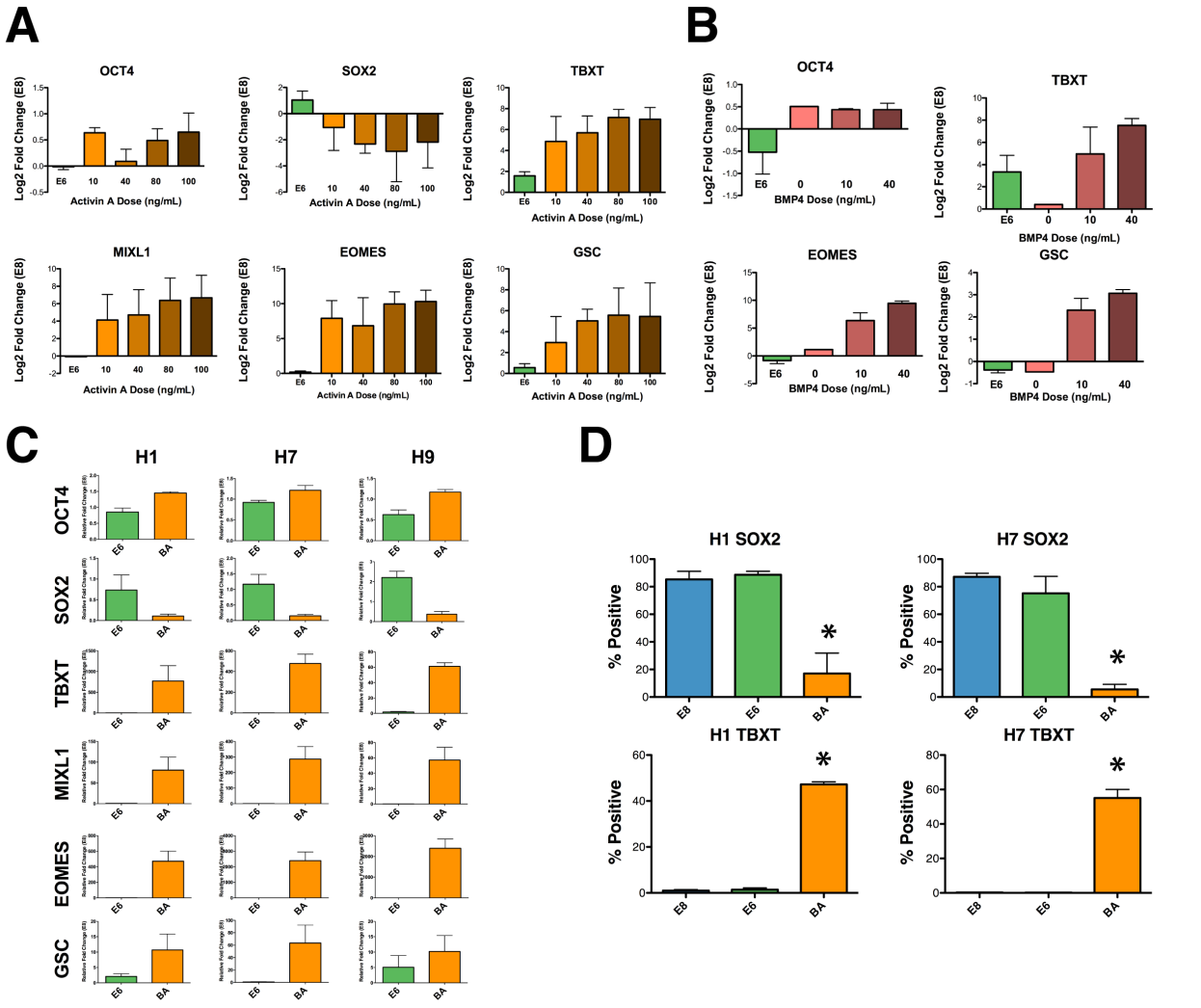
Stem Cell Reports, Volume *13*

Supplemental Information

Transcriptomically Guided Mesendoderm Induction of Human Pluripotent Stem Cells Using a Systematically Defined Culture Scheme

