



**Karolinska  
Institutet**

Karolinska Institutet

<http://openarchive.ki.se>

---

This is a Peer Reviewed Published version of the following article, accepted for publication in *Stem Cells*.

2021-02-17

# Trajectory mapping of human embryonic stem cell cardiogenesis reveals lineage branch points and an ISL1 progenitor-derived cardiac fibroblast lineage

Mononen, Mimmi; Leng, Chuen Yan; Xu, Jiejia; Chien, Kenneth R

---

*Stem Cells*. 2020 Oct 1;38(10):1267-1278.

<http://doi.org/10.1002/stem.3236>

<http://hdl.handle.net/10616/47525>

*If not otherwise stated by the Publisher's Terms and conditions, the manuscript is deposited under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.*

## **Supplementary Materials and Methods**

### **Maintenance and cardiac differentiation of hESCs**

The cardiac differentiation protocol was adapted from Lian et al. [9]. In brief, ES03 hESCs were maintained in Essential 8 (E8) Flex Medium (Thermo Fisher Scientific, USA) on Matrigel (Corning, Germany) coated 6-well plates and passaged in Versene (Thermo Fisher Scientific). Cells were regularly tested for mycoplasma. Chromosome analysis of the ES03 cell line indicated normal 46, XX female karyotype (Fig. S1A). Upon differentiation on day -2, one confluent 6-well plate was split into one Matrigel coated 12-well plate, at a density of about 1 million cells per well, in E8 Flex with 5 $\mu$ M Y-27632 (Sigma, Germany). The medium was changed to E8 Flex without Y-27632 on the following day. On day 0 of differentiation, the media was replaced with 12  $\mu$ M CHIR99021 (Sigma) in RPMI media (Thermo Fisher Scientific) supplemented with B-27 minus insulin (Thermo Fisher Scientific). The medium was replaced with RPMI/B-27 minus insulin exactly 24h later. On day 3 of differentiation, combined medium was prepared by mixing medium from the 12-well plate with fresh RPMI/B-27 minus insulin in 1:1 ratio and adding 5  $\mu$ M IWP2. Medium was replaced with RPMI/B-27 minus insulin on day 5 of differentiation, and to RPMI supplemented with B-27 (Thermo Fisher Scientific) on day 7 and every three days thereafter.

### **Single cell library preparation and sequencing**

Cells undergoing cardiac differentiation were collected for single cell RNA sequencing at eight time points, day 3, 4, 5, 6, 7, 8, 9 and 15 from two batches of differentiation. One well per time point and differentiation batch was dissociated into single cells using Trypsin-EDTA (0.25%) (Thermo Fisher Scientific), and 64 cells were manually picked from each batch using a hand-pulled capillary needle. Single cells were transferred into 8-tube PCR strips containing Smart-Seq2 lysis buffer with 1:10 000 ERCC RNA spike-in mix (Ambion, ThermoFisher

Scientific), snap frozen on dry ice, and thereafter kept at -20 °C. Each single cell was processed into a cDNA library by following the Smart-Seq2 protocol [18, 19]. Fragment sizes were examined by Agilent 2100 Bioanalyzer (Agilent, USA) and the concentrations of the cDNA libraries were measured using Qubit (Thermo Fisher Scientific) according to the manufacturers' instructions. Barcoded single-cell Illumina libraries were prepared using Illumina Nextera XT DNA Sample preparation kit (Illumina, USA) with dual indexes. The libraries were pooled into batches of 128 cells and sequencing was performed on Illumina HiSeq 2500 for 50 bp single-end reads.

