Stem Cell Reports

Article

The multi-lineage transcription factor ISL1 controls cardiomyocyte cell fate through interaction with NKX2.5

Bonnie E.J. Maven,[1,2,3] Casey A. Gifford,[1,3] Melanie Weilert,[4] Barbara Gonzalez-Teran,[1,3] Ruth Hüttenhain,[1,5,6] Angelo Pelonero,[1,3] Kathryn N. Ivey,[1,3] Kaitlen Samse-Knapp,[1,3] Wesley Kwong,[1,3] David Gordon,[1,5,6] Michael McGregor,[1,5,6] Tomohiro Nishino,[1,3] Eyuche Okorie,[1,3] Sage Rossman,[1,3] Mauro W. Costa,[1,3] Nevan J. Krogan,[1,5,6] Julia Zeitlinger,[4,7] and Deepak Srivastava[1,3,8,9,*]

[1]Gladstone Institutes, San Francisco, CA, USA
[2]Developmental and Stem Cell Biology PhD Program, University of California, San Francisco, San Francisco, CA, USA
[3]Roddenberry Center for Stem Cell Biology at Gladstone, San Francisco, CA, USA
[4]Stowers Institute for Medical Research, Kansas City, MO, USA
[5]Department of Cellular and Molecular Pharmacology, University of California San Francisco, San Francisco, CA, USA
[6]Quantitative Biosciences Institute (QBI), University of California San Francisco, San Francisco, CA, USA
[7]Department of Pathology and Laboratory Medicine, University of Kansas School of Medicine, Kansas City, KS, USA
[8]Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA, USA
[9]Department of Pediatrics, UCSF School of Medicine, San Francisco, CA, USA
*Correspondence: deepak.srivastava@gladstone.ucsf.edu
https://doi.org/10.1016/j.stemcr.2023.09.014

SUMMARY

Congenital heart disease often arises from perturbations of transcription factors (TFs) that guide cardiac development. ISLET1 (ISL1) is a TF that influences early cardiac cell fate, as well as differentiation of other cell types including motor neuron progenitors (MNPs) and pancreatic islet cells. While lineage specificity of ISL1 function is likely achieved through combinatorial interactions, its essential cardiac interacting partners are unknown. By assaying ISL1 genomic occupancy in human induced pluripotent stem cell-derived cardiac progenitors (CPs) or MNPs and leveraging the deep learning approach BPNet, we identified motifs of other TFs that predicted ISL1 occupancy in each lineage, with NKX2.5 and GATA motifs being most closely associated to ISL1 in CPs. Experimentally, nearly two-thirds of ISL1-bound loci were co-occupied by NKX2.5 and/or GATA4. Removal of NKX2.5 from CPs led to widespread ISL1 redistribution, and over-expression of NKX2.5 in MNPs led to ISL1 occupancy of CP-specific loci. These results reveal how ISL1 guides lineage choices through a combinatorial code that dictates genomic occupancy and transcription.

INTRODUCTION

Cardiac malformations are typically caused by abnormal specification or morphogenetic events related to specific subsets of developing cardiac cells. Transcription factors (TFs) essential for cardiogenesis have been identified, and mutations in these often underlie human cardiac malformations (Yasuhara and Garg, 2021). However, the precise mechanism by which they guide the gene expression necessary for proper differentiation and morphogenesis remains an active area of investigation.

ISL1 is a TF transiently expressed during emergence and expansion of multipotent progenitor second heart field (SHF) cells, before being downregulated upon further differentiation (Black, 2007). ISL1 activates cardiac transcriptional networks by directly inducing essential cardiac TFs such as *MEF2C* and *GATA* factors (Srivastava, 2006). In mice that lack *ISL1*, the SHF derivative structures are severely reduced because of defects in proliferation, survival, and migration, leading to lethality by embryonic day (E) 10.5 (Cai et al., 2003). In humans, truncating mutations in *ISL1* have been associated with SHF defects, including double outlet right ventricle (Wang et al., 2019). A microdeletion of the entire *ISL1* allele has been

associated with tetralogy of Fallot (Osoegawa et al., 2014), a related outflow tract defect. Thus, ISL1 is critical for establishing and maintaining mammalian SHF cardiac progenitor (CP) populations.

ISL1 is also expressed more broadly, including in neuronal and pancreatic progenitor cells, where it is essential for normal development, suggesting a combinatorial code involving ISL1-associated factors that leads to cell type-specific gene expression and cell fate determination. Within neuronal progenitors, ISL1 interacts with TFs such as the neuronal LHX3 and the ubiquitously expressed LDB1 to regulate genes associated with this fate (Seo et al., 2015). Loss of ISL1 leads to complete absence of motor neurons (MNs) in mice (Pfaff et al., 1996). Similarly, in endodermal progenitors that give rise to the pancreas, ISL1 interacts with the TFs SSBP3 and LDB1 and is required for formation of insulin-producing beta cells (Ediger et al., 2014; Galloway et al., 2015). ISL1 is known to interact with several factors during cardiac differentiation, including LDB1, but this interaction is not cardiac specific (Galloway et al., 2015; Narkis et al., 2012), so is unlikely to be sufficient to guide cardiac-specific gene expression. ISL1 also interacts with the ubiquitously expressed chromatin modifier KDM6B to activate cardiac gene expression
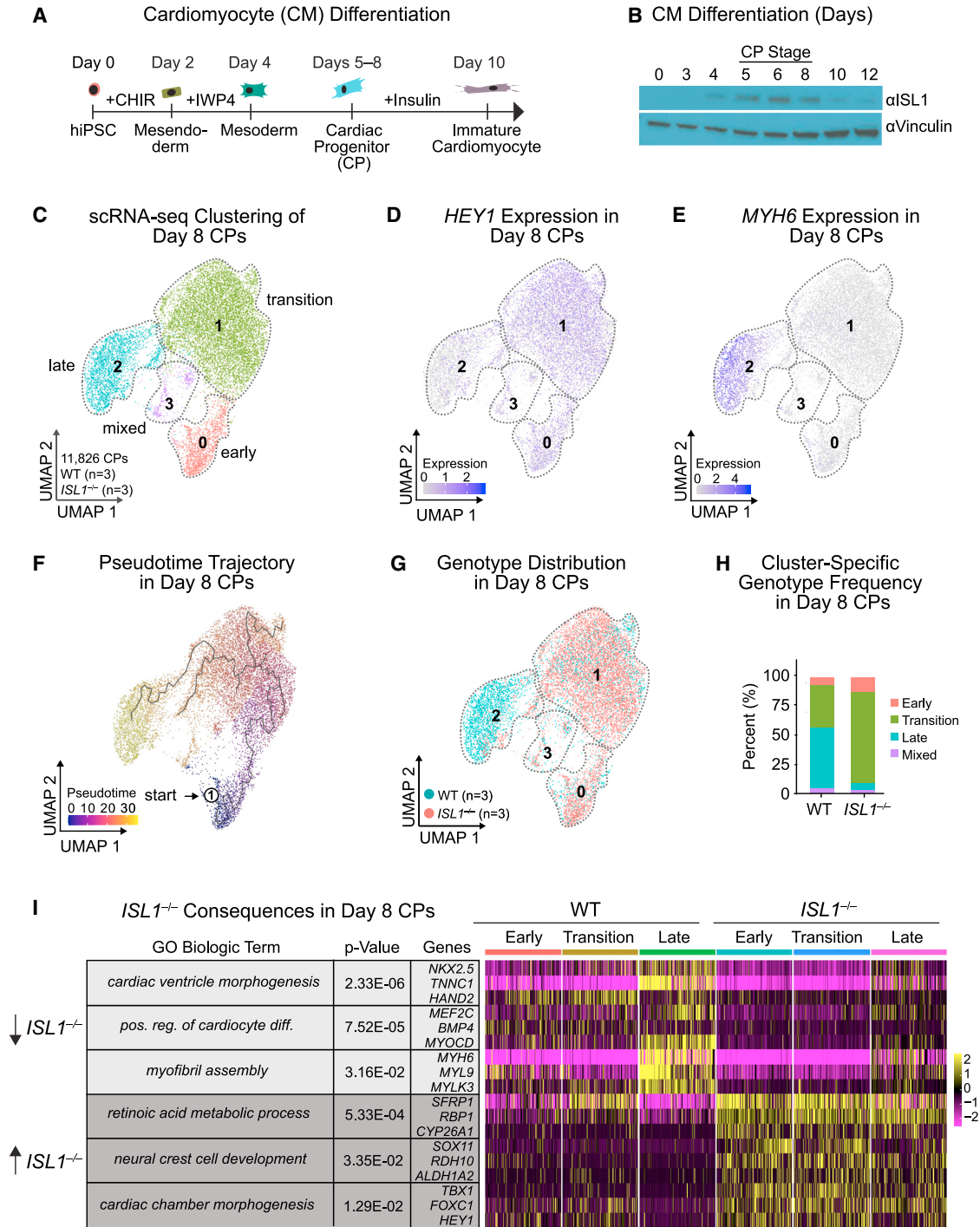
Figure 1. ISL1 is required for cardiac cell-fate-specific gene regulation

(A) Schematic of CM differentiation from hiPSCs.

(B) Western blot of ISL1 expression during hiPSC-CM differentiation (days 0–12). Vinculin is the loading control.

(C) Hierarchical UMAP clustering of WT (n = 3 independent experiments) or *ISL1*$^{-/-}$ (n = 3 independent experiments) day 8 CPs after scRNA-seq.

(D and E) Expression of *HEY1* (D) and *MYH6* (E) in day 8 CPs, superimposed on the UMAP from (C).

(F) Monocle pseudotime analysis of day 8 CPs. Colors represent arbitrary units of pseudotime.

(Wang et al., 2016), but the ISL1-associated TFs that lead to cardiac-specific gene expression and subsequent differentiation remain unknown.

Here, we used human induced pluripotent stem cells (hiPSCs) to capture transient ISL1-expressing CP and MN progenitor (MNP) populations (NeuroLINCS Consortium et al., 2021) and investigated how cardiac and neuronal gene expression are differentially regulated by ISL1. We identify NKX2.5 as a critical binding partner for ISL1 to direct cardiac-specific gene expression, and demonstrate how a single TF can have distinct regulatory roles in directing fate decisions based on cellular context.

RESULTS

ISL1 is necessary for instructing cell-fate-specific gene regulation

We differentiated hiPSCs into MYH6$^+$:TNNT2$^+$ CPs and subsequently into cardiomyocytes (CMs) using a WNT modulation protocol (Figures 1A and S1A) (Lian et al., 2013). These cells pass through a transient CP-like stage associated with ISL1 enrichment between days 5 and 8 (Figure 1B), followed by beating foci concurrent with ISL1 downregulation, and ultimately a beating monolayer of CMs by day 10 (Video S1).

We generated an $ISL1^{-/-}$ hiPSC line using two Cas9-gRNAs to excise a portion of the gene starting in exon 2 and ending in exon 5 (Figure S1B). The excision abolished $ISL1$ RNA and protein expression (Figures S1C and S1D). CMs differentiated from $ISL1^{-/-}$ hiPSCs displayed decreased sarcomeric gene expression at the CP stage (Figure S1A) and a delay in beating induction (Figure S1E and Video S1), similar to previous results (Quaranta et al., 2018).

We performed single-cell RNA sequencing (scRNA-seq) on wildtype (WT) or $ISL1^{-/-}$ progenitor populations at day 8 (Figure 1C). Within the cardiac lineage, we identified four transcriptional signatures (Figures 1C and S1F; Table S1). Expression of more immature markers (FOXC1, HEY1, and SOX4) was higher in clusters 0 and 1, while expression of more differentiated gene markers (e.g., TNNT2, ACTN2, and MYH6) was high in cluster 2 (termed "late" cluster) (Figures 1D, 1E, and S1F). In addition, cluster 0 expressed higher levels of cell replication genes (NUF2, PRC1, and KIF4A) than cluster 1, indicating that cells from cluster 0 are more immature and proliferative; we therefore refer to cluster 0 as the "early" cluster and cluster 1 as the "transition" cluster (Figure S1F). Cluster 3 seems to be a mixture of early

and late differentiating cells (termed "mixed") (Figure S1F). Pseudotime analysis confirmed this cell trajectory in silico (Figure 1F).

In the $ISL1^{-/-}$ line, a greater proportion of CPs were in the early and transition clusters compared with the WT line, and fewer cells were in the late cluster (Figures 1G, 1H, and S1G). Within each of the early, transition, and late clusters, we found significant differences between genotypes (Table S1). For example, $ISL1^{-/-}$ CPs were depleted for genes associated with the Gene Ontology (GO) terms of cardiac ventricle morphogenesis (e.g. NKX2.5, TNNC1) and myofibril assembly (e.g. MYH6, MYL9), and enriched for GO terms related to retinoic acid metabolic process (e.g. CYP26A1, ALDH1A2, RBP1) and cardiac chamber morphogenesis, specifically atrial markers (e.g. HEY1, NR2F1) (Figures 1I and S1H; Table S1). This is in agreement with previous findings that $ISL1$ promotes ventricular cell fate and suppresses atrial identity in vitro (Quaranta et al., 2018).

Complete deletion of $ISL1$ revealed its early role in cardiogenesis (Figures 1A and 1B), but may have concealed functions in differentiation after cell fate determination. Therefore, we used small interfering RNAs (siRNA) to knockdown (KD) $ISL1$ specifically during the CP stage, after CP commitment (Figure S1I). Rhodamine-labeled siRNAs targeting $ISL1$ or a scrambled (Scra) control were introduced on day 5 of CM differentiation, and isolated rhodamine$^+$ cells with significant downregulation of $ISL1$ were analyzed on day 8 (Figures S1I and S1J). scRNA-seq showed that $ISL1$ KD at the CP stage did not substantially change cell fate, with $ISL1$ KD cells clustering with controls (Figure S1K). Next, we generated a high-confidence list of genes potentially regulated by ISL1 in CPs by intersecting the significantly downregulated genes in both the $ISL1^{-/-}$ and $ISL1$ KD conditions (Table S1). This list included cardiac genes such as MYL9 and TNNT2 (Figures S1L and S1M), confirming our previous $ISL1^{-/-}$ results, suggesting that many dysregulated genes are likely a direct consequence of the absence of $ISL1$.

To identify cardiac-specific consequences of $ISL1$ removal, we differentiated the same $ISL1^{-/-}$ hiPSC line into SMI32- and NKX6.1-expressing MNPs, which form axonal projections by day 18, followed by scRNA-seq (Figures 2A–2C) (NeuroLINCS Consortium et al., 2021). $ISL1$ deletion led to decreased expression of MN markers NKX6.1 and SMI32, consistent with observations in an $ISL1$ hypomorphic mouse model (Figure 2C) (Kim et al., 2016).

(G) FeaturePlot display of the distribution of WT and $ISL1^{-/-}$ day 8 CPs across clusters depicted in (C).

(H) Percentages of each genotype in each cluster. Cell numbers for each genotype are normalized to the total number of cells in each cluster.

(I) Heatmap of genes and accompanying biological GO terms significantly dysregulated in scRNA-seq of $ISL1^{-/-}$ as compared with WT day 8 CPs.
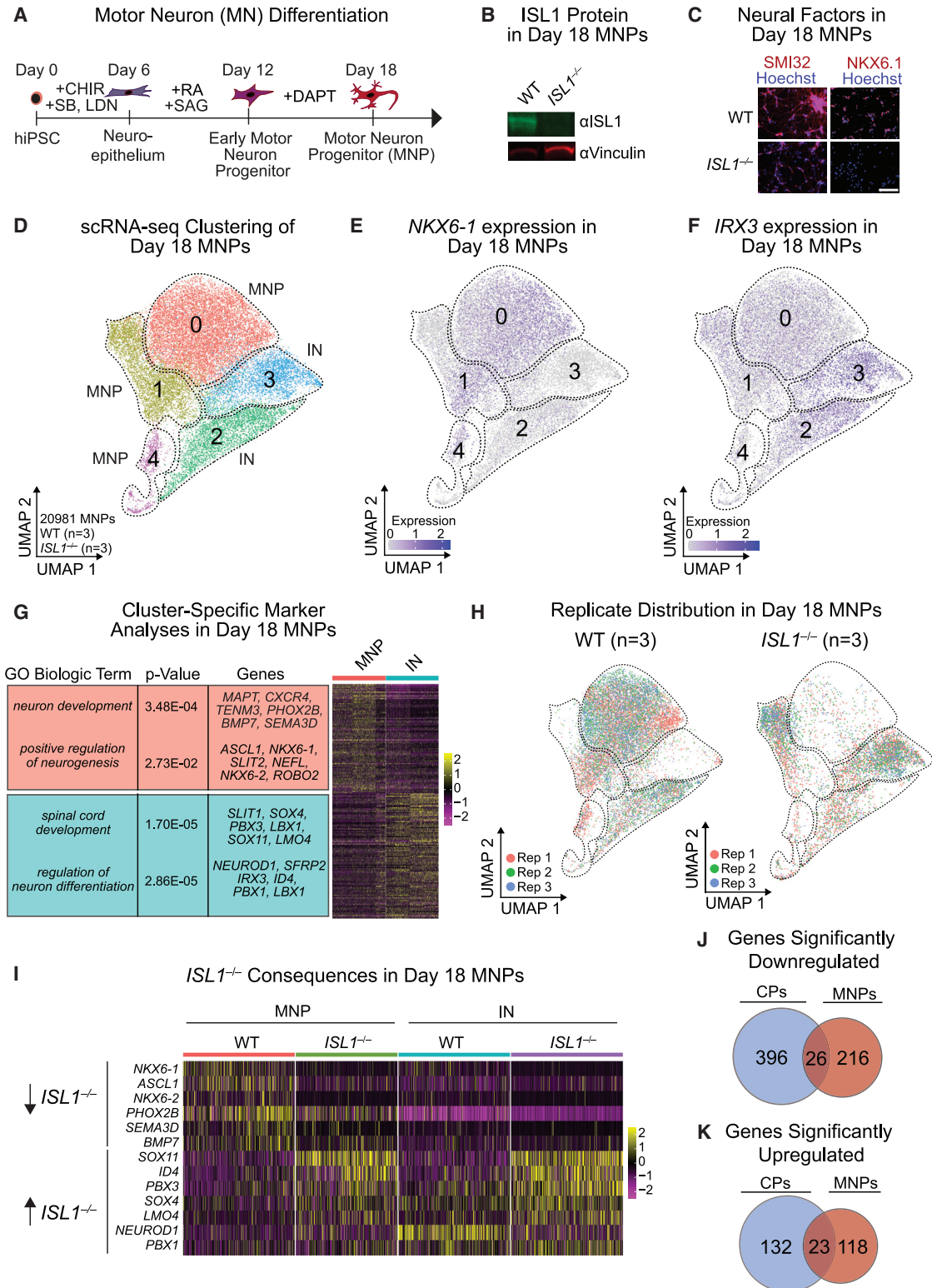
Figure 2. scRNA-seq analyses of ISL1 function in MNPs

(A) Schematic of MN differentiation from hiPSCs.

