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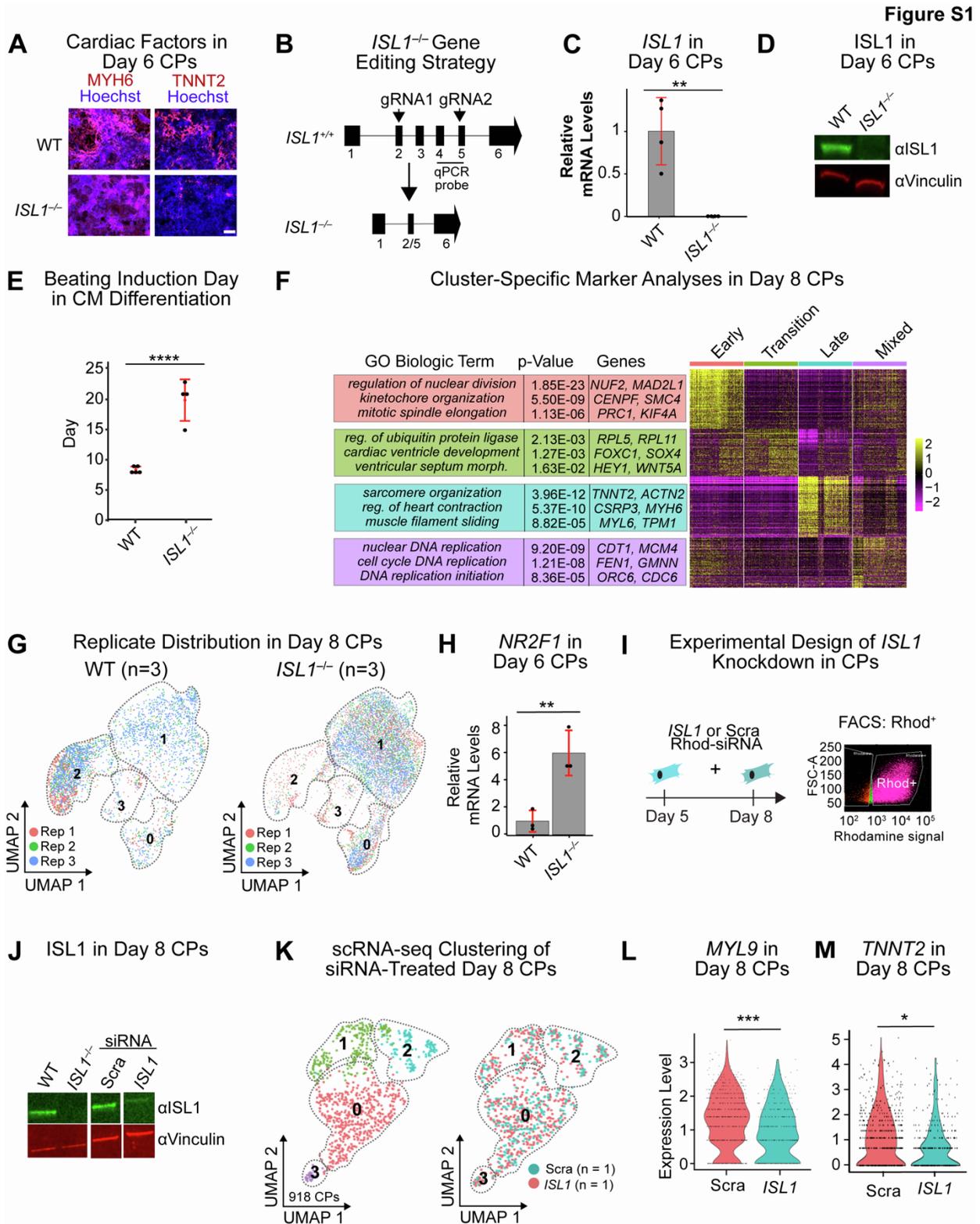
Figure S1. Additional analyses of the absence of *ISL1* in CPs, related to Figure 1.