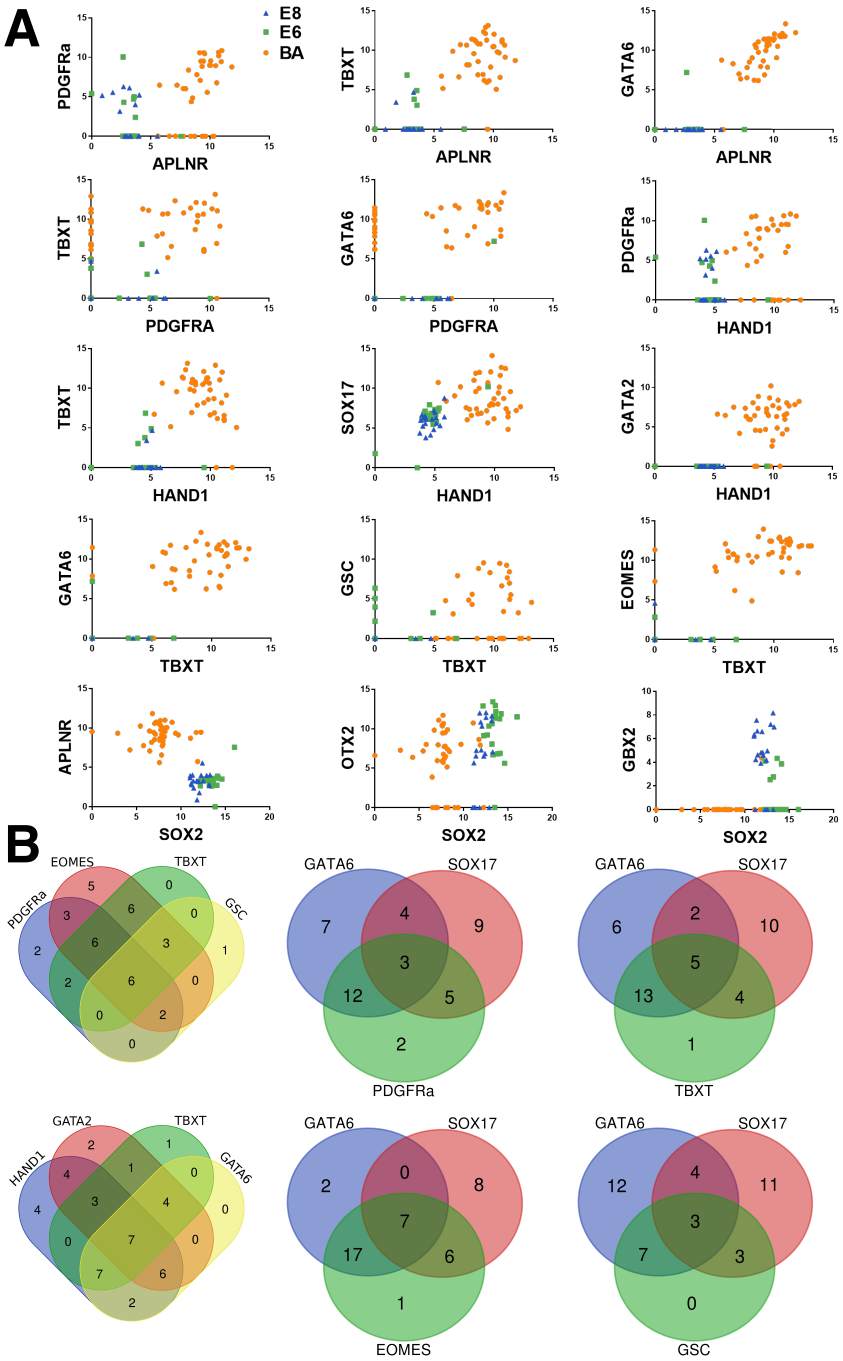
Richard L. Carpenedo, Sarah Y. Kwon, R. Matthew Tanner, Julien Yockell-Lelièvre, Chandarong Choey, Carole Doré, Mirabelle Ho, Duncan J. Stewart, Theodore J. Perkins, and William L. Stanford

Figure S1. BMP4 and Activin A dose-response and mesendoderm differentiation of H1 and H7 hESCs. Related to Figure 1.