(B) Western blot of ISL1 in WT and $ISL1^{-/-}$ day 18 MNPs. Vinculin served as the loading control.

In the MNP lineage, we observed five distinct transcriptional clusters (Figure 2D). All clusters expressed neuron-specific markers, with clusters 0, 1, and 4 also expressing MNP markers such as *NKX6.1*, *NKX6.2*, and *PHOX2A* (termed "MNP" clusters), whereas clusters 2 and 3 had higher expression of interneuron (IN) markers such as *IRX3* and *PAX2* (termed "IN" clusters) (Figures 2D–2G; Table S2). In the absence of *ISL1*, cells were more likely to be clustered in the IN clusters rather than MNP clusters (Figure 2H). In addition, we found downregulation of MNP genes, such as *NKX6.1* and *PHOX2B*, and upregulation of IN genes such as *PBX1* and *LMO4* in *ISL1*$^{-/-}$ cells (Figure 2I), consistent with observations in knockout *ISL1* mouse models (Song et al., 2009). The overlap in differentially regulated genes in CPs and MNPs was small, indicating that ISL1 regulates a unique group of genes and regulatory networks in each cell type (Figures 2J and 2K).

ISL1 genomic localization is lineage specific

To determine which genes are directly regulated by ISL1 in CPs and MNPs, we used an antibody to endogenous ISL1 to perform chromatin immunoprecipitation followed by sequencing (ChIP-seq) in triplicate (Figure S2A). ISL1 consistently interacted with 4,715 genomic regions in day 6 CPs, and 10,767 regions in day 18 MNPs, with only 227 peaks shared between cell types (Figure 3A). In each cell type, ISL1-bound regions were most frequently distal (>50 kb) to gene transcription start sites (TSSs) (Figure 3B) and characterized as intronic or distal intergenic, within regions with active histone marks and a relative lack of repressive marks, consistent with ISL1 likely binding to distal enhancers (Figures S2B and S2C). Examples of cell-type-specific ISL1 genomic localization are illustrated by CP-specific peaks at the cardiac-related *MYL4* locus and MNP-specific peaks at the MN-related *PHOX2B* locus (Figures 3C and S2D). Some sites were shared between cell types, such as the ISL1 interaction near the *SETD5* locus, which encodes a ubiquitous histone methyltransferase (Figure 3C).

To interrogate the function of ISL1-occupied loci, we used the Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al., 2010). Significantly enriched GO terms for genes near ISL1-bound CP peaks were associated with cardiac differentiation, such as *Cardiac chamber development* and *Cardiac chamber morphogenesis*, whereas significant terms for ISL1 MNP peaks were related to neural differentiation, such as *Cranial nerve development* and *Cell differentiation in spinal cord*. GREAT did not detect any significant terms among peaks shared by CPs and MNPs (Figures S2E and S2F).

We expected that ISL1 occupancy might depend on cell-type-specific co-factors. To identify putative co-bound TFs, we visualized TF motifs in the vicinity of the ISL1-bound sites with Hypergeometric Optimization of Motif EnRichment (HOMER) (Li et al., 2009). We combined the top 16 motifs from each cell type (*sans* ISL1 motif) and compared motif enrichment of this set across CPs and MNPs (Figure S3A). ISL1-bound loci in CPs were enriched in the family of motifs for GATA (including GATA4), TEAD, NKX2 (including NKX2.5), and HAND2. In contrast, ISL1-bound loci in MNPs were enriched in motifs in the family of motifs for NEUROD, CUX, LHX (including NKX6.1), and PHOX2. The enrichment of LHX family motifs in ISL1-bound regions in MNPs is consistent with the known ISL1-LHX3 composite motifs identified in developing MNs (Seo et al., 2015). These data provide evidence for unique enrichment of known TF motifs near ISL1-bound sites in either cell type.

Next, we used the deep learning model BPNet to identify *de novo* motifs that help to predict ISL1 binding (Avsec et al., 2021). This model can learn complex rules by which combinations of *de novo* sequence motifs best predict the experimental binding data. BPNet was trained to predict the cell-type-specific binding of ISL1 at all 15,483 regions and optimized by hyperparameter tuning (see experimental procedures, Figure S3B). Interpretation tools were used to systematically extract the sequence motifs from the network (Figure 3D), which matched known motifs recognized by relevant TFs. As expected, the ISL1 motif was among the *de novo* motifs important for the binding in both cell types and included motifs for the neural TF families ONECUT, NEUROD, and EBF2, all of which have known roles in MN development (Son et al., 2011; Velasco et al., 2017) and an ISL1-containing composite motif that was previously identified to be important in MNPs (Seo et al., 2015). In addition, the NKX2 family motif identified may correspond with NKX2.2 or NKX2.8, both important for neuronal progenitor populations (Jarrar et al., 2015).

(C) Immunofluorescence of neural factors SMI32 and NKX6.1 in day 18 MNPs. Scale bar, 100 μM.

(D) Hierarchical UMAP clustering of WT (n = 3 independent experiments) and *ISL1*$^{-/-}$ (n = 3 independent experiments) scRNA-seq data in day 18 iPSCs differentiated toward MNPs.

(E and F) Expression levels of the MN-related TF *NKX6-1* (E) and the IN-related TF *IRX3* (F) in day 18 MNPs, superimposed on the UMAP from (D).

(G) Heatmap of genes enriched with accompanying GO terms in each of the MNP or IN clusters depicted in (D).

(H) Replicate (Rep) comparison in WT (n = 3; 14,609 cells) or *ISL1*$^{-/-}$ (n = 3; 6,372 cells) day 18 MNPs across the clusters depicted in (D).

(I) Heatmap clustering of genes significantly dysregulated in scRNA-seq of *ISL1*$^{-/-}$ compared with WT day 18 MNPs.

(J) Venn diagram of significantly downregulated genes shared between *ISL1*$^{-/-}$ day 8 CPs and *ISL1*$^{-/-}$ day 18 MNPs.

(K) Venn diagram of significantly upregulated genes shared between *ISL1*$^{-/-}$ day 8 CPs and *ISL1*$^{-/-}$ day 18 MNPs.
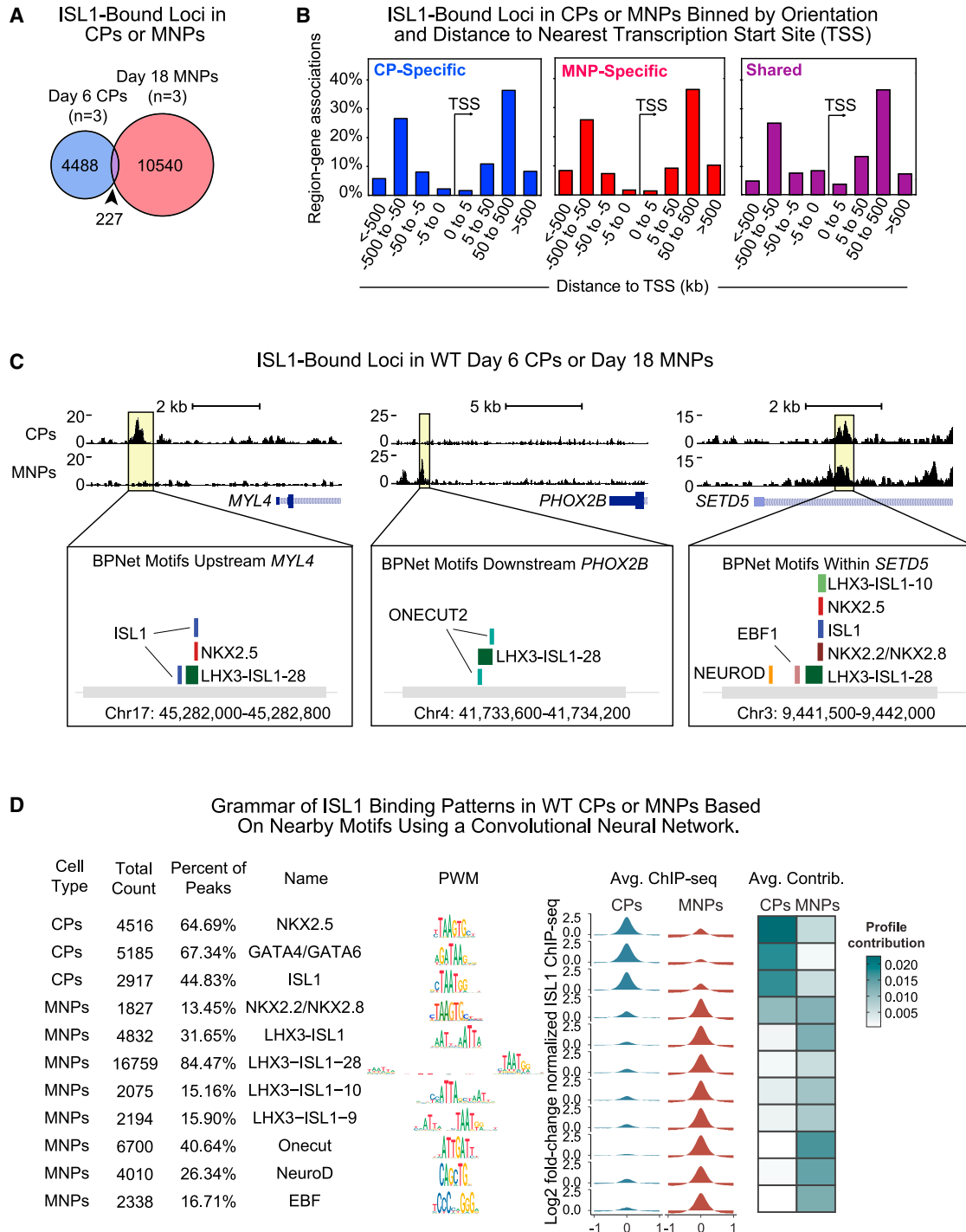
Figure 3. Grammar of distinct ISL1 binding patterns in cardiac or MNPs based on nearby motifs using deep learning

(A) Venn diagram of ISL1 peaks in CPs or MNPs. Overlap is statistically significant (Regione R) (p = 0.001).

(B) Histogram of ISL1-bound peaks based on distance from gene TSSs.

(C) Example ChIP-seq tracks showing ISL1-bound loci with CP-specific, MNP-specific, or shared peaks. Corresponding *de novo* motifs identified by BPNet that predict cell-type-specific ISL1 binding are shown below. Data shown from a single representative replicate.

(D) *De novo* motifs identified by BPNet after simultaneous training of the CP-specific or MNP-specific binding profiles. Displayed for each identified motif is: the total count of motifs and percentages, name based on the most likely TF(s) that binds them, the position weight matrices (PWMs), and average ChIP-seq intensity and average contribution to the ISL1 binding predictions as extracted from BPNet.

Predicted ISL1-associated motifs in CPs belonged to cardiac TF families NKX2 and GATA (Figure 3D), consistent with our HOMER analyses. The GATA motif likely corresponds with binding of the cardiac TFs GATA4 and/or GATA6, while the NKX2 family motif most likely corresponds with the cardiac TF NKX2.5, which is necessary for development of the outflow tract and right ventricle in mice (Lyons et al., 1995; Zhang et al., 2014), similar to ISL1. The average contribution to predicting ISL1 binding was greatest with the NKX2.5 motif. Examples of *de novo* motifs found in cell-type-specific and shared ISL1-bound regions are illustrated in the ChIP-seq tracks in Figure 3C.

We examined the percentage of ISL1-bound peaks with GATA and/or NKX2.5 motifs and the nature of those sites (Figure S3C) and found cardiac-related GO terms enriched for each category (Figure S3D). This was also the case when the motifs of GATA and NKX2.5 were found in combination or when neither were there. The regions with only the NKX2.5 motif were also enriched in terms related to epithelial-to-mesenchymal transition, which may reflect the role of NKX2.5 in cardiac cell migration. When the NKX2.5 motif was found in combination with the GATA motif, a majority of the terms were exclusively cardiac, demonstrating the importance of combinatorial TFs during cardiac differentiation (Figure S3D).

ISL1 forms a DNA-binding complex with NKX2.5

To experimentally identify TFs that cooperate with ISL1 for cardiac fate specification, we used a combination of targeted mass spectrometry (MS) based on selected reaction monitoring (SRM) (Calvo et al., 2011) and co-immunoprecipitation (co-IP) (Figure 4A). We selected TFs for which there were specific known and/or *de novo* motif enrichment in ISL1 ChIP-seq in WT CPs from our studies (Figures 3D and S3A) and designed targeted SRM assays for unique peptides representing them (Table S3). We immunoprecipitated ISL1 from WT or *ISL1*$^{-/-}$ CPs as a negative control, followed by SRM, and showed enrichment for ISL1 and LDB1, validating the ISL1 IP. We found no enrichment for PHOX2A or LHX3, indicating the specificity of this method (Figure 4B). Of the cardiac TFs analyzed, NKX2.5 showed the strongest enrichment for several peptides in the ISL1 IP compared with controls, while other TFs including GATA4, GATA6, and TEAD4 showed some enrichment (Figure 4B).

Since the deep learning model predicted the binding of ISL1 in CPs based on a combination of ISL1, NKX2.5, and GATA4/GATA6 motifs, we asked whether the cooperativity between the *de novo*-derived motifs was distance dependent by testing how *in silico* mutated motif patterns would be predicted to disrupt co-binding at specified distances. The strongest synergy was observed when the ISL1 and NKX2.5 motifs were in close proximity of under 35 bp,

while no detectable synergy was observed at nucleosome distance (70–150 bp) (Figure 4C). In addition, short-range synergy was observed between the GATA4/GATA6 motif and the ISL1 and NKX2.5 motifs.

To further investigate the cooperativity between ISL1 and NKX2.5 and ISL1 and GATA4, we asked whether these interactions could occur in an alternative cell type in the absence of additional putative cardiac co-factors. In co-IPs from COS-7 cells, exogenously expressed ISL1 interacted strongly with LDB1 (positive control), NKX2.5, and GATA4 (Figure S4A), confirming that these interactions do not require cardiac-specific co-factors. Given the similar cardiac expression pattern and loss-of-function effects between ISL1 and NKX2.5 (Zhang et al., 2014) and the evidence that they genetically interact (Dorn et al., 2015), we focused our investigation on understanding the function of ISL1 and NKX2.5 interaction. To this end, we validated the endogenous ISL1-NKX2.5 interaction in CPs (Figures S4B and S4C).
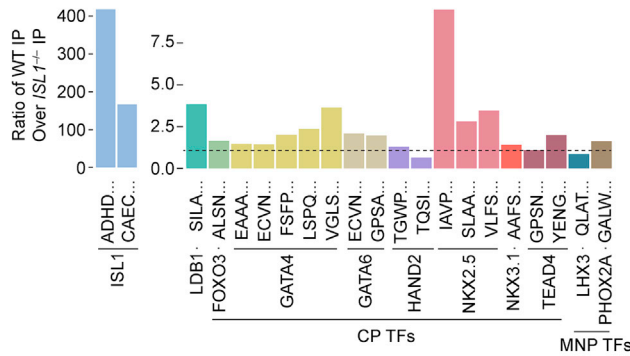
NKX2.5 co-occupies a majority of ISL1-bound loci in CPs and drives overlapping cardiac gene regulatory networks

We next tested if NKX2.5 co-localizes to ISL1-bound genomic regions in CPs by ChIP-seq using an antibody to endogenous NKX2.5 in day 6 CPs (Figure S4D). This analysis revealed 32,637 sites. NKX2.5 peaks overlapped with more than 60% of ISL1 peaks in CPs (2,911 loci) (Figures 5A and 5B), but with only 5% of the total MNP ISL1 peaks (483 loci) (Figure S4E). Loci bound only by NKX2.5 tended to be both proximal and distal to TSSs, in contrast with sites only bound by ISL1 or co-bound by both, which were primarily distal sites (Figure S4F). NKX2.5 tended to bind near promoters, yet ISL1-only or co-bound loci tended to be enriched in distal intergenic regions, consistent with a role in regulating enhancers when co-bound (Figure S4G). GREAT analysis showed that peaks co-bound by ISL1 and NKX2.5 were most enriched for cardiac developmental terms such as *Cardiac chamber development* and *Heart morphogenesis*, similar to ISL1-only bound loci, but distinct from NKX2.5-only loci (Figure S4H).

To determine the ISL1-binding dependency on NKX2.5, we generated an *NKX2.5*$^{-/-}$ hiPSC line using CRISPR-Cas9 gene editing (Figures S5A and S5B). As expected, because of the known role of NKX2.5 in repression of *ISL1*, deletion of *NKX2.5* led to upregulation of ISL1 protein (Figure S5C) (Dorn et al., 2015). Given that deletion of *NKX2.5* did not grossly disrupt morphological structures in CMs (Figure S5D), we investigated the degree to which NKX2.5 shares a regulatory network with ISL1 in developing CPs by comparing the scRNA-seq profiles of *NKX2.5*$^{-/-}$ and *ISL1*$^{-/-}$ CPs (Figure 5C). The day 8 CP cells clustered into the same broad clusters as before, indicated
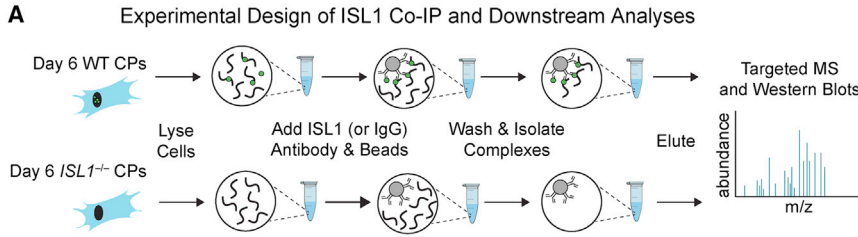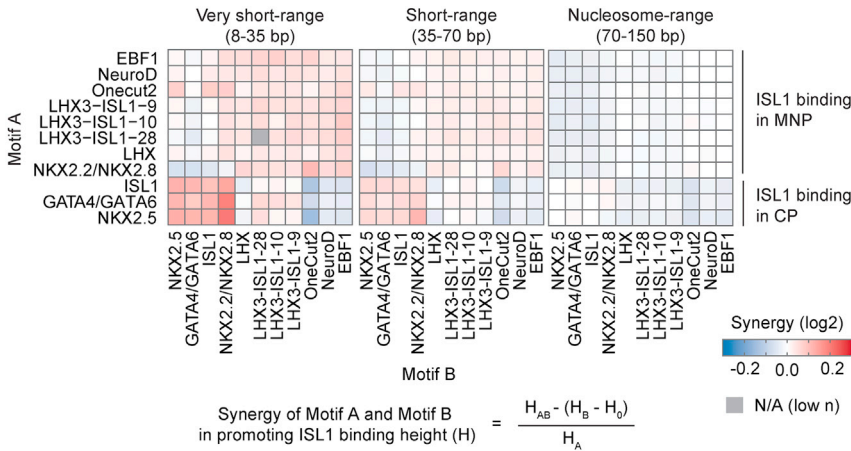
Figure 4. Proteomics and BPNet identify NKX2.5 and other TFs that guide ISL1 DNA-binding specificity in CPs

(A) Schematic for ISL1 co-IP and subsequent SRM targeted proteomics analyses.

