scientific reports

Check for updates

OPEN

An integrated transcriptomic approach for identifying enhancers in limb motor pools

In Seo Yeo[1], Dojin Jeong[1], Yunjeong Lee[1], Jae-Woo Ahn[1], Se-Eun Bae[2] & Mi-Ryoung Song[1,3]✉

Motor neurons (MNs), which control body movement, consist of distinct motor pools, that target specific muscles. The precise delineation of these motor pools necessitates an understanding of their specialized gene expression profiles. By integrating single-cell transcriptomics and epigenomics, using scATAC-seq, scRNA-seq and histone modification datasets, we identified cis-regulatory elements (CREs) specific to five MN-lineage clusters, which exhibit distinct gene ontology terms, transcription factor (TF) binding motifs, and disease-associated traits across them. Notably, LIM-homeodomain (LIM-HD) TF motifs and their footprints were differentially enriched in corresponding individual motor column CREs, including a combined Isl1 and Lhx3 motif consistent with their known role in MN identity specification. Within the genomic locus of *Etv4*, a gene essential for limb motor pools, we identified three enhancer candidates among CREs enriched in LMC neurons. Functional assays demonstrated that these enhancers exhibited robust and specific activity in ETV4-expressing motor pools when electroporated into chicken neural tubes. This study demonstrates that an integrated in silico analysis of epigenetic and transcriptomic data at the single-cell level can predict putative CREs governing the identity of distinct MN subgroups, providing insights into motor pool diversification and gene regulation.

Spinal motor neurons (MNs) innervate various target muscles throughout the body, organized into motor columns and motor pools within the spinal cord. During spinal cord development, MNs arise in the ventral region of the spinal cord and undergo differentiation driven by complex gene regulatory networks. MN progenitors (pMN) emerge in response to a gradient of morphogen sonic hedgehog (SHH), which activates transcriptional cascades to establish the pMN domain where pMNs reside. Newborn postmitotic MNs, originating from the pMN, first acquire a uniform pan-MN identity before undergoing a diversification process, forming distinct motor pools, each innervating a specific muscle. A variety of transcription factors (TFs) orchestrate the developmental progression of MNs. This diversification requires a precisely coordinated interplay between numerous cis- and trans-regulatory elements among individual differentiating MNs over time. For instance, LIM-homeodomain (LIM-HD) TFs ISL1 and LHX3 establish a LIM-code that initially provides nascent postmitotic MN fates and then later defines motor columnar identity, while Erythroblast Transformation Specific (ETS) class TFs such as ETV1/ER81 and ETV4/PEA3 assign motor pool identities[1,2]. Despite our knowledge of key trans-acting players for MN diversification, our understanding of their counterparts cis-regulatory elements (CREs), such as enhancers, promoters, silencers, and insulators in developing MNs, remains limited.

Recent advancements in single-cell transcriptomics, including single-cell RNA sequencing (scRNA-seq) and single-cell ATAC sequencing (scATAC-seq), have enabled the profiling of dynamic changes in gene expression and chromatin accessibility within individual cells on a genome-wide scale. Integrating single-cell transcriptional and epigenetic states has broadened our understanding of MN diversification and allowed us to even predict interactions between cis- and trans-regulatory elements under both developmental and pathological conditions. Nevertheless, the developmental mechanisms underlying the emergence of MN lineages in the spinal cord, particularly with respect to epigenetics, remain largely unknown. For instance, it is unclear whether the formation of motor pools relies on specific epigenetic mechanisms to regulate spatially and temporally restricted gene expression programs. Given the intricate functional specialization of motor pools in targeting distinct muscles, understanding how epigenetic landscapes, such as enhancer-promoter interactions and chromatin accessibility patterns, contribute to this diversification is critical. Motor pools represent a unique system to explore how CREs

[1]Department of Life Sciences, Gwangju Institute of Science and Technology, Oryong-dong, Buk-gu, Gwangju 61005, Republic of Korea. [2]KAIST stem cell center, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea. [3]GIST InnoCORE AI-Nano Convergence Initiative for Early Detection of Neurodegenerative Diseases, Gwangju Institute of Science and Technology, Cheomdangwagiro 123, Gwangju 61005, Republic of Korea. ✉email: msong@gist.ac.kr

coordinate with TFs to establish precise neuronal identities, bridging developmental biology and epigenetics in the context of functional neural circuits.

In this study, we combined transcriptional and epigenetic data of developing MNs at single-cell resolution to identify a set of CREs specific to developing MNs. CREs differentially accessible in different motor columns were enriched with TF binding motifs and gene ontology (GO) terms consistent with individual MN clusters. These MN-specific CREs were precise enough to predict putative motor pool enhancers for representative marker genes, showing a strong correlation with other epigenetic and transcriptomic data from embryonic spinal cord tissues and stem cell-derived MNs. The activities of the candidate motor pool enhancers were validated both in vitro and in vivo, supporting the robustness of our approach. Together, our analyses provide compelling evidence that specialized MN enhancers play a crucial role in MN diversification. These enhancers can be accurately predicted in silico based on spatiotemporal dynamics of large-scale single-cell transcriptomic and epigenetic data. This study enhances our understanding of the gene regulatory networks involved in MN development and potentially progression of MN diseases as well.

Results

Single-cell analysis of chromatin accessibility in developing mouse MNs

Recent advancements in transcriptomics technology have enabled the simultaneous exploration of chromatin landscapes and transcription profiles at the single-cell level across a variety of tissues using scATAC-seq and scRNA-seq in identical samples. However, applying this technology to dynamic and rare populations, such as developing MNs in the spinal cord, has been challenging. As an alternative strategy, comparing and integrating scATAC-seq and scRNA-seq data from two different sources can provide a similar extent of information. To investigate into chromatin accessibility in developing MNs, we examined a previously published integrated scATAC-seq and scRNA-seq dataset from the cervical and thoracic regions of mouse neural tubes at E9.5 to E13.5 [3]. The scATAC-seq data obtained from 18,779 cells underwent dimensional reduction and clustering using Signac, resulting in a total of seven cell clusters (Supplementary Fig. 1, Supplementary Fig. 2a, Supplementary Data 1). We then adopted cell type labels from another previously published scRNA-seq dataset analyzed from E9.5 to E13.5 mouse cervical and thoracic spinal cords[4]. Integration of scATAC-seq with the scRNA-seq dataset using the Seurat package successfully defined individual cell clusters as neural progenitors, neurons, neural crest, mesoderm, dorsal root ganglia and blood cells[3,4]. Accordingly, scATAC-seq peaks around cell-type markers such as *Sox2* (neural progenitors), *Tubb3* and *Elavl3* (neurons) were specifically defined in certain clusters (Supplementary Fig. 2b, d). To focus on our study on developing spinal cords, we further divided the neuronal progenitors and neurons into 23 cell clusters and transferred cell subtype annotations obtained from scRNA-seq data to them, identifying 11 progenitors and 12 neuronal subtypes including pMNs and MNs within the developing spinal cords (Supplementary Fig. 2c, Supplementary Data 1, Supplementary Data 2). Comparisons between the scRNA-seq and scATAC-seq datasets revealed similar overall cell composition and cluster transcriptional and chromatin accessibility profiles. Notably, gene activity scores and imputed gene expression levels displayed a robust correlation (Supplementary Fig. 2e).

Further, we subdivided a total of 1,004 MNs into five clusters including pMN, early differentiated MN (early MN), medial motor column (MMC)/phrenic motor column (PMC), lateral motor column (LMC) and preganglionic motor column (PGC) identities[4–7] (Fig. 1a, b, Supplementary Data 1). We eliminated pMN progenitors from E12.5 and E13.5 mouse embryos for our analysis since this stage is beyond the peak period of MN production (E9.5 to E10.0) and these remnant progenitors mainly produce oligodendrocytes starting around E12.5 [1,8,9]. Thus, we successfully defined the chromatin landscapes of major developing cell types comprising MNs. The scaled imputed RNA expression of representative markers for each MN cluster was consistent with the corresponding scATAC-seq signal (Fig. 1c). Cell type markers examined were as follows: *Sox2*, *Olig2*, *Hes5*, *Hes6*, *Neurog2*, *Neurod4* for pMN; *Cldn3*, *Pou2f2* for early MN; *Foxp1*, *Aldh1a2*, *Hoxc4*, *Hoxc5* for LMC; *Foxp1*, *Hoxc9*, *Nos1* for PGC; *Alcam*, *Pou3f1*, *Lhx3*, *Mecom* for MMC/PMC (Supplementary Data 2). A UMAP of scATAC-seq data by developmental stages showed that MNs sequentially arose from pMN progenitors from E9.5 to E13.5 (Fig. 1d). We next examined the chromatin landscapes near major cell type markers (Fig. 1e). The CREs around *Olig2* were selectively accessible in pMNs. Similarly, cell type-specific accessibility was found in other cell type marker genes including *Cldn3* in early MNs, *Lhx3* in MMC/PMC, *Aldh1a2* in LMC, and *Nos1* in PGC.

Identification and characterization of putative MN-specific cis-regulatory elements in developing MNs

To investigate the gene regulatory mechanisms underlying MN development, we sought to identify MN-specific CREs. We began by analyzing a scRNA-seq dataset from Delile et al.[4] performing marker gene analysis to identify 3,509 genes with transcripts enriched in MN-lineage cells, including pMN and MN subtypes, as defined by differential gene expression analysis (Supplementary Data 3). We next identified 48,922 putative CREs located within the loci of these 3,509 MN-enriched genes. To define regulatory elements with MN-specific activity, we compared chromatin accessibility at these CREs between MN-lineage cells and other neural clusters[3]. This analysis uncovered a total of 19,269 MN-specific CREs (referred to as MN-CREs hereafter) summing all CREs enriched in pMN (6,252 CREs), early MN (3,601 CREs), LMC (2,157 CREs), PGC (3,422 CREs) and MMC/PMC (3,837 CREs) (Supplementary Fig. 1, Supplementary Data 4). Most MN-CREs (84.43%) were located distal to the promoter, more than 2 kb away from the nearest transcription start site (TSS), primarily within intronic and distal intergenic regions. This distal positioning suggests that MN-CREs are predominantly enhancers rather than proximally located promoter elements (Fig. 1f)[3,10]. These findings align with the established importance of enhancers in fine-tuning gene expression during MN differentiation and lineage specification.