### **Single cell RNA sequencing data analysis**

Raw sequencing reads were de-multiplexed using bcl2fastq (Illumina) and mapped to the human hg38 genome using TopHat. In total 1 792 transcriptomes were sequenced. Data from two hESC cardiac differentiation batches was combined to create one single-cell dataset with Seurat R package. Filtering, normalization and subsequent analysis including calculating principal components and t-SNE and UMAP dimensionality reduction were performed as described in the Seurat 3.0 R package tutorial (<http://satijalab.org/seurat/>). Briefly, cells with low levels of detected transcripts or expressing less than 200 genes, as well as genes that are expressed in less than 3 cells, were excluded from downstream analysis. We proceeded with the Seurat 3.0 package in R [21-23] for downstream analysis. To identify variable genes, the average expression level and dispersion for each gene were calculated, and the genes were divided into 20 bins based on their average expression and Z-scores were calculated for each bin. Following the Seurat 3.0 pipeline, we selected the 2000 genes with the highest standardized variance for downstream analyses.

The RunTSNE function in Seurat was used to compute a t-SNE dimensional reduction and clustering. The FindNeighbors considering nine first principal components and FindClusters function with resolution 0.5 in the Seurat package were used for clustering.

### **Identification of marker genes**

The FindMarkers function in Seurat was used to identify positive marker genes for each cluster, and the gene list was filtered for genes having a significant adjusted p value ( $< 0.05$ ). Differential gene expression analysis between wild-type and ISL1 knockout cells was also performed using FindMarkers function.

### **Functional enrichment analysis**

50 top ranked marker genes for each cluster were selected for functional enrichment analyses. Gene ontology (GO) analysis was performed using STRING version 11.0 statistical overrepresentation test against *homo sapiens* genome [24]. PANTHER 14.1 statistical overrepresentation test against *homo sapiens* genome applying Fisher's exact test and calculation of Benjamin-Hochberg false discovery rate [35] was used to find GO terms for the integrated dataset, and PANTHER Pathways [36] were called for specific populations as indicated in the text.

### **Monocle pseudotime analysis**

Monocle 2 package was used for pseudotime analysis of the cells belonging to the pluripotent and myogenic populations (clusters 0-5) following the online tutorial (<http://cole-trapnell-lab.github.io/monocle-release/docs/>) [25-27]. In addition, pseudotime values were determined for all hESC-derived cells, and the values were imported to the Seurat object and visualized as t-SNE, FDG and UMAP-based feature plots (Fig. S4B).



### **Force-directed graphing**

We performed force-directed graphing using the igraph R package, and the nearest neighbour matrix was calculated using the method described by Kee et al. [20]. Briefly, we first performed t-SNE analysis on the entire dataset using the Seurat 3.0 package, supplemented by graphing analysis developed by Kee et al. to map out differentiation axes in pseudotime. We found that calculating the 14 nearest neighbors on the t-SNE produced a visually useful graph drawing (Fig. 3A). Calculating a range of number of nearest neighbors produced similar results (Fig. S5A-C).

### **Cell cycle analysis**

In order to study the cell cycle heterogeneity between cell populations, we sought to analyze the overall cell cycle-related gene expression. Each cell was assigned a score for cell cycle phases S and G2/M using the CellCycleScoring function in Seurat and a published list of cell cycle genes, following the cell cycle regression tutorial ([https://satijalab.org/seurat/v3.1/cell\\_cycle\\_vignette.html](https://satijalab.org/seurat/v3.1/cell_cycle_vignette.html)). The cells with positive scores for G2/M phases were considered as being in a proliferative state, whereas cells with negative cell cycle scores were likely not cycling. We visualized the cell cycle scores using UMAP dimensional reduction indicating differences between and within different clusters. We also ran an alternative UMAP analysis by regressing out cell cycle scores during data scaling using S.Score and G2M.score as input for vars.to.regress argument in ScaleData. We used WhichCells function in Seurat to calculate the percentage of cells with G2M.Score greater than 0 in each cluster to estimate the percentage of proliferative cells in each cluster.