(B) Barplot results from targeted affinity purification followed by MS, depicting enrichment of individual peptides corresponding to cardiac factors, positive controls (ISL1, LDB1), and negative controls (LHX3, PHOX2A) (n = 1 independent experiment).

(C) Cooperative relationship among all *de novo* identified motifs binned by distance from ISL1 binding predicted by *in silico* motif mutation.

$$\text{Synergy of Motif A and Motif B in promoting ISL1 binding height (H)} = \frac{H_{AB} - (H_B - H_0)}{H_A}$$

by expression of markers in each cluster and a progression of differentiation validated by pseudotime analysis (Figures 5C, 5D, and S6E; Table S4).

We also observed *ISL1* expression throughout all clusters in WT day 8 CPs, yet *NKX2.5* expression was enriched in the more differentiated late cluster (Figures S5F and S5G). Therefore, we focused our analyses on the transition and late clusters (Figures 5E, 5F, and S5H; Table S4). Both *ISL1* and *NKX2.5* were required for proper expression of genes related to *Muscle cell development* (e.g. *CSRP3*, *MYL9*) and *Regulation of muscle contraction* (e.g. *NPPA*, *CNN1*) (Figure 5F). These findings for CPs devoid of *NKX2.5⁻/⁻* are in line with previous bulk

RNA-seq findings (Anderson et al., 2018). Without *ISL1* or *NKX2.5*, we observed upregulation of genes related to *Glial cell differentiation* (e.g. *SOX4*, *SOX11*) and *Regulation of binding* (e.g. *NFIB*, *TMSB4X*). The finding of upregulated glial-related markers suggests *ISL1* and *NKX2.5* might be necessary to repress alternative cell fates; on the other hand, *SOX4* and *SOX11* are critical for early outflow tract development (Paul et al., 2014) and their upregulation could indicate a critical role for *ISL1* and *NKX2.5* in suppressing them as CPs proceed through differentiation. Among genes specifically upregulated in the absence of *NKX2.5* were markers related to *Positive regulation of cardiomyocyte differentiation* (e.g. *GATA4*,
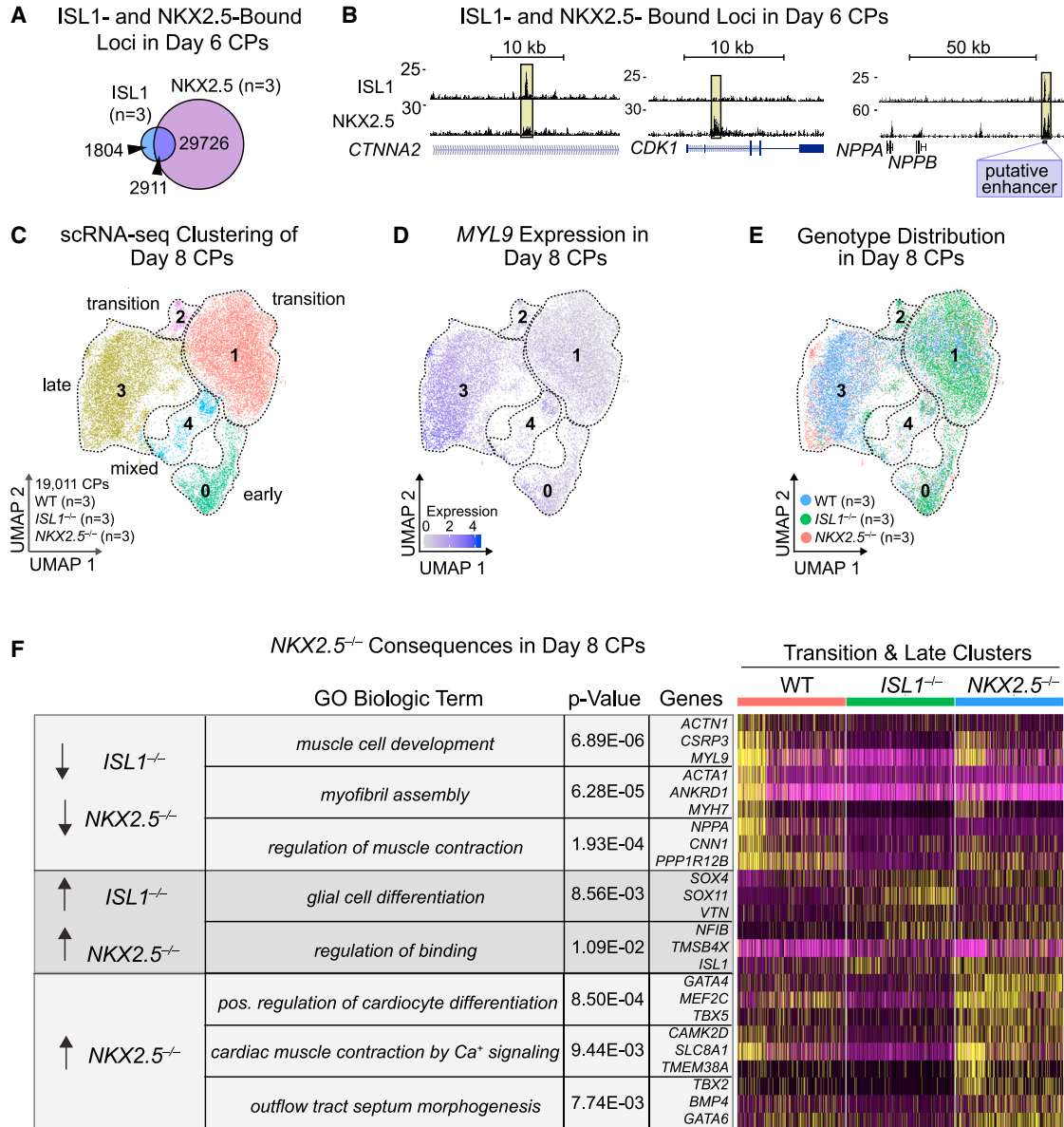
Figure 5. NKX2.5 binds a majority of ISL1 peaks and cooperates with ISL1 to regulate gene expression in CPs

(A) Venn diagram of NKX2.5 and ISL1 ChIP-seq peaks in WT day 6 CPs.

(B) Example ChIP-seq tracks showing loci bound by ISL1, NKX2.5, or both in day 6 CPs. Putative *NPPA/NPPB* enhancer labeled as related to Figure S5I. Data shown from single representative replicate.

(C) Hierarchical UMAP scRNA-seq clustering of WT, *ISL1*$^{-/-}$, and *NKX2.5*$^{-/-}$ day 8 CPs (n = 3 independent experiments).

(D) Expression of *MYL9* in day 8 CPs, superimposed on the UMAP from (C).

(E) FeaturePlot display of the distribution of WT and *ISL1*$^{-/-}$ day 8 CPs across the clusters depicted in (C).

(F) Heatmap of significantly dysregulated genes with biological GO terms in day 8 *ISL1*$^{-/-}$ or *NKX2.5*$^{-/-}$ CPs compared with WT.

TBX5), which could indicate a role for *NKX2.5* in blocking differentiation until the correct time in development. In addition, we observed upregulated genes related to *Cardiac muscle contraction by Ca*$^+$ *signaling* (e.g. *CAMK2D, SLC8A1*), in agreement with *NKX2.5*'s role in suppressing pacemaker cell fate (Dorn et al., 2015).

To test the hypothesis that some cardiac genes are co-regulated by ISL1 and NKX2.5, we evaluated the regulation of *NPPA* and *NPPB*, which were significantly downregulated in *ISL1*$^{-/-}$ and *NKX2.5*$^{-/-}$ CPs (Figure 5F; Table S4). We assessed activation of a putative *NPPA/NPPB* regulatory region occupied by both TFs cloned upstream of a luciferase gene

(Figure S5I). In COS-7 cells, expression of either ISL1 or NKX2.5 activated the reporter, but the presence of both TFs had an additive effect, indicating that NKX2.5 is likely important for regulating a subset of CP-specific genes when co-bound with ISL1.

NKX2.5 is necessary and sufficient to localize ISL1 to a subset of genomic regions

We next tested the contribution of NKX2.5 to the genomic localization of ISL1 by performing ISL1 ChIP-seq in $NKX2.5^{-/-}$ CPs (Figure S6A). In the absence of NKX2.5, we observed 1,539 fewer ISL1-bound peaks ("lost" peaks), while 3,176 peaks remained ("maintained" peaks), and 2,319 additional peaks were detected ("gained" peaks) compared with the WT setting (Figures 6A and 6B). GREAT analysis of the gained peaks showed enrichment of cardiac terms such as *Cardiac chamber morphogenesis* and *Heart morphogenesis*, similar to the maintained peaks, suggesting that, in the absence of *NKX2.5*, ISL1 binds ectopically to alternative cardiac loci (Figure S6B). For example, there is increased ISL1 binding at the *TBX5* locus in the absence of NKX2.5, which might drive the aberrant increase of *TBX5* expression shown previously (Figures 5F and S6C). In addition to the cardiac terms, *Artery development* was also enriched, which could indicate ISL1 is also redistributed to other loci regulating genes involved in endothelial or smooth muscle cells, alternative fates of the CPs (Figure S6B). Lost loci were enriched for terms associated with cell polarity, such as *Apical constriction* and *Regulation of Maintenance of Cell Polarity* (Figure S6B).

We compared ISL1-bound peaks with GATA4-bound loci in day 6 CPs (Figure S6D) (Gonzalez-Teran et al., 2022). Co-bound loci were enriched in terms such as *Striated muscle cell differentiation* and *cardiomyocyte differentiation* (Figure S6E). Among the lost, gained, and maintained ISL1-bound loci in $NKX2.5^{-/-}$ CPs for GATA4 binding, a greater portion of the ISL1-bound gained and maintained loci were co-bound with GATA4 compared with the ISL1-bound lost loci (25.7% and 38.1%, compared with 9.75%, respectively) (Figure 6C). Thus, GATA4 may aid in maintaining ISL1 presence at cardiac loci, as well as contribute to redistribution of ISL1 upon loss of *NKX2.5*.

As a complementary approach, we compared the intensity of the ISL1 peaks identified in WT CPs with those found in $NKX2.5^{-/-}$ CPs (Figure 6D), identifying peaks that had increased intensity, decreased intensity, or no change. Among those with decreased signal, the NKX2.5 motif was proportionally more common than the ISL1 or GATA4/GATA6 motif (Figure S6F), while among the peaks with increased signal, the NKX2.5 motif was proportionally less common than the ISL1 or GATA4/GATA6 motifs. Peaks with increased signal were enriched for GO terms related to the nervous system such as the *Facial nerve struc-*

tural organization and *Regulation of axon guidance* (Figure S6G). Some genes within these terms are typical of specialized CMs within the sinus node, the electrical pacemaker of the heart. Interestingly, the sinus node myocardium is *ISL1* positive yet *NKX2.5* negative, and *ISL1* is critical for its development (Espinoza-Lewis et al., 2009; Liang et al., 2015). Another locus with increased signal was located within *NRP1* (Figure 6E), which is necessary for proper sinus node development (Maden et al., 2012). *NRP1* expression was increased in $NKX2.5^{-/-}$ CPs, consistent with increased ISL1 binding (Figure 5F) and supporting the conclusion that NKX2.5 is necessary for proper ISL1 localization to a large number of its binding sites in WT CPs.

Last, we tested if ectopic introduction of NKX2.5 in MNPs could override the effect of ISL1's native neuronal partners at some loci, redistributing ISL1 to new sites. We overexpressed either *EGFP* (control) or *NKX2.5* in day 15 MNPs and performed ChIP-seq on both ISL1 and NKX2.5 3 days later (Figures 6G, S6H, and S6I). We observed a large overlap in the NKX2.5-bound peaks with ISL1-bound peaks (75%) (Figure S6J). Indeed, despite a mild decrease in ISL1 expression (Figure S6I), ISL1 redistributed to more than 300 new sites typically occupied in CPs but not in MNPs (Figure S6K, p = 0.001), including the *MYH11* locus (Figure 6H). Overall, our data indicate that NKX2.5 is both necessary and sufficient for ISL1 genomic localization to a subset of cardiac-enriched loci, although it does not alone alter fate of the MNPs.

DISCUSSION

Here we used hiPSC models, scRNA-seq, ChIP-seq, and the deep learning model BPNet to determine how ISL1 executes cell-type-specific functions in distinct cell types through differentially interacting partners. We found that ISL1 functions in a complex with NKX2.5 to broadly co-occupy cardiac-specific genes, with binding modulated by GATA4 co-occupancy. NKX2.5 is believed to direct ISL1's localization to hundreds of cardiac-specific loci, and the introduction of NKX2.5 is sufficient to re-localize ISL1 even in neural progenitors. These analyses reveal how tissue-enriched TFs can cooperatively regulate the differentiation of specific cell types.

ISL1 interacts with NKX2.5 to drive cardiac regulatory networks

ISL1 forms transcriptional complexes with TFs LHX3 and LDB1 to facilitate MNP differentiation (Lee et al., 2012) and has been shown to interact with LDB1 in a mouse CP model to facilitate long-range looping (Caputo et al., 2015). However, no cardiac-specific ISL1 interactor had previously been identified.
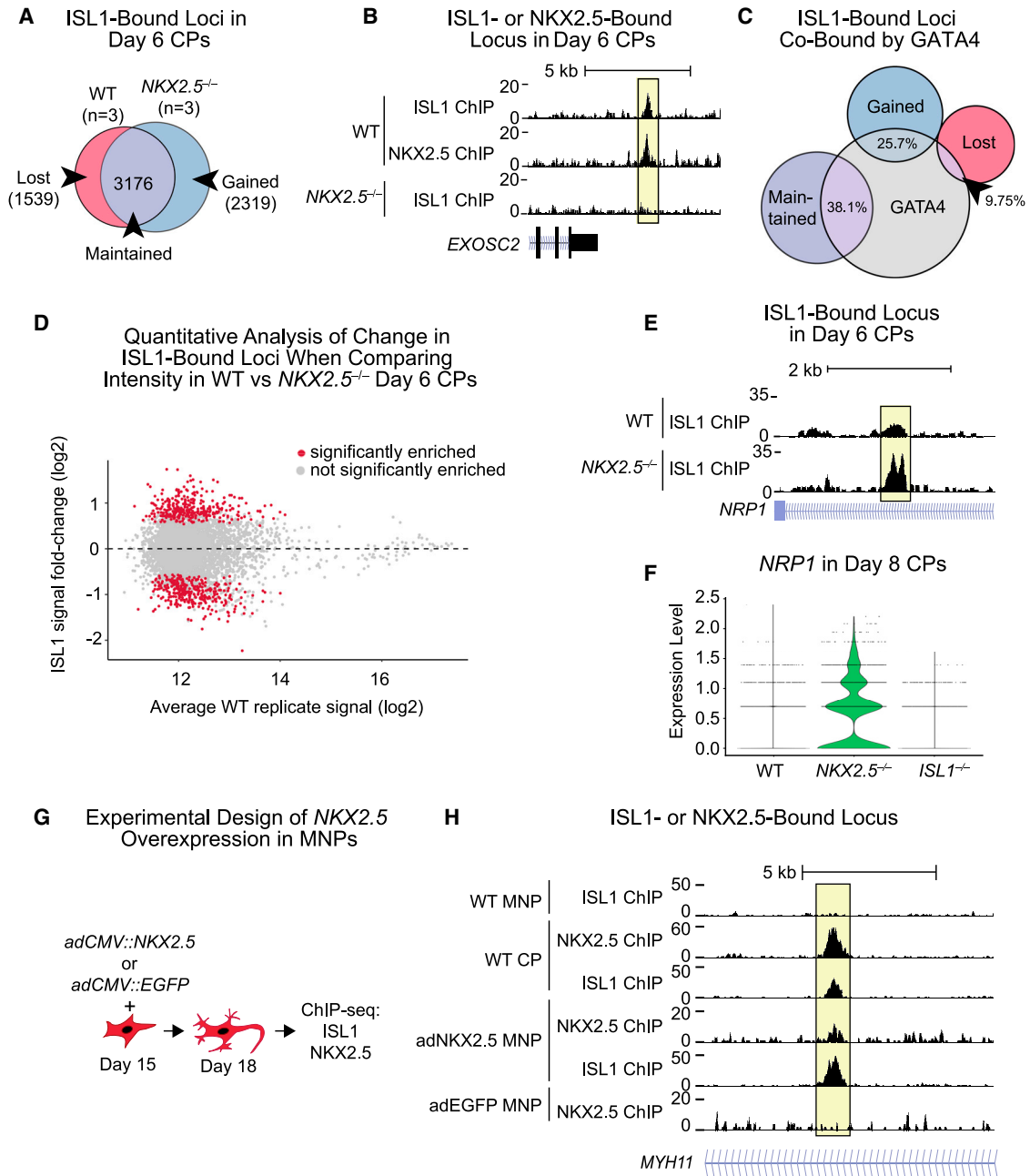
Figure 6. ISL1 depends on NKX2.5 for DNA localization at a subset of cardiac loci

(A) Venn diagram of ISL1-bound DNA regions that are lost, maintained or gained in $NKX2.5^{-/-}$ day 6 CPs.