GO analysis and the enriched motif analysis revealed significant enrichment of functions and TF motifs related to MN development in these clusters (Fig. 2a, Supplementary Data 5, 6). The TF motifs identified

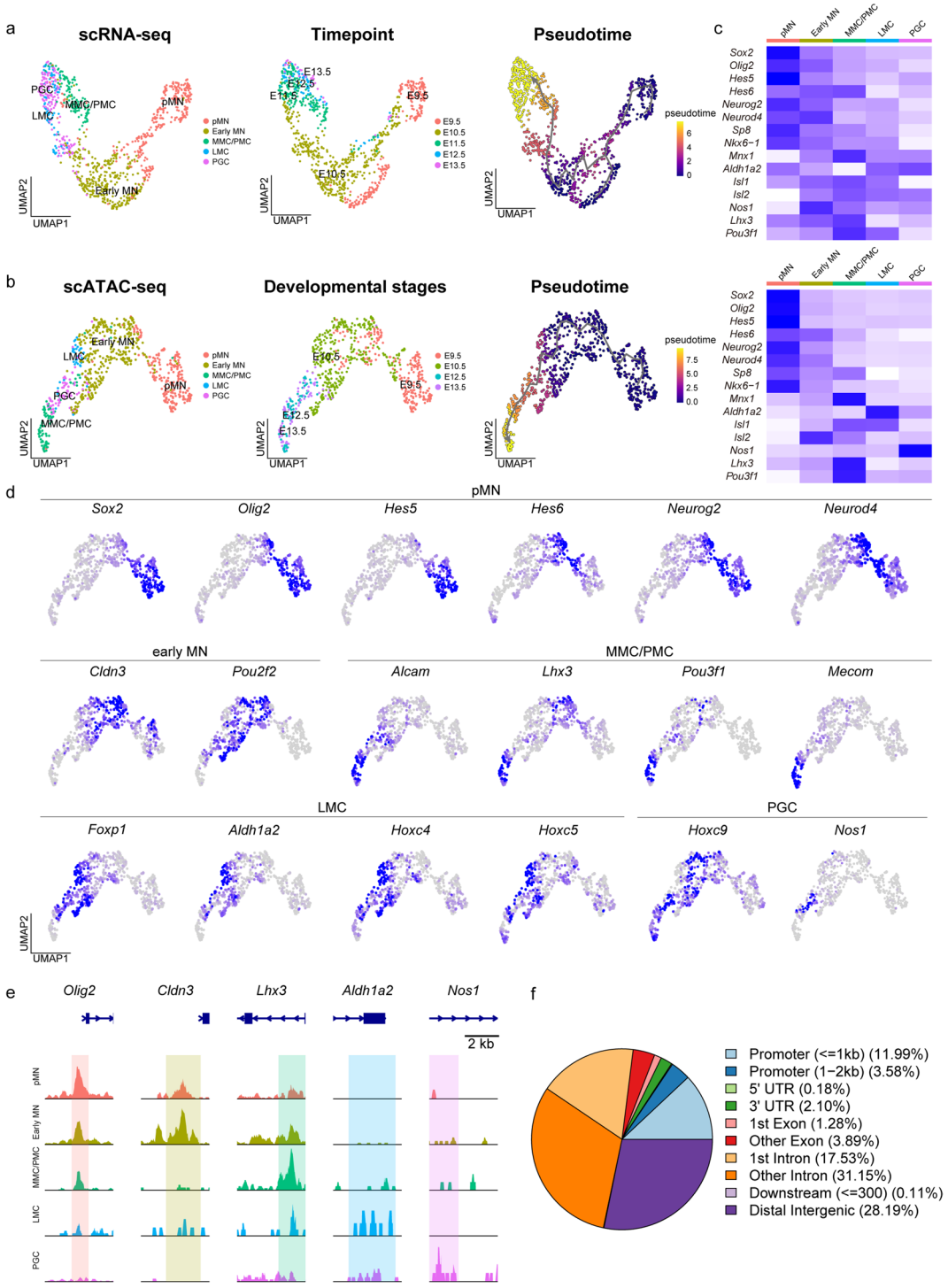


Fig. 1. Single-cell transcriptional and chromatin accessibility profiling on the developing mouse MNs. (**a**) UMAP plots of scRNA-seq dataset showing timepoint and pseudotime trajectory, identifying 5 clusters (1,004 cells). (**b**) UMAP plots of integrated scATAC-seq dataset showing developmental stages and pseudotime trajectory with gene activities-based cell type assignments from 1,004 MNs, identifying 5 clusters (726 cells). (**c**, **d**) The gene activity score, scaled imputed RNA expression and feature plots showing gene expression of marker genes are colored in UMAP. (**e**) Aggregated scATAC-seq tracks showing chromatin accessiblility peaks around the marker genes for each cluster. The scale bar (2 kb) applies to all gene panels. (**f**) Distribution of primary MN-CRE peaks in genomic features and around transcription start sites (TSS).

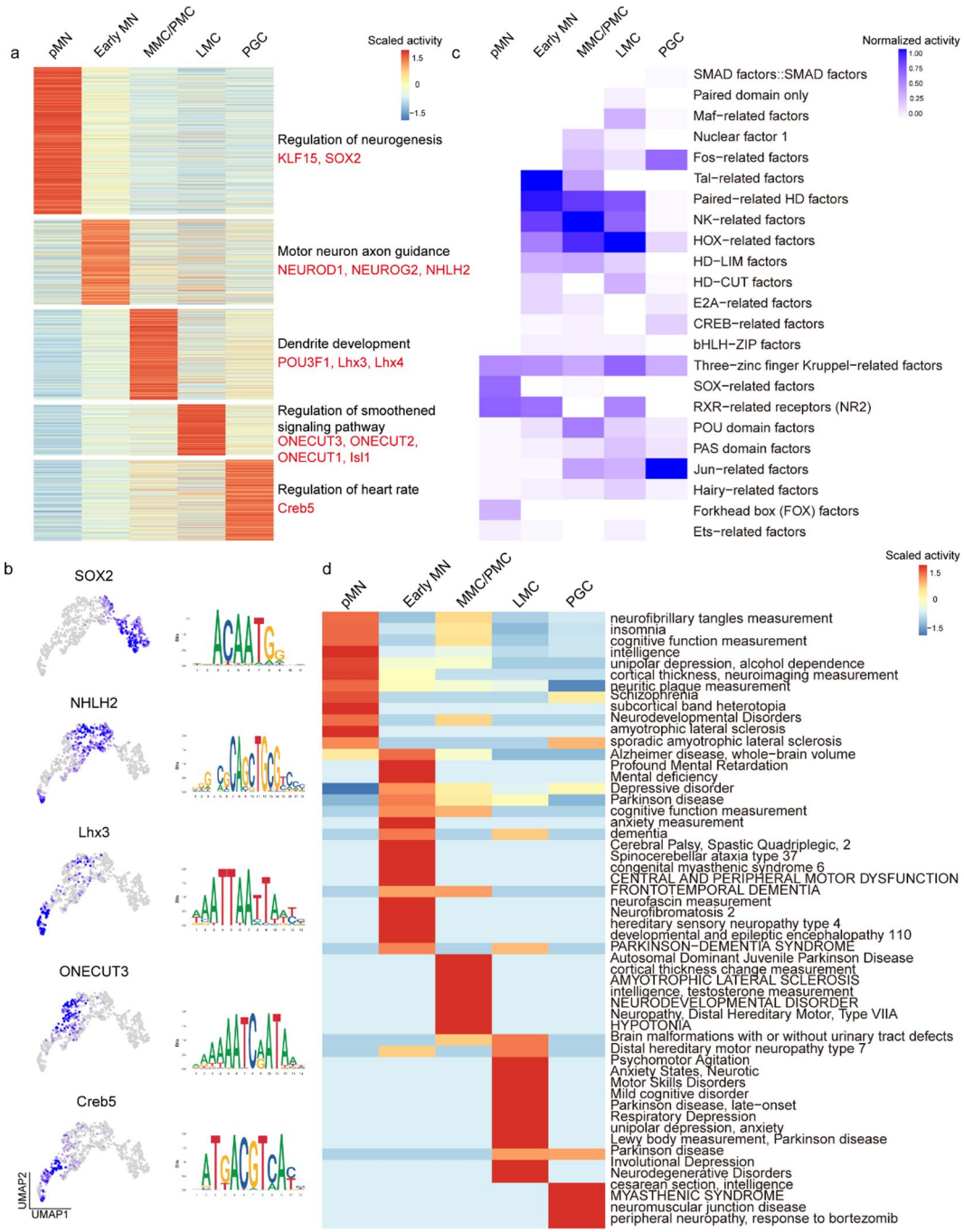


Fig. 2. Identification and characterization of cis-regulatory elements (CREs) associated with MN genes. (**a**) Heatmap showing the association of five MN clusters (columns) with corresponding MN-CRE modules (rows). The representative GO terms and TF motifs highly enriched in each MN-CRE module are shown on the right. (**b**) Most enriched TF motifs in the MN-CRE peak set were determined through analysis using TFBSTools, with motif data obtained from the JASPAR database. ChromVAR was used to visualize motif activity across single cells, and representative motif logos are shown. (**c**) Heatmap showing enriched TF binding motifs grouped by TF families in MN-CRE modules. (**d**) Heatmap showing scaled enrichment of sequence variants associated with the indicated neurological traits/diseases in the human orthologs of CREs.

in individual CRE clusters corresponded closely with the motif activities of TFs in each cluster (Fig. 2a, b). In addition, GO terms and TF motifs for motor column-CREs suggested roles in specifying motor column identity. For example, pMN-specific CREs were enriched for KLF15 and SOX2 motifs, which are linked to biological processes (BPs) such as the regulation of neurogenesis[11–14]. Early MN-specific CREs were enriched for NEUROG2 and NEUROD1 motifs, which are key TFs involved in neuronal identity and MN axon guidance[15–17] (Fig. 2a). MMC/PMC-specific CREs were enriched for Lhx3, Lhx4 and POU3F1 motifs, which are essential for

MMC and PMC differentiation and the associated BP was the regulation of dendrite development[18–20]. LMC-specific CREs were enriched for ONECUT factors and Isl1 motifs, which are critical for assigning LMC fates and the associated BP was regulation of smoothened signaling pathway[21–23]. PGC-specific CREs were enriched for Creb5 motif and the associated BP was the regulation of heart rate, consistent with its role in regulating internal organs. Furthermore, each cluster displayed unique TF motif family enrichment, emphasizing the diverse regulatory mechanisms governing MN differentiation. The enrichment patterns of TF motif families highlight their distinct roles in MN differentiation. For instance, SOX2-related factors were predominantly enriched in the pMN cluster, consistent with their role in maintaining progenitor identity[13,14]. HD-LIM factors showed notable enrichment in the early MN, MMC/PMC and LMC clusters, emphasizing their importance in motor neuron differentiation and dendrite development[23,24]. HD-CUT factors were specifically enriched in the LMC cluster, supporting their role in LMC[22] (Fig. 2c). Moreover, MN-CREs across motor column clusters were linked to various central nervous system-related parameters including MN diseases, such as amyotrophic lateral sclerosis, neuropathy and neuromuscular junction disease (Fig. 2d, Supplementary Data 7). Taken together, these results highlight the role of distinct TFs in regulating MN-CREs and their contribution to biological functions and disease mechanisms.

Assessing LIM-HD TF binding among MN-CREs

To investigate how MN-CREs mediate gene regulation, we examined their interaction with trans-acting elements, especially LIM-HD TF family members ISL1, ISL2, LHX3, and LHX4, which are known to drive MN identity and differentiation[1,18,19,25–27]. scRNA-seq and scATAC-seq analyses revealed distinct expression patterns of LIM-HD TFs across motor columns, consistent with their known developmental roles (Fig. 3a). For instance, *Isl1* expression was barely detected in pMNs but increased during MN differentiation and motor column specification. Similar to *Isl1*, *Isl2* was highly expressed in early MN and maintained in other motor columns. *Lhx3* and *Lhx4* showed modest expression in the pMN and the highest level in the MMC, aligning with their roles in specifying MMC identity[1,19]. Thus, we successfully identified column-specific expression patterns of LIM-HD TFs using scATAC-seq data, which were consistent with their known in vivo expression and suggested functions.

