, a subset of our ATAC-seq peaks may also be in close proximity to actively repressed genes, indicating repression of alternative fate genes.

In order to define unique regulators of each MN subtype we performed motif analysis restricted to column-enriched genes. Our analysis identified prominent TA-rich homeodomain motifs in the regulatory regions of MNs belonging to all three columns, which closely match Hox binding sites

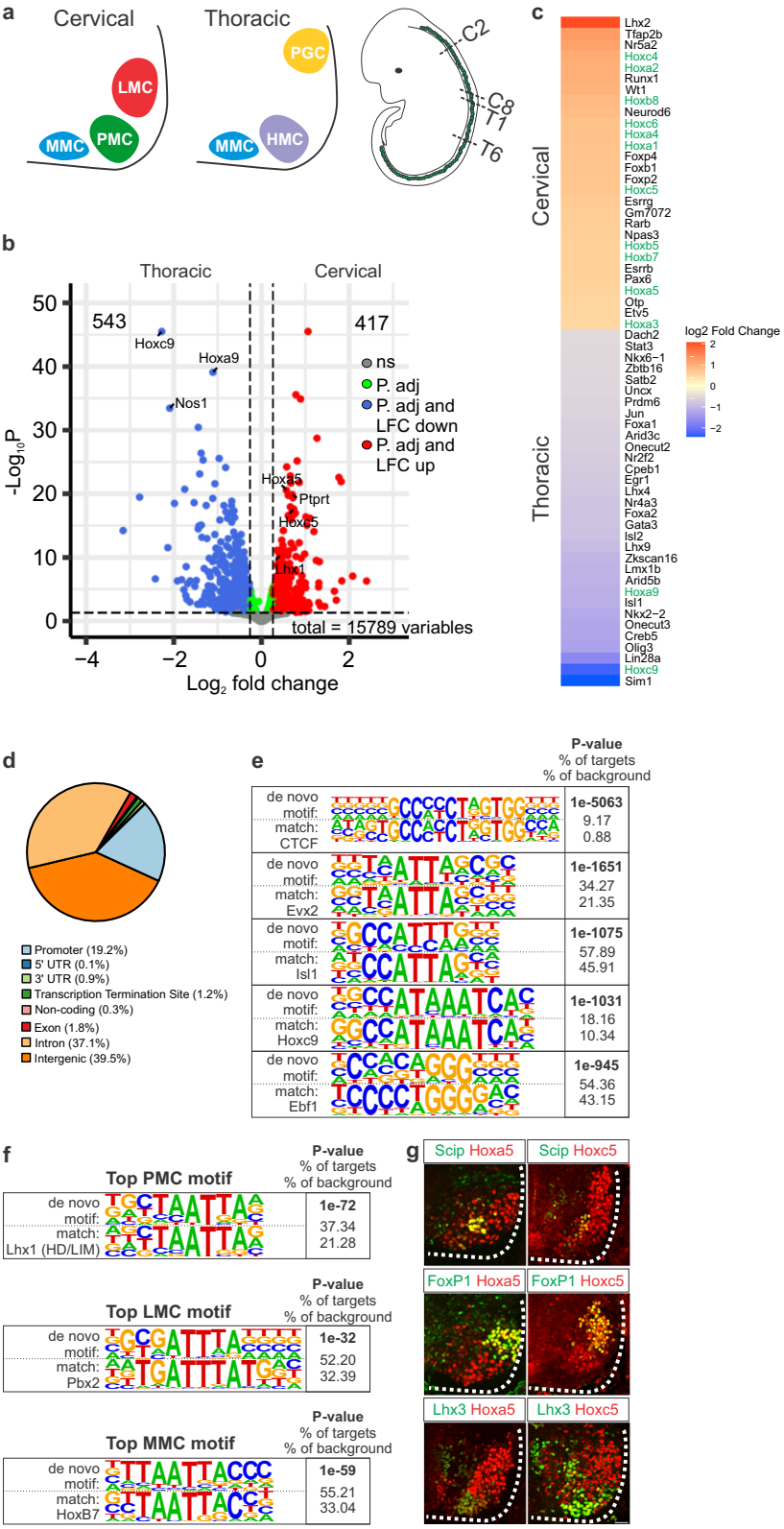



**Fig. 1 | Gene expression and chromatin accessibility profiles in cervical MN subtypes. a** MN subtypes at cervical and thoracic levels of the spinal cord. PMC=Phrenic Motor Column, LMC=Lateral Motor Column, MMC=Medial Motor Column, HMC=Hypaxial Motor Column and PGC=Preganglionic Motor Column. Cervical and thoracic MNs were sorted for RNA-seq. **b** Volcano plot showing differential gene expression between cervical and thoracic MNs, determined by DESeq2, with fold change cutoff of 1.2-fold and significance cutoff of FDR < 0.05. **c** Heat map showing differential TF expression between cervical and thoracic MNs. Top 60 TFs are shown. Hox TFs are highlighted in green. **d** Cervical MNs were sorted for ATAC-seq analysis. Distribution of ATAC-seq peak location relative to the nearest transcription start site (TSS). **e** HOMER output of top motifs enriched in ATAC-seq peaks. Both the de novo motif (top) and the best matched known TF motif (bottom) are shown, along with p-value and prevalence. **f** Top HOMER motif identified for motor column-specific genes after intersection with either scRNA-seq (LMC and MMC) or microarray (PMC) data. De novo motifs match known homeodomain motifs for all columns. **g** Hox5 paralog expression in cervical motor columns. Both PMC (Scip + ) and LMC (FoxP1 + ) neurons show high expression of Hoxa5 and Hoxc5. MMC (Lhx3 + ) neurons express low levels of both Hoxa5 and Hoxc5, while Hoxb5 is not expressed in MNs (Figure S1c, d). Scale bar = 100 μm.

(Fig. 1f). At cervical levels of the spinal cord, several Hox paralogs contribute to LMC specification, while Hox5 paralogs are the major Hox proteins driving PMC specification[6,16,17,30]. We were surprised to identify Hox motifs in MMC neuron-enriched genes, as MMC development is thought to be Hox-independent[21,22], although these motifs may reflect binding of other homeodomain proteins with similar recognition sites. We confirmed that both Scip+ PMC and FoxP1+ LMC neurons express high levels of Hoxa5 and Hoxc5, while also observing Hoxa5 and Hoxc5 expression in Lhx3+ MMC neurons at low, but detectable levels (Fig. 1g, S1c)[16]. Hoxb5 is not expressed in MN populations at e12.5 (Figure S1d).

## Hoxa5 regulates cervical MN chromatin accessibility

We previously found that MN-specific *Hox5* deletion results in severe defects in phrenic MN development, including a dramatic loss of motor

innervation at the diaphragm[16,36]. These defects are primarily due to *Hoxa5* deletion, as *Hoxc5* single mutants show normal diaphragm innervation (Figure S2a), indicating that Hoxa5 is the predominant Hox5 paralog controlling PMC development. To test whether Hoxa5 may exert some of its functions through regulating chromatin accessibility, we performed ATAC-seq on sorted *Hoxa5*-deleted (*Hoxa5flox/flox; Olig2::Cre*, referred to as $Hoxa5^{MN\Delta}$) cervical MNs. Principal component analysis (PCA) showed a high degree of concordance between replicates, with *Hoxa5* deletion accounting for the majority of variance (Figure S2b).

To define chromatin accessibility changes induced by the loss of *Hoxa5*, we performed differential analysis using DESeq2. We identified a total of 8819 and 7326 peaks that were either gained or lost in $Hoxa5^{MN\Delta}$ MNs, respectively (q-value < 0.05, ±1.5×) (Fig. 2a). To identify molecular pathways impacted by chromatin changes after *Hoxa5* loss, we performed Gene Ontology (GO) term enrichment analysis using the nearest annotated neighboring genes for individual chromatin accessibility peaks. We found that genes with decreased accessibility are associated with developmental processes such as axonogenesis, regionalization, axon guidance, and cell fate commitment, consistent with the phenotypes observed in phrenic MNs in mice lacking *Hox5* genes (Figure S2c)[16,36]. Genes with increased accessibility show similar classification, suggesting that Hoxa5 may repress these processes in other MN subtypes (Figure S2d). We also saw overrepresentation of genes related to regulation of synapse structure, organization and assembly, indicating potential changes in synaptic connectivity (Figure S2d).

We next performed motif analysis for differentially accessible peaks and found that a Sox TF motif is the top motif enriched in peaks that are lost in *Hoxa5*-deleted MNs (Fig. 2b). We previously found that Sox5 and Hox5 are co-expressed at cervical levels of the spinal cord[16]. We observed that Sox5 levels are reduced after *Hox5* deletion (*Hoxa5flox/flox; Hoxc5-/-; Olig2::Cre* mice, referred to as $Hox5^{MN\Delta}$), suggesting that Hoxa5 may regulate chromatin accessibility indirectly through Sox TFs (Figure S2e). Both lost and gained peaks show enrichment for homeodomain motifs (Fig. 2b, c), indicating that Hoxa5 may also bind directly to regions of altered chromatin accessibility. Notably, the binding motif for Pou3f1 (Scip) is enriched in peaks that are lost after Hox5 deletion, suggesting co-operative function of Hoxa5 and Scip in chromatin opening.

To further define the TFs that are enriched in differentially accessible regions and thus may, in addition to Hoxa5, control MN gene regulatory programs, we performed footprinting analysis in control and $Hoxa5^{MN\Delta}$ ATAC-seq peaks, using TOBIAS[37] with motifs from the Jasper databases (Fig. 2d)[38]. During ATAC-seq, open chromatin is detected by exposure of genomic DNA to Tn5, a highly active transposase, which preferentially inserts into open chromatin sites, and adds sequencing primers. Regions of open chromatin are identified as genomic intervals with high sequencing coverage. However, TF binding to DNA prevents Tn5-mediated cleavage in an otherwise nucleosome-free region, leaving small regions, referred to as footprints, which show a reduction of read coverage within high-read regions. The TOBIAS computational approach uses transposase insertion sites to identify motifs that are protected from transposition, hence likely bound by a TF. We utilized this approach to identify TF motifs that were differentially enriched in control and *Hoxa5*-deleted MN chromatin. Differential footprinting analysis showed that motifs for Klf TFs (Klf5, Klf15, Klf10) showed a higher footprinting score in control peaks, whereas motifs for homeobox TFs such as Lhx showed a higher footprinting score in $Hoxa5^{MN\Delta}$ peaks. The higher prevalence of Klf motifs in control peaks suggests that Hoxa5-mediated chromatin reorganization may expose previously inaccessible Klf binding sites. Overall, the differential footprints found between control and $Hoxa5^{MN\Delta}$ MNs support the idea that Hoxa5 may regulate the binding ability of downstream TFs, and suggest that a stepwise transcriptional cascade initiated by Hox5 proteins may underlie MN development and maintenance.