(B) Example of an NKX2.5-dependent ISL1-bound peak. Data shown from a single representative replicate.

(C) Venn diagram of ISL1-bound loci from (A) compared with GATA4-bound loci.

(D) Quantitative analysis of change in ISL1 ChIP peak intensity in WT CPs compared with ISL1 binding in $NKX2.5^{-/-}$ day 6 CPs. Regions with statistically significant change using DESeq are in red ($p \leq 0.05$); points in gray are not significant. Fold change of ISL1 binding ($NKX2.5^{-/-}/NKX2.5^{+/+}$) is graphed across ISL1 ChIP signal intensity in WT CPs.

(E) ChIP-seq track of ISL1 detailing increased intensity of ISL1 binding in the absence of NKX2.5 at the $NRP1$ locus. Data shown from single representative replicate.

(F) Violin plot of $NRP1$ expression in WT, $NKX2.5^{-/-}$, and $ISL1^{-/-}$ day 8 CPs (n = 3 independent experiments for each).

(G) Schematic of $NKX2.5$ experimental overexpression in MNPs.

(H) ChIP-seq tracks of an ISL1 peak gained within the $MYH11$ locus when $NKX2.5$ was overexpressed in MNPs. Data shown from a single representative replicate.

To analyze ISL1-containing complexes predicted by BPNet analysis, we used SRM and demonstrated that ISL1 forms putative protein complexes with NKX2.5, GATA4, GATA6, and TEAD4. There is previous evidence of complexes that form between several of these factors, such as NKX2.5 and GATA4 (Durocher et al., 1997). In addition, mapping of several cardiac TFs in developing mice showed co-bound loci between GATA4, NKX2.5, and a TEAD protein, suggesting co-regulatory roles for these complexes in the heart (Akerberg et al., 2019). Our data shows that ISL1 is capable of forming proteomic complexes with several of these factors, and we specifically found that the ISL1-NKX2.5 complex co-binds to DNA at hundreds of genomic loci. GATA4 is also often co-localized with ISL1 and NKX2.5 at numerous loci and seems to affect ISL1 localization, particularly in the absence of NKX2.5. Moreover, exogenous expression of NKX2.5 in MNPs was sufficient to re-localize ISL1 to a subset of cardiac loci. Thus, NKX2.5 is necessary for proper ISL1 binding in CPs, although other factors present in cardiac cells likely also contribute.

Genes co-bound by ISL1 and NKX2.5 were largely involved in early cardiac chamber morphogenesis, suggesting the combination dictates early CM gene expression. Comparison of dysregulated genes in NKX2.5$^{-/-}$ CPs and ISL1$^{-/-}$ CPs suggests that early cardiac genes are co-regulated by these factors, and genes uniquely regulated by NKX2.5 are associated with later cardiogenic events. Thus, the cooperativity between ISL1 and NKX2.5 may shift the multipotent progenitor to a CM fate, and downregulation of *ISL1* allows progression to a more mature cell type guided by NKX2.5 with its other binding partners.

ISL1 displays lineage-specific regulation of a network of genes

Our data indicate that, in CPs or MNPs, loss of ISL1 results in cells predominantly adopting different cell fates along the same lineage, rather than alternative cell fates. Specifically, CPs lacking *ISL1* were more atrial like than ventricular, and MNP-directed differentiation led to more IN-like cells in the absence of *ISL1*. In addition, there was no significant upregulation of atrial factors upon deletion of *ISL1* in MNPs or significant upregulation of IN genes in the *ISL1*$^{-/-}$ CPs. This suggests there is either redundancy in the repression of such genes, or that an *ISL1*-independent mechanism represses those transcriptional programs.

In summary, we identified and characterized a cardiac-specific protein complex involving ISL1 that directs CM fate in human CP cells. These results demonstrate how a single TF can have divergent functions in different cell types based on its interaction with cell-type-specific factors, and help to broaden our understanding of how complex cell-specific cues direct TF function.

EXPERIMENTAL PROCEDURES

Resource availability

Corresponding author

Further information and requests for resources, reagents, and materials should be directed to and will be fulfilled by the lead contact, Deepak Srivastava (Deepak.srivastava@gladstone.ucsf.edu).

Materials availability

Requests for materials or resources should be directed to the lead contact, Deepak Srivastava.

Data and code availability

The scRNA-seq have been deposited at GEO:GSE195476. The ChIP-seq datasets have been deposited at GEO:GSE195476. The TSQ datasets have been deposited in the Panorama database at PMID: 29487113. Code for the scRNA-seq data analysis is available at https://github.com/bejmaven/Maven_at_al_2023_SrivastavaLab. Code for the BPNet analyses is available at https://github.com/zeitlingerlab/Maven_ISL1_2022. Additional information required to reanalyze the data reported here is available from the lead contact.

Experimental model and subject details

iPSC culture and differentiation

hiPSCs-CPs and hiPSC-MNPs were generated using previously published methods (Lian et al., 2013; NeuroLINCS Consortium et al., 2021). Full culture conditions and medium formulations can be found in the Supplemental information.

Generation of CRISPR-Cas9 genome-edited hiPSC lines

WTC11 iPSCs were targeted using CRISPR-Cas9 ribonucleic proteins (RNPs), as described (Hultquist et al., 2016; Kim et al., 2014) with minor modifications. Single guide RNAs (sgRNAs) were generated by Dharmacon and Cas9 protein was obtained from QB3 MacroLab (Cas9-nls), and complexed *in vitro* to form RNPs. Briefly, we electroporated the RNP/sgRNA using the Lonza Nucleofector X-unit 4D-Nucleofector with the Lonza P3 Primary Cell 96-well Nucleofector. Each selected colony was genotyped to confirm homozygosity. To generate the *NKX2.5*$^{-/-}$ hiPSC line, we used a single sgRNA targeting strategy while the *ISL1*$^{-/-}$ hiPSC line used two sgRNAs to excise a portion of the ISL1 gene. sgRNAs and oligonucleotides used for genotyping are listed in Table S5.

Adenovirus

Adenoviral human type 5 (dE1/E3) viral particles expressing Ad-NKX2-5 or negative control ad-EF1a-eGFP.

Method details

Immunostaining

Cells were fixed methanol-free diluted paraformaldehyde with water 1:4 (w/v) at room temperature (RT) for 15 min with gentle shaking. Fixed cells were washed twice with Dulbecco's PBS (DPBS) then exposed to a blocking solution consisting of 5% normal donkey serum and 0.2% Triton X-100 in DPBS for 1 h at RT. Cells were incubated with primary antibody (see the supplemental experimental procedures) at 4°C overnight, washed three times in DPBS, followed by incubation with secondary antibody at RT for 1 h, and washed three times using DPBS containing NucBlue Live ReadyProbes dye. Imaging was performed via a Zeiss

S1 Confocal Microscope and image processing done using ImageJ (Schneider et al., 2012).

Whole cell lysate extraction

CP or MNP cell pellets were thawed on ice then resuspended in cold RIPA buffer supplemented with protease and phosphatase inhibitors. Samples were incubated on ice for 10 min, followed by 10 min of gentle rocking at 4°C, then centrifuged at 14,000×g for 20 min at 4°C to isolate whole cell lysate (supernatant).

Western blotting

Western blotting was performed with Bis-Tris Mini Protein Gels following the manufacturer's protocols and subsequently transferred to nitrocellulose membranes and blocked. Primary antibodies were added and incubated overnight at 4°C. Next, the membrane was washed three times with PBS-T and incubated with appropriate secondary antibodies. Protein detection as performed using the Odyssey Imager or Western Blotting Detection Reagent followed by exposure to film.

RNA extraction, RT-qPCR reaction, and analysis

Cell pellets were lysed in Trizol and RNA extracted with the RNeasy kit and treated with DNase according to manufacturer's instructions. We converted 1 μg RNA to cDNA with SuperScript III First-Strand Synthesis System. Wes used 1 μL of a 1:50 w/v dilution per qRT-PCR reaction with TaqMan Universal PCR Master Mix. All reactions were run on the 7900HT Fast Real-Time system (Applied Biosystem). TaqMan probes are listed in the Supplemental information. Data shown are averages of at least three biological replicates and three technical replicates, normalized to *GAPDH* RNA levels.

Single cell RNA-seq

Day 8 CPs or day 18 MNPs were used to generate single cell libraries with the Chromium Single Cell 3′ Library and Gel Bead Kit v3 and Chromium i7 Multiplex Kit. Libraries were sequenced on the Illumina NovaSeq 6000 System or the Illumina NextSeq500 based on the Chromium Single Cell v3 specifications. See the supplemental experimental procedures.

ChIP-seq

Cells were collected from tissue culture plates, cross-linked with paraformaldehyde followed by glycine quenching. For immunoprecipitation, the pellets were thawed and lysed, and the resulting nuclei pellet was sheared. Following incubation with appropriate antibodies, protein complexes were immunoprecipitated with dynabeads and bound proteins eluted. The samples were then processed for sequencing with NEBNext Ultra II DNA Library Kit for Illumina. After sequencing on the Illumina NextSeq500 or the Illumina HiSeq 4000 (UCSF Center for Advanced Technology), samples were analyzed for mapping, filtering, peak calling, and visualization. See the supplemental experimental procedures.

BPNet model training and motif identification

We trained a convolutional neural network using the BPNet code, architecture and loss features consistent with the published approach (Avsec et al., 2021) to explain WT ISL1 ChIP-seq data. See the supplemental experimental procedures.

Analysis of protein complexes

Cells were resuspended in IP buffer, sheared, and then incubated with primary antibodies. Protein complexes were immunoprecipitated using antibody-conjugated beads, followed by washing. Elu-

ates were prepared for western blotting and MS. See the supplemental experimental procedures.

Cloning of overexpression constructs

We used a *CMV::3x-FLAG* construct (Jäger et al., 2011). Cloning was performed to insert *EGFP* or the human *NKX2.5* downstream of *CMV*. For *CMV::hLDB1-HA*, the coding sequence of human *LDB1* with a downstream human influenza hemagglutinin ("HA") tag was assembled into the *pcDNA4/TO* Vector. For *CMV::hISL1-Bio*, we cloned in a TEV biotinylation site ("Bio") and then inserted the human coding sequence of *ISL1*. Sanger sequencing confirmed all constructs.

Mapped motif pair synergy

To test synergy effects between the TF-MoDISco mapped motifs, we adopted the *in silico* genomic motif interaction approach described (Avsec et al., 2021). See the supplemental experimental procedures.

Transfection of siRNA

We attached a fluorescent probe to each siRNA with the Label IT siRNA Tracker Reagent Intracellular Localization Kit, CX-Rhodamine, and transfected with Lipofectamine RNAiMAX. To begin, we replated D5 CPs 1:1 into 12-well plates pre-coated with fibronectin (diluted 1:80 with PBS) and added siRNA-lipofectamine complexes into each well. Media were refreshed 24 h later and cells analyzed by flow cytometry 72 h after transfection.

Fluorescence-activated cell sorting

CPs were harvested from plates using accutase, then washed twice with PBS. Cells were strained through a 40-μM mesh (BD Falcon) and sorted on the BD Aria II 5 flow cytometer. Populations with shifted PE Texas Red-A signal, corresponding to Rhodamine, were collected and processed for downstream analysis.

Luciferase assay

The putative *NPPA/NPPB* regulatory region was synthesized and inserted into the pNL1.1[Nluc] plasmid and co-transfected together with *renilla luciferase* control pGL4.73[hRluc/SV40], *CMV::hISL1-Bio* and/or *CMV::hNKX2.5-3xFLAG* into COS-7 cells using Fugene HD. Luminescence activity was quantified with Dual-Glo Luciferase Assay System. Fold-change shown are averages of at least three biological replicates and four technical replicates, compared with the empty vector controls and then normalized to the *renilla luciferase* control.

Measuring enrichment of ISL1 peaks

DESeq2 (fitType = local) was used to compare enrichment differences of WT and *NKX2.5*$^{-/-}$ ISL ChIP-seq signal across peaks from WT ISL1 ChIP-seq in day 6 CPs. Total raw count pileup of WT and *NKX2.5*$^{-/-}$ ISL1 ChIP-seq in day 6 CPs were collected across every WT ISL1 peak in day 6 CPs across a 1000 bp centered window. Using both an adjusted p value cutoff of p ≤ 0.05 and the reported log2 fold-change enrichment of WT versus *NKX2.5*$^{-/-}$ ISL1 ChIP-seq signal intensity, peaks were classified as having significantly increased, significantly decreased, or not changed in enrichment upon removal of NKX2.5.

Mapped motif enrichment

Motif enrichment was tested using Bonferroni-corrected chi-square tests across classified ISL1 ChIP-seq peaks (classification described above) measured across WT and *NKX.5*$^{-/-}$ CPs.

Adenovirus infection of neural cells

Media were changed on day 15 MNPs to stage-appropriate media, and cells were infected drop-wise at an multiplicity of infection (MOI) of 100 viral units per cell. Media was removed and replaced with stage-appropriate media 48 h later. Infection was confirmed by detection of EGFP by fluorescence microscopy 72 h post infection.

Quantification and statistical analysis

Details pertaining to statistical significance and value of n are reported in the methods details, figures and the accompanying figure legends. The level of significance in all figures is represented as follows: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.stemcr.2023.09.014.

AUTHOR CONTRIBUTIONS

DECLARATION OF INTERESTS

REFERENCES

Akerberg, B.N., Gu, F., VanDusen, N.J., Zhang, X., Dong, R., Li, K., Zhang, B., Zhou, B., Sethi, I., Ma, Q., et al. (2019). A reference map of murine cardiac transcription factor chromatin occupancy identifies dynamic and conserved enhancers. Nat. Commun. *10*, 4907.

Anderson, D.J., Kaplan, D.I., Bell, K.M., Koutsis, K., Haynes, J.M., Mills, R.J., Phelan, D.G., Qian, E.L., Leitoguinho, A.R., Arasaratnam, D., et al. (2018). NKX2-5 regulates human cardiomyogenesis via a HEY2 dependent transcriptional network. Nat. Commun. *9*, 1373. https://doi.org/10.1038/s41467-018-03714-x.

Avsec, Ž., Weilert, M., Shrikumar, A., Krueger, S., Alexandari, A., Dalal, K., Fropf, R., McAnany, C., Gagneur, J., Kundaje, A., and Zeitlinger, J. (2021). Base-resolution models of transcription-factor binding reveal soft motif syntax. Nat. Genet. *53*, 354–366.

Black, B.L. (2007). Transcriptional pathways in second heart field development. Semin. Cell Dev. Biol. *18*, 67–76.

Cai, C.-L., Liang, X., Shi, Y., Chu, P.-H., Pfaff, S.L., Chen, J., and Evans, S. (2003). Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. Dev. Cell *5*, 877–889.

Calvo, E., Camafeita, E., Fernández-Gutiérrez, B., and López, J.A. (2011). Applying selected reaction monitoring to targeted proteomics. Expert Rev. Proteomics *8*, 165–173.

Caputo, L., Witzel, H.R., Kolovos, P., Cheedipudi, S., Looso, M., Mylona, A., van IJcken, W.F.J., Laugwitz, K.-L., Evans, S.M., Braun, T., et al. (2015). The Isl1/Ldb1 Complex Orchestrates Genomewide Chromatin Organization to Instruct Differentiation of Multipotent Cardiac Progenitors. Cell Stem Cell *17*, 287–299.

Dorn, T., Goedel, A., Lam, J.T., Haas, J., Tian, Q., Herrmann, F., Bundschu, K., Dobreva, G., Schiemann, M., Dirschinger, R., et al. (2015). Direct nkx2-5 transcriptional repression of isl1 controls cardiomyocyte subtype identity. Stem Cell. *33*, 1113–1129.

Durocher, D., Charron, F., Warren, R., Schwartz, R.J., and Nemer, M. (1997). The cardiac transcription factors Nkx2-5 and GATA-4 are mutual cofactors. EMBO J. *16*, 5687–5696.

Ediger, B.N., Du, A., Liu, J., Hunter, C.S., Walp, E.R., Schug, J., Kaestner, K.H., Stein, R., Stoffers, D.A., and May, C.L. (2014). Islet-1 Is essential for pancreatic β-cell function. Diabetes *63*, 4206–4217.

Espinoza-Lewis, R.A., Yu, L., He, F., Liu, H., Tang, R., Shi, J., Sun, X., Martin, J.F., Wang, D., Yang, J., and Chen, Y. (2009). Shox2 is essential for the differentiation of cardiac pacemaker cells by repressing Nkx2-5. Dev. Biol. *327*, 376–385.

Galloway, J.R., Bethea, M., Liu, Y., Underwood, R., Mobley, J.A., and Hunter, C.S. (2015). SSBP3 Interacts With Islet-1 and Ldb1 to Impact Pancreatic β-Cell Target Genes. Mol. Endocrinol. *29*, 1774–1786.

Gonzalez-Teran, B., Pittman, M., Felix, F., Thomas, R., Richmond-Buccola, D., Hüttenhain, R., Choudhary, K., Moroni, E., Costa, M.W., Huang, Y., et al. (2022). Transcription factor protein interactomes reveal genetic determinants in heart disease. Cell *185*, 794–814.e30.

Hultquist, J.F., Schumann, K., Woo, J.M., Manganaro, L., McGregor, M.J., Doudna, J., Simon, V., Krogan, N.J., and Marson, A. (2016). A Cas9 ribonucleoprotein platform for functional genetic studies of HIV-host interactions in primary human T cells. Cell Rep. *17*, 1438–1452.

Jäger, S., Cimermancic, P., Gulbahce, N., Johnson, J.R., McGovern, K.E., Clarke, S.C., Shales, M., Mercenne, G., Pache, L., Li, K., et al. (2011). Global landscape of HIV-human protein complexes. Nature *481*, 365–370.

Jarrar, W., Dias, J.M., Ericson, J., Arnold, H.-H., and Holz, A. (2015). Nkx2.2 and Nkx2.9 are the key regulators to determine cell fate of branchial and visceral motor neurons in caudal hindbrain. PLoS One *10*, e0124408.

Kim, K.-T., Kim, N., Kim, H.-K., Lee, H., Gruner, H.N., Gergics, P., Park, C., Mastick, G.S., Park, H.-C., and Song, M.-R. (2016). ISL1-based LIM complexes control Slit2 transcription in developing cranial motor neurons. Sci. Rep. *6*, 36491.