## **RNA velocity analysis**

RNA velocity was performed according to the Velocyto tutorial (<http://velocyto.org>) [17]. In brief, a .loom file with counts divided in spliced/unspliced/ambiguous was produced in Python using the human hg38 genome annotation file (.gtf) and an expressed repeat annotation file from the UCSC genome browser. Subsequent RNA velocity analysis was performed in R according to the tutorial (<http://pklab.med.harvard.edu/velocyto/notebooks/R/chromaffin2.nb.html>). After filtering out reads with very low counts, we detected 8 911 genes that had both spliced and unspliced reads, of which 5 027 also had reads that spanned intron-exon junctions. Due to the relatively high number of genes detected with spanning reads, we used the spanning reads to produce an estimate of RNA velocity. We also compared several variants (with or without nearest cell pooling, and excluding spanning reads) of velocity estimates provided in Le Manno et al, and found very similar results. R scripts are available upon request.

## **Integration of single cell datasets**

The RPKM (Reads Per Kilobase Million) read counts of the hESC-derived single cell dataset were converted to TPM (Transcripts Per Kilobase Million) to match with the dataset generated by Cui et al. The human fetal dataset had lower average number of detected genes per cell compared to the hESC-differentiation dataset (Fig. S2A).

We proceeded with Seurat version 3.0 workflow for integration and data transfer [21, 23], which essentially “anchors” diverse datasets together, enabling integration of datasets across scRNA-seq technologies and modalities. The workflow applies a correction vector for each individual cell in the query dataset, which is determined by a cell similarity score and an anchor score. For detailed workflow please see

(<https://satijalab.org/seurat/v3.1/integration.html>).

### **Immunocytochemistry**

For imaging with confocal microscope, ES03 cells were seeded on coverslips on 12-well plates for differentiation. Cells were fixed with 4% paraformaldehyde for 20 minutes, permeabilized with 0.5% Triton X-100 for 20 minutes and blocked with 3% BSA in PBS for 30 minutes at room temperature. Coverslips were incubated with primary antibodies (with simultaneous incubation of antibodies) in blocking buffer overnight at 4°C, followed by secondary antibody incubation for one hour at RT and nuclear counterstaining with DAPI. Coverslips were washed three times with PBST for 5 minutes between each step.

### **Human fetal heart tissues**

Human fetal hearts at 6.5 and 10 weeks of the gestation stages were obtained from authorized sources in Karolinska University Hospital with an approved ethical permission (Dnr 2015/1369-31/2) and appropriate informed consents. The fetal hearts were snap-frozen and cryosectioned at 10- $\mu$ m thickness for immunostaining.

### **Immunofluorescence**

Cardiac tissue sections were fixed with 4% paraformaldehyde, permeabilized in 0.5% Triton X-100, blocked (3% BSA, 3% normal donkey serum in PBS) and incubated with primary antibodies in blocking buffer overnight at 4°C. Sections were incubated with secondary antibodies in blocking buffer for one hour at RT and counterstained with DAPI.

Primary antibodies			Dilution
anti-HHEX	Atlas Antibodies	HPA051894	1:500
anti-ISL1	DSHB	39.4D5	1:100

anti-NKX2-5	Santa Cruz	sc-14033	1:1000
anti-TFAP2A	Abcam	ab108311	1:500
anti-TGFBI	Abcam	ab66957	1:100
anti-TNNT2	ThermoFisher	MS-295-P1	1:500
anti-Vimentin	Millipore	AB5733	1:2000