We next examined whether LIM-HD TFs exhibited distinct motif activity in MN-CREs, potentially driving column-specific gene regulation. Using suggested TF motifs provided from JASPAR[28], CIS-BP[29], and others, we observed differential enrichment of LIM-HD TF motifs across MN clusters (Fig. 3a–c). The Isl1 and ISL2 motifs showed higher activity in early and other postmitotic MNs, while the Lhx3 and Lhx4 motif activities were most enriched in MMC-specific CREs. We performed TF footprint analysis on scATAC-seq profiles to assess the differential occupancy of LIM-HD TF across MN clusters. While Isl1 and ISL2 did not exhibit preferential binding to LMC-CREs compared to pMNs, Lhx3 and Lhx4 showed the strongest binding in MMC/PMC relative to pMNs. In contrast, Isl1 and ISL2 binding to CREs showed no significant differences between MMC/PMC and pMNs (Fig. 3d). These findings highlight the distinct regulatory roles of LIM-HD TFs in MN-CREs, linking their activity to specific motor neuron subtypes and developmental stages.

We also explored whether the combinatorial action of the LIM-HD TFs Isl1 and Lhx3 could be identified in MN-CREs. Previous studies have suggested tethered Isl1-Lhx3 binding motifs to DNA including HxRE-long by ChIP-seq analyses and Isl1:Lhx3-RE by SELEX[30,31]. Consistent with these findings, we found enrichment of these two motifs in early MN and MMC/PMC-CREs, supporting the role of Isl1-Lhx3 complexes in establishing pan-MN and MMC identities[1] (Fig. 3e). We noticed that the Isl1-Lhx3 motifs feature 3-bp gap between their core AT-rich binding sites. When we modeled the tethered motifs by combining the individual Isl1 and Lhx3 motifs with a similar spacing between them, similar MMC/PMC-enriched TF activity was observed indicating that the LIM-HD complex activity could be deduced from individual TF motif (Fig. 3e). Thus, the motif configurations of paralogs ISL2 and Lhx4, whose expression and function overlap with Isl1 and Lhx3 in developing MNs, was assessed using the similar strategy[15,19,32]. Position weight matrix (PWM) analysis of various LIM-HD TF combinations revealed that all other possible combinations including Isl1 and Lhx4, Isl2 and Lhx3, Isl2 and Lhx4 were all enriched in early MN and MMC/PMC-CREs. Similarly, TF footprinting analyses demonstrated enhanced binding of these TF pairs in MMC/PMC-CREs compared to pMN-CREs. These results align with previous observations that Isl2 and Lhx4 can also drive MN identity, similar to Isl1 and Lhx3[19,33,34].

To further quantify the distribution of Isl1 and Lhx3 motif activities across clusters, we analyzed the proportion of Isl1-Lhx3 motifs, including HxRE-long and Isl1:Lhx3-RE motifs, in each cluster. The HxRE-long motifs showed higher proportions in MMC/PMC (18.79%) and early MN clusters (15.52%) compared to pMNs (10.40%), while Isl1:Lhx3-RE motifs exhibited similarly high proportions in MMC/PMC clusters (14.43%) (Supplementary Fig. 3a, b, Supplementary Data 8). In contrast, TF footprinting analyses of LMC-CREs compared to pMN-CREs revealed no significant differences, suggesting that LIM-HD TFs may not function as a complex within LMC neurons. Together, this analysis provides insight into motif-level differences among MN clusters, reflecting the transcriptional specificity and regulatory roles of LIM-HD TFs in motor neuron development.

Spatiotemporal dynamics of the limb motor pool enhancers during MN development

A LIM-HD TF complex composed of Isl1 and Lhx3 is critical for establishing motor columnar identity, as demonstrated in stem cell-derived MNs and animal models[30,33,35,36]. Thus, we next investigated whether the MMC/PMC-CREs we identified are selectively active in the MMC population. Isl1 is a well-characterized gene whose expression begins early in MN development, coinciding with the emergence of MMCs, and plays a pivotal role in specifying MMC identity[27,37]. Several *Isl1* enhancers containing Isl1-Lhx3 binding motifs have been identified in our previous work and others[5,33,35,36]. DA analysis between MMC and the remaining MN subtypes revealed that 3 out of 64 CREs around the *Isl1* gene were significantly enriched in the MMC population (Supplementary Fig. 4, Supplementary Data 4). In particular, this included *Crest1*, an *Isl1* enhancer originally

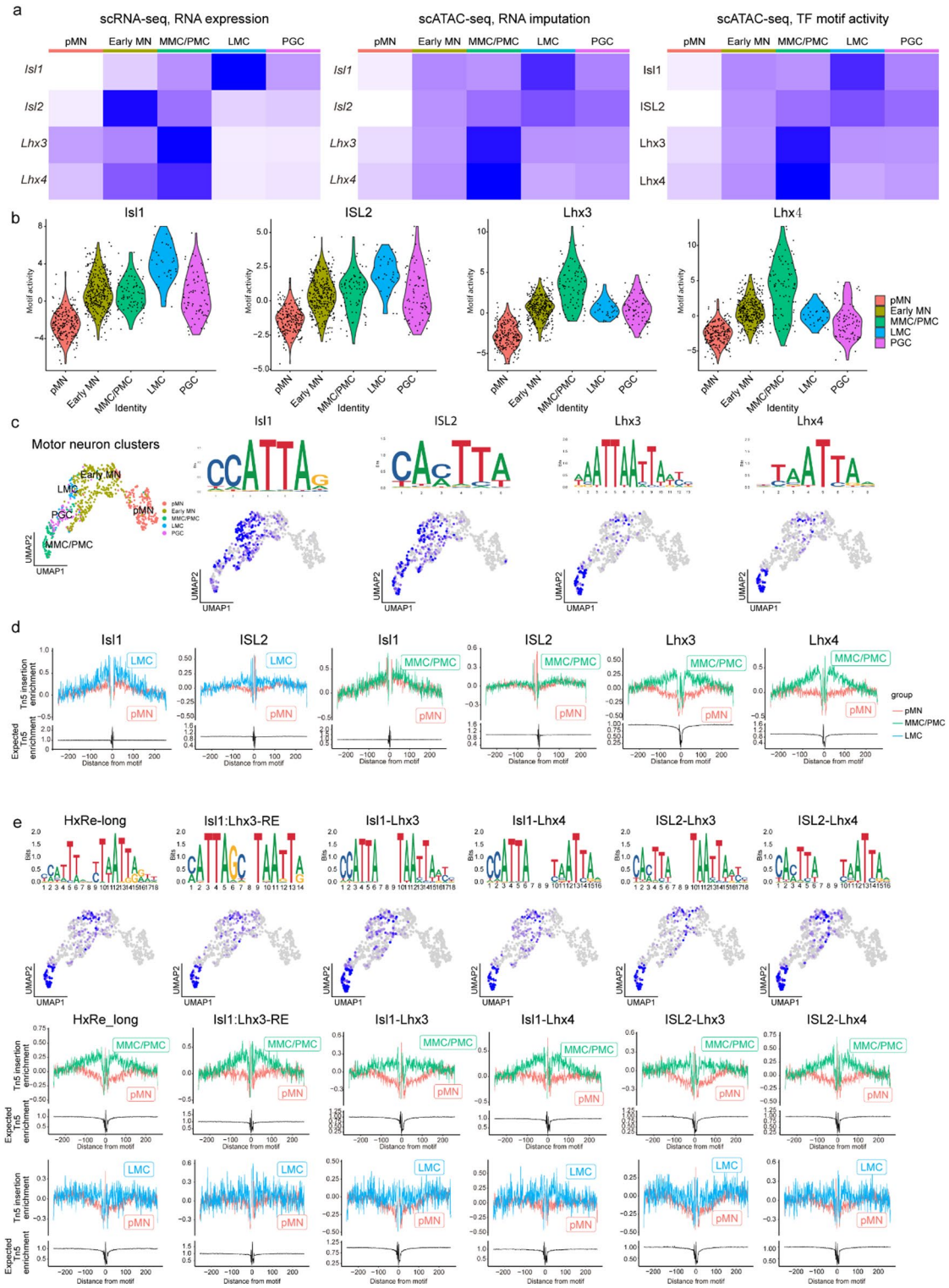


Fig. 3. Analysis of LIM-HD TF binding motifs among MN-CREs. (**a**) The heatmap of RNA expression, imputed RNA expression, and TF motif activity of LIM-HD TFs in various MN clusters. (**b**) Violin plots showing TF motif activity of LIM-HD TFs. (**c**) Suggested TF motifs and UMAP plot showing TF motif activity of individual LIM-HD TFs measured by chromVAR *Z* scores (**d**) TF footprinting analysis for LIM-HD TFs. (**e**) Representative TF motifs, TF motif activity measured by chromVAR Z scores and TF footprint for various combinations of LIM-HD TFs.

characterized in zebrafish, located 160 kb away from the *Isl1* locus and known for its activity restricted to MMCs[36]. Within *Crest1*, ChIP-seq data from stem cell-derived MNs showed enrichment of both Isl1 and Lhx3 binding, supporting the in vivo occupancy of the Isl1-Lhx3 complex at this enhancer[35]. Consistent with this, we previously demonstrated that co-electroporation of Isl1 and Lhx3 together with a *Crest1*-GFP reporter construct led to enhanced reporter activity, further confirming their cooperative activation of *Crest1* enhancer in vivo[37]. Together, these findings support our in silico prediction that MMC-enriched CREs are likely direct targets of LIM-HD TFs and highlight their relevance in regulating genes essential for MMC specification.

Having established that our approach can identify functionally relevant CREs in the MMC lineage, we next turned to the LMC population to explore the spatiotemporal dynamics of enhancer activity during limb-innervating motor pool development. To this end, we focused on validating candidate enhancers associated with a representative gene linked to LMC-derived motor pools. A UMAP plot of cells along the trajectory path from pMN to LMC revealed that many TF genes known to be expressed in limb motor pools, or implicated in their development, showed their distinct gene expression during the progression of LMC (Fig. 4a, Supplementary Fig. 5a). Some of these LMC-specific TFs such as *Foxp1*, *Irx6*, *Isl2*, *Onecut3*, *Isl1*, and *Onecut2* had CREs selectively accessible in LMCs (Fig. 4b, Supplementary Fig. 5b, Supplementary Data 9). Among them, we focused on *Etv4* as a representative LMC motor pool marker[2,38]. Differential accessibility (DA) analysis identified three enhancers (E3, E4, and E5) that were significantly enriched in LMC clusters compared to other MN subtypes (Fig. 4c). To assess their specificity among MN subtypes, we examined the chromatin landscape surrounding the *Etv4* locus across different MN clusters. While the promoter (P) region exhibited high accessibility across all MN population, E3, E4 and E5 showed preferential accessibility in the LMC cluster. Violin plots and pseudotime analyses of CRE activity further confirmed that E3, E4, and E5 were specifically accessible in the LMC cluster, coinciding with active *Etv4* transcription (Fig. 4d, e). Motif analysis revealed that various TF motifs, including those of LIM-HD proteins, were located within the three enhancer candidates, suggesting that these enhancers are potential regulatory targets of LIM-HD TFs involved in LMC identity (Supplementary Fig. 6, Supplementary Data 10).