Kim, S., Kim, D., Cho, S.W., Kim, J., and Kim, J.-S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res. *24*, 1012–1019.

Lee, S., Cuvillier, J.M., Lee, B., Shen, R., Lee, J.W., and Lee, S.-K. (2012). Fusion protein Isl1-Lhx3 specifies motor neuron fate by inducing motor neuron genes and concomitantly suppressing the interneuron programs. Proc. Natl. Acad. Sci. USA *109*, 3383–3388.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Lian, X., Zhang, J., Azarin, S.M., Zhu, K., Hazeltine, L.B., Bao, X., Hsiao, C., Kamp, T.J., and Palecek, S.P. (2013). Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/β-catenin signaling under fully defined conditions. Nat. Protoc. *8*, 162–175.

Liang, X., Zhang, Q., Cattaneo, P., Zhuang, S., Gong, X., Spann, N.J., Jiang, C., Cao, X., Zhao, X., Zhang, X., et al. (2015). Transcription factor ISL1 is essential for pacemaker development and function. J. Clin. Invest. *125*, 3256–3268.

Lyons, I., Parsons, L.M., Hartley, L., Li, R., Andrews, J.E., Robb, L., and Harvey, R.P. (1995). Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. Genes Dev. *9*, 1654–1666.

Maden, C.H., Gomes, J., Schwarz, Q., Davidson, K., Tinker, A., and Ruhrberg, C. (2012). NRP1 and NRP2 cooperate to regulate gan-gliogenesis, axon guidance and target innervation in the sympathetic nervous system. Dev. Biol. *369*, 277–285.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Narkis, G., Tzchori, I., Cohen, T., Holtz, A., Wier, E., and Westphal, H. (2012). Isl1 and Ldb co-regulators of transcription are essential early determinants of mouse limb development. Dev. Dyn. *241*, 787–791.

NeuroLINCS Consortium, Li, J., Lim, R.G., Kaye, J.A., Dardov, V., Coyne, A.N., Wu, J., Milani, P., Cheng, A., Thompson, T.G., et al. (2021). An integrated multi-omic analysis of iPSC-derived motor neurons from C9ORF72 ALS patients. iScience *24*, 103221.

Osoegawa, K., Schultz, K., Yun, K., Mohammed, N., Shaw, G.M., and Lammer, E.J. (2014). Haploinsufficiency ofinsulin gene enhancer protein 1(ISL1) is associated with d-transposition of the great arteries. Mol. Genet. Genomic Med. *2*, 341–351. https://doi.org/10.1002/mgg3.75.

Paul, M.H., Harvey, R.P., Wegner, M., and Sock, E. (2014). Cardiac outflow tract development relies on the complex function of Sox4 and Sox11 in multiple cell types. Cell. Mol. Life Sci. *71*, 2931–2945.

Pfaff, S.L., Mendelsohn, M., Stewart, C.L., Edlund, T., and Jessell, T.M. (1996). 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.

Quaranta, R., Fell, J., Rühle, F., Rao, J., Piccini, I., Araúzo-Bravo, M.J., Verkerk, A.O., Stoll, M., and Greber, B. (2018). Revised roles of ISL1 in a hES cell-based model of human heart chamber specification. Elife *7*, e31706. https://doi.org/10.7554/elife.31706.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Seo, S.Y., Lee, B., and Lee, S. (2015). Critical roles of the LIM domains of Lhx3 in recruiting coactivators to the motor neuron-specifying Isl1-Lhx3 complex. Mol. Cell Biol. *35*, 3579–3589.

Son, E.Y., Ichida, J.K., Wainger, B.J., Toma, J.S., Rafuse, V.F., Woolf, C.J., and Eggan, K. (2011). Conversion of mouse and human fibroblasts into functional spinal motor neurons. Cell Stem Cell *9*, 205–218.

Song, M.-R., Sun, Y., Bryson, A., Gill, G.N., Evans, S.M., and Pfaff, S.L. (2009). Islet-to-LMO stoichiometries control the function of transcription complexes that specify motor neuron and V2a interneuron identity. Development *136*, 2923–2932.

Srivastava, D. (2006). Making or breaking the heart: from lineage determination to morphogenesis. Cell *126*, 1037–1048.

Velasco, S., Ibrahim, M.M., Kakumanu, A., Garipler, G., Aydin, B., Al-Sayegh, M.A., Hirsekorn, A., Abdul-Rahman, F., Satija, R., Ohler, U., et al. (2017). A multi-step transcriptional and chromatin state cascade underlies motor neuron programming from embryonic stem cells. Cell Stem Cell *20*, 205–217.e8.

Wang, Y., Li, Y., Guo, C., Lu, Q., Wang, W., Jia, Z., Chen, P., Ma, K., Reinberg, D., and Zhou, C. (2016). ISL1 and JMJD3 synergistically

control cardiac differentiation of embryonic stem cells. Nucleic Acids Res. *44*, 6741–6755.

Wang, Z., Song, H.-M., Wang, F., Zhao, C.-M., Huang, R.-T., Xue, S., Li, R.-G., Qiu, X.-B., Xu, Y.-J., Liu, X.-Y., and Yang, Y.Q. (2019). A New ISL1 Loss-of-Function Mutation Predisposes to Congenital Double Outlet Right Ventricle. Int. Heart J. *60*, 1113–1122.

Yasuhara, J., and Garg, V. (2021). Genetics of congenital heart disease: a narrative review of recent advances and clinical implications. Transl. Pediatr. *10*, 2366–2386.

Zhang, L., Nomura-Kitabayashi, A., Sultana, N., Cai, W., Cai, X., Moon, A.M., and Cai, C.-L. (2014). Mesodermal Nkx2.5 is necessary and sufficient for early second heart field development. Dev. Biol. *390*, 68–79.

Supplemental Information

The multi-lineage transcription factor ISL1 controls cardiomyocyte cell fate through interaction with NKX2.5

Bonnie E.J. Maven, Casey A. Gifford, Melanie Weilert, Barbara Gonzalez-Teran, Ruth Hüttenhain, Angelo Pelonero, Kathryn N. Ivey, Kaitlen Samse-Knapp, Wesley Kwong, David Gordon, Michael McGregor, Tomohiro Nishino, Eyuche Okorie, Sage Rossman, Mauro W. Costa, Nevan J. Krogan, Julia Zeitlinger, and Deepak Srivastava

SUPPLEMENTARY FIGURE TITLES AND LEGENDS

A Cardiac Factors in Day 6 CPs
MYH6 Hoechst · TNNT2 Hoechst
WT
ISL1⁻/⁻

B *ISL1⁻/⁻* Gene Editing Strategy
gRNA1 gRNA2
ISL1⁺/⁺
1 2 3 4 5 6
qPCR probe
ISL1⁻/⁻
1 2/5 6

C *ISL1* in Day 6 CPs
Relative mRNA Levels
WT *ISL1⁻/⁻*
**

D ISL1 in Day 6 CPs
WT *ISL1⁻/⁻*
αISL1
αVinculin

E Beating Induction Day in CM Differentiation
Day
WT *ISL1⁻/⁻*
****

F Cluster-Specific Marker Analyses in Day 8 CPs
Early · Transition · Late · Mixed

GO Biologic Term	p-Value	Genes
regulation of nuclear division	1.85E-23	NUF2, MAD2L1
kinetochore organization	5.50E-09	CENPF, SMC4
mitotic spindle elongation	1.13E-06	PRC1, KIF4A
reg. of ubiquitin protein ligase	2.13E-03	RPL5, RPL11
cardiac ventricle development	1.27E-03	FOXC1, SOX4
ventricular septum morph.	1.63E-02	HEY1, WNT5A
sarcomere organization	3.96E-12	TNNT2, ACTN2
reg. of heart contraction	5.37E-10	CSRP3, MYH6
muscle filament sliding	8.82E-05	MYL6, TPM1
nuclear DNA replication	9.20E-09	CDT1, MCM4
cell cycle DNA replication	1.21E-08	FEN1, GMNN
DNA replication initiation	8.36E-05	ORC6, CDC6

G Replicate Distribution in Day 8 CPs
WT (n=3) *ISL1⁻/⁻* (n=3)
Rep 1, Rep 2, Rep 3
UMAP 2 / UMAP 1

H *NR2F1* in Day 6 CPs
Relative mRNA Levels
WT *ISL1⁻/⁻*
**

I Experimental Design of *ISL1* Knockdown in CPs
ISL1 or Scra Rhod-siRNA
Day 5 + Day 8
FACS: Rhod⁺
FSC-A / Rhodamine signal
Rhod⁺

J ISL1 in Day 8 CPs
siRNA
WT *ISL1⁻/⁻* Scra *ISL1*
αISL1
αVinculin

K scRNA-seq Clustering of siRNA-Treated Day 8 CPs
918 CPs
Scra (n = 1) *ISL1* (n = 1)
UMAP 2 / UMAP 1

L *MYL9* in Day 8 CPs
Expression Level
Scra *ISL1*
***

M *TNNT2* in Day 8 CPs
Expression Level
Scra *ISL1*
*

Figure S1. Additional analyses of the absence of *ISL1* in CPs, related to Figure 1.

(A) Immunofluorescence of cardiac factors MYH6 and TNNT2 in wild type (WT) and *ISL1*$^{-/-}$ day 6 cardiac progenitors (CPs). Scale bar, 100 µM. (B) Schematic of *ISL1* knockout using two guide RNAs (gRNAs) to excise large portion in each locus. Numbers denote exons. Quantitative PCR (qPCR) probe as related to (C) noted. (C) qPCR confirmation of lack of *ISL1* RNA in day 6 CPs. Data represented as means ± 1.96SD from independent experiments (n = 4 biological replicates, ** p-value ≤ 0.01). (D) Western blot of ISL1 in WT and *ISL1*$^{-/-}$ day 6 CPs. Vinculin served as loading control. (E) Timing of beating induction in WT and *ISL1*$^{-/-}$ cardiomyocytes (CMs). Data represented as means ± 1.96SD from independent experiments (n = 5 biological replicates, **** p-value ≤ 0.0001). Also see **Supplemental Video S1**. (F) Heatmap of genes enriched with accompanying GO terms in each of the 4 clusters depicted in **Figure 1C**. (G) UMAP of scRNA-seq replicate ("Rep") comparison in WT (n = 3 biological replicates, 3636 cells) or *ISL1*$^{-/-}$ (n = 3 biological replicates, 8190 cells) day 8 CPs. (H) qPCR of *NR2F1* in *ISL1*$^{-/-}$ day 6 CPs. Data represented as means ± 1.96SD from independent experiments (n = 3 biological replicates, ** p-value ≤ 0.01). (I) Schematic of *ISL1* siRNA knockdown and sorting strategy in CPs. (J) Western blot of ISL1 in whole cell lysates of CPs when transfected with *ISL1* siRNA or scrambled ("Scra") control siRNA. Vinculin served as loading control. (K) Hierarchical clustering analysis of CPs transfected with either Scrambled ("Scra") control or *ISL1* siRNA, showed alongside a FeaturePlot display indicating distribution of cells treated with Scra (n = 1 biological replicate, 600 cells) or *ISL1* (n = 1 biological replicate, 318 cells) siRNA. (L) ViolinPlots of *MYL9* expression in control or *ISL1* siRNA-treated day 8 CPs (p = 1.18E-05). (M) ViolinPlots of *TNNT2* expression in control or *ISL1* siRNA-treated day 8 CPs (p = 0.01).

A Replicate Comparison of ISL1 ChIP-seq in
WT Day 6 CPs and Day 18 MNPs

B Genomic Annotation and Distribution
in ISL1-Bound Loci

ISL1-Bound Loci in CPs

ISL1-Bound Loci in MNPs

ISL1-Bound Loci in Both

Percentage (%)

Promoter (<=1kb)
Promoter (1-2kb)
Promoter (2-3kb)
5' UTR
3' UTR
1st Exon
Other Exon
1st Intron
Other Intron
Downstream (<=300)
Distal Intergenic

C Distribution of Histone Marks
in ISL1-Bound Loci in Day 6 CPs

ISL1
H3K27Ac
H3K36Me3

$\log_2$ normalized read density

-2.0 kb   center   2.0 kb

D Control for ISL1-Bound Loci in Day 6 CPs or Day 18 MNPs

$ISL1^{-/-}$

CPs   MNPs

MYL4   PHOX2B   SETD5

E GREAT Analysis of
ISL1-Bound Loci in Day 6 CPs

-log10 (Binomial p-value)

0 5 10 15 20 25

Cardiac chamber dev.
Cadiac chamber morph.
Cardiac septum dev.
Cardiac muscle tissue dev.
Cardiac ventricle dev.
Cardiac septum morph.
Outflow tract septum morph.
Outflow tract morph.
Cardiomyocyte diff.
Reg. of cardiac muscle tissue dev.

F GREAT Analysis of
ISL1-Bound Loci in Day 18 MNPs

-log10 (Binomial p-value)

0 4 8 12 16 20

Cranial nerve dev.
Cell diff. in spinal cord
MN axon guidance
Sympathetic ganglion dev.
Sympathetic nervous system dev.
Parasympathetic nervous system dev.
Ganglion dev.
Retinal ganglion cell axon guidance
Noradrenergic neuron diff.
Ventral interneuron fate commitment

Figure S2. ChIP-seq analyses of ISL1 function in CPs and MNPs, related to Figure 3.

(A) Replicate comparison of ISL1 ChIP-seq in day 6 CPs (n = 3 biological replicates) and day 18

MNPs (n = 3 biological replicates) using DiffBind. (B) Distribution of ISL1-bound loci at annotated

genomic features in day 6 CPs, day 18 MNPs and ISL1-bound loci in both. (C) ISL1-bound peaks

(n = 3 biological replicates) in day 6 CPs overlayed with binding of histone marks (H3K27Ac, n =

1 biological replicate; H3K36Me3, n = 2 biological replicates) from cardiac progenitor data sets.

Data shown from merged replicates as applicable (see Supplemental Experimental Procedures).

(D) Tracks of ISL1 ChIP in *ISL1*$^{-/-}$ day 6 CPs or day 18 MNPs displaying antibody specificity, as

related to **Figure 3C**. Data shown from single representative replicate. (E) GREAT GO Biological

Process terms for ISL1-bound peaks in day 6 CPs. (F) GREAT GO Biological Process terms for

ISL1-bound peaks in day 18 MNPs.

A Enrichment of Motifs in ISL1-Bound Loci in Day 6 CPs and Day 18 MNPs

B Hyperparameter Tuning Results

C NKX2.5 and GATA Motif Distribution in ISL1 Peaks in Day 6 CPs

D GREAT Analyses of Motif-Specific ISL1-Bound Loci in Day 6 CPs
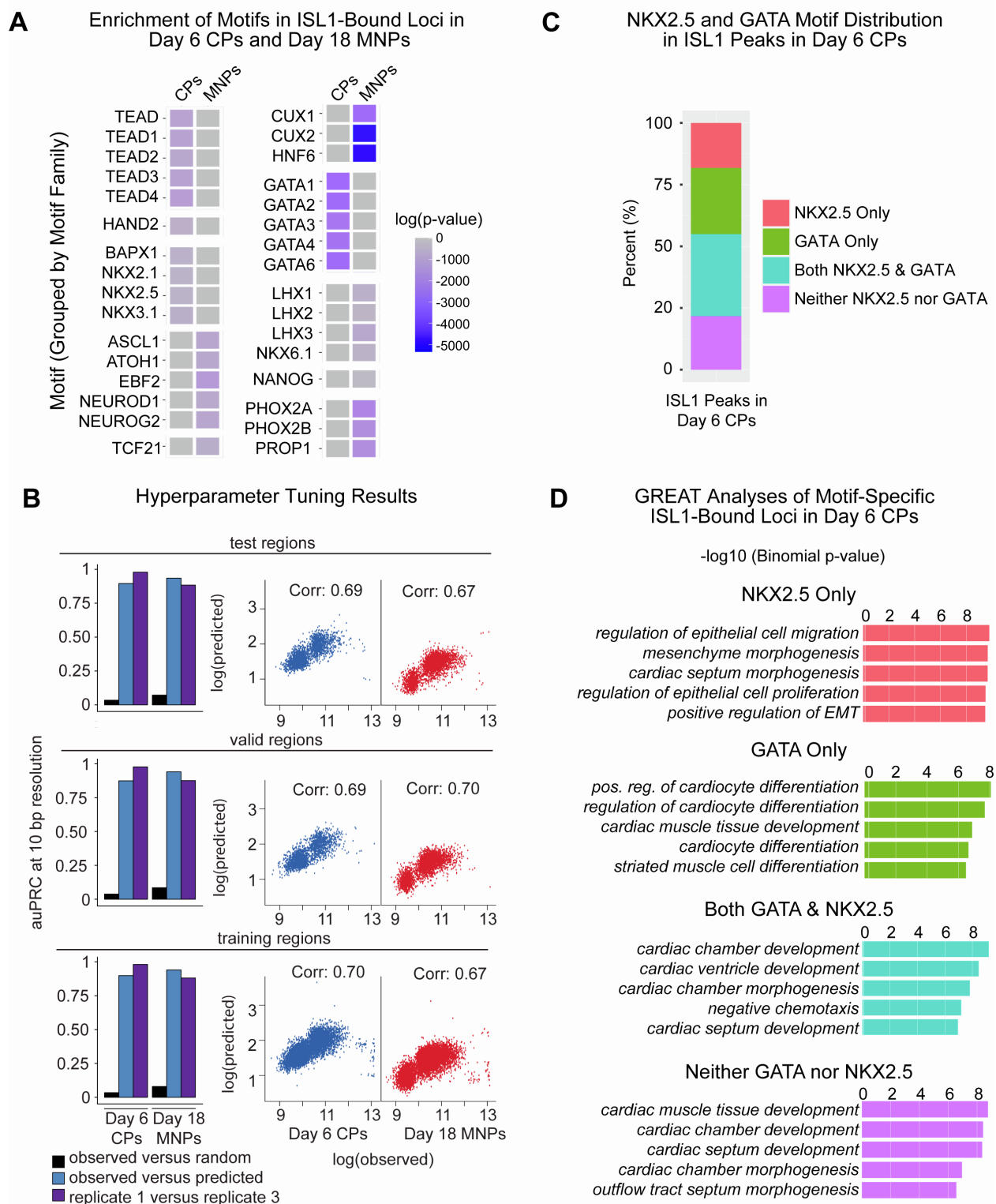
Figure S3. ISL1 ChIP-seq motif analyses in CPs and MNPs, related to Figure 3.