### Genetic manipulation of hESCs

The genetic manipulation of hESCs was carried out as described in Xu et al. 2019. Briefly, two gRNAs targeting exon 4 of *ISL1*, with sequences CGATGTGATACACCTTGGAG and GGCGGCTGCAGGCCACACAG, were cloned into PB-CRISPR vector. Two million hESCs, 1 $\mu$ g pCyL43 and 4 $\mu$ g pPB-CRISPR-ISL1-gRNA DNA were mixed in 100 $\mu$ l nucleofection solution and nucleofection was performed with program B-16 of Nucleofector 2b device (Lonza). After two weeks of 0.5  $\mu$ g/mL Puromycin selection, cells were singularized and sparsely seeded to generate single cell derived clones. Genomic DNA was extracted from putative *ISL1* mutated hESC clones using Quick-gRNA MiniPrep kit (Zymo Research, D3007). The targeted genome region was amplified using forward primer TCTGTGGTACAAGCTAGGTGTT and reverse primer ACACTACGACCACCCCTGTA. The Amplified PCR fragment was inserted into pCR4 TOPO vector via the TOPO-TA method (ThermoFisher Scientific). Plasmids from individual *E. coli* clones were Sanger sequenced using the T7 primer. One hESC clone with induced stop codons within the targeted exon was chosen for further analysis. *ISL1* deletion was validated by quantitative PCR and western blot and immunostaining (Fig. S6B).

## Quantitative PCR

Total RNA was purified using ZYMOGEN Direct-zol RNA MiniPrep kit using DNase I treatment to remove any genomic DNA contamination. 500 ng of total RNA was used for reverse transcription with SuperScript IV First- Strand Synthesis System (Invitrogen). Quantitative PCR was performed using the reagents and workflow as recommended by PowerUp SYBR Green (Applied Biosystems) guidelines. Gene expression levels were normalized to *GAPDH*. Results are represented as fold change compared to the wild-type  $\pm$  SEM, and are averaged from four biological repeats, with three technical repeats each.

qPCR primers		
Target	Forward	Reverse
<i>APOA2</i>	GCAGCAACTGTGCTACTCCT	TCTCCGAACCAAAGCTCCTTC
<i>COL3A1</i>	GTTGCACGAAACACACTGGG	ACAGCCTTGCGTGTTTCGATA
<i>COL1A1</i>	CCCCGAGGCTCTGAAGGTC	GGAGCACCATTGGCACCTTT
<i>GAPDH</i>	CGACAGTCAGCCGCATCTT	CCCAATACGACCAAATCCGTTG
<i>HHEX</i>	CTCAGCGAGAGACAGGTCAA	TCTCCTCCATTTAGCGCGTC
<i>HNF4A</i>	TGACGATGGGCAATGACACG	CTCGAGGCACCGTAGTGTTT
<i>ISL1</i>	TCAGTATTTGGACGAGAGCTG	CATTTGATCCCGTACAACCTGA
<i>MYL2</i>	AGGCTCCGGGTCCAATTAAC	AGGGTCCGCTCCCTTAAGTT
<i>MYL7</i>	AAGCCATCCTGAGTGCCTTC	AACATCTGCTCCACCTCAGC
<i>NKX2-5</i>	CAAGTGTGCGTCTGCCTTTC	CACAGCTCTTTCTTTTCGGCTC

# Figure S1

© 2020 The Authors. This is article is distributed under the terms of the Creative Commons Attribution Non-Commercial License <http://creativecommons.org/licenses/by-nc/4.0/>, which permits non-commercial reuse, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, whether in whole or in part, see Oxford University Press' Rights and Permissions.