In addition to CREs identified through DA analysis, we also examined aggregated chromatin accessibility tracks in the LMC population to identify CREs that did not show significant differential accessibility but overlapped with epigenetic marks indicative of enhancer activity. Through this approach, we identified additional candidates (Fig. 4c). The E1 and E7 regions were supported by previously published ChIP-seq datasets for H3K27Ac, H3K4me1, and H3K4me2 in embryonic stem cell-derived MNs[35,39] as well as bulk-ATAC-seq data from E13.5 embryonic mouse MNs[40], all of which are associated with active enhancers. By contrast, the E2 and E6 regions displayed comparatively modest enrichment of epigenetic marks. Nevertheless, they were included as exploratory candidates in our analyses. Altogether, by integrating CRE accessibility with epigenetic data, we identified seven putative enhancers for *Etv4*. These findings underscore the value of combining single-cell analyses with chromatin landscape profiling to elucidate gene regulatory mechanisms in MN development.

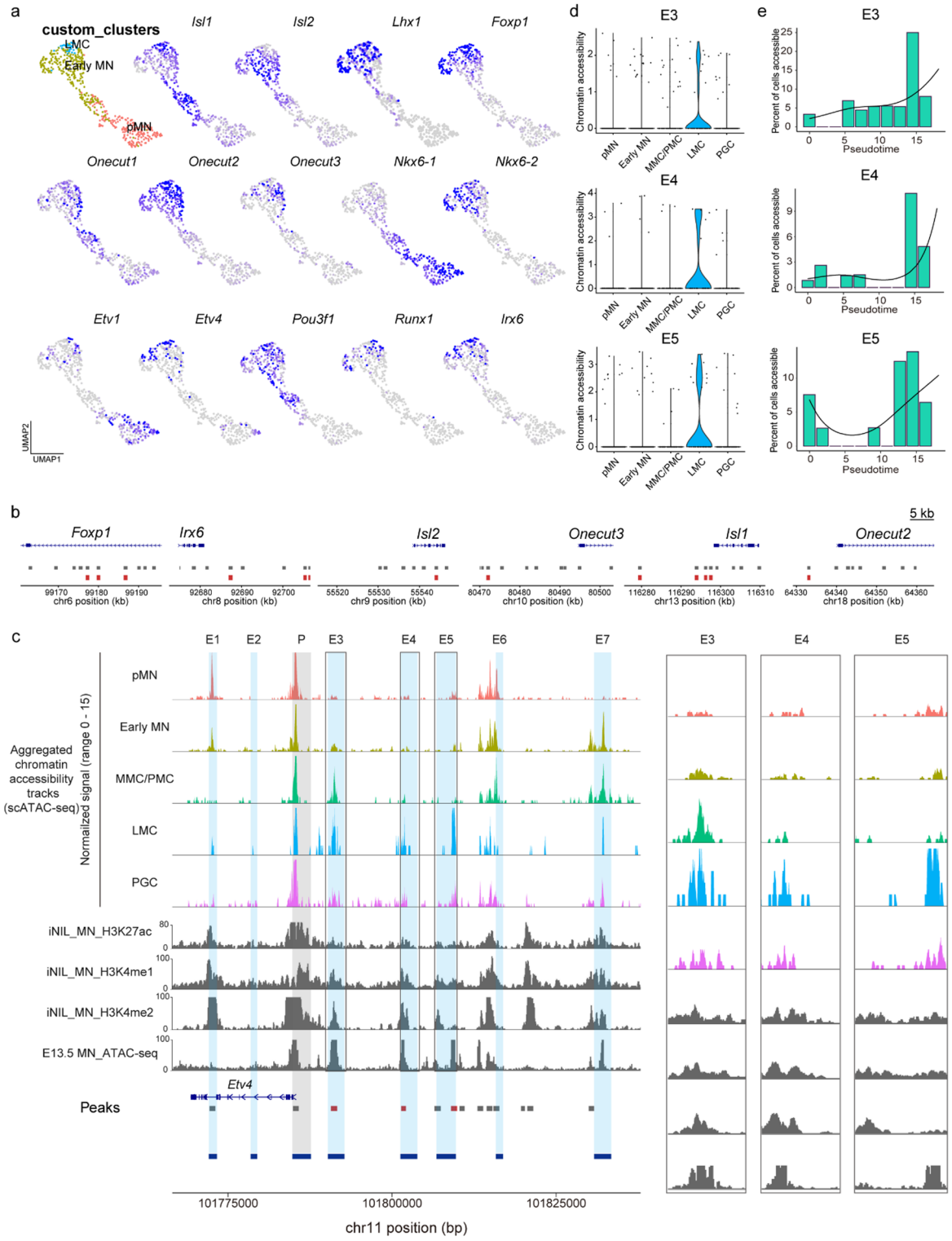
In silico assessment of *Etv4* expression among limb motor pool clusters

Recent advancements in transcriptomics technology have enabled comprehensive profiling of gene transcription at the single-cell level, even within dynamic and rare populations such as developing MNs. Thus, we first identified motor pool clusters using a previously published scRNA-seq dataset from Liau et al.,[6] which analyzed MNs collected from the mouse spinal cord at E13.5, a stage when motor pool identity is settled[4] (Fig. 5a). A total of 5,162 and 4,699 MNs from the rostral (C4-T3) and caudal (L1-S5) segments of the mouse spinal cord were classified into 6 or 7 different MN clusters, respectively. Cells within the LMC cluster were further divided into 16 groups at the brachial level and 10 groups at the lumbar level[6]. Using cell annotation from Liau et al. and representative marker gene expressions reported in previous literature, we identified several major brachial motor pools[5,6,41]. These include cutaneous maximus (CM), pectoralis (Pec), flexor carpi ulnaris (FCU), anterior latissimus dorsi (ALD) and scapulohumeralis posterior (Sca) motor pool clusters[6] (Fig. 5a). Similarly, we identified major lumbar motor pools, including tibialis anterior (Ta), adductor/gracilis (A/G), vastus (V), rectus femoris (Rf), tensor fasciae latae (Tfl) and gluteus (Gl) based on the gene expression of *Etv4*, *Isl1*, *Lhx1*, *Etv1*, *Nkx6.1*, *Nkx6.2*, *Aldh1a2*, *Hoxd9,* and *Pou3f1* [2,5,6,42] (Fig. 5b). Consequently, we successfully located major LMC motor pools within the scRNA-seq dataset.

Given that Etv4 is distinctively expressed in LMC motor pools and plays various roles in motor pool organization, we further investigated the expression of Etv4 in developing LMC motor pools in vivo. We examined the expression of ETV4 along with other motor neuronal markers, including ISL1 (LMCm marker), ISL2 (MN marker), LHX1 (LMCl marker), FOXP1 (LMC marker), and MNR2 (MN marker). In E13.5 mouse embryos, ETV4 was expressed in the CM, Pec, ALD motor pools at the brachial level, and in the Gl motor pools at the lumbar level[2,38] (Fig. 5c). Similarly, in HH30 chicken embryos, *ETV4* transcripts were found in the Pec, iliotrochanterici (ITR) and anterior iliotibialis (ITB) motor pools, corresponding to the muscles in mice[2,42] (Fig. 5d). Serial transverse sections of E13.5 mouse spinal cords revealed that ETV4 expression persists within the LMC, mostly in ISL1^{-}LMCl from L1 to L4 and ISL1^{+}LMCm from L5, suggesting additional Etv4 motor pools in the lumbar segment (Supplementary Fig. 7a). Consistent with this, UMAP analysis of *Etv4* showed its expression localized in cm1 and crl4 clusters, which belong to caudal LMCm (Isl1^{+} Hoxd9^{-}) and rostral LMCl (Lhx1^{+}Hoxd9^{+}), respectively (Fig. 5b, Supplementary Fig. 7a). Collectively, we identified ETV4 motor pools among limb-innervating MNs both in silico and in vivo.

In vivo validation of *Etv4* enhancer candidates in chicken embryonic spinal cords

Next we sought to validate the in vivo activity of these putative CREs in developing MNs. To assess the selective activity of putative CREs in ETV4-expressing motor pools in vivo, we generated GFP reporter constructs containing a mini CMV promoter under the control of candidate enhancer sequences. The reporters were



electroporated into the early chicken embryonic neural tube, with an mCherry CMV reporter, which was co-electroporated as an internal control to assess electroporation efficiency (Fig. 6a,b, Supplementary Data 11). At HH30, among the electroporated putative CREs, E4 and E5 exhibited GFP activity in the region corresponding to the Pec motor pool, whereas E3 and E4 displayed specific GFP activity in the regions where ITR and ITB motor pools were located. The remaining reporters either exhibited no GFP activity or showed non-specific activity within the neural tube. To assess whether conventional LIM-HD TFs, known for driving MN identity, also induce *Etv4* expression, we conducted luciferase assays in vitro. Neuro2A cells, chosen for their neuronal origin, provided a cellular context similar to undifferentiated neurons. The E3, E4, and E5 reporters also became activated by LIM-HD TFs Isl1 and Isl2 when their luciferase reporter activities were tested after co-transfection (Fig. 6c, Supplementary Data 12). Thus, these results demonstrate that the *Etv4* enhancers are selectively active

◀**Fig. 4**. Spatiotemporal dynamics of the *Etv4* enhancers in silico. (**a**) Feature plots showing gene expression of LMC-enriched TFs along the differentiation trajectory from pMN to LMC. Cells along the LMC lineage trajectory in Supplementary Figure 4a were shown. (**b**) Genomic regions spanning 50 kb around some representative TF genes identified as part of LMC-CREs from DA analysis. Grey bars denote scATAC-seq peaks. Peaks corresponding to LMC-CREs are displayed as red bars. Each region shows the locus of a gene, including *Foxp1*, *Irx6*, *Isl2*, *Onecut3*, *Isl1*, and *Onecut2*. The scale bar (5 kb) applies to all gene panels. (**c**) Genome browser tracks showing aggregated chromatin accessibility profiles for MN clusters along MN differentiation around the *Etv4* gene. Light grey shading highlights the promoter and light blue shading highlights putative enhancers. Representative tracks of H3K27ac, H3K4me1, and H3K4me2 ChIP-seq in mouse embryonic stem cell (ESC)-driven MNs[35], bulk ATAC-seq in E13.5 MNs[40] at the *Etv4* gene enhancers. Y-axis ranges are indicated for each track to allow comparison of signal intensities. For clarity, we re-rendered and magnified the scATAC-seq tracks to better visualize peak overlaps with epigenetic marks (right). (**d, e**) Violin plots and pseudotime plots showing chromatin accessibility of putative *Etv4* enhancers.

in ETV4-expressing motor pools and activated by LIM-HD TFs. These findings shed light on the regulatory influence of LIM-HD TFs on *Etv4* enhancer activity in a neuronal context.