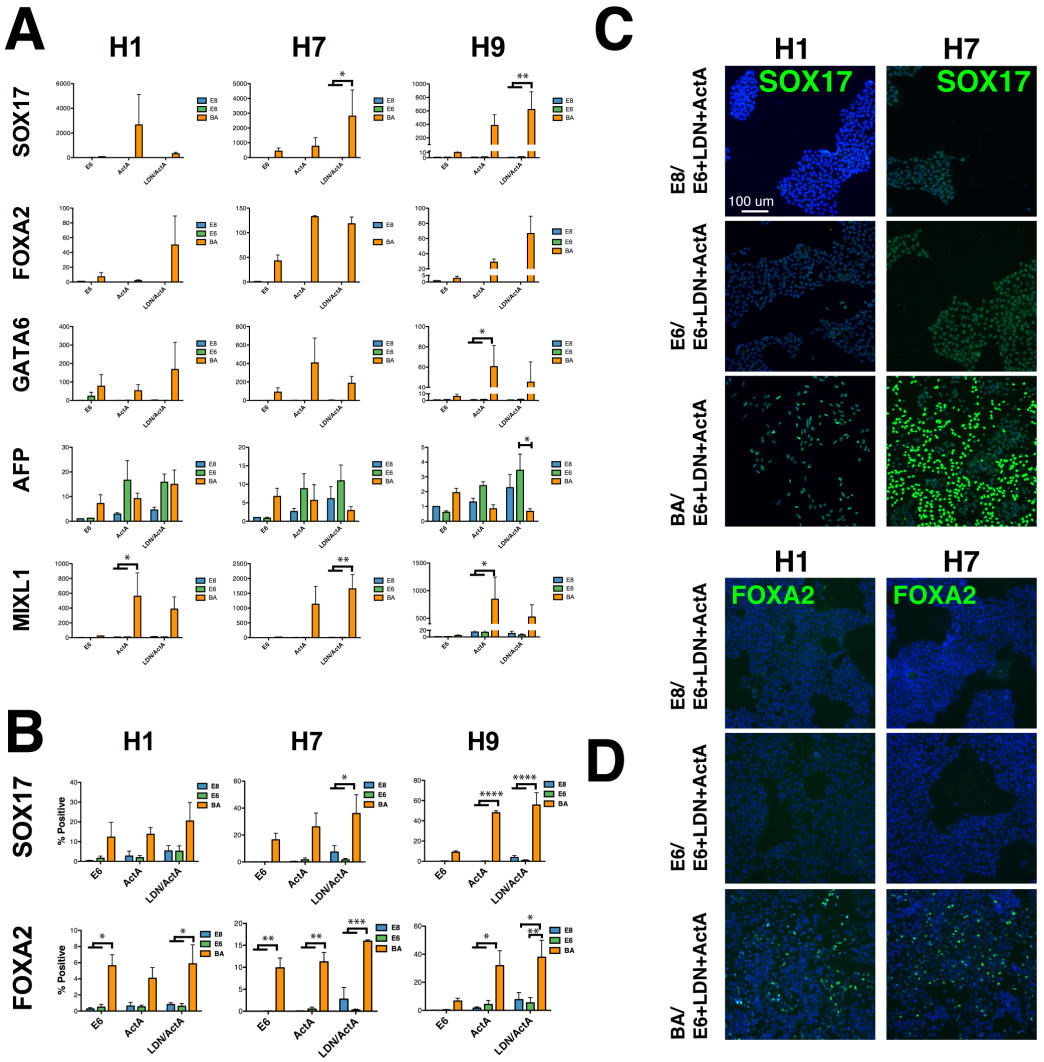
(A). Activin A dose was varied from 10-100 ng/mL with a constant BMP4 dose of 40 ng/mL. No significant difference in mesendoderm gene expression was observed at increasing Activin A dosage. (B). BMP4 dose was varied from 0-40 ng/mL with Activin A concentration of 40 ng/mL. (C). Single-cell seeding, BA treatment was performed for H1 and H7 hESCs as well as H9s, demonstrating similar trends in gene expression for all three lines. (D) Analysis of protein expression by immunofluorescence and high-content imaging likewise demonstrated similar loss of SOX2 and up-regulation of TBXT among all hESCs examined. Error bars represent SEM, n=3, *=$p<0.05$.

Figure S2. Co-expression of genes in 48-hour BA cells by single-cell gene expression analysis. Related to Figure 4.