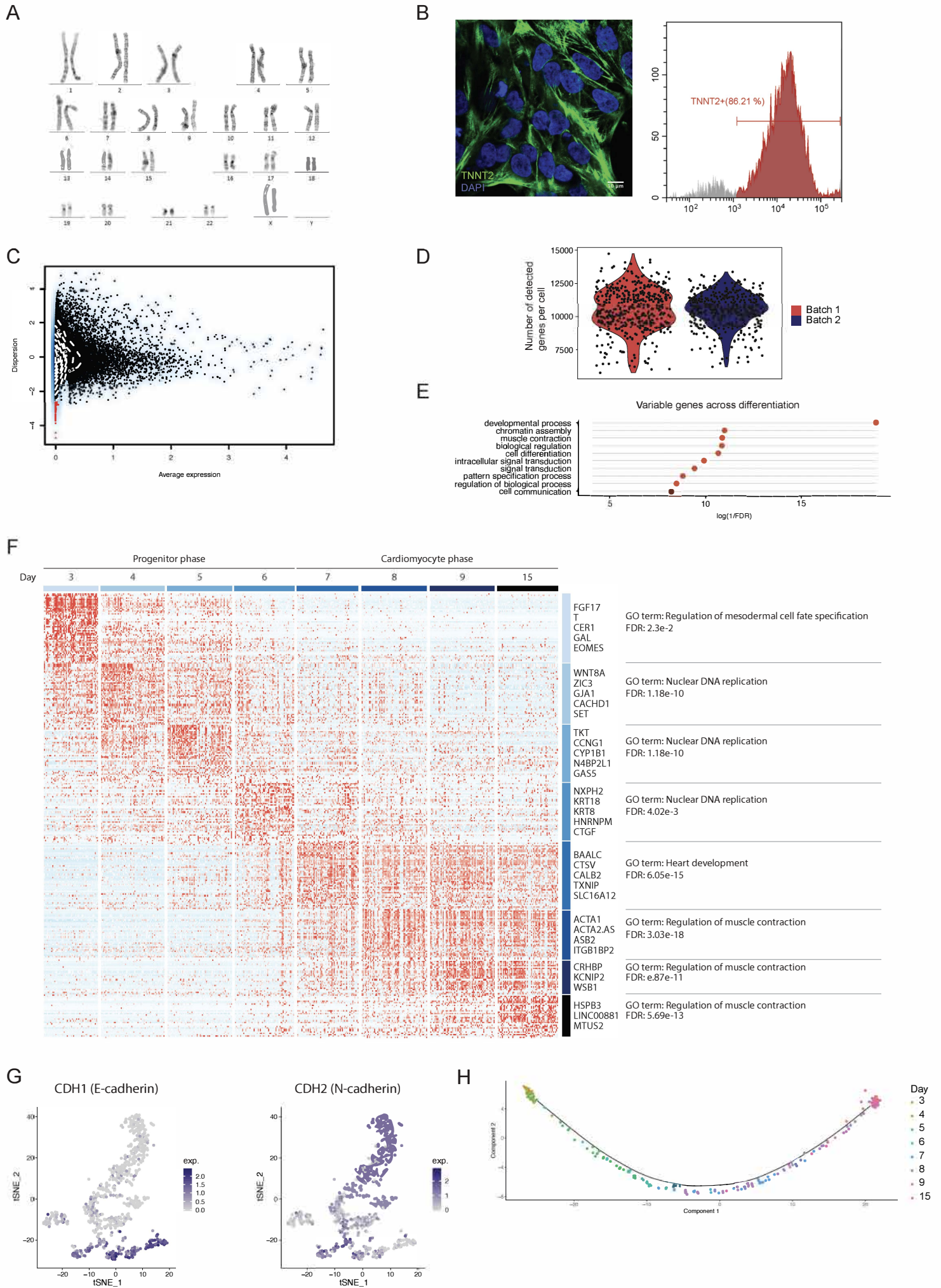


Figure S2