(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 (%)

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

C

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

- ISL1
- H3K27Ac
- H3K36Me3

D

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

$ISL1^{-/-}$

MYL4   PHOX2B   SETD5

E

GREAT Analysis of
ISL1-Bound Loci in Day 6 CPs

-log10 (Binomial p-value)

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)

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

C

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

B

Hyperparameter Tuning Results

D

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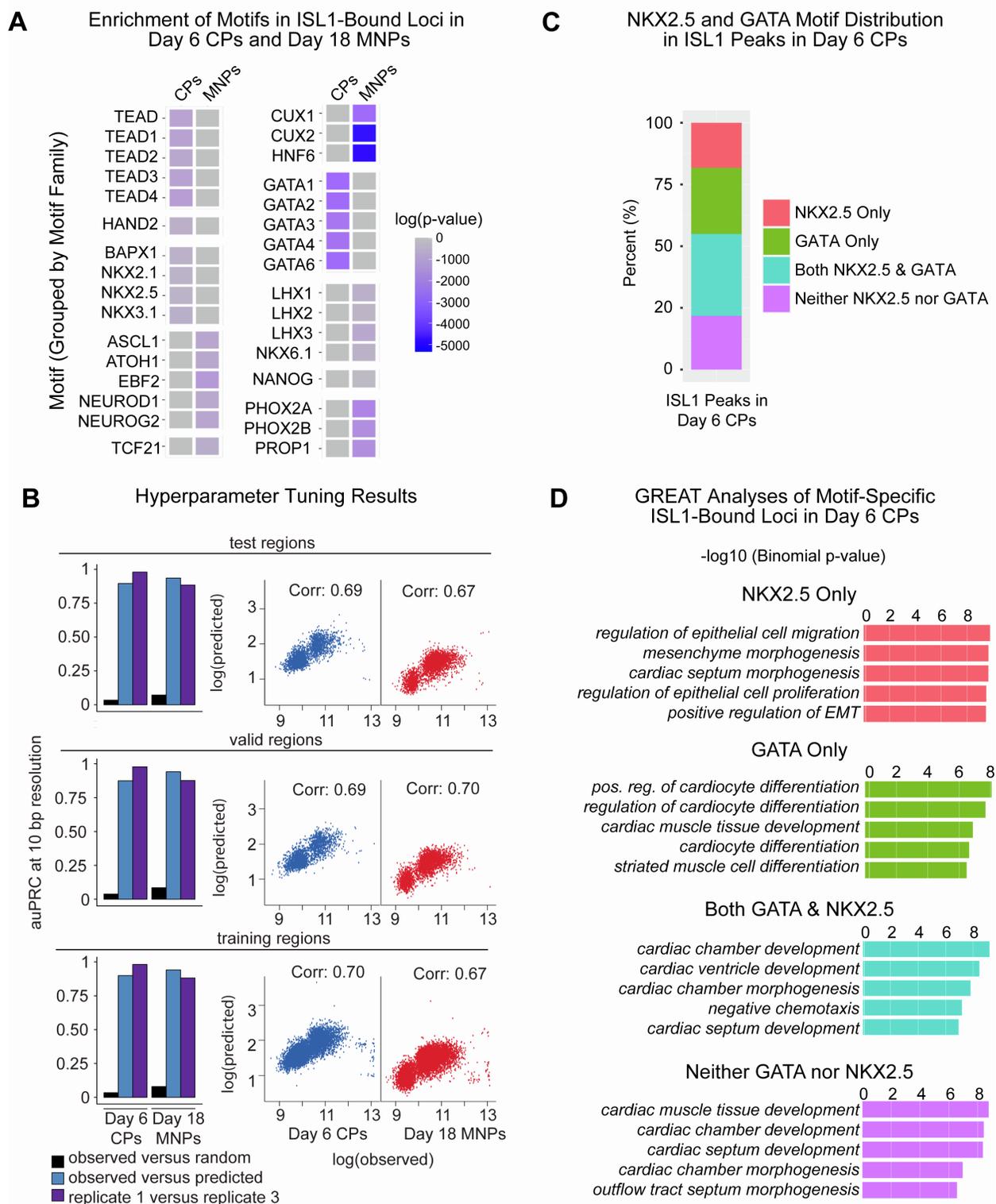
Figure S3. ISL1 ChIP-seq motif analyses in CPs and MNPs, related to Figure 3.