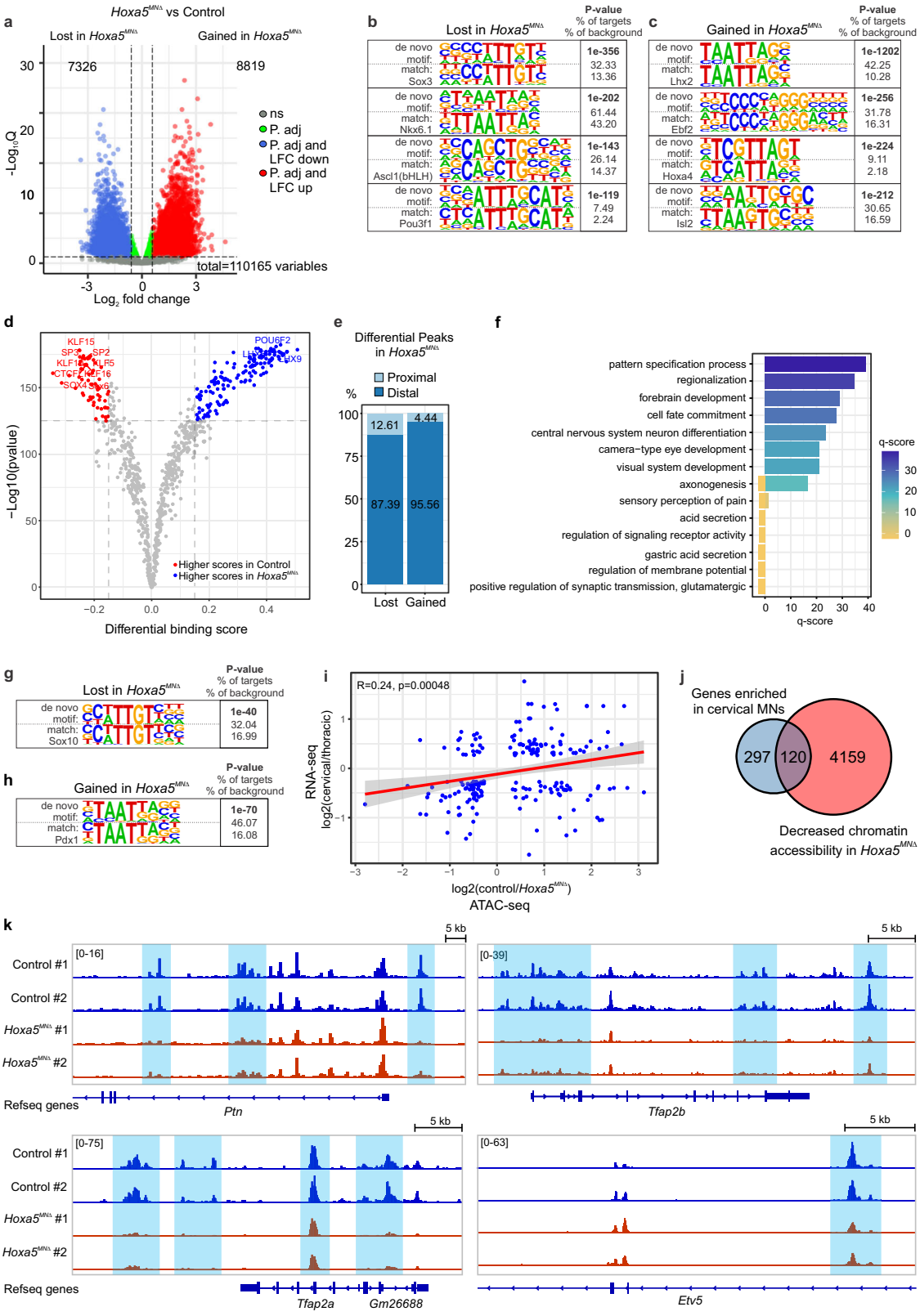
To test whether Hoxa5 differentially alters chromatin accessibility at promoter or enhancer regions, we analyzed the distribution of differential ATAC-seq peaks at proximal (≤3 kb) and distal (>3 kb) regions from an annotated TSS. While the majority of peaks that are gained or lost in $Hoxa5^{MN\Delta}$ MNs are distributed at distal enhancers, there is a higher percentage of peaks with decreased accessibility in $Hoxa5^{MN\Delta}$ MNs located at proximal regions, suggesting that Hoxa5 may have a different impact on chromatin accessibility at proximal and distal regulatory elements (Fig. 2e). We performed GO analysis restricted to the genes assigned to proximal regions, and found that genes associated with decreased chromatin accessibility largely correspond to genes related to neuron differentiation and development (Fig. 2f). Motif analysis for proximal regions reveals motifs similar to total peak analysis, with Sox and homeodomain motifs being the top motifs enriched in lost and gained $Hoxa5^{MN\Delta}$ peaks, respectively (Fig. 2g, h).

To assess whether the ability of Hoxa5 to modify chromatin opening might contribute to MN-specific gene expression, we performed correlation analysis between our RNA-seq dataset and peaks that were either gained or lost in gene proximal regions (≤3 kb) in *Hoxa5*-deficient MNs. We found that genes preferentially expressed in cervical MNs correlate with open chromatin in control MNs, as compared to $Hoxa5^{MN\Delta}$ MNs (Fig. 2i). We also found that approximately 30% of genes enriched in cervical MNs overlap with genes showing decreased accessibility in $Hoxa5^{MN\Delta}$ MNs (Fig. 2j). For example, the transcription factors Tfap2a, Tfap2b and Etv5 are enriched in cervical MNs and show decreased chromatin accessibility in *Hoxa5*-deleted MNs (Fig. 2k). In addition, we find decreased chromatin accessibility in phrenic-specific genes, such as *Ptn*, which are downregulated in $Hox5^{MN\Delta}$ mice (Fig. 2k)[16]. Altogether, these results, together with our previously published RNA-sequencing data showing changes in gene expression after *Hox5* deletion[36], suggest that Hox5 TFs may regulate MN-specific gene expression programs partially by altering the MN chromatin landscape.

### Hoxa5 and Pbx1 modules directly control MN genes

While Hoxa5 may regulate a subset of genes through chromatin remodeling, it is likely that its major mode of action involves either direct binding to its transcriptional targets or indirect induction through intermediate TFs. To understand how Hoxa5 induces distinct MN subtype identities, we wanted to identify direct Hoxa5 transcriptional targets in the spinal cord. Since TALE cofactors bind DNA cooperatively with Hox proteins and are essential for many Hox actions, including phrenic MN development and diaphragm innervation, we also investigated targets of Pbx1, which is strongly expressed in all cervical MN columns[14]. To identify both unique and shared transcriptional targets of Hoxa5 and Pbx1, we performed chromatin immunoprecipitation followed by sequencing (ChIP-seq) from e12.5 mouse cervical spinal cord chromatin, and identified a total of 4499 Hoxa5 peaks and 14875 Pbx1 peaks. To test if Hoxa5 and Pbx1 co-regulate a subset of cis regulatory elements, we intersected Hoxa5 with Pbx1 peaks and found that 30% of Hoxa5 (1357) peaks co-occur with Pbx1 peaks (Fig. 3f). The majority of Pbx1 peaks were not bound by Hoxa5, indicating that many Pbx functions are likely Hoxa5-independent[39]. Notably, we also identified a significant portion of Hoxa5 peaks not bound by Pbx1 indicating either that other Pbx proteins, such as Pbx3, may form distinct complexes with Hoxa5, or suggesting Pbx-independent Hoxa5 DNA binding. Further analysis of the genomic distribution of Hoxa5 and Pbx1 bound regions showed that the majority of Hoxa5 peaks were located within the promoter region of annotated genes, while Pbx1 peaks were distributed between promoters, intronic and intergenic regions (Fig. 3a–c, S3a, b). Regions co-occupied by both Hoxa5 and Pbx1 were predominantly associated with promoter regions, mostly mirroring Hoxa5 peak distribution (Fig. 3f).

To test whether Hoxa5 binding leads to chromatin opening, we intersected Hoxa5 ChIP-seq peaks with ATAC-seq peaks changed in $Hoxa5^{MN\Delta}$ MNs. We find a greater overlap between gene proximal regions (≤3 kb) with decreased chromatin accessibility and Hoxa5 binding (~17%), as compared to those that show increased chromatin accessibility after *Hoxa5* deletion (~7%), indicating that direct binding of Hoxa5 may lead to chromatin opening in a subset of genes (Figure S3c, d). However, it appears




that the majority of chromatin changes occur independently of Hoxa5 binding and may be due to changes in the expression of downstream chromatin remodelers.

GO term enrichment analysis revealed that the peaks bound by either Hoxa5 or Pbx1 or both are associated with genes that regulate axon guidance, pattern specification process, and neuron differentiation (Fig. 3i, S3e-f), consistent with known Hoxa5/Pbx1 functions in MNs. To investigate whether certain DNA motifs were enriched in the Hoxa5-bound, Pbx1-bound, and combined Hoxa5-Pbx1 bound sites, we applied HOMER de novo motif search. Surprisingly, the top motif identified in Hoxa5-bound sites was a bHLH motif, although canonical Hox motifs, as well as a POU/Homeobox motif were also enriched (Fig. 3d). Pbx1-bound sites were

**Fig. 2 | Hoxa5 contributes to chromatin accessibility in cervical MN subtypes.** **a** Volcano plot showing differential chromatin accessibility between control and *Hoxa5*$^{MN\Delta}$ MNs, determined by DESeq2, with fold change cutoff of 1.5-fold and FDR < 0.05. 8819 peaks were significantly gained, while 7326 peaks were significantly lost in *Hoxa5*$^{MN\Delta}$ MNs. **b**, **c** Top HOMER motifs for peaks that are lost (**b**) and gained (**c**) in *Hoxa5*$^{MN\Delta}$ MNs. **d** Comparison of TF activities between control and *Hoxa5*$^{MN\Delta}$ MNs. Volcano plot showing the TOBIAS differential binding score on the x-axis and -log10 (*p* value) on the y-axis; each dot represents one TF. **e** Distribution of differential ATAC-seq peaks in *Hoxa5*$^{MN\Delta}$ MNs. **f** Gene ontology (GO) enrichment analysis of biological pathways of differentially accessible genes with peaks in proximal regions. Positive q-scores represent categories enriched in peaks lost in *Hoxa5*$^{MN\Delta}$ MNs. **g**, **h** Top HOMER motif for proximal peaks that are lost (**g**) and gained (**h**) in *Hoxa5*$^{MN\Delta}$ MNs. **i** Correlation analysis of differentially expressed genes (cervical vs thoracic RNA-seq) and differential chromatin accessibility (control vs *Hoxa5*$^{MN\Delta}$ ATAC-seq) at proximal regions. **j** Overlap between genes enriched in cervical MNs and genes with decreased accessibility in *Hoxa5*$^{MN\Delta}$ MNs. **k** Genome browser views of ATAC-seq signals from genes with decreased accessibility in *Hoxa5*$^{MN\Delta}$ MNs.

enriched for Hox, Pbx and Meis motifs (Fig. 3e). A previously established Hox-Pbx composite motif (Fig. 3g) was identified in the top enriched motifs in both Pbx1-bound and Hoxa5-Pbx1 shared peaks (Fig. 3e, h). Hoxa5-Pbx1 bound sites showed a high prevalence of homeobox motifs, indicating Hox/Pbx cooperative binding to canonical homeobox sequences (Fig. 3h).