(A) Enrichment analysis of motifs found in ISL1-bound loci in day 6 CPs and day 18 MNPs,

grouped by motif families. (B) Hyperparameter tuning results, related to **Figure 3D.** (C) Frequency of ISL1 peaks in day 6 CPs containing only the NKX2.5 motif, only the GATA motif, both motifs, or neither motif. (D) GREAT GO Biological Process terms of ISL1 peaks in day 6 CPs containing only the NKX2.5 motif, only the GATA motif, both motifs, or neither motif.

A   FLAG IP in COS-7 Cells

ISL1-Bio	+	−	+	−	+	−	+
LDB1-3xFLAG	−	+	+	−	−	−	−
GATA4-3xFLAG	−	−	−	+	+	−	−
NKX2.5-3xFLAG	−	−	−	−	−	+	+

Input — αFLAG / αISL1

αVinculin

FLAG IP Eluate — αFLAG / αISL1

B   ISL1 IP Eluate in Day 6 WT CPs

ISL1 IP   IgG IP

αISL1

αNKX2.5

C   ISL1 IP Input and Output in Day 6 WT CPs

Input   IgG output   ISL1 IP output

αISL1

αH3K36Me3

D   Replicate Comparison of NKX2.5 ChIP-seq in WT Day 6 CPs

R1, R3, R2

Histogram
Count 0 to 3
Correlation 0 to 1

E   ISL1-Bound Loci in MNPs vs NKX2.5-Bound Loci in Day 6 CPs

ISL1 in MNPs   NKX2.5 in CPs

10284   32154

483

F   ISL1- or NKX2.5-Bound Loci in Day 6 CPs Binned by Orientation and Distance to Nearest Transcription Start Site (TSS)

ISL1-bound only — Region-gene associations 0%–40%, TSS

NKX2.5-bound only — 0%–30%, TSS

Co-bound — 0%–40%, TSS

x-axis bins: <-500, -500 to -50, -50 to -5, -5 to 0, 0 to 5, 5 to 50, 50 to 500, >500

Distance to TSS (kb)

G   Genomic Annotation and Distribution in ISL1- and NKX2.5-Bound Loci

ISL1-Bound Loci Only — Percentage (%) 0 25 50 75 100

NKX2.5-Bound Loci Only — Percentage (%) 0 25 50 75 100

Co-Bound Loci — Percentage (%) 0 25 50 75 100

Legend:
- Promoter (<=1kb)
- Promoter (1-2kb)
- Promoter (2-3kb)
- 5' UTR
- 3' UTR
- 1st Exon
- Other Exon
- 1st Intron
- Other Intron
- Downstream (<=300)
- Distal Intergenic

H   GREAT Analysis of ISL1- and NKX2.5-Bound Loci in Day 6 CPs

-log10 (Binomial p-value)

ISL1-bound only peaks (0 2 4 6 8)
- Reg. of cell polarity
- Reg. of maint. of cell polarity
- Cardiac ventricle morph.
- Transc. regulatory DNA binding
- Glomerulus dev.

NKX2.5-bound only peaks (0 10 20 30)
- Nuc-transcribed mRNA proc.
- Cotransl. memb. targeting
- SRP-dep memb. targeting
- Protein targeting to ER
- Cytoplasmic translation

ISL1-NKX2.5 co-bound peaks (0 10 20)
- Cardiac chamber dev.
- Heart morph.
- Cardiac chamber morph.
- Cardiac septum dev.
- Cardiac septum morph.

Figure S4. Characteristics of the ISL1-interacting partner NKX2.5, related to Figures 4 and 5.

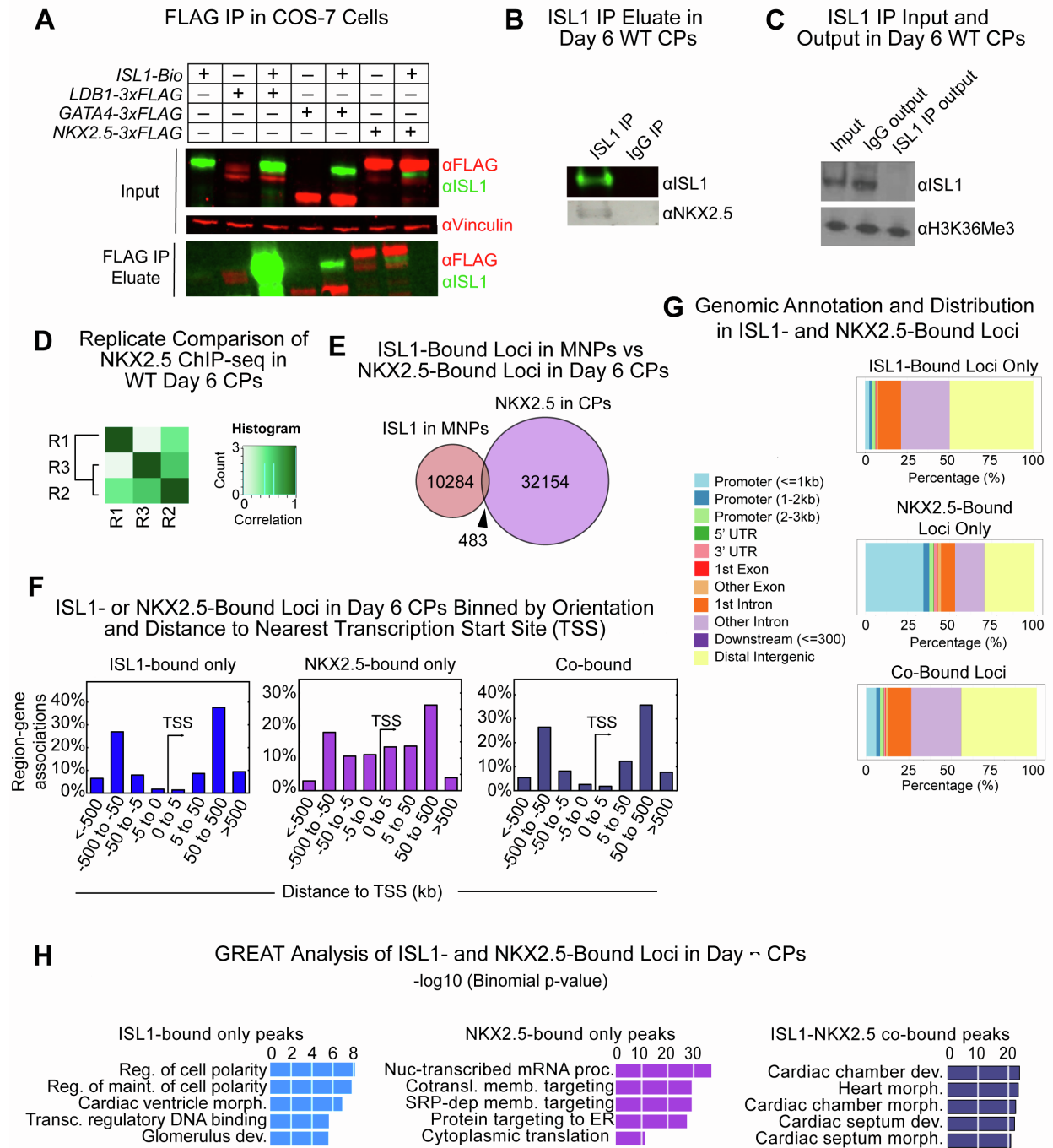(A) Western blot with indicated antibodies after IP with ISL1 antibodies in COS-7 cells expressing

indicated constructs. (B) Western blot of ISL1 or NKX2.5 after immunoprecipitation (IP) with ISL1 antibodies in day 6 CPs. IgG IP shown as control. (C) Input and output western blot for co-IP related to western blot of ISL1 IP in (B). (D) Replicate comparison of NKX2.5 ChIP-seq (n = 3 biological replicates) in day 6 CPs using DiffBind. (E) NKX2.5-bound peaks in day 6 CPs in comparison to ISL1-bound peaks in day 18 MNPs. (F) Histogram of ISL1-bound peaks based on distance from gene transcription start sites (TSSs) for peaks that are unique to ISL1 or NKX2.5, or shared by both, in CPs. (G) Distribution of ISL1-bound, NKX2.5-bound or co-bound loci at annotated genomic features in day 6 CPs. (H) The 5 most significant GREAT GO Biological Process terms for ISL1-specific bound CP peaks, ISL1-NKX2.5 shared CP peaks, or peaks bound by NKX2.5 only.

A   Gene Editing Strategy to Generate *NKX2.5⁻/⁻*

D   Cardiac Factors in Day 15 CMs

B   NKX2.5 in Day 6 CPs

C   ISL1 in Day 6 CPs

E   Pseudotime Trajectory in Day 8 CPs

F   *ISL1* Expression in WT Day 8 CPs

G   *NKX2.5* Expression in WT Day 8 CPs

H   Replicate Distribution in Day 8 CPs

I   Regulation of *NPPA/NPPB* Putative Enhancer
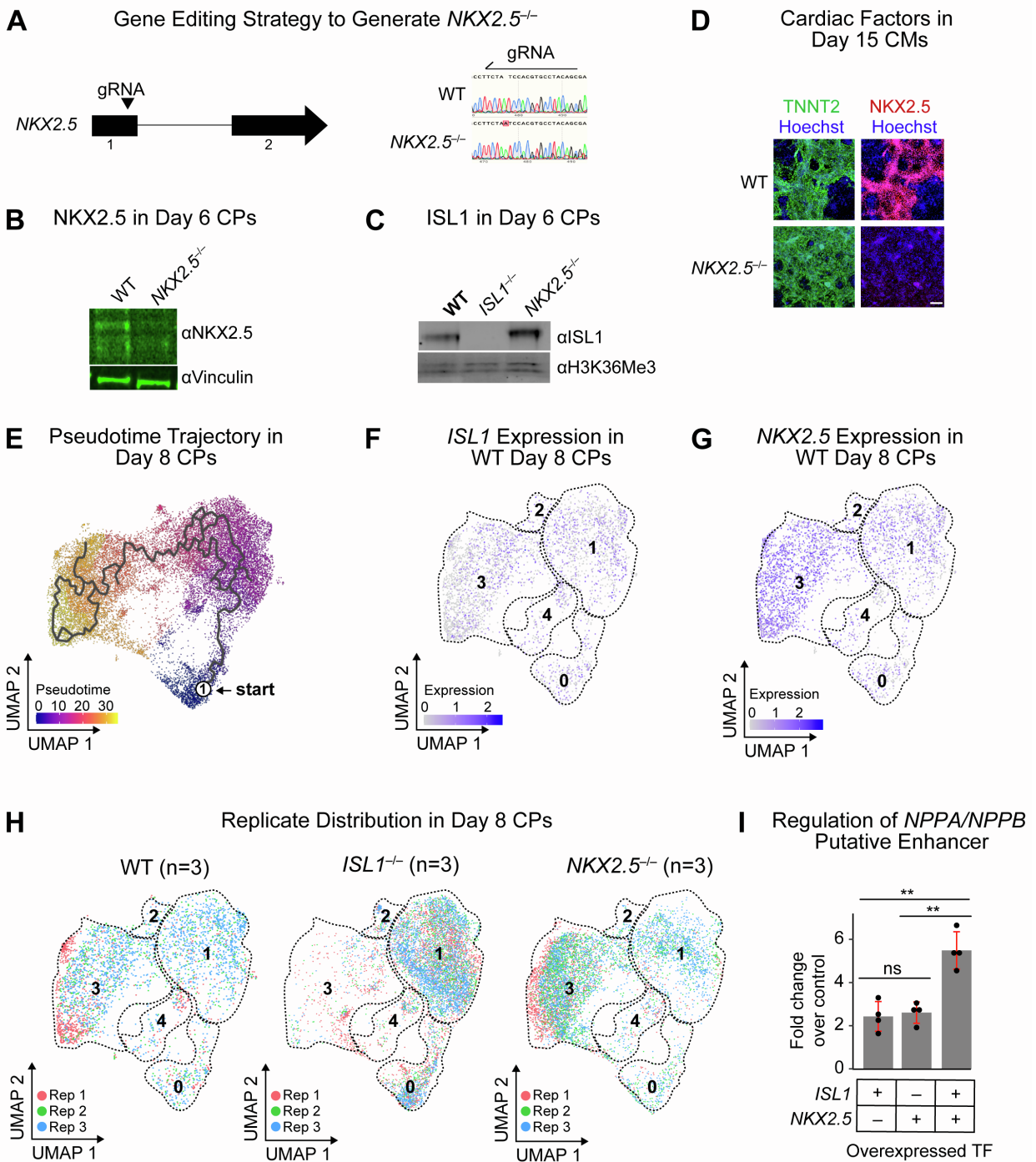

Figure S5. Additional analyses of the absence of *NKX2.5* in CPs, related to Figure 5.

(A) *NKX2.5* genomic CRISPR/Cas9 targeting strategy to introduce a frameshift mutation to both

loci. Numbers denote exons. (B) Western blot of NKX2.5 in WT or *NKX2.5⁻/⁻* day 6 CPs. (C)

Western blot of ISL1 expression in WT, *ISL1*$^{-/-}$ or *NKX2.5*$^{-/-}$ day 6 CPs using anti-ISL1 antibody. H3K36Me3 served as a loading control. (D) Immunofluorescence of TNNT2 and NKX2.5 in WT or *NKX2.5*$^{-/-}$ day 15 CMs. Scale bar, 100 µM. (E) Monocle pseudotime analysis of day 8 CPs. Colors represent arbitrary units of pseudotime. (F, G) Expression levels of *ISL1* (F) and *NKX2.5* (G) in WT day 8 CPs, superimposed on the UMAP from **Figure 1C**. (H) Replicate ("Rep") comparison in WT (n = 3 biological replicates; 3636 cells), *ISL1*$^{-/-}$ (n = 3 biological replicates; 8090 cells), or *NKX2.5*$^{-/-}$ (n = 3 biological replicates; 7185 cells) day 8 CPs. (I) Luciferase activity of a *NPPA/NPPB* putative enhancer regulatory region upstream of luciferase in response to ISL1 and NKX2.5 expression, relative to empty vector in COS7 cells. Data represented as means ± 1.96SD from independent experiments (n = 4 biological replicates, ** p-value ≤ 0.01).
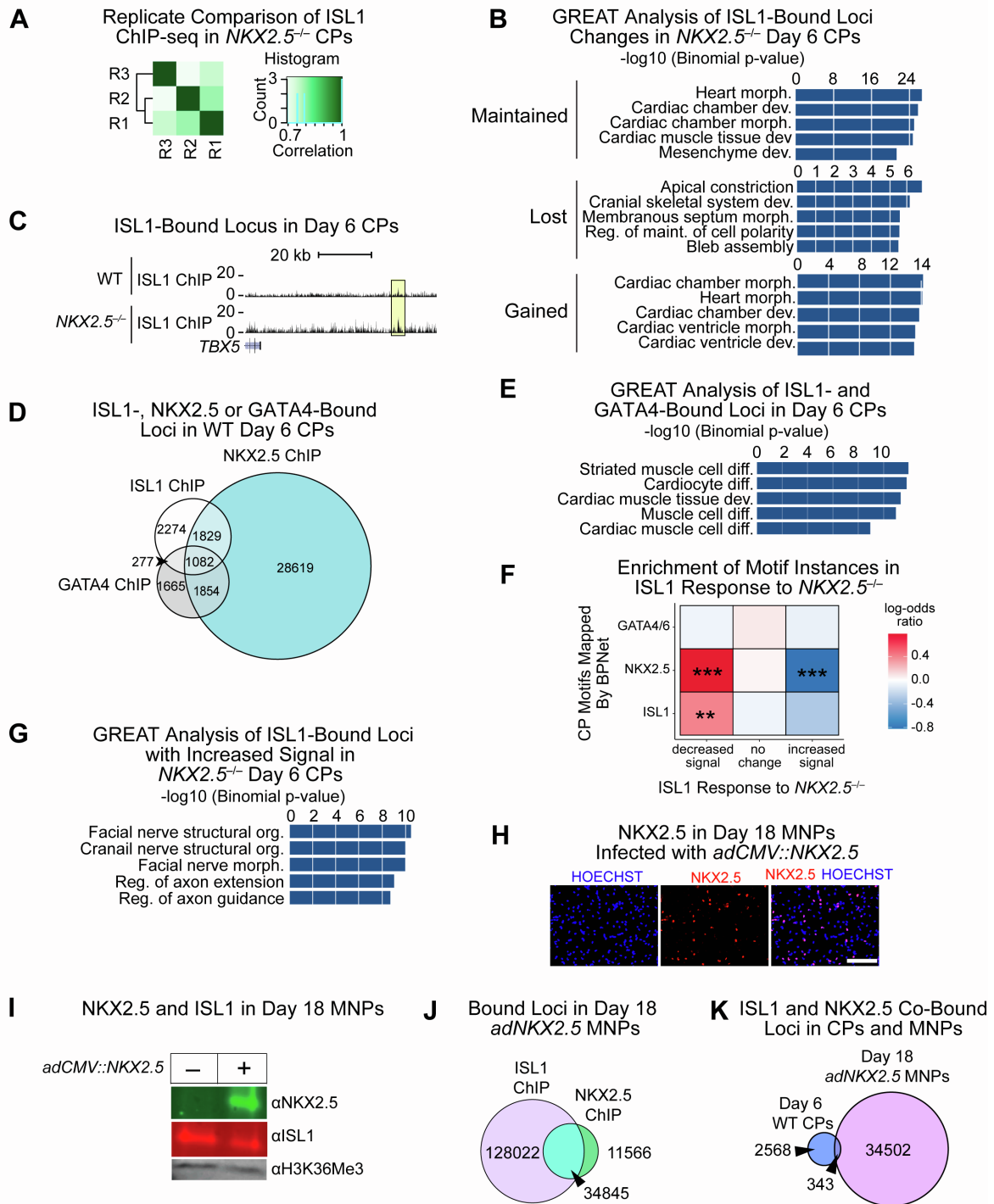
A Replicate Comparison of ISL1 ChIP-seq in *NKX2.5⁻/⁻* CPs

B GREAT Analysis of ISL1-Bound Loci Changes in *NKX2.5⁻/⁻* Day 6 CPs
-log10 (Binomial p-value)

Maintained
- Heart morph.
- Cardiac chamber dev.
- Cardiac chamber morph.
- Cardiac muscle tissue dev
- Mesenchyme dev.