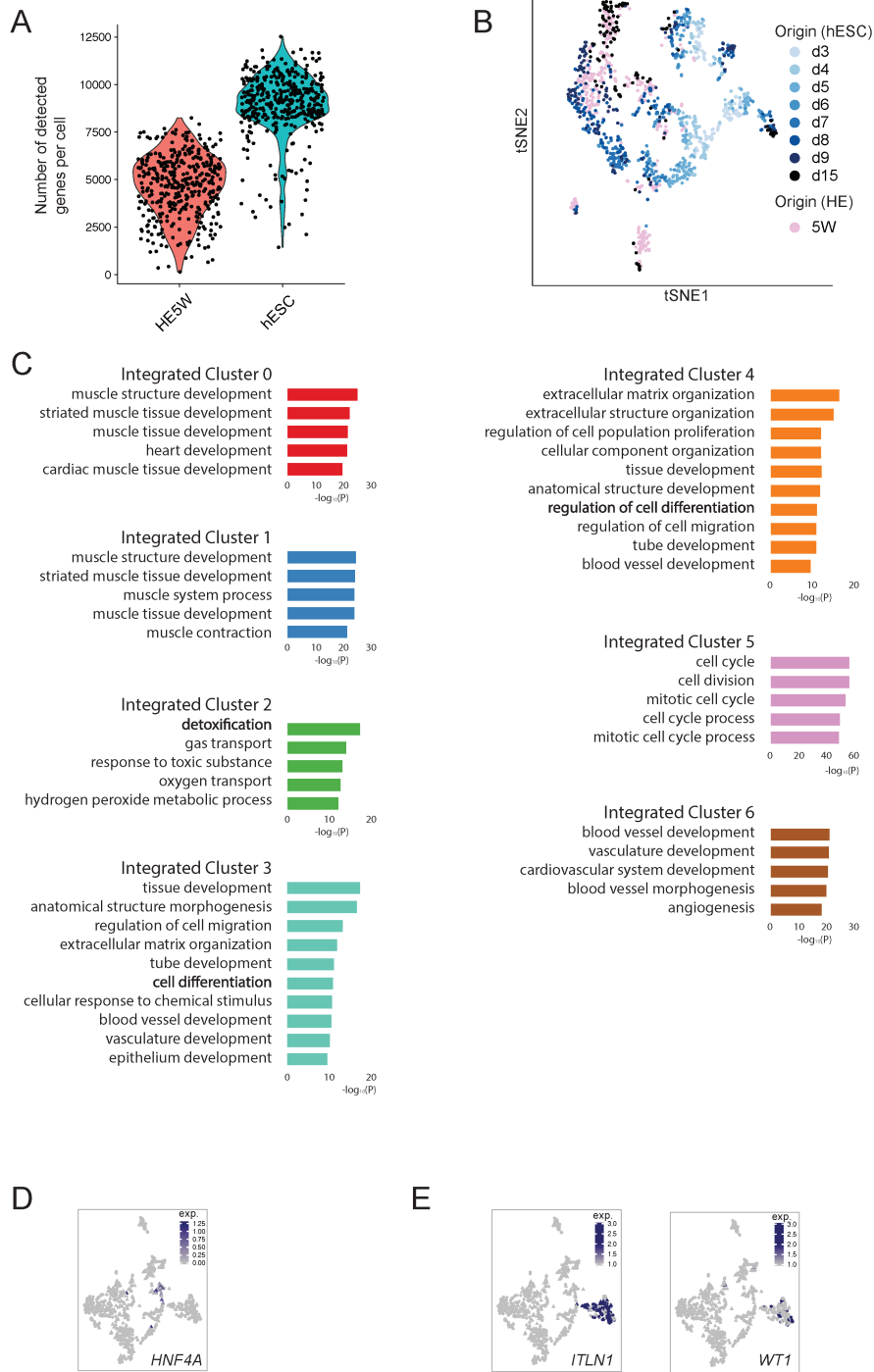
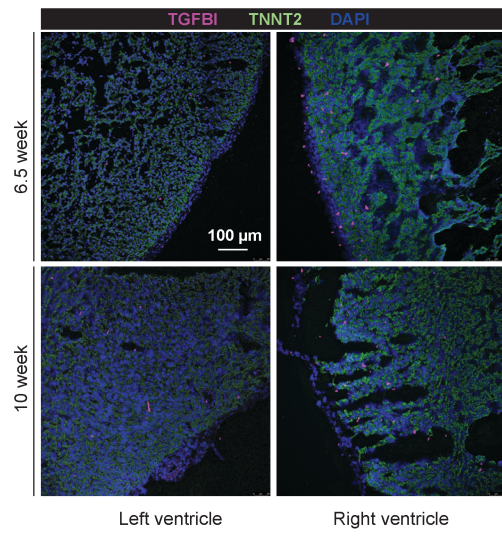