In order to define how Hox/Pbx TFs regulate gene expression at cervical levels of the spinal cord, we intersected our ChIP-seq datasets with genes that were found to be preferentially expressed in cervical MNs in our RNA-seq experiments. We find that about 30% and 50% of MN genes enriched at cervical levels show direct binding of Hoxa5 and Pbx1, respectively, while 13% of genes, including the TFs Tfap2a and Tshz3, show binding of both Hoxa5 and Pbx1 (Fig. 3j, l). In addition, we find that the phrenic-specific genes Negr1 and Pcdh10, which are down-regulated in the absence of *Hoxa5*, also include peaks for both Hoxa5 and Pbx1 (Fig. 3k)[36]. Next, we associated Hoxa5, Pbx1 and Hoxa5-Pbx1 intersected peaks to MN column-specific genes (Fig. 3m). Our analysis identified that more than 50% of PMC, LMC and MMC genes show enrichment of either Hoxa5 or Pbx1 or both, underscoring the overarching function of Hox and Pbx-mediated transcriptional programs in these MN populations. Surprisingly, we found that a significant portion of PMC genes (45%) do not show enrichment of either Hoxa5 or Pbx1, despite the critical role of these TFs in PMC development, likely reflecting the existence of intermediate TFs downstream of Hox-mediated transcriptional programs. Assessment of TF motifs present in Hoxa5-bound PMC, LMC, and MMC genes using HOMER revealed enrichment of distinct motifs for each column, indicating that the specific cellular context in each MN subtype might alter Hoxa5 binding specificity (Fig. 3n). De novo motif analysis of Pbx1-bound peaks in all motor columns revealed enrichment of either Pbx motifs or motifs for the TALE cofactor Meis1. This suggests that Pbx1 may bind specific motor column loci in a Hoxa5-independent manner, in a complex with other TALE factors such as Meis1 and Meis2. Together, these results suggest that Hoxa5 and Pbx1 either individually or collaboratively target cis-regulatory modules that orchestrate different aspects of MN development.

### Scip cooperates with Hox/Pbx programs to induce PMC identity

Our ChIP-seq data, together with published RNA-seq datasets[14,36], indicate that Hox/Pbx TFs directly bind and regulate genes that are essential for MN specification and development in multiple motor columns in the cervical spinal cord. However, it is unclear how these TFs can induce specific MN identities given their broad expression patterns. We previously showed by retrograde labeling that the expression of Scip, a POU domain TF, is restricted to PMC neurons and that overexpression of FoxP1, which is required for the establishment of LMC identity[21,22], suppresses Scip expression[16]. Similar to *Hox5* knockout mice, *Scip* mutant mice also die at birth due to respiratory dysfunction and show a reduction and disorganization of phrenic MNs[20]. In addition, co-expression of Hoxa5 and Scip in mouse embryonic stem cell (ESC)-derived MNs induces expression of phrenic-specific genes[19], and we identified enrichment for POU motifs in Hoxa5 ChIP-seq peaks (Fig. 3d). Therefore, we next asked whether context-specific functions of Hoxa5 are achieved via interactions or cooperativity with other MN-specific TFs, such as Scip.

To test whether Hoxa5 and Scip associate with each other during PMC specification, we created tagged constructs (Fig. 4a) and performed protein co-immunoprecipitation (co-IP) assays using transiently transfected 293T cells. To validate our findings, we also tested the interaction between Hoxa5 and Pbx1, which has been previously established[40,41]. The co-elution of Hoxa5 and Pbx1 and Hoxa5 and Scip in the same IP fraction suggests that these proteins can form a complex (Fig. 4b, c). The hexapeptide (YPWM) domain of Hox proteins is critical for their interaction with Pbx cofactors[40,42,43]. To test whether the same domain is required for the Hoxa5 interaction with Scip, we mutated the YPWM domain of Hoxa5 to AAAA (Hoxa5$^{YPWM>AAAA}$, Fig. 4a) and performed co-IP assays. While we found a decreased association of Hoxa5$^{YPWM>AAAA}$ with Pbx1 as expected, we did not find any changes in its interaction with Scip, suggesting that the Hoxa5-Scip interaction is independent of the hexapeptide motif (Fig. 4d–e, S4a). These data support a model where Hoxa5, Pbx1, and Scip form a complex to induce phrenic-specific programs and both Pbx1 and Scip bind to Hoxa5 through non-competitive interactions.

To understand whether the Hoxa5 interaction with Scip is paralog-specific, we tested the ability of Scip to interact with Hoxc9, a Hox paralog required for the generation of thoracic respiratory MN subtypes that is ~36% identical to Hoxa5[44]. We found that Hoxc9 does not form a complex with Scip (Fig. 4g), suggesting that Scip does not broadly associate with Hox proteins, but rather exhibits paralog-dependent specificity. Due to the absence of a canonical hexapeptide motif, Hoxc9 also shows decreased interaction with Pbx1 (Fig. 4f)[43].

Outside of the homeodomain and the YPWM motif, N-terminal domains of Hox protein sequences diverge substantially. To identify the region of Hoxa5 necessary for complex formation with Scip, we serially deleted the N-terminal end of Hoxa5 and created three HA-tagged N-terminal deletion constructs: HA-Hoxa5-ΔN30, HA-Hoxa5-ΔN86 and HA-Hoxa5-ΔN144 (Fig. 4h) and performed co-IP experiments. 293T cells were co-transfected with expression constructs encoding HA-tagged Hoxa5 deletion constructs and V5-tagged Scip. Pull-down experiments with an antibody against the V5 epitope showed that HA-Hoxa5-ΔN86 and HA-Hoxa5-ΔN144 do not co-IP with Scip (Fig. 4i), suggesting aminoacids 30-86 at the N-terminal region of Hoxa5 are essential for complex formation with Scip. All N-terminal deletion constructs retain the ability to interact with Pbx1 (Figure S4b-c), indicating that other Hoxa5 functions are maintained and that this N-terminal Hoxa5 domain is specifically required for the interaction with Scip.

To test whether the Hoxa5/Scip interaction can also be observed in vivo, we prepared whole tissue lysate from the cervical spinal cord of e12.5 mouse embryos and performed co-IP. Similar to transiently transfected 293 T cells, we were able to IP Hoxa5 using a goat anti-Scip antibody. Further probing the blot with a rabbit-anti-Scip antibody, we were also able to detect Scip in the same IP fraction (Fig. 4j). However, we were unable to detect Hoxa5 or Scip in whole-cell lysate, likely due to lower endogenous expression.

To test if Hoxa5 and Scip expression is sufficient to suppress LMC identity, we co-electroporated constructs expressing mouse Hoxa5 and Scip under a pCAGGs promoter in chicken embryos, which lack phrenic MNs. We found that the overexpression of Hoxa5 and Scip did not affect the number of MNs generated, as electroporated cells still expressed normal levels of Isl1/2, but suppressed the expression of Foxp1 (Fig. 4k). Our data collectively indicate that Hoxa5 and Scip cooperate to induce phrenic and suppress limb-innervating MN identity.

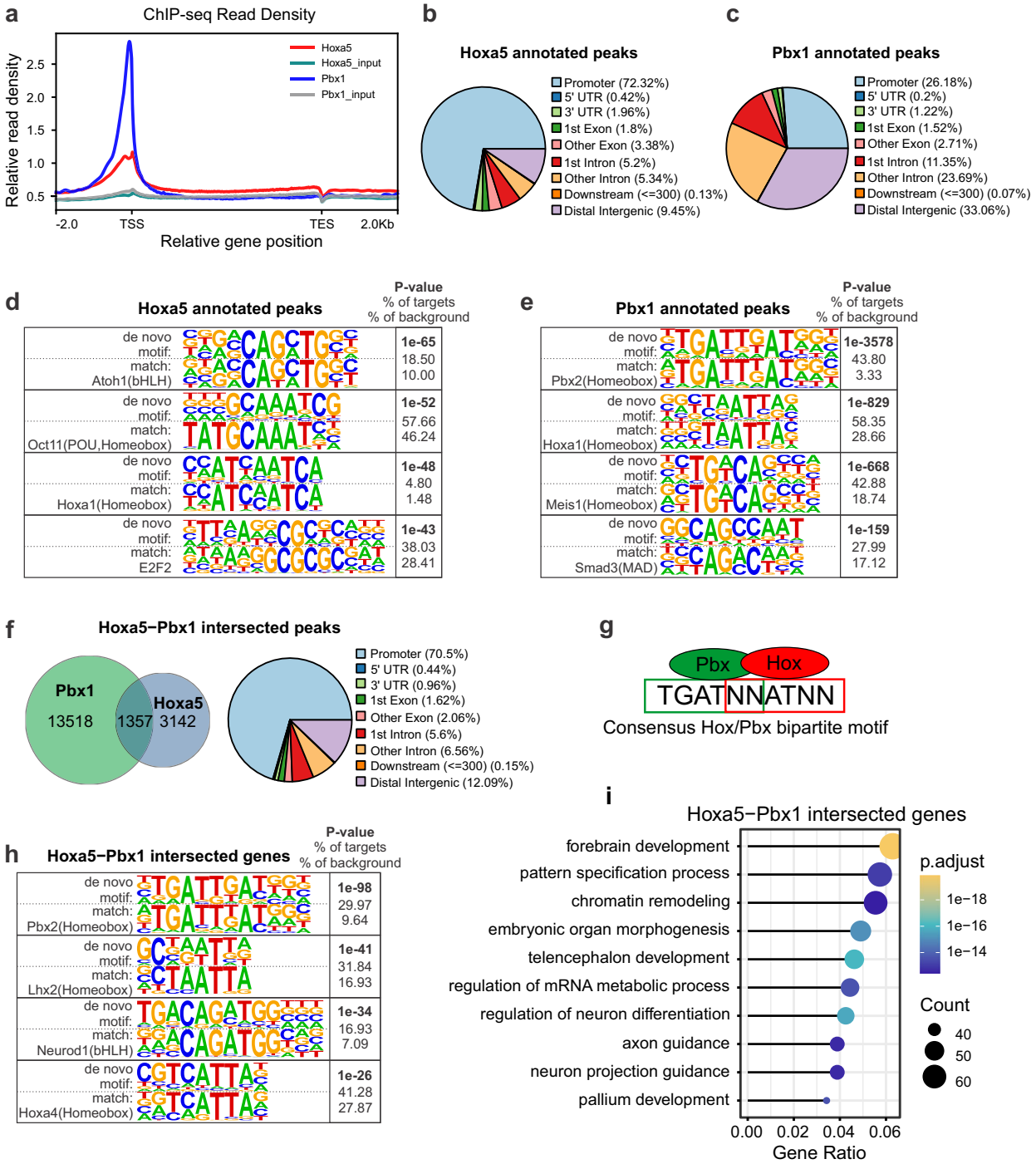



**Fig. 3 | Direct regulation of MN-specific genes by Hoxa5 and Pbx1. a** Average distribution around the TSS of Hoxa5 and Pbx1 target genes. **b**, **c** Pie chart illustrating peak location relative to the nearest TSS for Hoxa5 (**b**) and Pbx1 (**c**) enriched peaks. The distribution of Hoxa5-bound peaks is enriched in promoters compared to Pbx1. **d**, **e** HOMER output of top motifs enriched in Hoxa5 (**d**) and Pbx1 (**e**)-bound peaks. **f** Overlap of Hoxa5 and Pbx1 enriched peaks and pie chart illustrating peak location relative to the nearest TSS for Hoxa5 and Pbx1 enriched peaks. **g** Consensus Hox/Pbx bipartite motif. **h** HOMER output of top motifs enriched in Hoxa5/Pbx1-bound peaks. **i** GO enrichment analysis of genes with Hoxa5-Pbx1 intersected ChIP-seq peaks. **j** Overlap of genes enriched in cervical MNs and Hoxa5-bound, Pbx1-bound and Hoxa5/Pbx1-bound ChIP-seq peaks. **k**, **l** Genome browser views of example genes bound by both Hoxa5 and Pbx1. Phrenic-specific genes *Pcdh10* and *Negr1* (**k**) and transcription factors *Tfap2b* and *Tszh3* (**l**) are shown. Hoxa5 and Pbx1 input tracks are scaled to their respective ChIP-seq tracks. **m** Analysis of column-specific genes for Hoxa5 and Pbx1 binding. **n** Top HOMER motif identified for Hoxa5 and Pbx1 enriched peaks in motor column-specific genes.