Lost
- Apical constriction
- Cranial skeletal system dev.
- Membranous septum morph.
- Reg. of maint. of cell polarity
- Bleb assembly

Gained
- Cardiac chamber morph.
- Heart morph.
- Cardiac chamber dev.
- Cardiac ventricle morph.
- Cardiac ventricle dev.

C ISL1-Bound Locus in Day 6 CPs

D ISL1-, NKX2.5 or GATA4-Bound Loci in WT Day 6 CPs

E GREAT Analysis of ISL1- and GATA4-Bound Loci in Day 6 CPs
-log10 (Binomial p-value)

- Striated muscle cell diff.
- Cardiocyte diff.
- Cardiac muscle tissue dev.
- Muscle cell diff.
- Cardiac muscle cell diff.

F Enrichment of Motif Instances in ISL1 Response to *NKX2.5⁻/⁻*

G GREAT Analysis of ISL1-Bound Loci with Increased Signal in *NKX2.5⁻/⁻* Day 6 CPs
-log10 (Binomial p-value)

- Facial nerve structural org.
- Cranial nerve structural org.
- Facial nerve morph.
- Reg. of axon extension
- Reg. of axon guidance

H NKX2.5 in Day 18 MNPs Infected with *adCMV::NKX2.5*

HOECHST | NKX2.5 | NKX2.5 HOECHST

I NKX2.5 and ISL1 in Day 18 MNPs

adCMV::NKX2.5 − / +

αNKX2.5
αISL1
αH3K36Me3

J Bound Loci in Day 18 *adNKX2.5* MNPs

ISL1 ChIP — 128022
NKX2.5 ChIP — 11566
34845

K ISL1 and NKX2.5 Co-Bound Loci in CPs and MNPs

Day 6 WT CPs — 2568
Day 18 *adNKX2.5* MNPs — 34502
343

Figure S6. Additional analyses of the ISL1-bound loci lost or gained in *NKX2.5⁻/⁻* CPs, related to Figure 6.

(A) Replicate comparison of ISL1 ChIP-seq in *NKX2.5⁻/⁻* day 6 CPs using DiffBind. (B) GREAT

GO Biological Process terms of ISL1-bound loci that were classified as gained, maintained or lost in *NKX2.5*$^{-/-}$ day 6 CPs. (C) ChIP-seq track of ISL1 detailing increased intensity of ISL1 binding in the absence of NKX2.5 at the *TBX5* locus. Data shown from single representative replicate. (D) Venn Diagram of loci bound by ISL1, NKX2.5 or GATA4 in day 6 CPs. (E) GREAT GO Biological Process terms of ISL1- and GATA4-cobound loci in WT day 6 CPs. (F) Log-odds ratio with Bonferroni p-value correction of enrichment of the *de novo* identified GATA, NKX2.5 and ISL1 motifs among the increased, decreased or maintained signal intensities of the ISL1 ChIP peaks in *NKX2.5*$^{-/-}$ day 6 CPs. ** denotes p-value ≤ 0.01 and ** denotes p-value ≤ 0.001. (G) GREAT GO Biological Process terms of ISL1-bound loci that had increased intensity in *NKX2.5*$^{-/-}$ day 6 CPs, as determined in **Figure 6D**. (H) Immunofluorescence confirmation of NKX2.5 overexpression in day 18 MNPs. Scale bar, 200 μM. (I) Western blot of NKX2.5 and ISL1 expression in *adCMV::NKX2.5*-infected day 18 MNPs. H3K36Me3 served as loading control. (J) Venn diagram comparison of ISL1-bound and NKX2.5-bound loci in *NKX2.5*-overexpressed day 18 MNPs ("adNKX2.5 MNPs") (n = 1 biological replicate). (K) Venn diagram of ISL1- and NKX2.5-cobound loci in Day 6 WT CPs compared to *NKX2.5*-overexpressed day 18 MNPs ("adNKX2.5 MNPs").

SUPPLEMENTAL TABLES

Supplemental Table 1. Differentially expressed markers in WT, *ISL1*−/− Day 8 CPs, and in Day

8 CPs treated with *ISL1* siRNA, related to **Figures 1** and **S1**.

Supplemental Table 2. Differentially expressed markers in WT and *ISL1*−/− Day 18 MNPs,

related to **Figure 2**.

Supplemental Table 3. SRM peptides used for quantification of proteins from ISL1 IPs in WT or

ISL1−/− Day 6 CPs, related to **Figure 3**.

Transcription Factor	Peptide	Expressed in CPs	Expressed in MNPs
ISL1	ADHDVVER	Yes	Yes
ISL1	CAECNQYLDESCTCFVR	Yes	Yes
LDB1	SILAMHAQDPQMLDQLSK	Yes	Yes
FOXO3	ALSNSVSNMGLSESSSLGSAK	Yes	No
GATA4	EAAAYSSGGGAAGAGLAGR	Yes	No
GATA4	ECVNCGAMSTPLWR	Yes	No
GATA4	FSFPGTTGSLAAAAAAAAR	Yes	No
GATA4	LSPQGYASPVSQSPQTSSK	Yes	No
GATA4	VGLSCANCQTTTTTLWR	Yes	No
GATA6	ECVNCGSIQTPLWR	Yes	No
GATA6	GPSADLLEDLSESR	Yes	No
HAND2	TGWPQHVWALELK	Yes	No
HAND2	TQSINSAFAELR	Yes	No
NKX2-5	IAVPVLVR	Yes	No
NKX2-5	SLAAAGELSAR	Yes	No
NKX2-5	VLFSQAQVYELER	Yes	No

NKX3-1	AAFSHTQVIELER	Yes	No
TEAD4	GPSNAFFLVK	Yes	No
TEAD4	YENGHYSYR	Yes	No
LHX3	QLATGDEFYLMEDSR	No	Yes
PHOX2A	GALWAGVAGGGGGGPGAGAAELLK	No	Yes

Supplemental Table 4. Differentially expressed markers in WT, *ISL1*$^{-/-}$, and *NKX2.5*$^{-/-}$ Day 8 CPs, related to **Figure 5**.

Supplemental Table 5. Resources used in this paper, related to Experimental Procedures.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Cell lines. The WTC11 hiPSC line (male) was obtained from the Gladstone Stem Cell Core. All hiPSC lines were regularly checked for chromosomal abnormalities via karyotyping by Cell Line Genetics. COS-7 cells were from ATCC.

Cell culture conditions. hiPSCs were grown on tissue culture-treated polystyrene plates coated with hESC-qualified matrigel with Essential 8™ Medium. Cells were passaged every 3-4 days at 1:10 split ratio with accutase in the presence of 5 μM ROCK inhibitor, Y-27632. COS-7 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) with high-glucose GlutaMax Supplement supplemented with 10% fetal bovine serum. All cultures were maintained in humidified incubators at 37°C and 5% CO_2.

Generation of cardiac cells. We followed a previously published protocol (Lian et al. 2013) with minor modifications. In brief, we exposed hiPSC cells at 70% confluency to 6 μM CHIR for 48

hours, then subsequently 5 µM IWP4 for 48 hours. We continued with RPMI 1640 Medium with B27™ Supplement, minus insulin until day 10, when the media was switched to RPMI 1640 Medium with B27™ Supplement. We confirmed their quality in part by assessing their morphology and the expression of markers via immunofluorescence at day 6 and day 15. Day 6 markers: rabbit-anti-MYH6, mouse-anti-MYL3. Day 15 markers: mouse-anti-TNNT2, rabbit-anti-MYL2, mouse-anti-MYL3. Robust beating by day 10 further confirmed the quality of the WT CM differentiation.

Generation of motor neuron cells. hiPSCs were differentiated into MNPs in a three-step process as previously published (NeuroLINCS Consortium et al. 2021). In short, iPSCs were cultured with neural differentiation media composed of IMDM and F12, supplemented with non-essential amino acids, B27™ Supplement minus insulin, N-2, LDN 193189 dihydrochloride, SB 431542 and CHIR. On day 6 of differentiation, cells were passaged and reseeded with MN precursor media composed of IMDM and F12 supplemented with NEAA, B27™ Supplement minus insulin, N-2, LDN 193189 dihydrochloride, SB431542, CHIR, retinoic acid and Smoothened receptor agonist. On day 12 of differentiation, cells were passaged and reseeded with MN precursor expansion media composed of IMDM and F12 supplemented with NEAA, B27™ Supplement minus insulin, N-2, Gamma-secretase inhibitor (Compound E), γ-secretase inhibitor (DAPT), dibutyryl CAMP (db-cAMP), retinoic acid, SAG, ascorbic acid, recombinant human BDNF protein, and recombinant human GDNF protein. On day 18 of differentiation, cells were collected for assays. We confirmed the quality of each differentiation by assessing synapse morphology and the presence of known MN markers (NKX6.1, SMI32, ISL1).

Single Cell RNA-Sequencing
Single-cell transcriptome library preparation and sequencing. Day 8 CPs or day 18 MNPs were collected via accutase and washed twice with cold PBS, each time with spinning (800 rpm,

3 min.). Cells were then passed through a 70 µM cell strainer and centrifuged for 3 min. at 150g. They were resuspended with cold DPBS and quantified with the Countess Cell Counter. When relevant, fluorescent activated cell sorting (FACS) was used next to isolate the population of interest before continuing. Single-cell droplet libraries were then generated with the 10X Genomics Chromium controller according to the Chromium Single Cell 3' Reagent Kit v3 User Guide.

Next, the Chromium Single Cell 3' Library and Gel Bead Kit v3 and Chromium i7 Multiplex Kit were used according to the manufacturer's specifications to generate the libraries. Libraries were sequenced on the Illumina NovaSeq 6000 System or the Illumina NextSeq500 based on the Chromium Single Cell v3 specifications. Samples from the same time point were pooled and sequenced together.

Processing of the raw sequencing reads. Reads were demultiplexed with the Cell Ranger v3.1.0 pipeline (10X Genomics) and aligned to the human hg19 genome. UMI counts were then quantified on a per gene and per cell basis to generate a gene-barcode matrix. Data from each dataset (WT D18 MNP Replicates 1-3 and $ISL1^{-/-}$ D18 MNP Replicates 1-3; WT D8 CP Replicates 1-3, $ISL1^{-/-}$ D8 CP Replicates 1-3, $NKX2.5^{-/-}$ D8 CP Replicates 1-3, Scrambled control siRNA D8 CP, $ISL1$ siRNA D8 CP) were then aggregated and normalized according to the sequencing depth, which generated a combination of gene-barcode matrices for all the samples.

Cell filtering and cell-type clustering analysis. Following sequencing of the pooled samples, we filtered and performed cluster analyses as described previously (de Soysa et al. 2019), with modifications for use with the Seurat v3 R package. For each aggregated dataset, cells were normalized for total gene expression and genes expressed per cell. Low quality or likely doublet cells were excluded from analysis. A linear regression was then performed to eliminate technical variability. Cell cycle phase scores were calculated and regressed out according to the Seurat

cell cycle vignette (https://satijalab.org/seurat/) using a list of canonical cell cycle markers. Highly variable genes were computed and then used for Principal Component analysis. The output from this analysis was used to distribute the clustering into distinct populations (clusters) and Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction was performed to visualize these clusters. Harmony batch correction (Korsunsky et al. 2019) was then performed and UMAP dimensionality reduction was performed again. In CPs, non-cardiac endoderm and endothelial cell clusters were identified in CP datasets based on marker gene expression. These populations were subclustered out, and harmony batch correction and UMAP dimensionality reduction was performed again to generate the final dataset to analyze. The MNP aggregated dataset did not have non-neuronal populations that were equally shared between $ISL1^{-/-}$ and WT cells, so all neural populations were kept for further analyses. To identify differentially expressed genes in each of the clusters, we used the FindAllMarkers function of Seurat with the options: return.thresh (p-value cut off) = $1x10^{-2}$ and logfc.threshold = 0.25. To identify differentially expressed genes in the $ISL1^{-/-}$ populations of each aggregated dataset, we used the FindMarkers function of Seurat with the options: return.thresh (p-value cut off) = $1x10^{-2}$ and logfc.threshold = 0.25. When necessary, cells in each group were downsampled (seed = 8). To determine overall biological GO of differentially expressed genes, we employed the Panther Statistical Overrepresentation GO Biological Process and displayed the most enriched GO categories that were significant (Bonferroni adjusted p-value ≤ 0.05) over the reference dataset of *homo sapiens* genes (Mi et al. 2019).

ChIP-Sequencing.
Crosslinking and immunoprecipitation of protein complexes (ChIP). CPs ($10x10^6$ cells), MNPs ($30x10^6$ cells), or transfected COS-7 ($3x10^6$ cells) were collected from tissue culture plates via accutase and pelleted (1000 rpm, 3 min.) After a DPBS wash, cells were quantified on a Countess Cell Counter and then cross-linked with a 1% paraformaldehyde solution (made fresh) with rotation for 10 min. at RT. Glycine was added at a final concentration of 0.125 M to quench

the cross-linker, with rotation for 10 min. at RT. Samples were pelleted, washed twice with cold DPBS and snap-frozen before being stored at -80 °C.

On the day of immunoprecipitation, frozen pellets were thawed on ice, then spun briefly. Cells were lysed with 1 mL per 10^6 cells of Cell Lysis Buffer (20 mM Tris-HCl, pH 8, 85 mM KCl, 0.5% NP-40) supplemented with PhosSTOP phosphatage inhibitors and cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail and incubated on ice for 10 min., then with rotation for 10 min. at 4 °C. Nuclei were pelleted by centrifugation (2500g, 5 min., 4 °C). The nuclear pellet was resuspended in Nuclear Lysis Buffer (50 mM Tris-HCl, pH8, 10 mM EDTA, pH8, 1% SDS) supplemented with protease and phosphatase inhibitors and incubated for 30 min. with rotation at 4 °C. The solution was then transferred to milliTUBE 1mL AFA Fiber tubes (Covaris) and sheared using a Covaris S2 sonicator for 15 min. (60 s cycles, 5% duty cycle, 200 cycles/burst, intensity = 6) to produce DNA sheared in the 200-700 bp range. Samples were then 3-fold diluted with dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl, Ph8, 167 mM NaCl) supplemented with protease and phosphatase inhibitors and an aliquot was removed for input control. Primary antibody was then added and samples were rotated overnight at 4 °C. The following day, protein complexes were immunoprecipitated using Dynabeads Protein A/Protein G for 2 hours with rotation at 4 °C. After incubation, samples were placed on a magnetic stand (DynaMag-2 Magnet). Samples were washed five times with RIPA buffer (50 mM HEPES-KOH, pH 7.5, 500 mM LiCl, 1 mM EDTA, 1% NP-40, 0.7% Na-deoxycholate), followed by two washes with final wash buffer (1xTE, 50 mM NaCl). Bound proteins were then eluted with agitation for 30 min. at 65 °C in elution buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS). To the eluted and input samples, a reverse cross-linking solution (250 mM Tris-HCl, pH 7.5, 32.5 mM EDTA, pH 8, 1.25 M NaCl) was added. If continuing with proteomic analyses (transfected COS-7 cells only), Benzonase was added overnight with no rotation at 65°C. For ChIP, proteinase K was added overnight with no rotation at 65 °C. The following morning, the ChIP samples were treated with RNase A. Primary antibody used for proteomic analyses was anti-human-ISLET-1 antibody.

Primary antibodies used for ChIP were: anti-human-ISLET-1 antibody and anti-NKX-2.5 Antibody (N-19).

Sample preparation and sequencing following ChIP. After immunoprecipitation of protein-digested complexes, DNA was purified with AMPure XP beads. Samples were then end-repaired, 5'-phosphorylated and dA-tailed with NEBNext Ultra II DNA Library Kit for Illumina. Diluted adaptor oligos were ligated for multiplex sequencing, and PCR amplified (98 °C 30 sec, 12 cycles of 98 °C for 10 sec and 65 °C for 75 sec, 65 °C for 2 min.). Samples were cleaned via AMPure XP beads and then analyzed for quantity with the Qubit 4 Fluorometer and the accompanying Qubit dsDNA HS Assay Kit. Samples were further assayed for quantity and quality using the 2100 Bioanalyzer Instrument and accompanying Bioanalyzer High Sensitivity DNA Analysis kit. Following this quality check, samples were sequenced single-end (SE) on the Illumina NextSeq500 or the Illumina HiSeq 4000 (UCSF Center for Advanced Technology, CAT) and then subsequently demultiplexed.

Analysis of ChIP-seq data. When applicable, these analyses were performed with The Galaxy Project interface (https://usegalaxy.org/) unless otherwise stated. ChIP-seq FASTQ sequences were mapped to the built-in human hg19 genome using Map with BWA-MEM (Li and Durbin 2009) and the following options: Set read groups (SAM/BAM specification), Auto-assign read group identifier, Auto-assign read group sample name, Auto-assign library name, and Analysis mode set to Simple Illumina Mode. The BWA-aligned data were filtered using Filter SAM or BAM (Li et al. 2009) with a minimum MAPQ quality score of 20 and BAM filetype output. The BAM files were processed into BigWig files using BamCoverage (Ramírez et al. 2016) for visualization on the UCSC genome browser (Kent et al. 2002) with the following options: 25 bp bin size, normalize coverage to 1X of the hg19 genome, output as a bigwig filetype. Replicate correlation analyses were done with DiffBind (Stark, Brown, and Others 2011) in R Studio. We calculated peaks with

the TF-specific MACS2 callpeak algorithm (Feng et al. 2012) using input as the control sample and a q-value of 0.05. Peakset overlaps were calculated via bedops (Neph et al. 2012). To build a consensus peak set for each condition, we first employed the bedops -element-of command with the overlapping criteria set to 10 bp to find overlapping peaks between each replicate. Then, we used bedops -merge to combine each of these comparisons into the final consensus peak set. These overlaps were visualized with Eulerr (https://CRAN.R-project.org/package=eulerr). RegioneR was used to determine significance of overlapping peak sets (Gel et al. 2016). First, MACS bed files were converted into GRange datasets, then permTest was used to determine significance of overlaps with the following conditions: 1000 permutations, non.overlapping set to "True", per.chromosome set to "True", alternative set to "auto" and verbone set to "True". The GREAT tool was used to determine biological significance of each consensus peak set in addition to distance to TSS for each peakeset (McLean et al. 2010). To determine distribution of peaks at genomic annotations, we used plotAnnoBar from ChIPseeker (Yu, Wang, and He 2015). To determine overlap of histone marks, we compared our data to the active H3K27Ac histone modification ChIP from cardiac progenitor (Lee et al. 2018). In addition, we compared our data to ChIP for the H3K36Me3 repressive modification in cardiac progenitors (Gonzalez-Teran et al. 2022). After merging replicate datasets with MergeSamFiles, we used BamCompare to scale each sample and its corresponding input control by read count. We then normalized these reads by comparing the log2 of the number of reads ratio. These were then plotted along ISL1 ChIP consensus peaks using computeMatrix, and visualized with plotProfile.