(A). Relative expression values of mesoderm and endoderm genes were plotted to assess co-expression in individual cells in BA (orange), E6 (green) and E8 (blue) cells, demonstrating the co-expression of many of these in the BA cells. Co-expression of *SOX2* with *OTX2* or *GBX2* separates E6 and E8 cells from BA cells (bottom row). (B) Venn diagrams depicting the overlap in cells with positive expression of mesoderm and endoderm genes in the BA population. Notable overlap was observed for expression of all genes examined, and expression of endoderm genes (*SOX17, GATA6*) was not able to identify a unique population that did not also express mesoderm genes.

Figure S3. Definitive endoderm differentiation of H1 and H7 hESC lines. Related to Figure 5.



(A). Gene expression analysis in response to definitive endoderm conditions for cells pre-differentiated in E8 (blue bars), E6 (green bars) or BA (orange bars). Similar trends were observed in H1and H7 hESCs compared to H9s. (B). Analysis of SOX17 and FOXA2 protein expression, based on immunofluorescence, with representative images in panel (C). Expression of SOX17 and FOXA2 was highest in BA pre-differentiated cells in Activin A/LDN secondary media. Errors bars represent SEM, n=3, *=p<0.05, **=p<0.01, ***=p<0.001, ****=p<0.0001. Scale bar represents 100 µm.

Table S2. List of probes used for single-cell qPCR analysis. Related to Figure 4.

ACTB	CHAT	FGF5	GBX2	ITGB4	NANOG	POU4F2	SLC2A2
ALB	COL10A1	FOXA1	GDF3	KRT10	NESTIN	POU5F1	SLC32A1
APLNR	COMP	FOXD3	GFAP	KRT14	NEUROD1	PROM1	SMTN
APOH	CPA1	FOXG1	GSC	KRT19	NEUROG2	PTCRA	SOX17
AQP1	CTSK	G6PC	HAND1	LEFTY1	NKX2-2	RCVRN	SOX2
B2M	DCN	GAD1	HAND2	MAP3K12	NKX2-5	RPLP0	SOX7
BMP4	DCX	GAD2	HES5	MIOX	NPPA	RUNX1	T
CCR5	DNMT3B	GALC	HNF4A	MIXL1	OLIG2	RYR2	TAT
CD34	DPP4	GAPDH	HPRT1	MSLN	OTX2	SFTPB	TUBB3
CD3E	ENO1	GATA1	IBSP	MYH1	PAX6	SFTPD	TYR
CD79A	EOMES	GATA2	IGF2	MYH7	PDGFRA	SLC17A6	ZFP42
CER1	FABP7	GATA6	INS	MYL3	PODXL	SLC17A7	ZIC1

Table S3. Primer sequences used in qPCR gene expression analysis. Related to Figures 1, 5 and 6.

Gene	Forward Sequence	Reverse Sequence
OCT4	TCAGCCAAACGACCATCTGCCG	AGCAAGGGCCGCAGCTTACA
SOX2	TACAGCATGTCCTACTCGCAG	GAGGAAGAGGTAACCACAGGG
Nanog	ACGCAGAAGGCCTCAGCACCTA	AGGTTCCCAGTCGGGTTCACCA
T	ACCTGTGTCGCCACCTTCCA	ACCACTGGCTGCCACGACAA
MIXL1	TCCTCAACCACTGTGCTCCTGG	AACCCCGTTTGGTTCGGGCA
EOMES	AGGCGCAAATAACAACAACACC	ATTCAAGTCCTCCACGCCATC
GSC	CGCGGGACACTTGCCCGTATTA	AAGGCAGCGCGTGTGCAAGA
PAX6	CCAGAAAGGATGCCTCATAAA	TCTGCGCGCCCCTAGTTA
Nestin	CCGCATCCCGTCAGCTGGAAAA	GCTTGGGCACAAAAGCCAGCA
OTX2	CTTAAGCAACCGCCTTACGC	AGGGGTGCAGCAAGTCCATA
AFP	AGCTGACCTCGTCGGAGCTGAT	TCCCTCGCCACAGGCCAATAGT
KDR	ACCGTTAAGCGGGCCAATGGA	ACCACGGCCAAGAGGCTTACCT
SOX17	CGCACGGAATTTGAACAGTA	GGATCAGGGACCTGTCACAC
FOXA2	GGGAGCGGTGAAGATGGA	TCATGTTGCTCACGGAGGAGTA
GATA6	AGGCTGCAGTTTTCCGGCAGAG	CGCCGCGCTGCTGGTGAATAAA
GAPDH	TTCTTTTGCGTCGCCAGCCG	TGACCAGGCGCCCAATACGA
EF1a	GCTGGCTTCACTGCTCAGGTGATT	TGCAATGTGAGCCGTGTGGCA