## Postnatal maintenance of phrenic MN identity

Our co-IP experiments indicate that Hoxa5 and Scip likely form a complex to induce transcriptional programs that control the establishment of embryonic phrenic MNs during development[16,19,20,36]. However, it is not clear whether expression of these two TFs is continuously required for phrenic MN maintenance at postnatal and adult stages. Since a number of Hox proteins show maintained expression at postnatal stages in brachial MNs[45], we evaluated the expression of Hoxa5 and Scip at different stages. Both Hoxa5 and Scip were strongly expressed in phrenic MNs at postnatal day (P)5.5, but their expression became weaker at P10.5 and undetected by P16.5 (Fig. 5a).

During development, Hox5 proteins control the expression of phrenic-specific cell adhesion molecules, such as *ALCAM, Negr1 and Pcdh10*[36]. To

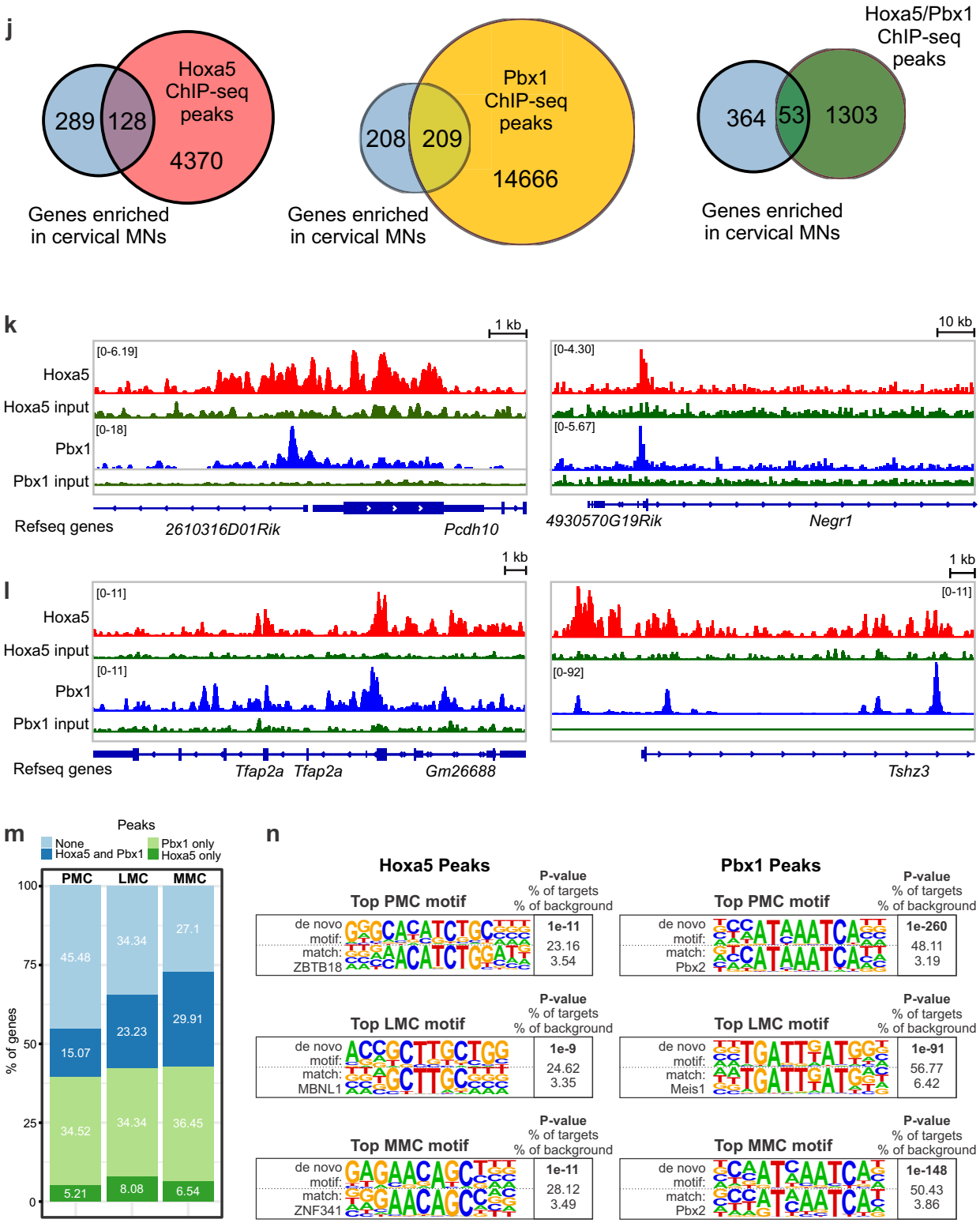



**Fig. 3** | continued

test whether Hox5 downstream genes are downregulated in a similar temporal fashion as Hoxa5 or are maintained postnatally, we performed in situ hybridization for the pan-MN marker *Vesicular Acetylcholine Transporter (VAchT), Alcam, Negr1, and Pcdh10*. Surprisingly, we observed maintained expression of these genes at P16.5, despite Hoxa5 downregulation, suggesting that additional gene regulatory mechanisms may control the maintenance of these early Hox5 target genes (Fig. 5a). In order to explore potential maintenance factors of phrenic MN identity downstream of early Hox/Pbx programs, we intersected Hoxa5 and Pbx1-enriched ChIP-seq peaks and differential ATAC-seq peaks with a curated list of mouse TFs[46,47]. We selected several TFs that either showed particularly high enrichment in ChIP-seq or ATAC-seq datasets or have known functions in MN development for further downstream analysis, including Ebf and Tshz factors, Neurod1, Onecut2, and Stat3. We examined expression of

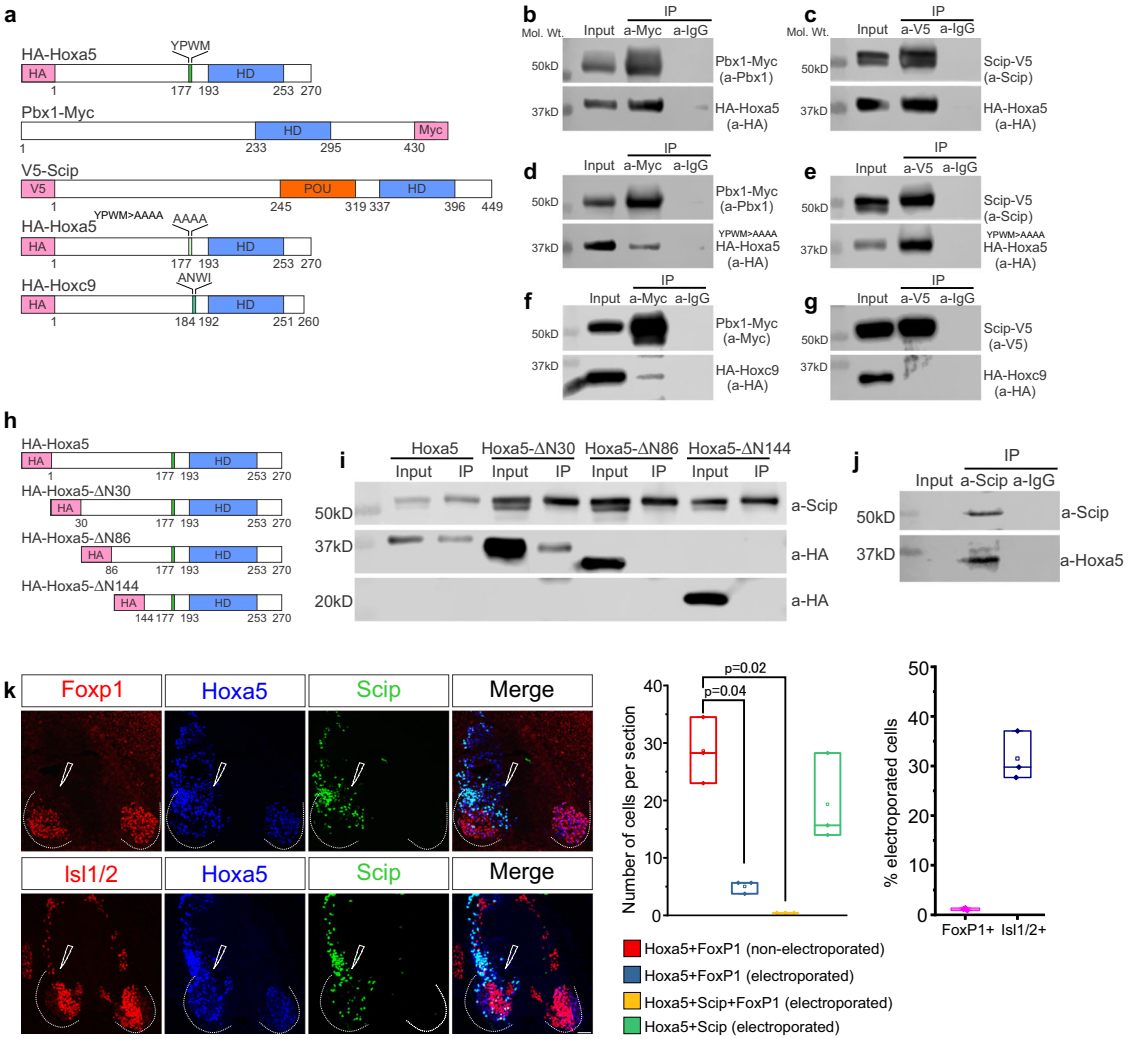



**Fig. 4 | Paralog-specific Hoxa5/Scip interaction promotes PMC identity.** **a** Overview of tagged constructs used for co-immunoprecipitation (co-IP) in transiently transfected 293T cells. **b**, **c** HA-Hoxa5 co-immunoprecipitates with Pbx1-Myc and V5-Scip. **d**, **e** HA-Hoxa5$^{YPWM>AAAA}$ co-immunoprecipitates with V5-Scip (**e**) while co-immunoprecipitation with Pbx1-Myc (**d**) is reduced. **f**, **g** HA-Hoxc9 does not interact with V5-Scip (**g**) and weakly interacts with Pbx1-Myc (**f**). **h** Overview of Hoxa5 N-terminal serial truncation constructs. **i** Transiently transfected 293T cells with HA-Hoxa5 N-terminal serial deletion constructs and Scip-V5 were subjected to co-immunoprecipitation assay using antibodies against V5. Hoxa5 and ΔN30 Hoxa5 co-immunoprecipitate with V5-Scip, while Hoxa5-ΔN86 and Hoxa5-ΔN144 do not. **j** Scip and Hoxa5 co-immunoprecipitation from whole cell lysate of e12.5 embryonic mouse spinal cord tissue. **k** Co-electroporation of Hoxa5 and Scip in chick embryos leads to a reduction in the number of FoxP1 positive cells in the cervical spinal cord, but does not affect overall MN identity, as seen by Isl1/2 expression ($n = 3$ embryos, one-way ANOVA with Dunnett's multiple comparisons test). Scale bar = 50 μm.

these TFs at cervical levels of the spinal cord at e12.5 and found either broad MN expression (e.g. *Stat3*) or no MN expression (e.g. *Pou6f2*, figure S5a). While we did not observe phrenic-specific enrichment of these TFs at e12.5, it is possible that they are upregulated in phrenic MNs at later embryonic or postnatal stages, consistent with their roles in MN maturation[48,49]. We performed in situ hybridization at P30 and found broad MN expression of *Stat3*, *Neurod1*, *Tshz3* and *Pou6f2*, indicating that these TFs may contribute to MN maintenance programs independent of their subtype identity (Figure S5b).