Discussion

Integrating transcriptome and epigenome data offers a powerful approach to dissect cell-specific regulatory landscapes, surpassing the limitations of relying solely on scRNA-seq data and enabling a more comprehensive view of gene expression dynamics that drive cell diversity. Nevertheless, a detailed understanding of cellular heterogeneity and complexity within the central nervous system based on single-cell genomics remains incomplete. In this study, we leveraged scRNA-seq and scATAC-seq data to characterize the developmental process of MNs in the spinal cord from E9.5 to E13.5, during which pMN progenitors differentiate into MNs with distinct motor column identity, examining both transcriptomic and chromatin accessibility profiles. With precise lineage-specific master regulator TFs guiding individual MN subtypes, anatomical positions, and muscle-specific innervation, MNs provide an optimal model to study fundamental regulatory principles underpinning cellular diversity. Motor pools consist of uniquely specialized populations, controlling the contraction of one specific muscle, and exhibit spatially and temporally dynamic transcriptomic and regulatory chromatin landscapes. Their clear anatomical positions also make them suitable for integration with emerging spatial transcriptomics tools. Here we developed a strategy to predict motor pool-specific enhancers by tracking transcriptomic and epigenomic changes through MN differentiation. We concentrated our analysis on CREs associated with genes highly enriched in MNs, enhancing specificity and making these CREs practical for further applications in tracing and manipulating specific motor pools. Among these CREs, we focused on identifying enhancers associated with key motor pool-specific TFs, whose precisely timed and localized expression patterns play a pivotal role in shaping the gene regulatory networks that define and maintain functional motor pool identities.

Our DA analysis identified 19,269 MN-CREs exhibiting increased chromatin accessibility in MNs compared to non-MN cells. These CREs were enriched with binding motifs for MN-specific TFs, including NEUROG2, ONECUT factors, and LIM-HD TFs. GO enrichment analysis was associated these MN-CREs with MN-related biological processes, including neurogenesis regulation, MN axon guidance, dendrite development, and regulation of the Smoothened signaling pathway. Additionally, MN-CREs were significantly enriched in neuronal disease-related genes. These results suggest that certain MN subtypes, defined by specific CRE enrichments, may have greater vulnerability to neurodegenerative diseases.

LIM-HD TFs are critical for motor columnar identity, with their combinatorial actions extensively characterized in structural, biochemical and genetic studies[1,18,25]. For instance, ISL1 and LHX3 together specify MN fate and MMC identity, whereas LHX3 alone promotes V2 interneuronal fate[1]. ISL1 and LHX1 drive differentiation of LMCm and LMCl, respectively[23]. However, the distribution of the 'LIM-code' across the genome at the single-cell level remains unexplored. Here, we mapped the LIM code distribution within CREs across MN clusters. MMC/PMC-CREs prominently featured motifs for LIM-HD TFs such as ISL1, ISL2, LHX3, and LHX4. Importantly, the ISL1-LHX3 tandem motif, HxRE, previously defined by SELEX and ChIP-seq experiments, was enriched in both MMC/PMC and early MNs, consistent with their roles in specifying early MN and MMC identities[30,31]. Notably, more than 82% of MMC-associated scATAC-seq peaks contained both ISL1 and LHX3 motifs, implying a cooperative interaction in open chromatin regions to drive MN-specific gene expression. TF footprinting revealed higher ISL1, ISL2, and LHX3 binding across MNs than in pMNs, with ISL1 and LHX3 binding specifically enriched in MMC, aligning with previous findings that ISL1-LHX3 fusion protein activity is sufficient to induce MN fate or synergistic binding of ISL1 and LHX3 occurs in developing MNs[1,30,35]. Other LIM-HD interactions, such as those between ISL1 and LHX4 or ISL1 and LHX8, similarly demonstrate context-specific cooperative functions in MN specification and broader neuronal differentiation roles[34,43]. Together, these findings indicates the universal usage of combinatorial LIM code for neural specification.

This study demonstrated that integrating scRNA-seq and scATAC-seq can effectively define CREs across large populations, such as MNs, and even classify CREs into much smaller functional units such as motor pools. We first successfully validated our approach by focusing on the MMC population, where we confirmed that Isl1 and Lhx3 bind to a previously characterized *Isl1* enhancer, supporting the predictive power of our method for MN subtype-specific CREs. Further, we turned our attention to the LMC population to define CREs associated with distinct motor pools. Within LMC, TFs with motor pool-specific expression were among the highest ranked based on imputed RNA levels and chromatin accessibility. We focused on identifying *Etv4*

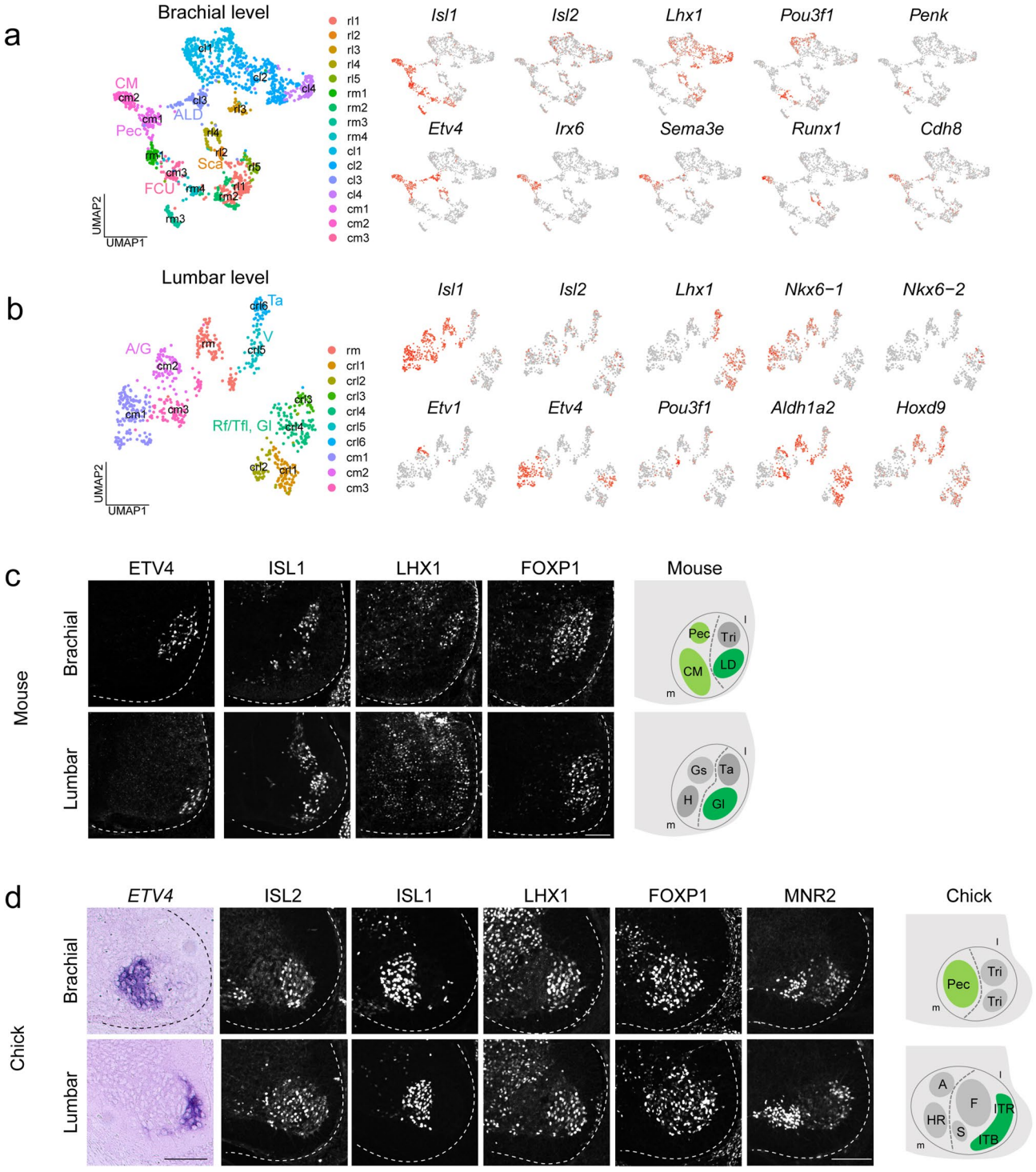


Fig. 5. Visualization of limb motor neuron clusters in silico and conserved distribution of limb motor pools across multiple species. (**a**,**b**) Clustering analysis of the single cell transcriptomes of LMC motor neurons isolated from E13.5 brachial and lumbar spinal cords[6]. Representative LMC motor pools are labeled. UMAP plots display the expression of the representative MN markers in each plot. (**c**) Expression of ETV4, ISL1, LHX1 and FOXP1 in E13.5 mouse embryonic spinal cords. (**d**) Expression of *ETV4*, ISL2, ISL1, LHX1, FOXP1, and MNR2 in HH30 chicken embryonic spinal cords. Muscle abbreviations: cutaneous maximus (CM), pectoralis (Pec), flexor carpi ulnaris (FCU), anterior latissimus dorsi (ALD), scapulohumeralis posterior (Sca), tibialis anterior (Ta), adductor/gracilis (A/G), vastus (V), rectus femoris (Rf), tensor fasciae latae (Tfl), gluteus (Gl), tricepts (Tri), gastrocnemius (Gs), hamstring (H), adductor (A), hip-retractor (HR), femorotibialis (F), sartorius (S), iliotrochanterici (ITR), anterior iliotibialis (ITB). Scale bars, 100 μm.