Known motif enrichment. We identified enriched known motifs in WT CP and MNP peak sets using the Hypergeometric Optimization of Motif EnRichment algorithm (HOMER, http://homer.ucsd.edu). We combined the position weight matrices (PWMs) for the ten most enriched motifs for each cell type, excluding the ISL1 motif. We again employed HOMER to quantify motif enrichment in each peak set of the combined PWM motifs, while using the peak set

of the reciprocal cell type as the background.

Analysis of protein complexes.

Immunoprecipitation of protein complexes (IP). For co-IPs of ISL1 in CPs for use in downstream proteomic analyses (western blot or mass spectrometry), we began with 10x10^6 WT or ISL1$^{-/-}$ day 6 CPs, previously snap-frozen without crosslinking. We resuspended each thawed pellet of cells with 2 mL of IP Buffer (0.5 M Tris-HCl, pH 7.4, 0.15 M NaCl, 0.001 M EDTA; pH adjusted to 7.4 at 4 °C) supplemented with 1:20 of 10% NP-40 and protease and phosphatase inhibitors. The resuspended pellet was then put on rotation for 30 minutes at 4 °C. At this point, Dynabeads Protein G beads were conjugated with either 2 µg of ISL1 antibody or Rabbit IgG per 50 µL of beads slurry. Next, samples were loaded into milliTUBE 1mL AFA Fiber Tubes and sheared using a Covaris S2 sonicator. Sheared lysates were transferred to Protein LoBind Tubes and spun at 3500 g at 4 °C. The supernatant was transferred to a new tube, and these were quantified with the Pierce™ BCA Protein Assay Kit. For each IP, 2 mg of protein lysate was used. An aliquot was saved for input control, and then conjugated antibody-beads were added and samples were placed on rotation for four hours at 4 °C. After incubation, tubes were placed on magnetic stand (DynaMag-2 Magnet), and an aliquot was saved as output control. Two washes were done with IP Buffer supplemented with 1:200 of 10% NP-40, followed by three washes with IP Buffer without NP-40. To remove protein eluate from the antibody-conjugated beads, one half of the sample was aliquoted and denatured in sample buffer to prepare for western blot analyses. The remaining half was prepared for mass spectrometry. Primary antibodies used for the IPs were: anti-ISL1 and anti-IgG (Rabbit).

IP eluate preparation for mass spectrometry. For continuing to mass spectrometry analyses, on-bead digestion was performed to release eluate proteins from the protein A/G beads. The beads were resuspended in one bead slurry volume of reduction buffer (2 M Urea, 50 mM Tris,

pH 8.0, 1 mM Dithiothreitol (DTT)) and incubated for 30 min. at 37 °C followed by addition of 3 mM iodoacetamide and an incubation period of 45 min. in the dark at RT with 600 rpm shaking to ensure bead suspension. Following this, an additional 3mM DTT was added to the suspension. Then 1875 ng of Trypsin per 50 µL bead slurry was added and incubated overnight at 37 °C with 600 rpm shaking. The following morning, approximately 940 ng of Trypsin was added and the tubes were incubated at 37 °C with 600 rpm shaking for one hour. Following this, beads were pelleted for 2000 rpm for 4 min. and tubes were placed on magnetic tray. Eluate supernatant was transferred to a Protein LoBind Tube. We continued with desalting of the samples with OMIX Tips following manufacturer instructions. After lyophilization, samples were resuspended in 0.2% acetonitrile/0.2% formic acid before loading onto mass spectrometer.

Targeted proteomics measurements. Once a list of target proteins was identified, peptides corresponding to each protein were chosen from the PeptideAtlas database (Desiere et al. 2006) by prioritizing peptides that are unique for the protein of interest and have been detected in previous experiments with high frequency or were predicted to have favorable characteristics for MS analysis (such as peptide length below 25 amino acids). PEPotecHeavy Grade 1 peptides were synthesized from Life Technologies Corporation for 4-5 candidate peptides per protein containing a heavy isotope labeled C-terminal arginine or lysine. Synthesized peptides were resuspended in 2% acetonitrile/0.1% formic acid, pooled, then desalted with OMIX Tips following manufacturer instructions, and resolubilized in 0.2% acetonitrile/0.2% formic acid. To generate targeted proteomics assays, the peptide mixture was analyzed on a Q-Exactive Plus mass spectrometer operated in data-dependent acquisition (DDA) mode coupled to an Easy nLC 1200 nano-flow ultra high-pressure liquid chromatography interfaced via a Nanospray Flex nanoelectrospray source. Samples were loaded onto a C18 column (25 cm x 75 µm I.D. packed with ReproSil Pur C18 AQ 1.9 µm particles). Mobile phase A consisted of 0.1% FA, and mobile phase B consisted of 0.1% FA/80% ACN. Peptides were separated at a flow rate of 300 nl/min.

using a gradient from 4.5% to 32% acetonitrile over 53 min. All MS spectra were collected with Orbitrap detection, while the 20 most abundant ions were fragmented by HCD and detected in the Orbitrap. Resulting data was searched against the Uniprot Human protein database (downloaded on February 1, 2017) for peptide identifications using the MaxQuant data analysis algorithm (version v1.5.8.0) (Cox and Mann 2008) with the default parameters other than the following: group-specific parameter set to Arg10 and Lys8 for the heavy-labeled modifications. All peptide and protein identifications were filtered to a 1% false discovery rate. SRM assay generation was performed using Skyline (Pino et al. 2020). For all peptides optimal transitions for identification and quantification were selected based on a spectral library generated from the DDA MS experiments. The Skyline spectral library was used to extract optimal coordinates for the SRM assays, e.g., peptide fragments and peptide retention times. For the targeted proteomics measurements, heavy labeled peptides were spiked in equal amounts into each IP sample, which were analyzed by LC-SRM on a Thermo Scientific TSQ Quantiva MS system equipped with a Proxeon Easy nLC 1200 ultra high-pressure liquid chromatography and autosampler system. Samples were injected onto a C18 column (25 cm x 75 μm I.D. packed with ReproSil Pur C18 AQ 1.9 μm particles) in 0.1% formic acid and then separated with an 80 min. gradient from 5% to 40% Buffer B (90% ACN/10% water/0.1% formic acid) at a flow rate of 300 nl/min. SRM acquisition was performed operating Q1 and Q3 at 0.7 unit mass resolution. For each peptide the best minimum of 3 transitions were monitored in a scheduled fashion with a retention time window of 10 min. and a cycle time fixed to 2 sec. We removed any peptides that did not produce high-quality transitions. The most robust peptides passing this selection are displayed in **Supplemental Table 3**. Argon was used as the collision gas at a nominal pressure of 1.5 mTorr. Collision energies were calculated by, $CE = 0.0348 * (m/z) + 0.4551$ and $CE = 0.0271 * (m/z) + 1.5910$ (CE, collision energy and m/z, mass to charge ratio) for doubly and triply charged precursor ions, respectively. RF lens voltages were calculated by, $RF = 0.1088 * (m/z) + 21.029$ and $RF = 0.1157 * (m/z) + 0.1157$ (RF, RF lens voltage and m/z, mass to charge ratio) for doubly

and triply charged precursor ions, respectively. The resulting data was analyzed with Skyline for identification and quantification of peptides (Pino et al. 2020). To calculate the enrichment of each peptide, Skyline first produced values of enrichment of target peptide over the heavy-labeled peptide. The overall enrichment score displayed in **Figure 4B** was calculated as the ratio of the peptide enrichment values in the ISL1 IP in WT CPs over those in *ISL1*$^{-/-}$ CPs.

BPNet Model Training and Analyses

BPNet model training and motif identification. The model outputs were two ChIP-seq datasets showing WT ISL1 binding across (1) CPs collected at day 6 and (2) MNPs collected at day 18. The model inputs were 1 kb sequences across consensus MACS2 peaks. The model was controlled by a bias transformation step, which used merged coverage of the WT ChIP-seq controls to distinguish uninformative coverage during model training. Peaks occurring across chromosomes 1, 8, and 9 were withheld as validation datasets peaks, and peaks across chromosomes 5, 10, and 12 were withheld as test datasets for model performance assessment. Model performance was assessed based on (1) the Pearson and Spearman correlation of predicted counts between the observed and predicted profiles and (2) the ability of the model to correctly predict ChIP-seq summit positions, as measured by the area under the precision-recall curve (auPRC) with positive and negative class probability thresholds set to 0.0025 and 0.001, respectively, to accommodate the typical distribution of ChIP-seq peak coverage. These performance metrics are consistent with the published approach. Model parameters were optimized by assessing the influence of the following parameter combinations on model performance: (1) learning rates of 0.01, 0.004, 0.001, and 0.0004; (2) convolutional filter depths of 16, 64, 128, 256, and 512; (3) number of convolutional layers of 5, 7, 9, and 11; and (4) counts loss scaling (lambda) weights of 1, 10, 100, 250, 500, and 1000. The final model architecture contained 9 convolutional layers, a filter depth of 64, a learning rate of 0.001, and a lambda value of 100. Upon acquisition of an optimized model, DeepLIFT (Shrikumar, Greenside, and Kundaje

2017) and TF-MoDISco (Shrikumar et al. 2018) were used to assign sequence contribution to the input sequence regions, cluster and aggregate high-contribution features into motif representations, and map those motif representations back to the ChIP-seq peak regions, consistent with the published BPNet approach (Avsec et al. 2021). To further analyze CP-specific subsets of these data, we used the bedops toolset to capture peaks harboring motifs of NKX2.5 (and not GATA), GATA (and not NKX2.5), both or neither and displayed percentages of these. The GREAT tool was then used to determine biological significance of each motif-specific peak set.

Mapped motif pair synergy. Genomic instances of two consensus motifs within 150 bp of one another were mutated individually and simultaneously and the predicted ISL1 binding effects were recorded. Synergy was assessed using the formula:

$$synergy = log2 \left(\frac{H_{AB} - (H_B - H_\emptyset) + p_{AB}}{H_A + p_B} \right)$$

where H_{AB} is the effect when both motifs were present (WT), H_B is the effect when motif A is mutated, H_A is the effect when motif B is mutated, $H_\emptyset$ is the effect when both motifs are mutated, p_{AB} is the 20th percentile of the considered effect in the WT sequence, and p_B is the 20th percentile of the considered effect in the sequence with motif A mutated. Effects (H) were measured as the ISL1 predicted maximum binding across a 500 bp window, centered on motif A. Pseudocounts (p) were measured to control for predicted background effects of binding. Mutation of a motif involved replacing the genomic sequence with random sequence and measuring effects. Each mutation effect was averaged over 64 trials with different random sequences to ensure correct mutation behavior was measured.

Avsec, Žiga, Melanie Weilert, Avanti Shrikumar, Sabrina Krueger, Amr Alexandari, Khyati Dalal, Robin Fropf, et al. 2021. "Base-Resolution Models of Transcription-Factor Binding Reveal Soft Motif Syntax." *Nature Genetics* 53 (3): 354–66.

Cox, Jürgen, and Matthias Mann. 2008. "MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification." *Nature Biotechnology* 26 (12): 1367–72.

Desiere, Frank, Eric W. Deutsch, Nichole L. King, Alexey I. Nesvizhskii, Parag Mallick, Jimmy Eng, Sharon Chen, James Eddes, Sandra N. Loevenich, and Ruedi Aebersold. 2006. "The PeptideAtlas Project." *Nucleic Acids Research* 34 (Database issue): D655-8.

Feng, Jianxing, Tao Liu, Bo Qin, Yong Zhang, and Xiaole Shirley Liu. 2012. "Identifying ChIP-Seq Enrichment Using MACS." *Nature Protocols* 7 (9): 1728–40.

Gel, Bernat, Anna Díez-Villanueva, Eduard Serra, Marcus Buschbeck, Miguel A. Peinado, and Roberto Malinverni. 2016. "RegioneR: An R/Bioconductor Package for the Association Analysis of Genomic Regions Based on Permutation Tests." *Bioinformatics (Oxford, England)* 32 (2): 289–91.

Gonzalez-Teran, Barbara, Maureen Pittman, Franco Felix, Reuben Thomas, Desmond Richmond-Buccola, Ruth Hüttenhain, Krishna Choudhary, et al. 2022. "Transcription Factor Protein Interactomes Reveal Genetic Determinants in Heart Disease." *Cell* 185 (5): 794-814.e30.

Jäger, Stefanie, Peter Cimermancic, Natali Gulbahce, Jeffrey R. Johnson, Kathryn E. McGovern, Starlynn C. Clarke, Michael Shales, et al. 2011. "Global Landscape of HIV-Human Protein Complexes." *Nature* 481 (7381): 365–70.

Kent, W. James, Charles W. Sugnet, Terrence S. Furey, Krishna M. Roskin, Tom H. Pringle, Alan M. Zahler, and David Haussler. 2002. "The Human Genome Browser at UCSC." *Genome Research* 12 (6): 996–1006.

Korsunsky, Ilya, Nghia Millard, Jean Fan, Kamil Slowikowski, Fan Zhang, Kevin Wei, Yuriy

Baglaenko, Michael Brenner, Po-Ru Loh, and Soumya Raychaudhuri. 2019. "Fast, Sensitive and Accurate Integration of Single-Cell Data with Harmony." *Nature Methods* 16 (12): 1289–96.

Lee, Jaecheol, Ning-Yi Shao, David T. Paik, Haodi Wu, Hongchao Guo, Vittavat Termglinchan, Jared M. Churko, et al. 2018. "SETD7 Drives Cardiac Lineage Commitment through Stage-Specific Transcriptional Activation." *Cell Stem Cell* 22 (3): 428-444.e5.

Li, Heng, and Richard Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform." *Bioinformatics* 25 (14): 1754–60.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25 (16): 2078–79.

Lian, Xiaojun, Jianhua Zhang, Samira M. Azarin, Kexian Zhu, Laurie B. Hazeltine, Xiaoping Bao, Cheston Hsiao, Timothy J. Kamp, and Sean P. Palecek. 2013. "Directed Cardiomyocyte Differentiation from Human Pluripotent Stem Cells by Modulating Wnt/β-Catenin Signaling under Fully Defined Conditions." *Nature Protocols* 8 (1): 162–75.

McLean, Cory Y., Dave Bristor, Michael Hiller, Shoa L. Clarke, Bruce T. Schaar, Craig B. Lowe, Aaron M. Wenger, and Gill Bejerano. 2010. "GREAT Improves Functional Interpretation of Cis-Regulatory Regions." *Nature Biotechnology* 28 (5): 495–501.

Mi, Huaiyu, Anushya Muruganujan, Xiaosong Huang, Dustin Ebert, Caitlin Mills, Xinyu Guo, and Paul D. Thomas. 2019. "Protocol Update for Large-Scale Genome and Gene Function Analysis with the PANTHER Classification System (v.14.0)." *Nature Protocols* 14 (3): 703–21.

Neph, Shane, M. Scott Kuehn, Alex P. Reynolds, Eric Haugen, Robert E. Thurman, Audra K. Johnson, Eric Rynes, et al. 2012. "BEDOPS: High-Performance Genomic Feature Operations." *Bioinformatics* 28 (14): 1919–20.

NeuroLINCS Consortium, Jonathan Li, Ryan G. Lim, Julia A. Kaye, Victoria Dardov, Alyssa N. Coyne, Jie Wu, et al. 2021. "An Integrated Multi-Omic Analysis of IPSC-Derived Motor Neurons from C9ORF72 ALS Patients." *IScience* 24 (11): 103221.

Pino, Lindsay K., Brian C. Searle, James G. Bollinger, Brook Nunn, Brendan MacLean, and Michael J. MacCoss. 2020. "The Skyline Ecosystem: Informatics for Quantitative Mass Spectrometry Proteomics." *Mass Spectrometry Reviews* 39 (3): 229–44.

Ramírez, Fidel, Devon P. Ryan, Björn Grüning, Vivek Bhardwaj, Fabian Kilpert, Andreas S. Richter, Steffen Heyne, Friederike Dündar, and Thomas Manke. 2016. "DeepTools2: A next Generation Web Server for Deep-Sequencing Data Analysis." *Nucleic Acids Research* 44 (W1): W160-5.

Shrikumar, Avanti, Peyton Greenside, and Anshul Kundaje. 2017. "Learning Important Features through Propagating Activation Differences." *ArXiv [Cs.CV]*. arXiv. http://arxiv.org/abs/1704.02685.

Shrikumar, Avanti, Katherine Tian, Žiga Avsec, Anna Shcherbina, Abhimanyu Banerjee, Mahfuza Sharmin, Surag Nair, and Anshul Kundaje. 2018. "Technical Note on Transcription Factor Motif Discovery from Importance Scores (TF-MoDISco) Version 0.5.6.5." *ArXiv [Cs.LG]*. arXiv. http://arxiv.org/abs/1811.00416.

Soysa, T. Yvanka de, Sanjeev S. Ranade, Satoshi Okawa, Srikanth Ravichandran, Yu Huang, Hazel T. Salunga, Amelia Schricker, Antonio Del Sol, Casey A. Gifford, and Deepak Srivastava. 2019. "Single-Cell Analysis of Cardiogenesis Reveals Basis for Organ-Level Developmental Defects." *Nature* 572 (7767): 120–24.

Stark, Rory, Gordon Brown, and Others. 2011. "DiffBind: Differential Binding Analysis of ChIP-Seq Peak Data." *R Package Version* 100 (4.3). https://bioconductor.statistik.tu-dortmund.de/packages/2.13/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf.

Yu, Guangchuang, Li-Gen Wang, and Qing-Yu He. 2015. "ChIPseeker: An R/Bioconductor Package for ChIP Peak Annotation, Comparison and Visualization." *Bioinformatics*

(Oxford, England) 31 (14): 2382–83.