We also identified several Klf family members in our intersected dataset and previously noticed that the footprinting score of multiple Klf TFs was reduced in *Hoxa5*-deficient ATAC-seq peaks (Fig. 2d), suggesting Hoxa5 may regulate both the expression and DNA-binding of Klf family members. We tested Klf expression at e12.5 and found that *Klf5* and *Klf6*, but not *Klf3*, *Klf7*, or *Klf15*, are highly expressed in phrenic MNs (Fig. 5b, S5c). We also found that *Klf5* and Klf6 expression is maintained in phrenic MNs at 1 month of age, although Klf6 is broadly expressed throughout the spinal cord (Fig. 5c). Since Klf5 appears to be specific to phrenic MNs both at embryonic and postnatal stages, we examined whether *Klf5* expression is dependent on Hox5, and found that *Klf5* expression is lost after *Hox5* gene deletion (Fig. 5d-e). Together these findings suggest that a subset of Klf TFs show continuous expression from embryonic to postnatal stages in phrenic MNs and may regulate a gene regulatory network required for phrenic MN identity maintenance (Fig. 5f).

## Discussion

Combinatorial TF expression and changes in chromatin accessibility underlie the development, diversification and maturation of MN subtypes[48]. Hox proteins are at the core of early transcriptional programs that diversify MNs along the rostrocaudal axis of the spinal cord[3]. At cervical levels of the spinal cord, MN columns show a differential requirement for Hox5 proteins- PMC neurons are largely dependent on Hox5 proteins for their survival and specification, LMC neurons show an intermediate requirement for the axonal pathfinding of a subset of pools, while MMC neurons appear to be resistant to *Hox5* loss. Here, we sought to address how Hox5 proteins can serve multiple functions in the development and specification of distinct MN subtypes. We find that Hox5 paralogs exert their functions through modulating chromatin accessibility, direct binding and regulation of

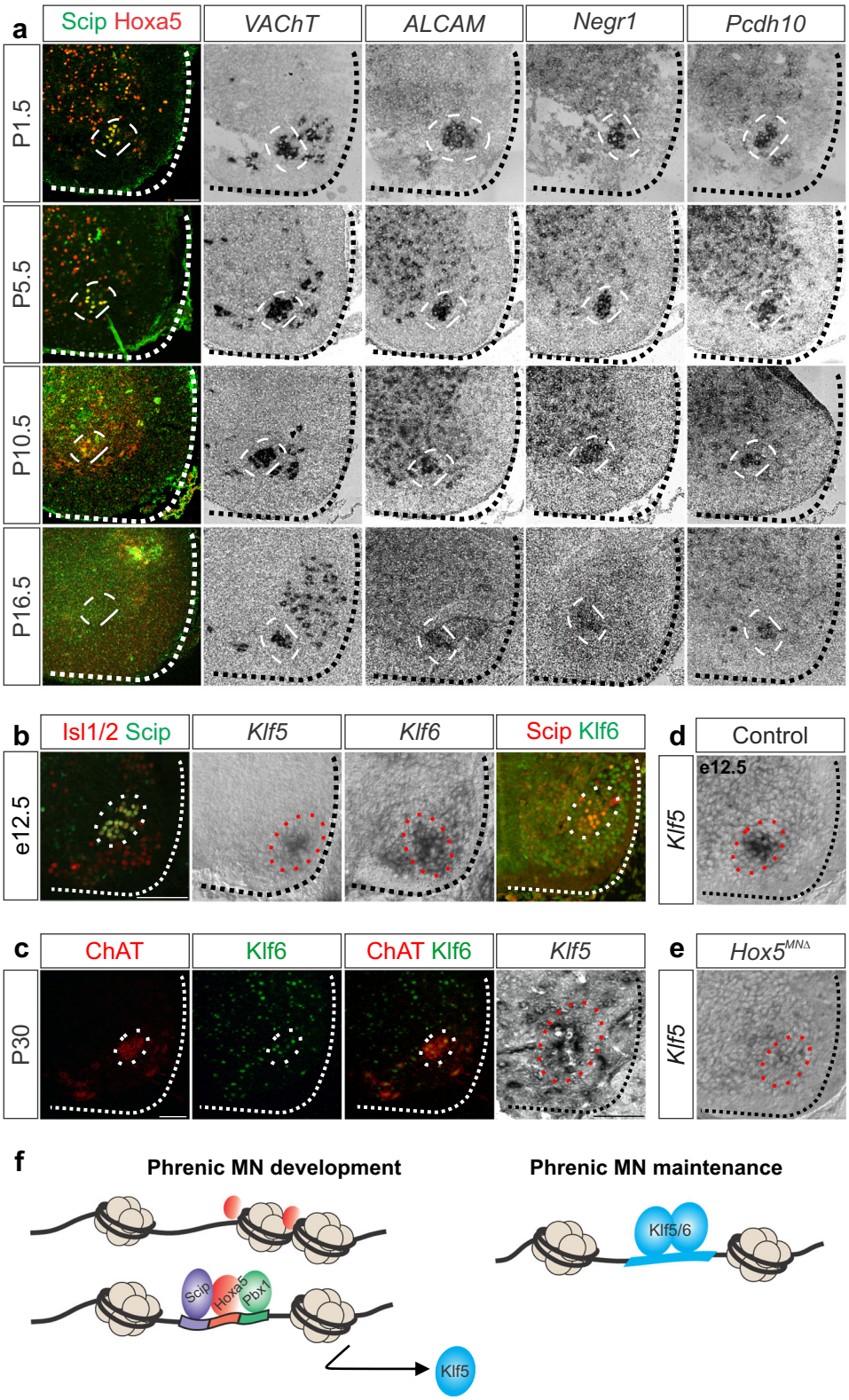



**Fig. 5 | Maintenance of phrenic MN identity at postnatal stages. a** Expression of Hoxa5, Scip, *VAChT*, *ALCAM*, *Negr1* and *Pcdh10* in postnatal phrenic MNs. Phrenic MNs are shown inside the dashed white line. Hoxa5 and Scip are down-regulated after P10.5, while in situ hybridization shows sustained expression of phrenic-specific genes *Alcam*, *Negr1*, and *Pcdh10*. **b**, **c** Expression of *Klf5* and Klf6 in phrenic MNs at e12.5 (**b**) and 1 month (**c**). **d**, **e** PMC-specific *Klf5* expression is lost in $Hox5^{MN\Delta}$ mice. **f** Proposed model of phrenic MN specification and maintenance. Hoxa5 can bind inaccessible chromatin and forms a complex with Pbx1 and Scip to induce PMC-specific genes, including Klf5, which may act to maintain phrenic MN properties in adulthood. Scale bar = 100 μm.

transcriptional targets, and associating with MN subtype-specific co-factors. Our findings provide insights into how Hox5 proteins can selectively control both PMC and LMC properties. The high incidence of Hox motifs in open chromatin and Hoxa5 binding in MMC-associated genes is surprising, given the lack of overt MMC phenotypes in *Hox5* mutants. While MMC columnar identity is thought to be Hox-independent, it is possible that Hox-mediated transcriptional programs may contribute to MMC properties downstream of columnar identity, similar to LMC neurons.

Several Hox proteins exert their functions partially through their ability to reorganize chromatin[25]. We find that loss of *Hoxa5* from MNs leads to changes in chromatin accessibility, and that Hoxa5 contributes to chromatin opening in genes that are enriched in cervical MNs. Given the absence of domains that indicate an intrinsic ability of Hox proteins to remodel chromatin, it is likely that this property arises from their interactions with additional binding partners[25]. Motif analysis of regions with decreased accessibility in *Hoxa5*-deleted MNs revealed enrichment of Sox and Pou3f1

motifs. Sox5 and Pou3f1 are co-expressed with Hoxa5 in cervical MNs and we observe a reduction in Sox5 expression after *Hox5* deletion. Sox, Oct and Klf family members have known pioneer activity[50], indicating that the ability of Hoxa5 to either recruit or induce these TFs could mediate its chromatin remodeling activity.

Our data indicate that Hoxa5 has the differential ability to recruit Scip (Pou3f1/Oct6) and that this interaction is mediated by sequences at the N-terminal domain of the protein, which are the most divergent among Hox paralogs and thus likely to mediate paralog-specific protein interactions[51,52]. The ability of Hoxa5 to interact with this novel binding partner may have led to the emergence of phrenic MN identity in mammals, as avian species express Hox5, but not Scip, with similar rostrocaudal boundaries in the spinal cord. In mouse ESC-derived MNs, co-expression of Hoxa5 and Scip induces a transcriptional profile corresponding to phrenic MNs[19]. Here, we show that Hoxa5 and Scip co-expression is also sufficient to suppress LMC identity, revealing that the Hox5/Scip complex has a dual role in inducing phrenic and suppressing limb MN programs. Similarly, we previously found that combinatorial expression of Hoxa5 and FoxP1 suppresses phrenic MN identity[16], indicating that cross-repressive interactions ensure the right balance of phrenic and limb-innervating MNs at cervical levels of the spinal cord. Interestingly, FoxP1 and Scip expression domains overlap at more caudal levels of the brachial spinal cord that are devoid of Hoxa5 expression, indicating that Hoxa5 is specifically required for Scip/FoxP1 cross-repression. Motif analysis of ATAC-seq and ChIP-seq data indicates a different top motif for Hoxa5 binding in PMC neurons. One possibility is that the interaction of Hoxa5 with Scip can bias its binding preference to regulatory regions on phrenic-specific targets, suggesting conserved strategies for Hox binding selectivity[53–55]. Future experiments utilizing scRNA-seq and scATAC-seq from isolated phrenic MNs will further test this possibility.

The transcriptional programs that control MN maturation and maintenance are just beginning to emerge. In *C. elegans* MNs, terminal selectors are necessary for inducing and maintaining cholinergic transmission and other core features of MN identity throughout the lifetime of the animal[56–58]. In mammalian serotonergic neurons, an adult stage transcriptional program maintains their synaptic connectivity and protects axons from neurotoxic injury[59]. It is unclear whether mammalian MNs express maintenance factors that safeguard their integrity in adulthood, and whether these factors are broadly expressed in all MNs or are unique to specific MN subtypes. We find that a subset of Klf TFs are induced and maintained in phrenic MNs after downregulation of early Hox transcriptional programs, suggesting that they may act to maintain phrenic MN properties. Despite convergence of transcriptional programs in the majority of MN subtypes as they progress from development to adulthood, phrenic MNs appear to sustain their unique identity, as they form a distinct cluster in adult scRNA-seq data[60]. While Klf6 is broadly expressed in all adult MNs[48], Klf5 expression appears to be more restricted, suggesting phrenic-specific functions. In addition to phrenic MNs, Klf5 is also expressed in hypaxial MNs in the thoracic spinal cord, indicating a conserved role in respiratory MN populations[19,61]. Klf family members differentially regulate the intrinsic ability of CNS axons to regenerate[62,63], raising the possibility that specific family members may be broadly involved in neuroprotection or degeneration in adulthood.

## Methods

### Mouse genetics

The loxP-flanked *Hoxa5*[64], *Hb9::GFP*[29], *Hoxc5-/-*[65], *Olig2::Cre*[66], and *ChAT(BAC)-eGFP* (*ChAT::GFP*)[31] lines were generated as previously described and maintained on a mixed background. Mouse colony maintenance and handling was performed in compliance with protocols approved by the Institutional Animal Care Use Committee of Case Western Reserve University. We have complied with all relevant ethical regulations for animal use. Mice were housed in a 12-hour light/dark cycle in cages containing no more than five animals at a time. Male and female mice and embryos from a mixed background were used interchangeably for all experiments.