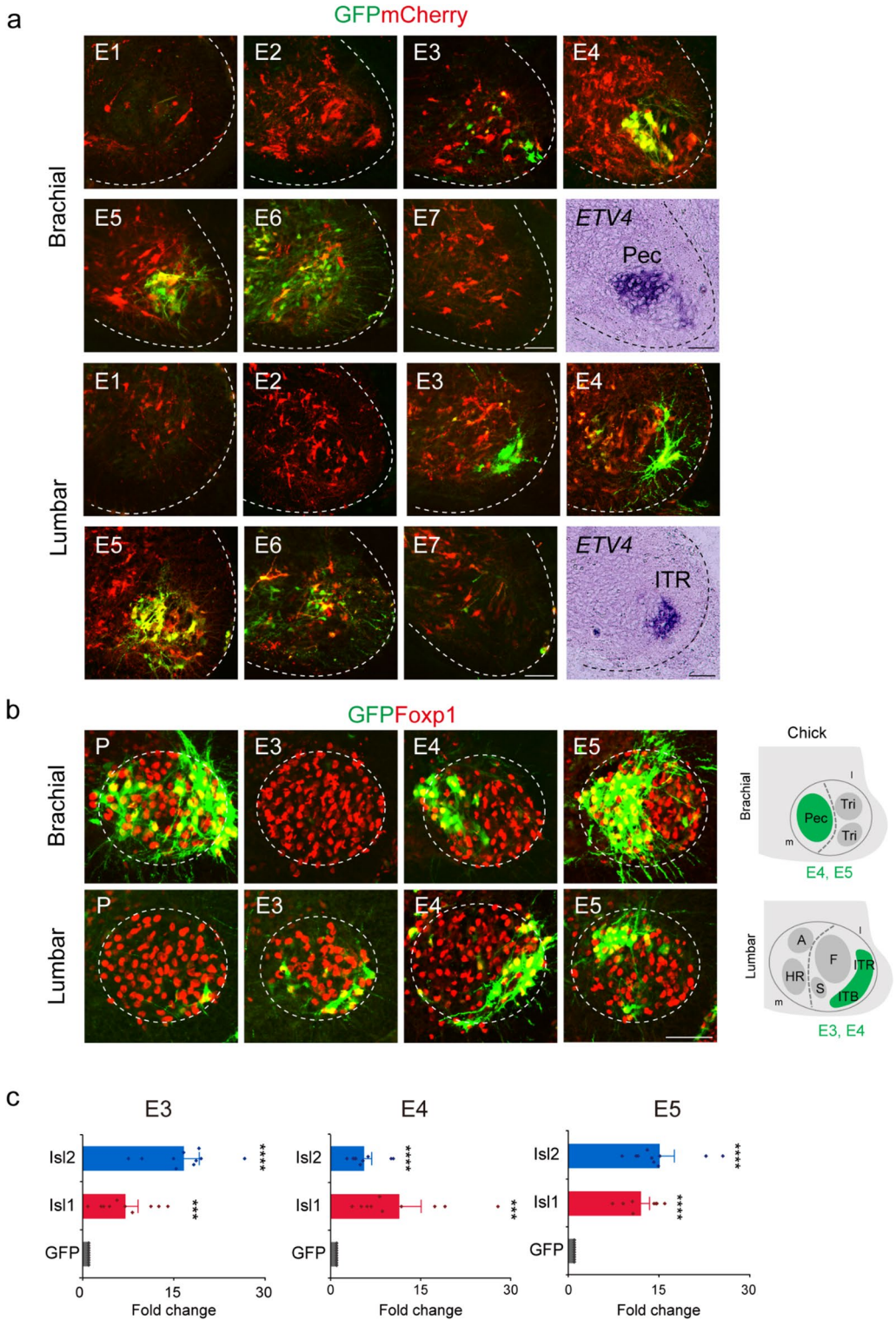


Fig. 6. Spatial activity of *Etv4* enhancers in chicken embryonic spinal cords. (**a**) GFP expression in HH30 chicken embryos electroporated with candidate *Etv4* enhancer reporters. The CMV::mCherry vector was co-electroporated as an internal control. GFP expression driven by E4 and E5 at the brachial level, and E3 and E4 at the lumbar level, overlaps with endogenous *ETV4* expression in Pec or ITR. (**b**) Immunostaining of GFP and FOXP1 reveals that E4 and E5 activities corresponds to LMCm, while E3 and E4 activity corresponds to LMCl. Schematic diagrams illustrate the activity of E3, E4 and E5 in chicken spinal cords. Scale bars, 100 µm. (**c**) Luciferase activity driven by the candidate *Etv4* enhancers, co-transfected with GFP, Isl1 or Isl2 expression vectors in Neuro2a cells. Individual fold change values are represented as dots. Error bars represent SEM. *** $p < 0.001$, **** $p < 0.0001$; unpaired Student's t-test ($n = 10$).

E1	Forward	5'- GGCCTCTTCGCTATTACGCCAGTCGACACCCCCTAAGACCAGAGTACA – 3'
	Reverse	5'- CCTCCCACCGTACACGCCTACGGATCCTTCCCGCCACTAAATCCCAGG – 3'
E2	Forward	5'- ATAGTCGACGTACGCTGTAGCTGTCTTCAG-3'
	Reverse	5'- ATAGGATCCACCCTACGTTCTGGATCTCACT-3'
E3	Forward	5'- ATAGTCGACGCCTTCTGCAGAACGTGC-3'
	Reverse	5'- ATAGGATCCGGAGCTCAAAGGACTCCTGAG-3'
E4	Forward	5'- ATAGTCGACGCCTGCGCCATGACAGCCTGAATGTTG – 3'
	Reverse	5'- ATAGGATCCCCTCTGCCTCTGTGAGTGCCTGG-3'
E5	Forward	5'- ATAGTCGACTCTATGCCATCTGAGGGGACC-3'
	Reverse	5'- ATAGGATCCATGTGCCAAGCTTCCTGCCCA-3'
E6	Forward	5'- ATAGTCGACGCCCATCTGTTTTGGAAGCT-3'
	Reverse	5'- ATAGGATCCATACAAACACCGGATCTGGC-3'
E7	Forward	5'- ATAGTCGACCCAGGCAGCCGCATTGGCAG-3'
	Reverse	5'- ATAGGATCCGCTTTCTTCCACACTATGCTTGC-3'
P	Forward	5'- ATAGTCGACCCCCCAATCTACTGGAGGAAGAG-3'
	Reverse	5'- ATAGGATCCGCTGGGGTTGGTGTGTCTG-3'

Table 1. Primers used in this study.

enhancers due to ETV4's well-characterized role in LMC-derived motor pools. ETV4-expressing motor pools undergo differentiation in response to limb-derived neurotrophic factor GDNF, which induces *Etv4* expression and promotes subsequent maturation processes, including motor pool positioning, dendrite arborization, axon innervation, and sensorimotor circuit formation[2,38,44]. Their cell bodies are located at stereotypical positions within the neural tube and can be easily defined by combinations of additional reliable marker expressions[5,18,25]. Additionally, the transcriptional cascade activating *Etv4* gene expression has been investigated in genetic models, highlighting the role of LIM-HD TF in regulating *Etv4* expression and motor pool specification[5]. Thus, identifying CREs associated with *Etv4* transcription broadens our understanding of transcriptional and epigenetic mechanisms that dictate motor pool development.

Our DA analysis revealed three potential Etv4 enhancer candidates, E3, E4, and E5, among 13 CREs near the *Etv4* locus. The genomic sequence containing all three enhancers showed elevated chromatin accessibility specifically in LMC, implying a cooperative enhancer network among them. In line with our in silico prediction, electroporation of GFP reporters for these enhancers into chicken neural tubes showed that all enhancers were active in Etv4 motor pools; E3 and E4 were active in brachial motor pools, while E4 and E5 were active in lumbar motor pools. Although E3, E4, and E5 are cis-regulatory elements regulating the same gene (*Etv4*), their distinct activation patterns in brachial and lumbar motor pools may result from several factors. First, the TFs likely differ between brachial and lumbar motor pools, leading to differential enhancer activation. Second, context-dependent interactions between enhancers and promoters or other regulatory elements could influence their activity in a spatially restricted manner. Lastly, epigenetic modifications, such as histone modifications or DNA methylation, further refine enhancer activity. Furthermore, all three enhancer activities increased in the presence of ISL1 and ISL2, suggesting that LIM-HD TFs may upregulate *Etv4* transcription, consistent with a previous study showing *Isl2* deletion resulted in the loss of *Etv4* expression[5].

Previous studies have reported that ETV4 marks major motor pools, including CM and ALD at the brachial level and Gl, Rf, and Tfl motor pools at the lumbar level[38,42]. Our scRNA-seq analysis also detected additional *Etv4* expression at the caudal lumbar levels which overlap with ISL1, implying their LMCm identity. While the exact identity of these motor pools requires further investigation, the caudal ETV4 expression within LMCm in mouse embryos aligns with the caudilioflexorius motor pool expressing ETV4 in chicken embryos, suggesting that motor pool identity and regulatory networks underlying *Etv4* expression are evolutionarily conserved[2,42,45,46].

In summary, this study underscores the power of integrating scRNA-seq and scATAC-seq to elucidate the cis-regulatory architecture of MNs, providing a high-resolution framework for understanding the transcriptional and epigenetic mechanisms that shape motor pool diversity. By focusing on CREs associated with motor pool-specific TFs, we identified key regulatory elements and demonstrated their functional activity using in vivo reporter assays. Our findings highlight the cooperative nature of enhancer networks and their reliance on context-specific TF interactions to drive precise gene expression programs. These insights not only advance our understanding of MN development but also provide a valuable resource for studying the mechanisms of neurodevelopmental and neurodegenerative disorders.

Materials and methods

DNA constructs

PCR fragments, amplified from genomic mouse DNA, were subcloned into the reporter vector pCS2 miniCMV-GFP. This vector contains a 30-bp TATA box and the transcription initiation site of the CMV promoter. Each putative CRE was cloned upstream of the miniCMV promoter. While most constructs (E2–E7) were cloned using SalI/BamHI sites, E1 was generated using the In-Fusion Cloning Kit (Takara Bio, USA). The primer information used in this study is provided in Table 1.

Luciferase assays
Neuro2A cells (ATCC) underwent transient transfection with both reporters and the TF expression plasmids pCAGGS1 mouse Isl1, pCAGGS1 mouse Isl2[37] utilizing the Lipofectamine 3000 reagent (Invitrogen). To ensure consistency, a CMV-β-galactosidase plasmid was co-transfected for normalization of transfection efficiency. After 48 h, cell extracts were subjected to luciferase assays and β-galactosidase assays. The presented data are the mean of triplicate measurements and have been repeated at least three times.

In ovo chick electroporation.
Fertilized chick eggs were purchased (Banseok's Farm, Hwasun-gun, Republic of Korea), stored at 16 °C for 0 to 10 days, and then incubated in a 38 °C humidified incubator. For *in ovo* electroporation, DNA constructs were introduced into HH12 chicken embryos. *Etv4* enhancer candidate constructs with the *GFP* reporter gene under miniCMV were injected into the chick spinal cord together with the internal control plasmid pmCherry-C1 (Clontech). Embryos were harvested at HH30 for subsequent immunohistochemical analyses. Electroporation was performed as described previously[37].

Mouse line
Hb9::GFP mice (MGI:3767834) used in this study were previously described, and were kindly provided by Dr. Samuel L. Pfaff[47]. All experiments were approved by the Animal Care and Ethics Committees of the Gwangju Institute of Science and Technology (Approval No. GIST-2023-037), and all experimental methods were performed in accordance with the relevant guidelines and regulations. Additionally, all animal experiments followed the ARRIVE guidelines.

Immunohistochemisty and in situ hybridization
Immunohistochemistry and in situ hybridization procedures followed previous protocols[5,37]. The antibodies utilized included rabbit anti-GFP (Invitrogen, A11122), mouse anti-GFP (Sigma, G6539), mouse anti-Isl2 (Santa Cruz, sc-390746), mouse anti-Isl2 (DSHB, 51.4H9), rabbit anti-Foxp1 (Abcam, ab16645), mouse anti-Isl1 (DSHB, 39.3F7), guinea pig anti-Isl2[1], rabbit anti-Lhx1 (Abcam, ab229474), rabbit anti-Etv4[48], mouse anti-MNR2 (DSHB, 81.5C10). For in situ hybridization, embryonic mouse cDNA (E12.5) orembryonic chicken cDNA (HH25) was used to generate riboprobes against *Etv4*[5].