(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 0–3

Correlation 0–1

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

ISL1 in MNPs — 10284

NKX2.5 in CPs — 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

NKX2.5-bound only

Co-bound

Region-gene associations (%)

Distance to TSS (kb): <-500, -500 to -50, -50 to -5, -5 to 0, 0 to 5, 5 to 50, 50 to 500, >500

TSS

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

ISL1-Bound Loci Only

NKX2.5-Bound Loci Only

Co-Bound Loci

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

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⁻/⁻*

B NKX2.5 in Day 6 CPs

C ISL1 in Day 6 CPs

D Cardiac Factors in Day 15 CMs

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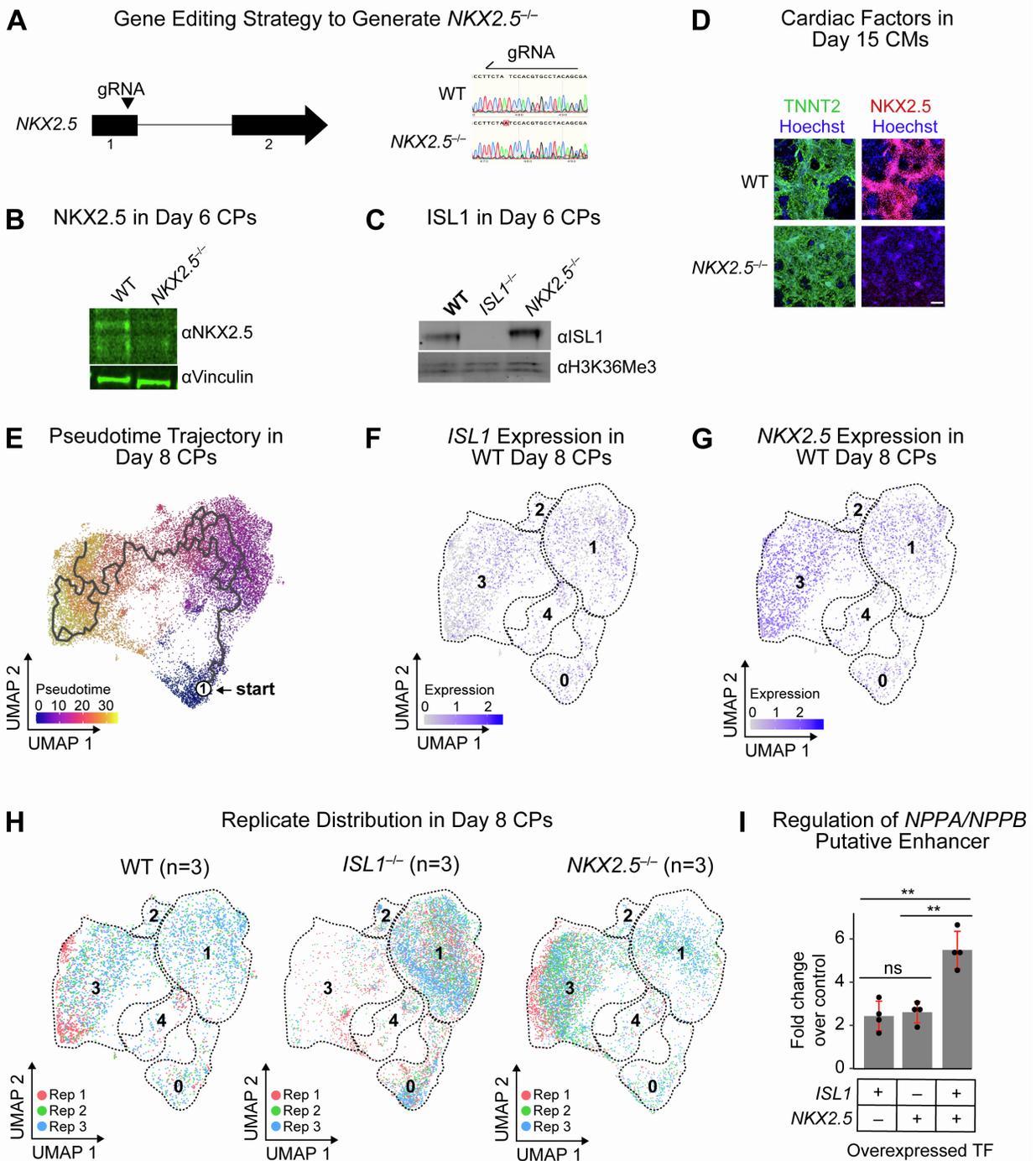
Figure S5. Additional analyses of the absence of *NKX2.5* in CPs, related to Figure 5.

(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⁻/⁻*
CP Motifs Mapped By BPNet
- GATA4/6
- NKX2.5 *** ***
- ISL1 **
decreased signal / no change / increased signal
ISL1 Response to *NKX2.5⁻/⁻*
log-odds ratio

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) = 1×10^{-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) = 1×10^{-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 (10×10^6 cells), MNPs (30×10^6 cells), or transfected COS-7 (3×10^6 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.