### Immunohistochemistry and in situ hybridization

In situ hybridization and immunohistochemistry were performed as previously described[16,36], on tissue fixed for 2 hours in 4% paraformaldehyde (PFA) and cryosectioned at 16μm. Postnatal mice (P5.5-P16.5) were perfused with a solution of phosphate-buffered saline (PBS) and 4% PFA, followed by a 2-hour post-fixation at 4°C. In situ probes were generated from e12.5 cervical spinal cord cDNA libraries using PCR primers with a T7 RNA polymerase promoter sequence at the 5' end of the reverse primer. All probes generated were 300-1000 bp in length. Primers used for making probes are listed in Table 1.

Whole mounts of diaphragm muscles from e18.5 mice were stained as described[16]. The following antibodies were used: guinea pig anti-Hoxa5, guinea pig anti-Hoxc5[30], goat-anti-Sox5[16], rabbit anti-Hoxb5[16], guinea pig anti-FoxP1[22], goat anti-Scip (1:5000; Santa Cruz Biotechnology, RRID:AB_2268536), mouse anti-Islet1/2 (1:1000, DSHB, RRID:AB_2314683), rabbit anti-Lhx3[67], rabbit anti-Klf6 (1:1000, Santa Cruz Biotechnology, Cat# SC-7158), goat anti-ChAT (1:200, Millipore, RRID:AB_2079751), rabbit anti-neurofilament (1:1000; Synaptic Systems, RRID:AB_887743), and α-bungarotoxin, Alexa Fluor 555 conjugate (1:1000; Invitrogen, RRID:AB_2617152). Images were obtained with a Zeiss LSM 800 confocal

**Table 1 | List of primers used in this study**

| | |
|---|---|
| **Pou6f2_F** | **TGGGACCATTCCACTGATGC** |
| Pou6f2_R | TAATACGACTCACTATAGGGGGCTTGATCTTCTGGGCACT |
| Stat3_F | GAGCTGCACCTGATCACCTT |
| Stat3_R | TAATACGACTCACTATAGGGGCACTACCTGGGTCGGCTTC |
| Klf6_F | GCCTGGAGTTGGAACGCTAT |
| Klf6_R | TAATACGACTCACTATAGGGCCTGTCACAGTGGGAGCATT |
| Klf15_F | CCATCTCGGAGGATGACAGC |
| Klf15_R | TAATACGACTCACTATAGGGGCACAAACTTTGAGGGCAGG |
| Klf3_F | CTCTGAAGTTCCCGTCCCAC |
| Klf3_R | TAATACGACTCACTATAGGGTAGCATGTGGCGTTTCCTGT |
| Klf7_F | CTTTGCCATCCCTGGAGGAG |
| Klf7_R | TAATACGACTCACTATAGGGATAAACTTTCCGGCACCCGT |
| Klf5_F | CCCTCCAGTTCCGATAATTT |
| Klf5_R | TAATACGACTCACTATAGGGCAGAAGTGGATACGTCGCTT |
| Tshz1_F | AAGGAAAGCAGCTGGTGTTG |
| Tshz1_R | TAATACGACTCACTATAGGGTCCCGGTCTCCTCTGTCTTA |
| Tshz2_F | CAGTGAGAGACGGAATTGCG |
| Tshz2_R | TAATACGACTCACTATAGGGGTCTTCAATGCCCTCTCCCT |
| Tshz3_F | AAGCAGTCCCGGATGAGAAA |
| Tshz3_R | TAATACGACTCACTATAGGGTTCCCTTGCTCGTCCATCTT |
| Ebf1_F | ACCCTGAAATGTGCCGAGTA |
| Ebf1_R | TAATACGACTCACTATAGGGCAATGTCGGCAGCTCTCTTC |
| Ebf2_F | AGAGTTCTGCTGACACACGA |
| Ebf2_R | TAATACGACTCACTATAGGGGTCCGCAGCTCTTTTCAACA |
| Ebf3_F | TTGGTGTGGAGTGAGCTGAT |
| Ebf3_R | TAATACGACTCACTATAGGGCATAGCTTGCAGTCCGTTCC |
| Negr1_F | TCGAGTTTCCATTTCCACAT |
| Negr1_R | TAATACGACTCACTATAGGGTATAGTTGCCGAAGTGCTCC |
| Pcdh10_F | TGCTGGACATCAATGACAAC |
| Pcdh10_R | TAATACGACTCACTATAGGGCAGCACTCTCACCAGAACCT |
| VAChT_F | TTCATTGATCGCATGAGCTA |
| VAChT_R | TAATACGACTCACTATAGGGTACGCAAGAGCAATAGGACC |

microscope and analyzed with Zen Blue, ImageJ (Fiji), and Imaris (Bitplane).

### MN dissociation and fluorescence-activated cell sorting

Spinal cords were dissected from e12.5 embryos in a *Hb9::GFP* or *ChAT::GFP* background in ice cold PBS and collected in PBS. After spinning down, the pellets were dissociated with Papain Dissociation System (Worthington, Cat# LK003176) following the manufacturer's instructions. Briefly, tissue was enzyme digested for 30 min at 37°C with DNase (117 units/mL) and gently triturated. The single cell solution was centrifuged and then resuspended in PBS with 1% BSA and DNase. Dissociated cells were filtered through a 70 μm filter and subjected to fluorescence-activated cell sorting (FACS) on a BD Aria-SORP digital cell sorter with 85 μm nozzle to enrich for GFP positive cells. The cells were collected in a microtube containing 100 μL of PBS with 1% BSA.

### RNA-sequencing sample preparation and data analysis

Bulk RNA-seq for each condition was performed with two biological replicates. Spinal cords of e12.5 embryos were dissected to isolate cervical (C3-C8) and thoracic (T1-T8) segments in ice cold PBS. RNA was extracted using the PicoPure RNA isolation system (Arcturus, #KIT0204) with RIN > 8 via Tapestation analysis (Agilent). rRNA depleted libraries were prepared from 10 to 20 ng of total RNA using the KAPA stranded RNA-seq kit with Riboerase (KAPA, #KK8483) and amplified by 15 cycles of PCR. Single end 75 bp sequencing was performed on the Illumina HiSeq 2500 and generated a total of over 50 million reads per sample after filtering. 75 bp raw reads were analyzed by FASTQC for quality control. Reads were then aligned to mm39 genome assembly via STAR (v2.7.5b) aligner. The aligned transcripts were quantitated based on features in the GENCODE annotation database (GRCm39, version 111) by featureCounts (v2.0.1). Output gene count files were formatted into an experimental read count matrix in R. The batch effect, corresponding to sequencing performed at different time points, was then removed using ComBat-seq[68]. PCA analysis was performed and two biological replicates per condition that were close in PC space were added to the model design for DESeq2. Low count genes were filtered (total 10 count in all samples) prior to DESeq2 (v1.42.0). Differentially expressed genes (DEGs) were detected using DESeq2 with a P-value cutoff of <0.05. To normalize and visualize gene tracks, individual BAM files were converted to bigwig using bamCoverage provided by deepTools (v3.3.1) with a bin size of 10 bp and normalized by Reads Per Kilobase per Million mapped reads (RPKM). To identify the TFs in DEGs, DEGs were intersected with a list of transcription factors curated in Lambert et al.[46].

### ATAC-seq library preparation

Bulk ATAC-seq for each condition was performed with at least two biological replicates as previously described[69] and scaled down to half. Briefly, 25,000 FAC-sorted cells were centrifuged at 500 g for 6 min in a chilled centrifuge to form a pellet. The pellet was washed once in 25 μL of ice cold PBS, resuspended in 25 μL of cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% Igepal CA-630) and centrifuged at 500 g for 10 min at 4°C. The cell pellet was resuspended in transposition reaction mix (12.5 μl TD-Buffer, 1.25 μl Tn5, 11.25 μl water) (Nextera DNA Library Prep Kit, Illumina, Cat# 15028212) and incubated for 30 min at 37 °C. Immediately following the transposition reaction, purification was carried out using mini elute PCR Purification Kit (Qiagen, Cat# 27104). The appropriate number of amplification cycles was determined using qPCR reaction as described[69]. The PCR cycles were carried out with Illumina Nextera adapter primers using the NEBNext High Fidelity 2x Master Mix (NEB, Cat# M0541S) using the following PCR program: (1) 5 min at 72°C, (2) 30 s at 98 °C, (3) 10 s at 98 °C, (4) 30 s at 63 °C, (5) 1 min at 72°C, and (6) repeat steps 3–5 with total cycles <12. Final PCR products were cleaned using PCRClean Dx beads (Aline Biosciences, Cat# C-1003) and assessed for quality using a Bioanalyzer. The libraries were sequenced on an Illumina NextSeq 550 (paired-end 75 bp) at the Genomics Core Facility at Case Western Reserve University.

### ATAC-seq data processing and analysis

ATAC-seq data were processed using the standardized uniform Encyclopedia of DNA Elements (ENCODE) pipeline from the ENCODE consortium[70]. Briefly, FastQC (v0.11.9) was used to check the pre-alignment read quality. FASTQ files from ATAC-seq reads were mapped to UCSC mm10 with Bowtie2 (v2.3.4.3). All unmapped reads, non-uniquely mapped reads, PCR duplicates and ChrM reads were removed using Samtools (v1.9). Peaks were called using MACS2 (v2.2.4) with parameters "--nomodel --shift 37 --ext 73 --pval 1e-2 -B --SPMR --call-summits". Peaks overlapping with the blacklist region defined by ENCODE were removed using Bedtools (v2.29.0). Next, replicated peaks in each condition were intersected using Bedtools (intersect) to identify open chromatin regions overlapping by at least 1 bp and defined as replicated peaks. Replicated peaks were annotated using HOMER (v4.10), which assigns each peak to the nearest gene transcriptional start site (TSS). Gene promoters are defined as ± 1 kb from the TSS. To identify differential peaks, FeatureCounts was used to obtain count data from the resulting ATAC-seq BAM files. Count data for all replicates and experimental conditions were combined into a single count matrix in R. The consensus peaks were identified as the peaks that were present in at least two samples. The count matrix was subsequently used to identify differential peaks with the R package DEseq2[71]. PCA was performed using plotPCA function within DEseq2 on Variance Stabilizing Transformation (VST)-transformed data. Proximal and distal peaks were defined by associating differential ATAC-seq peak distances to annotated TSS (ChIPseeker). Peaks that were at least 3 kb away from the annotated TSS were assigned as distal ATAC-seq peaks, while all others were assigned as proximal. To visualize the ATAC-seq signal in the UCSC genome browser, samples were normalized to 1x genomic coverage, also known as Reads per Genome Coverage (RPGC).

### Motif analysis

HOMER (v4.10) was used to perform de novo motif enrichment[32]. Motif analysis on ChIP-seq data was performed using a fixed 200 bp window around the peak center. Motif analysis on ATAC-seq data was performed using either a fixed 200 bp or 500 bp window around the peak center on differentially accessible chromatin. In both cases, the HOMER findMotifsGenome.pl command was used to perform de novo analysis against background sequences generated by HOMER that match the GC content. The top-scoring motifs, along with their p-value and enrichment, are shown.