Public transcriptomic datasets sources
The public scATAC-seq, scRNA-seq and ChIP-seq datasets used in this study were CRA005358[3], E-MTAB-7320[4], GSE183759[6], GSE80482[45], GSE80483[45,] and GSE198767[40].

scATAC-seq data re-analysis
scATAC-seq data analysis was performed as described in Shu et al.[3] (Supplementary Fig. 1). In brief, sequencing reads were aligned to the mm10 reference genome, and a cell-by-peak matrix was generated from fastq files using the software Cell Ranger ATAC (version 2.1.0). Cell barcode matrices obtained from different time points (E9.5, E10.5, E12.5, and E13.5) were aggregated into a single peak-barcode matrix using the Cell Ranger ATAC aggr pipeline. Quality control at the library level was performed by examining the periodicity of fragment size frequencies and TSS enrichment. After converting the cell-by-peak matrix into a Seurat object using Signac[49,] barcodes were filtered, requiring at least 3,500 fragments and a TSS enrichment score of >2, resulting in 18,779 high-quality barcodes. The R packages Seurat and Signac were used for dimensionality reduction, clustering and calculation of gene activity scores.

Integration of scATAC-seq and scRNA-seq data for cell type identity assignment
Integration analysis of the scRNA-seq and scATAC-seq data was performed using the R package Seurat, as described in Shu et al.[3]. In this analysis, labels from scRNA-seq data[4] were transferred to annotate cell types in scATAC-seq data using Seurat's label transfer workflow. The scRNA-seq data were pre-processed by identifying variable features, and canonical correlation analysis (CCA) was used to find transfer anchors between the scRNA-seq and scATAC-seq datasets. These anchors represented cells with similar expression and chromatin accessibility patterns. Cell type labels from the scRNA-seq data were transferred using the TransferData function, weighted by the LSI reduction of the scATAC-seq data. Transferred labels were incorporated into the metadata of the scATAC-seq dataset and visualized using UMAP to assess biological consistency.

Identification and characterization of MN-CREs
After determining the five MN clusters from the scATAC-seq data, we identified MN-specific cis-regulatory elements (MN-CREs) as follows. To focus on regulatory elements relevant to motor neuron development, we first selected a subset of peaks associated with pMN and MN marker genes curated from Delile et al.[4] Marker genes were identified using the `FindAllMarkers()` function on scRNA-seq data, and associated peaks were extracted based on gene activity correlations. We used the function FindMarkers() of Signac (version 1.14.0), which performed a DA test, comparing one MN cluster against the combined set of the other four MN clusters. The output of each test is $\log_2$ fold-change of the average accessibility (avg_$\log_2$FC) between the two groups and *p*-value based on Bonferroni correction using all peaks in the dataset. We selected the MN-CREs with a cutoff avg_$\log_2$FC > 0.25, identifying 19,269 MN-CREs (recommended by Signac). To predict putative target genes for each MN-CRE, we applied the LinkPeaks function in Signac (version 1.14.0), which calculates Pearson correlations between chromatin accessibility at distal peaks and gene expression across matched scATAC-seq and scRNA-seq profiles. Only positively correlated peak–gene pairs were retained, and associations were limited to genes located within 500 kb of each CRE center.

To associate MN-CREs with the biological processes (BP) of their putative target genes, we used the R package clusterProfiler (version 4.10.1) to identify enriched GO[50]. To determine the overlap of TF motifs with peaks, we then performed a hypergeometric test to determine the probability of observing the motif at the given frequency by chance using Signac (v1.1.1), by comparing with a background set of peaks matched for GC content or with the rest of the genome. To calculate the MN-CRE motif activity for each cell, we applied the Signac implementation of chromVAR using the RunChromVAR function with default parameters[51].

Curation of cell identity and neurological traits/diseases genes

Putative target genes of each CRE were determined using `LinkPeaks` as described above. Genes annotated to the CREs of each cluster were subjected to disease gene ontology (GO) analysis using the ToppGene Suite[52]. The enriched GO terms for each cluster were identified, and the -log10(*p*-values) of these terms were converted to z-scores for visualization. A heatmap was generated to visualize the enrichment of GO terms across clusters.

Identification of enriched TF family members in MN-CREs

To identify enriched TF motif families in MN-CREs, we conducted motif enrichment analysis using scATAC-seq profiles. Enriched motifs were identified using the FindMotifs function from the Signac package in R, with a motif list obtained from the JASPAR database. Each motif was annotated with its corresponding TF family based on the family names provided in the JASPAR database (https://jaspar.genereg.net)[28] and CIS-BP (http://cisbp.ccbr.utoronto.ca/)[29], enabling motifs to be grouped and analyzed at the family level. To compare enrichment patterns across clusters, motif family counts were normalized using Min-Max normalization method, rescaling the values between 0 and 1. Heatmaps were generated to visualize normalized enrichment levels across clusters, with a focus on biologically relevant TF families.

For more detailed analysis of motifs highly enriched in the E3, E4, and E5 enhancers, we performed motif scanning using FIMO (Find Individual Motif Occurrences)[53]. The genomic sequences of CREs in these clusters were extracted, and candidate motifs were selected based on the enriched TF families identified from earlier analyses. Position frequency matrices (PFMs) of these motifs were obtained from the JASPAR database, and FIMO was used to identify significant motif matches ($p < 1e\text{-}4$) within the CRE sequences.

Calculation of the CREs proportion, containing tethered motifs in each cluster

To investigate the distribution of motifs across motor neuron subtypes, cluster-specific CREs identified from DA analysis were used. The motifs HxRE-long and Isl1:Lhx3-RE were analyzed for their enrichment in each cluster. Position Weight Matrices (PWMs) for HxRE-long and Isl1:Lhx3-RE motifs were obtained from previously published studies[30,31]. These PWMs were used to scan cluster-specific CREs for motif matchs using the motifmatchr R package. The percentage of motif-containing CREs was calculated for each cluster by dividing the number of motif-containing CREs by the total number of cluster-specific CREs. The resulting data were visualized as bar plots using the ggplot2 package in R.

Trajectory analysis of MN differentiation

To investigate the differentiation trajectory from pMN to MN, we utilized Monocle3[54] to calculate pseudotime and generate trajectory plots. To define the starting point of the trajectory, the pMN cluster was designated as the root population based on their biological identity and marker gene expression. The trajectory plots were visualized with cells colored by pseudotime to show the progression from pMN to MN, and adjustments were made to improve clarity.

To investigate the enhancers regulated by TFs involved in MN differentiation, we first filtered cells from the pMN, early MN, and LMC clusters. The pMN cluster was designated as the root population for trajectory analysis, and a trajectory was constructed to examine the progression toward LMC differentiation. From the trajectory, only nodes leading to the LMC cluster were selected, and cells associated with this trajectory were further filtered. Subsequently, a new UMAP was generated using the filtered cells. Imputed RNA expression was visualized with FeaturePlot function to validate the trajectory-based analysis.

Visualization of chromatin accessibility and epigenetic landscapes for *Etv4* enhancer identification

The scATAC-seq data were visualized using the CoveragePlot function in the Signac package in R. For bulk histone modification data (H3K27ac, H3K4me1, H3K4me2) and ATAC-seq, visualization was performed using the Integrative Genomics Viewer (IGV). Chromatin accessibility for specific genomic regions was visualized as violin plots using the VlnPlot function in R. This approach allowed for the comparison of chromatin accessibility of prominent *Etv4* enhancer candidates across clusters. To analyze accessibility dynamics along pseudotime, Cicero was used[55]. Pseudotime analysis was performed to map the accessibility of selected regions as a function of pseudotime progression, and the percentage of cells with accessible chromatin was plotted alongside a smoothed trajectory curve.

Clustering analysis of scRNA-seq data

scATAC-seq data analysis was performed as described in Liau et al.[6]. All analyses were conducted using Seurat (v5.1.0)[56]. Quality control (QC) metrics, including mitochondrial gene percentage (percent.mt), were calculated, and cells with mitochondrial gene expression exceeding 10% were removed. Cells expressing MN markers (Mnx1, Chat, Slc18a3, Slc5a7) were retained for downstream analysis. Normalization was performed using the LogNormalize method with a scale factor of 10,000. Data scaling and dimensionality reduction were conducted using principal component analysis (PCA), and clusters were visualized using Uniform Manifold Approximation and Projection (UMAP).

Clusters were identified using the FindAllMarkers function to detect cluster-specific marker genes, and motor column identities were assigned to clusters based on the expression of known marker genes. Subsequently, LMC cells were isolated as a subset, and clustering was repeated with a resolution of 5. The expression patterns of marker genes were examined to validate the clustering results, and motor pool identities were assigned to each subcluster based on the expression of pool-specific markers.

Quantification and statistical analysis

DA analysis was conducted using a logistic regression (LR) framework to identify differentially accessible regions, with regions showing a p-value below 0.05 considered statistically significant. For TF or TF family enrichment analysis, significance was assessed based on an adjusted *p*-value (q-value) calculated using the Benjamini-Hochberg (BH) method to control for false discovery rate (FDR). Enriched features were defined as those with an adjusted p-value below 0.05. In GO analysis, both raw *p*-values and adjusted q-values were evaluated, with adjustments made using the Bonferroni correction and the BH method. Features with a q-value below 0.05 were deemed significant.

Statistical analysis for luciferase assay data was performed using GraphPad Prism software (version 10.4.1). Unpaired two-tailed t-tests were employed to compare differences between experimental groups as presented in Supplementary Data 12.

Data availability

All raw and analyzed data used in this study can be obtained from the Genomic Sequence Archive (GSA; https://ngdc.cncb.ac.cn/gsa/), EMBL-EBI (https://www.ebi.ac.uk/), or NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) using the accession codes provided in the 'Public transcriptomic datasets sources' section of the Methods. The code used for the analyses, as well as any additional data, is available upon reasonable request. For further information, please contact the corresponding author, Prof. Mi-Ryoung Song (email: msong@gist.ac.kr).