Figure S3

A



B

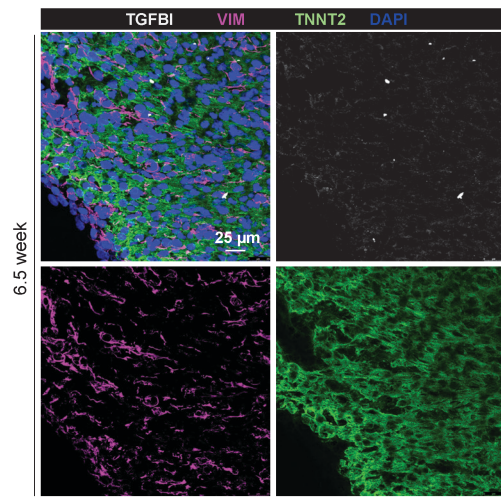




Figure S4

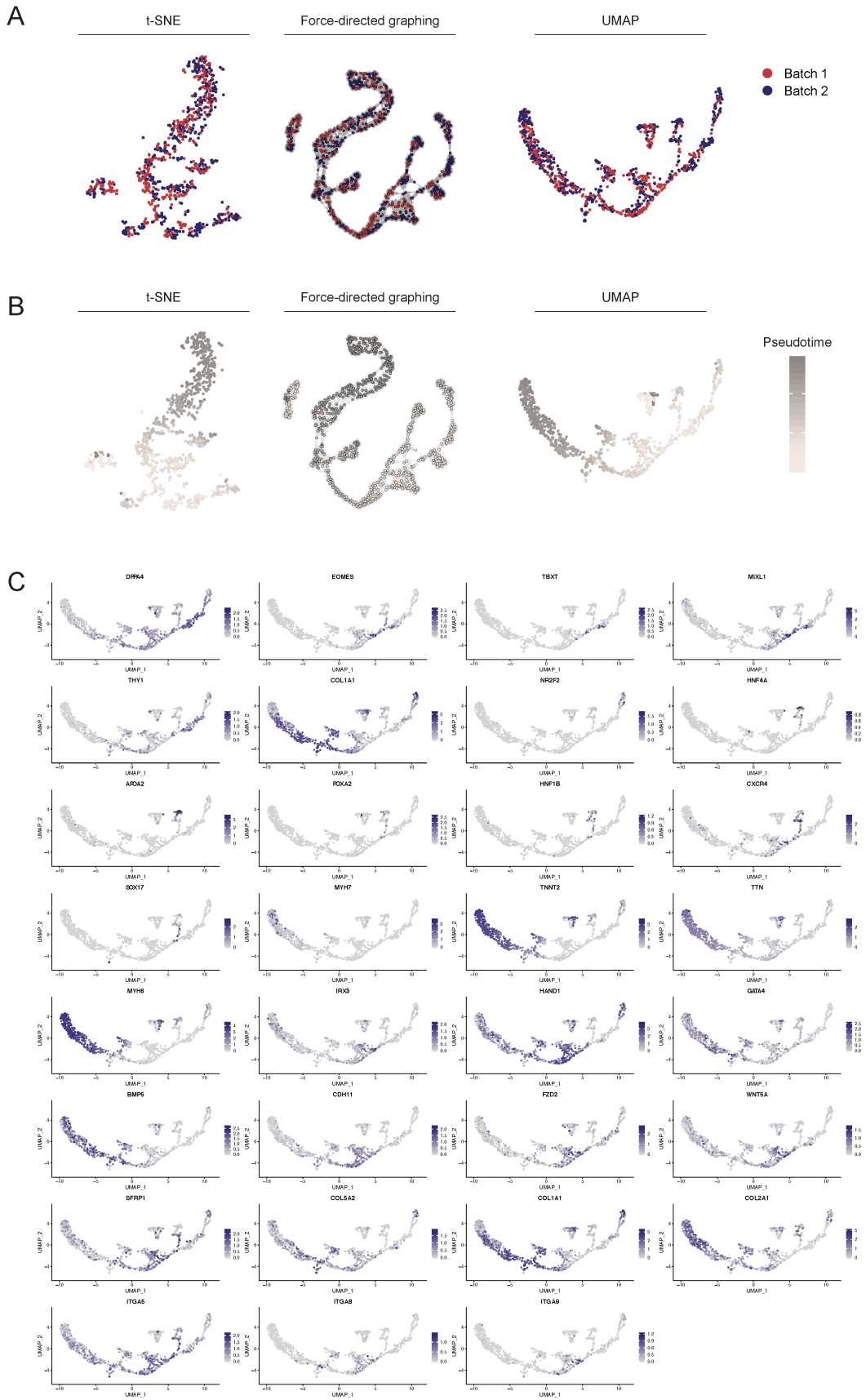


Figure S5

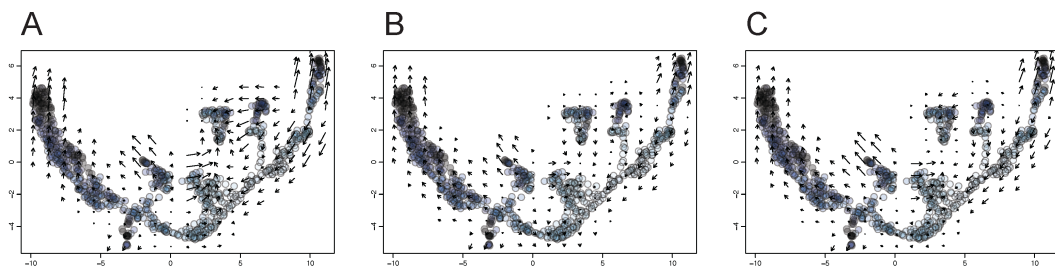