### Footprinting analysis

To analyze footprinting signatures in ATAC-seq data the TOBIAS package[37] was used. All replicates from each condition were merged into one .bam file using bedtools. Peaks were called using MACS2 with parameters "--nomodel --qvalue 0.01 --keep-dup all". Peak files were associated with motifs from JASPAR CORE Vertebrates collection 2022[72]. Merged BAM files were processed using ATACorrect to correct for Tn5 bias. Footprint scores were calculated using FootprintScores, and differential footprinting analysis was performed using BINDetect.

### Go enrichment

The enrichGO function from the clusterProfiler (v4.8.2) package was used to perform GO term analysis of enriched biological processes and generate the graphs with maximum of 500 genes set for each category. The top ten significant GO terms were plotted and ordered by the number of gene counts in each category.

### Single cell RNA-sequencing (scRNA-seq) re-analysis

The filtered matrix output from the Cell Ranger pipeline for rostral samples was obtained from the Gene Expression Omnibus repository with accession code GSE183759[33]. Seurat package (v4.4.0) was used to perform quality filtering, normalization, dimensionality reduction, and cell clustering. Briefly, cells were evaluated for quality, and those with gene counts between 1000 and 5300, UMI counts below 30500, and mitochondrial counts under 10% were

kept for further analysis. After filtering, 5460 cells were retained for downstream analysis. The resulting digital data matrices were then processed using a SCT transformation[73] to perform normalization, scaling, and identification of variable features with mitochondrial reads regressed out. MNs were separated by the expression of common MN markers such as Mnx1 or cholinergic markers such as ChAT or Slc18a3 or Slc5a7. Only the cells expressing MN markers were considered for downstream analysis leading to a total of 5011 cells. To identify cell clusters, Uniform Manifold Approximation and Projection (UMAP) was used with the first 30 principle components. Cells were clustered using FindClusters function (resolution = 0.3) and visualized using UMAP. Cell identities were assigned using known markers. Clusters that were close to each other in UMAP space expressing LMC (*FoxP1*, *Aldh1a2*) and MMC (*Mecom*, *Lhx3*) markers were merged to create a new cluster ID and defined as LMC and MMC clusters. Furthermore, conserved markers for LMC and MMC clusters were generated by using Findconservedmarker function with logFC thresholds of 0.25. To identify ATAC-seq peaks associated with MN clusters in the cervical spinal cord, the conserved marker genes obtained from scRNA-seq for LMC and MMC were intersected with the genes associated with the ATAC-seq peaks.

### ChIP-sequencing (ChIP-seq) sample preparation and analysis

e12.5 mouse cervical spinal cords were dissected and flash frozen in liquid nitrogen. Multiple embryonic spinal cords (~10 litters) were combined to perform one biological replicate due to the limitations of the starting material. The tissue samples, along with antibodies, rabbit anti-Hoxa5[30] and rabbit anti-Pbx1 (Cell Signaling Technology, RRID:AB_2160295) were sent to Active Motif for chromatin isolation and sonication, ChIP assay, library preparation, library QC, and Next-Generation sequencing on the Illumina platform and analysis. 75-nucleotide sequence reads generated by Illumina sequencing (NextSeq 500) were mapped to the mm10 genome using Bowtie2 (v2.3.5.1) with flag "--local". Reads were then sorted and indexed with samtools (v1.10) and filtered for blacklist regions defined by ENCODE. MACS2 was used to call narrowpeak with a flag "--nomodel --ext 220" against the input control. The extension length was estimated using deeptools' (v3.3.1) cross correlation plot. Macs2 peaks were further filtered using the adj P cutoff of e-7 and the signal value > 5. Peaks were annotated in R with the ChIPseeker package, which assigns each peak to the nearest gene's TSS and gene promoters were defined as ± 1.0 kb from the TSS. De novo motif enrichment was performed on genome-wide peak sets.

### Plasmid construction for co-immunoprecipitation (co-IP) and electroporation

To create expression vectors for co-IP experiments, RNA extracted from mouse spinal cord was converted to cDNA and used to amplify Hoxa5, Pbx1, and Scip using custom oligonucleotides with HA, Myc, and V5-tags. PCR amplified products and cloning vector (pcDNA3.1-myc-HisA) were digested to create compatible sites for ligation and transformed into NEB10 beta competent bacteria (NEB, Cat# C3019H). To create plasmids for chick electroporation, mouse Hoxa5 and Scip cloned into the pcDNA3.1 vector were used to amplify Hoxa5 and Scip and inserted into pCAG-tdTomato (Addgene, Cat #83029), a vector with the chick β-actin promoter/CMV enhancer. The complete length of cloned plasmids was sequenced at Eurofins and verified by mapping to the respective mRNAs using the UCSC mouse reference genome.

### Co-immunoprecipitation (co-IP) assays

HEK293 cells were transfected using Lipofectamine 3000 (Invitrogen, Cat# L3000008) according to the manufacturer's instructions. After 48 hours, cells were washed once in ice cold PBS and harvested in 1X RIPA buffer (Cell Signaling, Cat# 9806). Co-IP assay was carried out using protein A/G PLUS-Agarose beads (Santa Cruz, Cat# 2003). Briefly, 600 µg of total cell lysate was precleared with 20 µl of agarose beads for 30 min. For co-IP, 200 µg of precleared protein was incubated with 2 µg of anti-V5 (Santa Cruz Biotechnology, Cat# sc-271944, RRID:AB_10650278, Fig. 4c, e, g, i) or anti-Myc (Cell Signaling Technology, Cat# 2276, RRID:AB_331783, Figs. 4b, 4d, 4f, S4c) and incubated for 1 hour on a rocker at 4 °C. To conjugate beads with the antibodies bound to the protein, 20 µl of agarose beads were added and incubated at 4°C overnight. Protein complex bound beads were washed 3 times with RIPA and 2 times with PBS and the pellet was resuspended in 40 µl of 1x sample buffer and boiled for 3 min. 25 µL of the immunoprecipitated aliquots and 5% of total lysate (input control) were run on a standard SDS-PAGE gel. The gels were then transferred onto a PVDF membrane (BioRad, Cat# 1620177) using a wet transfer system and blocked by incubation with 3% BSA in TBST (TBS with 0.1% Tween-20). Membranes were probed with anti-HA, anti-V5, anti-Myc, anti-Pbx1 (Cell Signaling Technology, RRID:AB_2160295) or anti-Scip (Santa Cruz Biotechnology, RRID:AB_2268536). Blotted membranes were scanned using Odyssey infrared imaging system (Li-COR).

For in vivo co-IP, cervical tissue from e12.5 mouse embryos was washed once in ice cold PBS and homogenized in RIPA buffer (60 µL/embryo) using a disposable pestle. The lysate was incubated at 4 °C for 30 min and then clarified by spinning down at 4 °C for 10 min at 10,000 RPM. 200 µg of precleared lysate was incubated with 2 µg of goat anti-Scip (Santa Cruz Biotechnology, RRID:AB_2268536) and incubated for 1 h on a rocker at 4 °C. To conjugate beads with the antibodies bound to the protein, 20 µl of agarose beads were added and incubated at 4 °C overnight. Protein complex bound beads were washed 5 times in PBS and the pellet was resuspended in 40 µl of 1× sample buffer and boiled for 3 min. 20% of total lysate was used as input control for running a standard SDS-PAGE western blot. After transfer, the blot was blocked and probed with rabbit anti-Hoxa5, washed, and re-probed with rabbit anti-Scip (RRID:AB_2631304).

### In ovo electroporation

Electroporation was performed by introducing a DNA solution into the lumen of the neural tube of specific pathogen-free (SPF) chicken embryos (AVS Bio, Cat#10100326) at Hamburger-Hamilton stages 14–16[74] using 5 × 50 msec pulses at 25 V, with electrodes placed horizontally across the longitudinal axis of the embryo to achieve unilateral electroporation of the desired construct mixture. The DNA solution was composed of the relative ratios of each construct diluted in TE buffer with 0.5% Fast Green to aid injection visualization. The construct concentrations were adjusted to obtain a final ratio of 2:2:1 for Hoxa5:Scip:EGFP in which the total DNA electroporated per egg was 1.1 µg/µl. Electroporated embryos were incubated at 37°C for 3 days and analyzed at stages 25-26. We limited our analysis to rostrocaudal levels showing Hox5/FoxP1 co-expression and containing electroporated cells, as FoxP1 expression spans the entire brachial spinal cord.

### Statistics and reproducibility

The programs used for data analysis such as MACS2 for peak calling, DEseq2 for differential analysis, Homer for motif- enrichment analysis, clusterprofiler for GO term enrichment analysis, and Tobias for footprinting score analysis use algorithms that provide their own p values, q values, and/or FDR. The data was primarily analyzed in R (v 4.3.1) and the R scripts used for data analysis are freely available upon request. For electroporation experiments, data are presented as box and whisker plots with each dot representing data from one chick embryo. Small open squares in box and whisker plots represent the mean, while a horizontal line represents the median. P-values were calculated using a one-way ANOVA with Dunnett's multiple comparisons test. $p < 0.05$ was considered to be statistically significant, where * $p < 0.05$. For western blot quantification, experiments were performed independently in triplicates and band intensity was quantified by ImageJ software (v. 1.53). Protein interacting with the bait protein was quantified and normalized to the amount of bait protein and presented as a box and whisker plot. For co-IP experiments, statistical significance was determined using unpaired Student's t-test. For the correlation analysis of ATAC-seq and RNA-seq, Pearson correlation test was used.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

All sequencing data produced for this study are available at the Gene Expression Omnibus (GEO). Accession codes: GSE273362 (ATAC-seq), GSE273363 (ChIP-seq), GSE273364 (RNA-seq). Plasmids generated for electroporation experiments have been deposited to Addgene (ID numbers 225109-225110). All other plasmids are available from the corresponding author, P.P., upon request. Source data for graphs are provided as a supplementary data file. Uncropped blot images are provided (Figure S6). All other data is available from the corresponding author, PP, upon request.

## Code availability

The R scripts used for data analysis are freely available from the corresponding author, PP, upon request.