Received: 4 February 2025; Accepted: 3 September 2025
Published online: 06 October 2025

References

1. Thaler, J. P., Lee, S. K., Jurata, L. W., Gill, G. N. & Pfaff, S. L. LIM factor Lhx3 contributes to the specification of motor neuron and interneuron identity through cell-type-specific protein-protein interactions. *Cell* **110**, 237–249. https://doi.org/10.1016/s0092-8674(02)00823-1 (2002).
2. Lin, J. H. et al. Functionally related motor neuron pool and muscle sensory afferent subtypes defined by coordinate ETS gene expression. *Cell* **95**, 393–407. https://doi.org/10.1016/s0092-8674(00)81770-5 (1998).
3. Shu, M. et al. Single-cell chromatin accessibility identifies enhancer networks driving gene expression during spinal cord development in mouse. *Dev Cell* **57**, 2761–2775 e2766 (2022). https://doi.org/10.1016/j.devcel.2022.11.011
4. Delile, J. et al. Single cell transcriptomics reveals Spatial and Temporal dynamics of gene expression in the developing mouse spinal cord. *Development* **146** https://doi.org/10.1242/dev.173807 (2019).
5. Lee, Y. et al. Transcriptional control of motor pool formation and motor circuit connectivity by the LIM-HD protein Isl2. *Elife*. https://doi.org/10.7554/eLife.84596 (2023).
6. Liau, E. S. et al. Single-cell transcriptomic analysis reveals diversity within mammalian spinal motor neurons. *Nat. Commun.* **14**, 46. https://doi.org/10.1038/s41467-022-35574-x (2023).
7. 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. https://doi.org/10.1038/s41467-022-28636-7 (2022).
8. Nornes, H. O. & Carry, M. Neurogenesis in spinal cord of mouse: an autoradiographic analysis. *Brain Res.* **159**, 1–6. https://doi.org/10.1016/0006-8993(78)90105-1 (1978).
9. Richardson, W. D., Kessaris, N. & Pringle, N. Oligodendrocyte wars. *Nat. Rev. Neurosci.* **7**, 11–18. https://doi.org/10.1038/nrn1826 (2006).
10. Hintermann, A. et al. Developmental and evolutionary comparative analysis of a regulatory landscape in mouse and chicken. *Development* **149** https://doi.org/10.1242/dev.200594 (2022).
11. Ohtsuka, T. et al. Gene expression profiling of neural stem cells and identification of regulators of neural differentiation during cortical development. *Stem Cells*. **29**, 1817–1828. https://doi.org/10.1002/stem.731 (2011).
12. Ferri, A. L. et al. Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development* **131**, 3805–3819. https://doi.org/10.1242/dev.01204 (2004).
13. Graham, V., Khudyakov, J., Ellis, P. & Pevny, L. SOX2 functions to maintain neural progenitor identity. *Neuron* **39**, 749–765. https://doi.org/10.1016/s0896-6273(03)00497-5 (2003).
14. Zappone, M. V. et al. Sox2 regulatory sequences direct expression of a (beta)-geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells. *Development* **127**, 2367–2382. https://doi.org/10.1242/dev.127.11.2367 (2000).
15. Rhee, H. S. et al. Expression of terminal effector genes in mammalian neurons is maintained by a dynamic relay of transient enhancers. *Neuron* **92**, 1252–1265. https://doi.org/10.1016/j.neuron.2016.11.037 (2016).
16. Hand, R. & Polleux, F. Neurogenin2 regulates the initial axon guidance of cortical pyramidal neurons projecting medially to the corpus callosum. *Neural Dev.* **6**, 30. https://doi.org/10.1186/1749-8104-6-30 (2011).
17. Filova, I. et al. Combined Atoh1 and Neurod1 deletion reveals autonomous growth of auditory nerve fibers. *Mol. Neurobiol.* **57**, 5307–5323. https://doi.org/10.1007/s12035-020-02092-0 (2020).
18. Tsuchida, T. et al. Topographic organization of embryonic motor neurons defined by expression of LIM homeobox genes. *Cell* **79**, 957–970. https://doi.org/10.1016/0092-8674(94)90027-2 (1994).
19. Sharma, K. et al. LIM homeodomain factors Lhx3 and Lhx4 assign subtype identities for motor neurons. *Cell* **95**, 817–828. https://doi.org/10.1016/s0092-8674(00)81704-3 (1998).
20. 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. https://doi.org/10.1016/j.neuron.2008.06.025 (2008).
21. Francius, C. & Clotman, F. Dynamic expression of the onecut transcription factors HNF-6, OC-2 and OC-3 during spinal motor neuron development. *Neuroscience* **165**, 116–129. https://doi.org/10.1016/j.neuroscience.2009.09.076 (2010).

22. Roy, A. et al. Onecut transcription factors act upstream of Isl1 to regulate spinal motoneuron diversification. *Development* **139**, 3109–3119. https://doi.org/10.1242/dev.078501 (2012).
23. Kania, A., Johnson, R. L. & Jessell, T. M. Coordinate roles for LIM homeobox genes in directing the dorsoventral trajectory of motor axons in the vertebrate limb. *Cell* **102**, 161–173. https://doi.org/10.1016/s0092-8674(00)00022-2 (2000).
24. Palmesino, E. et al. Foxp1 and lhx1 coordinate motor neuron migration with axon trajectory choice by gating reelin signalling. *PLoS Biol.* **8**, e1000446. https://doi.org/10.1371/journal.pbio.1000446 (2010).
25. Jurata, L. W., Thomas, J. B. & Pfaff, S. L. Transcriptional mechanisms in the development of motor control. *Curr. Opin. Neurobiol.* **10**, 72–79. https://doi.org/10.1016/s0959-4388(99)00053-7 (2000).
26. Lee, S. K. & Pfaff, S. L. Transcriptional networks regulating neuronal identity in the developing spinal cord. *Nat. Neurosci.* **4 Suppl**, 1183–1191. https://doi.org/10.1038/nn750 (2001).
27. Pfaff, S. L., Mendelsohn, M., Stewart, C. L., Edlund, T. & Jessell, T. M. Requirement for LIM homeobox gene Isl1 in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation. *Cell* **84**, 309–320. https://doi.org/10.1016/s0092-8674(00)80985-x (1996).
28. Wasserman, W. W. & Sandelin, A. Applied bioinformatics for the identification of regulatory elements. *Nat. Rev. Genet.* **5**, 276–287. https://doi.org/10.1038/nrg1315 (2004).
29. Weirauch, M. T. et al. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443. https://doi.org/10.1016/j.cell.2014.08.009 (2014).
30. Lee, S. et al. A regulatory network to segregate the identity of neuronal subtypes. *Dev. Cell.* **14**, 877–889. https://doi.org/10.1016/j.devcel.2008.03.021 (2008).
31. Lee, S. et al. STAT3 promotes motor neuron differentiation by collaborating with motor neuron-specific LIM complex. *Proc. Natl. Acad. Sci. U S A.* **110**, 11445–11450. https://doi.org/10.1073/pnas.1302676110 (2013).
32. Briscoe, J., Pierani, A., Jessell, T. M. & Ericson, J. A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube. *Cell* **101**, 435–445. https://doi.org/10.1016/s0092-8674(00)80853-3 (2000).
33. Song, M. R. et al. Islet-to-LMO stoichiometries control the function of transcription complexes that specify motor neuron and V2a interneuron identity. *Development* **136**, 2923–2932. https://doi.org/10.1242/dev.037986 (2009).
34. Kim, K. T. et al. ISL1-based LIM complexes control Slit2 transcription in developing cranial motor neurons. *Sci. Rep.* **6**, 36491. https://doi.org/10.1038/srep36491 (2016).
35. Mazzoni, E. O. et al. Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity. *Nat. Neurosci.* **16**, 1219–1227. https://doi.org/10.1038/nn.3467 (2013).
36. Uemura, O. et al. Comparative functional genomics revealed conservation and diversification of three enhancers of the isl1 gene for motor and sensory neuron-specific expression. *Dev. Biol.* **278**, 587–606. https://doi.org/10.1016/j.ydbio.2004.11.031 (2005).
37. Kim, N., Park, C., Jeong, Y. & Song, M. R. Functional diversification of motor Neuron-specific Isl1 enhancers during evolution. *PLoS Genet.* **11**, e1005560. https://doi.org/10.1371/journal.pgen.1005560 (2015).
38. Livet, J. et al. ETS gene Pea3 controls the central position and terminal arborization of specific motor neuron pools. *Neuron* **35**, 877–892. https://doi.org/10.1016/s0896-6273(02)00863-2 (2002).
39. He, Y. et al. Spatiotemporal DNA methylome dynamics of the developing mouse fetus. *Nature* **583**, 752–759. https://doi.org/10.1038/s41586-020-2119-x (2020).
40. Patel, T. et al. Transcriptional dynamics of murine motor neuron maturation in vivo and in vitro. *Nat. Commun.* **13**, 5427. https://doi.org/10.1038/s41467-022-33022-4 (2022).
41. 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. https://doi.org/10.1016/j.celrep.2016.01.067 (2016).
42. De Garcia, M., Jessell, T. M. & N. V. & Early motor neuron pool identity and muscle nerve trajectory defined by postmitotic restrictions in Nkx6.1 activity. *Neuron* **57**, 217–231. https://doi.org/10.1016/j.neuron.2007.11.033 (2008).
43. Cho, H. H. et al. Isl1 directly controls a cholinergic neuronal identity in the developing forebrain and spinal cord by forming cell type-specific complexes. *PLoS Genet.* **10**, e1004280. https://doi.org/10.1371/journal.pgen.1004280 (2014).
44. Haase, G. et al. GDNF acts through PEA3 to regulate cell body positioning and muscle innervation of specific motor neuron pools. *Neuron* **35**, 893–905. https://doi.org/10.1016/s0896-6273(02)00864-4 (2002).
45. Misra, M., Shah, V., Carpenter, E., McCaffery, P. & Lance-Jones, C. Restricted patterns of Hoxd10 and Hoxd11 set segmental differences in motoneuron subtype complement in the lumbosacral spinal cord. *Dev. Biol.* **330**, 54–72. https://doi.org/10.1016/j.ydbio.2009.03.009 (2009).
46. Goulding, M. Specifying motor neurons and their connections. *Neuron* **21**, 943–946. https://doi.org/10.1016/s0896-6273(00)80612-1 (1998).
47. Thaler, J. P. et al. A postmitotic role for Isl-class LIM homeodomain proteins in the assignment of visceral spinal motor neuron identity. *Neuron* **41**, 337–350. https://doi.org/10.1016/s0896-6273(04)00011-x (2004).
48. 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. https://doi.org/10.1016/j.cell.2005.09.009 (2005).
49. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods.* **18**, 1333–1341. https://doi.org/10.1038/s41592-021-01282-5 (2021).
50. Wu, T. et al. ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innov. (Camb).* **2**, 100141. https://doi.org/10.1016/j.xinn.2021.100141 (2021).
51. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. ChromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods.* **14**, 975–978. https://doi.org/10.1038/nmeth.4401 (2017).
52. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37**, W305–311. https://doi.org/10.1093/nar/gkp427 (2009).
53. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018. https://doi.org/10.1093/bioinformatics/btr064 (2011).
54. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386. https://doi.org/10.1038/nbt.2859 (2014).
55. Pliner, H. A. et al. Cicero Predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol. Cell* **71**, 858–871 e858 (2018). https://doi.org/10.1016/j.molcel.2018.06.044
56. Stuart, T. et al. Comprehensive integration of Single-Cell data. *Cell* **177**, 1888–1902. https://doi.org/10.1016/j.cell.2019.05.031 (2019). e1821.

Acknowledgements

This work was supported by grants from NRF (NRF-2023R1A2C1002690, NRF-RS-2024-00346401, NRF-RS-2024-00411700 and NRF-RS-2025-02304360), SNUH Lee Kun-hee Child Cancer & Rare Disease Project (22B-001-0500), InnoCORE program of the Ministry of Science and ICT (GIST InnoCORE KH0830) and the GIST Research Institute in 2025. We thank Joyce Anne Diaz for technical support.

Author contributions

I.S.Y. and M.-R.S. conceived and directed the project; I.S.Y., D.J.J., Y.L, and J.-W.A. performed research; I.S.Y. and S.-E.B. performed data analyses; I.S.Y. and M.-R.S. wrote the manuscript. All authors reviewed and edited the article.

Declarations

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/s41598-025-18807-z.

Correspondence and requests for materials should be addressed to M.-R.S.

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

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