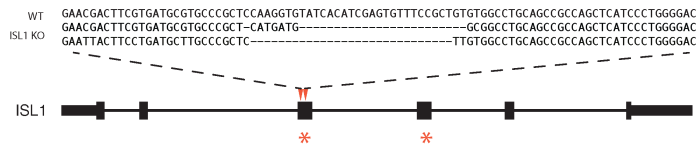
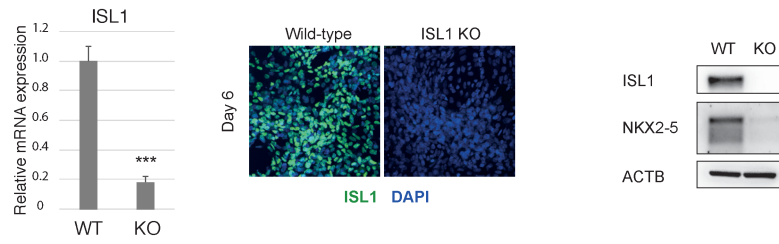


Figure S6

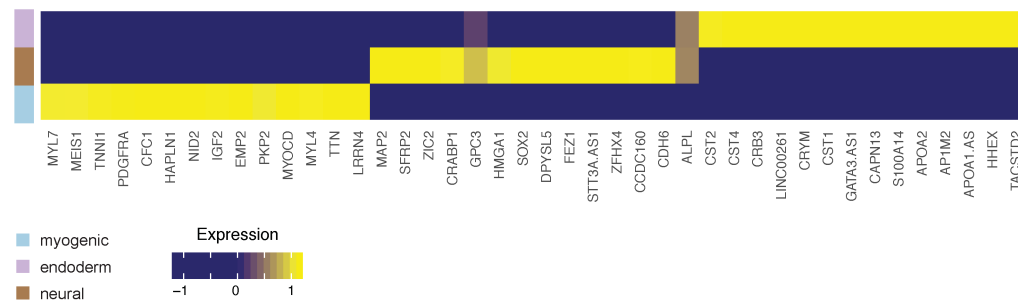
A



B



C



D