Received: 17 December 2023; Accepted: 3 September 2024;
Published online: 17 September 2024

## References

1. Catela, C., Shin, M. M. & Dasen, J. S. Assembly and function of spinal circuits for motor control. *Annu Rev. Cell Dev. Biol.* **31**, 669–698 (2015).
2. Shirasaki, R. & Pfaff, S. L. Transcriptional codes and the control of neuronal identity. *Annu Rev. Neurosci.* **25**, 251–281 (2002).
3. Philippidou, P. & Dasen, J. S. Hox genes: choreographers in neural development, architects of circuit organization. *Neuron* **80**, 12–34 (2013).
4. Rezsohazy, R., Saurin, A. J., Maurel-Zaffran, C. & Graba, Y. Cellular and molecular insights into Hox protein action. *Development* **142**, 1212–1227 (2015).
5. Zandvakili, A. & Gebelein, B. Mechanisms of Specificity for Hox Factor Activity. *J. Dev. Biol.* **4**, 16 (2016).
6. Lacombe, J. et al. Genetic and functional modularity of Hox activities in the specification of limb-innervating motor neurons. *PLoS Genet* **9**, e1003184 (2013).
7. Noyes, M. B. et al. Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell* **133**, 1277–1289 (2008).
8. Gehring, W. J. et al. Homeodomain-DNA recognition. *Cell* **78**, 211–223 (1994).
9. Berger, M. F. et al. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**, 1266–1276 (2008).
10. Affolter, M., Slattery, M. & Mann, R. S. A lexicon for homeodomain-DNA recognition. *Cell* **133**, 1133–1135 (2008).
11. Mann, R. S., Lelli, K. M. & Joshi, R. Hox specificity unique roles for cofactors and collaborators. *Curr. Top. Dev. Biol.* **88**, 63–101 (2009).
12. Hueber, S. D. et al. Comparative analysis of Hox downstream genes in Drosophila. *Development* **134**, 381–392 (2007).
13. Slattery, M. et al. Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* **147**, 1270–1282 (2011).
14. Hanley, O. et al. Parallel Pbx-Dependent Pathways Govern the Coalescence and Fate of Motor Columns. *Neuron* **91**, 1005–1020 (2016).
15. Carnesecchi, J. et al. Multi-level and lineage-specific interactomes of the Hox transcription factor Ubx contribute to its functional specificity. *Nat. Commun.* **11**, 1388 (2020).
16. Philippidou, P., Walsh, C. M., Aubin, J., Jeannotte, L. & Dasen, J. S. Sustained Hox5 gene activity is required for respiratory motor neuron development. *Nat. Neurosci.* **15**, 1636–1644 (2012).
17. Catela, C., Shin, M. M., Lee, D. H., Liu, J. P. & Dasen, J. S. Hox Proteins Coordinate Motor Neuron Differentiation and Connectivity Programs through Ret/Gfralpha Genes. *Cell Rep.* **14**, 1901–1915 (2016).
18. Joshi, R. et al. Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* **131**, 530–543 (2007).
19. Machado, C. B. et al. Reconstruction of phrenic neuron identity in embryonic stem cell-derived motor neurons. *Development* **141**, 784–794 (2014).
20. Bermingham, J. R. Jr. et al. Tst-1/Oct-6/SCIP regulates a unique step in peripheral myelination and is required for normal respiration. *Genes Dev.* **10**, 1751–1762 (1996).
21. Rousso, D. L., Gaber, Z. B., Wellik, D., Morrisey, E. E. & Novitch, B. G. Coordinated actions of the forkhead protein Foxp1 and Hox proteins in the columnar organization of spinal motor neurons. *Neuron* **59**, 226–240 (2008).
22. Dasen, J. S., De Camilli, A., Wang, B., Tucker, P. W. & Jessell, T. M. Hox repertoires for motor neuron diversity and connectivity gated by a single accessory factor, FoxP1. *Cell* **134**, 304–316 (2008).
23. Beh, C. Y. et al. Roles of cofactors and chromatin accessibility in Hox protein target specificity. *Epigenetics Chromatin* **9**, 1 (2016).
24. Porcelli, D., Fischer, B., Russell, S. & White, R. Chromatin accessibility plays a key role in selective targeting of Hox proteins. *Genome Biol.* **20**, 115 (2019).
25. Paul, R., Peraldi, R. & Kmita, M. The pioneering function of the hox transcription factors. *Semin Cell Dev. Biol.* **152-153**, 85–92 (2024).
26. Desanlis, I. et al. HOX13-dependent chromatin accessibility underlies the transition towards the digit development program. *Nat. Commun.* **11**, 2491 (2020).
27. Amandio, A. R., Lopez-Delisle, L., Bolt, C. C., Mascrez, B. & Duboule, D. A complex regulatory landscape involved in the development of mammalian external genitals. *Elife* **9**, https://doi.org/10.7554/eLife.52962 (2020).
28. Bulajic, M. et al. Differential abilities to engage inaccessible chromatin diversify vertebrate Hox binding patterns. *Development* **147**, dev194761 (2020).
29. Wichterle, H., Lieberam, I., Porter, J. A. & Jessell, T. M. Directed differentiation of embryonic stem cells into motor neurons. *Cell* **110**, 385–397 (2002).
30. Dasen, J. S., Tice, B. C., Brenner-Morton, S. & Jessell, T. M. A Hox regulatory network establishes motor neuron pool identity and target-muscle connectivity. *Cell* **123**, 477–491 (2005).
31. Tallini, Y. N. et al. BAC transgenic mice express enhanced green fluorescent protein in central and peripheral cholinergic neurons. *Physiol. Genomics* **27**, 391–397 (2006).
32. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).
33. Liau, E. S. et al. Single-cell transcriptomic analysis reveals diversity within mammalian spinal motor neurons. *Nat. Commun.* **14**, 46 (2023).
34. Allaway, K. C. et al. Genetic and epigenetic coordination of cortical interneuron development. *Nature* **597**, 693–697 (2021).
35. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 e3529 (2021).
36. Vagnozzi, A. N. et al. Phrenic-specific transcriptional programs shape respiratory motor output. *Elife* **9**, e52859 (2020).
37. Bentsen, M. et al. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat. Commun.* **11**, 4267 (2020).
38. Schoch, C. L. et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database (Oxf.)* **2020**, baaa062 (2020).
39. Bobola, N. & Sagerstrom, C. G. TALE transcription factors: Cofactors no more. *Semin Cell Dev. Biol.* **152-153**, 76–84 (2024).
40. Chang, C. P. et al. Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. *Genes Dev.* **9**, 663–674 (1995).
41. Chan, S. K., Jaffe, L., Capovilla, M., Botas, J. & Mann, R. S. The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. *Cell* **78**, 603–615 (1994).
42. Phelan, M. L., Rambaldi, I. & Featherstone, M. S. Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif. *Mol. Cell Biol.* **15**, 3989–3997 (1995).

43. Knoepfler, P. S. & Kamps, M. P. The pentapeptide motif of Hox proteins is required for cooperative DNA binding with Pbx1, physically contacts Pbx1, and enhances DNA binding by Pbx1. *Mol. Cell Biol.* **15**, 5811–5819 (1995).
44. Jung, H. et al. Global control of motor neuron topography mediated by the repressive actions of a single hox gene. *Neuron* **67**, 781–796 (2010).
45. Catela, C., Chen, Y., Weng, Y., Wen, K. & Kratsios, P. Control of spinal motor neuron terminal differentiation through sustained Hoxc8 gene activity. *Elife* **11**, e70766 (2022).
46. Lambert, S. A. et al. The Human Transcription Factors. *Cell* **175**, 598–599 (2018).
47. Hammelman, J., Patel, T., Closser, M., Wichterle, H. & Gifford, D. Ranking reprogramming factors for cell differentiation. *Nat. Methods* **19**, 812–822 (2022).
48. Patel, T. et al. Transcriptional dynamics of murine motor neuron maturation in vivo and in vitro. *Nat. Commun.* **13**, 5427 (2022).
49. Velasco, S. et al. A Multi-step Transcriptional and Chromatin State Cascade Underlies Motor Neuron Programming from Embryonic Stem Cells. *Cell Stem Cell* **20**, 205–217 e208 (2017).
50. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994–1004 (2012).
51. Baeza, M. *et al*. Inhibitory activities of short linear motifs underlie Hox interactome specificity in vivo. *Elife* **4**, https://doi.org/10.7554/eLife.06034 (2015).
52. Dard, A. et al. Human HOX Proteins Use Diverse and Context-Dependent Motifs to Interact with TALE Class Cofactors. *Cell Rep.* **22**, 3058–3071 (2018).
53. Sanchez-Higueras, C. et al. In vivo Hox binding specificity revealed by systematic changes to a single cis regulatory module. *Nat. Commun.* **10**, 3597 (2019).
54. Feng, S. et al. Transcription factor paralogs orchestrate alternative gene regulatory networks by context-dependent cooperation with multiple cofactors. *Nat. Commun.* **13**, 3808 (2022).
55. Bridoux, L. et al. HOX paralogs selectively convert binding of ubiquitous transcription factors into tissue-specific patterns of enhancer activation. *PLoS Genet* **16**, e1009162 (2020).
56. Li, Y. et al. Establishment and maintenance of motor neuron identity via temporal modularity in terminal selector function. *Elife* **9**, e59464 (2020).
57. Feng, W. et al. A terminal selector prevents a Hox transcriptional switch to safeguard motor neuron identity throughout life. *Elife* **9**, e50065 (2020).
58. Hobert, O. & Kratsios, P. Neuronal identity control by terminal selectors in worms, flies, and chordates. *Curr. Opin. Neurobiol.* **56**, 97–105 (2019).
59. Kitt, M. M. et al. An adult-stage transcriptional program for survival of serotonergic connectivity. *Cell Rep.* **39**, 110711 (2022).
60. Alkaslasi, M. R. et al. Single nucleus RNA-sequencing defines unexpected diversity of cholinergic neuron types in the adult mouse spinal cord. *Nat. Commun.* **12**, 2471 (2021).
61. Wang, W., Cho, H., Lee, J. W. & Lee, S. K. The histone demethylase Kdm6b regulates subtype diversification of mouse spinal motor neurons during development. *Nat. Commun.* **13**, 958 (2022).
62. Moore, D. L. et al. KLF family members regulate intrinsic axon regeneration ability. *Science* **326**, 298–301 (2009).
63. Wang, Z. et al. KLF6 and STAT3 co-occupy regulatory DNA and functionally synergize to promote axon growth in CNS neurons. *Sci. Rep.* **8**, 12565 (2018).
64. Tabaries, S., Lemieux, M., Aubin, J. & Jeannotte, L. Comparative analysis of Hoxa5 allelic series. *Genesis* **45**, 218–228 (2007).
65. McIntyre, D. C. et al. Hox patterning of the vertebrate rib cage. *Development* **134**, 2981–2989 (2007).
66. Dessaud, E. et al. Interpretation of the sonic hedgehog morphogen gradient by a temporal adaptation mechanism. *Nature* **450**, 717–720 (2007).
67. Tsuchida, T. et al. Topographic organization of embryonic motor neurons defined by expression of LIM homeobox genes. *Cell* **79**, 957–970 (1994).
68. Zhang, Y., Parmigiani, G. & Johnson, W. E. *ComBat-seq*: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform.* **2**, lqaa078 (2020).
69. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
70. Landt, S. G. et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22**, 1813–1831 (2012).
71. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
72. Castro-Mondragon, J. A. et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **50**, D165–D173 (2022).
73. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
74. Hamburger, V. & Hamilton, H. L. A series of normal stages in the development of the chick embryo. *J. Morphol.* **88**, 49–92 (1951).


## Acknowledgements

We thank Heather Broihier, Evan Deneris, Britton Sauerbrei, Ashleigh Schaffer and Helen Miranda for discussions and comments on the manuscript. We thank Clay Spencer and Evan Deneris for assistance and helpful discussions on bioinformatics analysis, Ayana Sawai, Jeremy Dasen and Lynn Landmesser for in ovo-electroporation protocols, Jeremy Dasen for Hox5, Sox5, FoxP1 and Lhx3 antibodies and Ganapati Mahabaleshwar for the Klf6 antibody. We also thank the Cytometry and Imaging core at Case Comprehensive Cancer Center (P30CA043703) for assistance with FACS and Active Motif for assistance performing ChIP-sequencing. This work was funded by NIH R01NS114510 to PP, F31NS120699 and T32GM008056 to RKC, F30HD096788 and T32GM007250 to ANV. PP is the Weidenthal Family Designated Professor in Career Development.

## Author contributions

R.K.C. and P.P. conceived and designed the study, R.K.C., R.L.B., A.N.V., M.L. and P.P. performed experiments and analyzed data, L.J. provided *Hoxa5 floxed* mice, R.K.C., R.L.B., and P.P. wrote the paper with input from all authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s42003-024-06835-w.

**Correspondence** and requests for materials should be addressed to Polyxeni Philippidou.

**Peer review information** *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Benjamin Bessieres. A peer review file is available.

**Reprints and permissions information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024