Supplemental Experimental Procedures

Spatial Analysis

Spatial analysis for cells seeded as colonies at different split ratios or single-cells at different densities was performed using the SpatStat package for R (Baddeley and Turner, 2005; Baddeley et al., 2015). The spatial coordinates of each cell within a field (i.e. each microscopic image) were converted into a spatial point pattern, and each field was divided into 5x5 grids (totaling 25 quadrats). The number of points (or cells) per quadrat was quantified using the 'quadratcount' feature in Spatstat for each quadrat in each image acquired for a given well. The Coefficient of Variation (CV; standard deviation divided by mean) was then calculated for the number of cells per quadrat for each seeding condition. For each well, the total number of cells was quantified, and normalization of CV values was performed by calculating the CV of a simulated random uniform distribution of points (equal to the cell number in that well) using the 'runifpoint' function in Spatstat.

Bioinformatics

FASTQ sequencing data for each sample was aligned to GENCODE version 23 human genome annotations (hg38), using the HISAT2 alignment tool (version 2.0.1) (Kim et al., 2015). The 'featureCounts' function from the Rsubread package (version 1.22.3) (Liao et al., 2019)was used to count reads per gene, which were passed onto DESeq2 (version 1.12.4) (Love et al., 2014)for library size normalization and detection of differentially expressed genes (FDR <= 0.05). Hierarchical clustering was done with the R 'heatmap.2' function from the gplots package (version 2.17.0) for differentially expressed genes. Principal component analysis was done using an in-house R script and the built-in R PCA function 'prcomp'. Temporal expression path clustering was generated by unsupervised hierarchical clustering a log2 fold (log2FC) change matrix using R's built-in 'hclust' function. This was done for all genes with an absolute log2FC of at least 2 in any single timepoint. Clusters which were visibly similar were then manually combined, and cluster trajectories were then plotted using an in-house R script. Gene set enrichment analyses (Subramanian et al., 2005) were conducted by comparing differentially expressed genes from RNA-seq data to custom Gene Sets produced in-house. The -log10(p-value) for each gene from RNA-seq data was used as a custom weighting when GSEA analysis was conducted (version 2.1.0). Gene ontology analysis was performed using DAVID version 6.8 (Huang et al., 2009) and the BINGO plugin for Cytoscape (Maere et al., 2005).

Single-cell Gene Expression Analysis

Cycle numbers to enrichment for each gene in each cell were normalized to housekeeping genes between cells. Normalized cycle numbers were converted to a cycle difference measurement, subtracting observed cycle number from the mean cycle number of the corresponding gene across all cells. An aggregate vector of all cycle differences was used to produce Z-scores for each gene in each cell; these Z-scores were then used to plot heatmaps, using the R 'heatmap.2' function from the gplots package (version 2.17.0). t-SNE plots were also produced from cycle differences using the Rtsne package (version 0.10) and an in-house script for plotting the results. Tissue and cell type enrichment for single cell expression data was done using Enrichr (Chen et al., 2013; Kuleshov et al., 2016).

Chondrocyte Micromass Differentiation and Analysis

Following 48-hour pre-differentiation in E8, E6 or BA, cells (H9 hESCs or iPSCs) were dissociated using TrypLE Express and resuspended in chondrogenic media supplemented with 10 μM Y-27632. Chondrogenic media consisted of high glucose DMEM (Life Technologies) containing 1% KOSR, 1% ITS+ premix (BD Biosciences), 1% Sodium Pyruvate, 1% non-essential amino acids, 1% Penicillin-Streptomycin (Life Technologies), 100 μg/mL ascorbic acid 2-phosphate, 10^{-7} M dexamethasone, and 40 μg/mL L-proline (Sigma). Cells were resuspended at a density of $2x10^7$ cells/mL and plated as 15 μL drops in Matrigel-coated 12-well plates for 2 hours, after which 1 mL of chondrogenic media containing Y-27632 was added. Fresh chondrogenic media was added daily for course of the 7-day protocol.

To assess differentiation by matrix production, 4-5 micromass cultures were pooled and digested overnight at 65°C using 40 μg/mL papain enzyme in digestion buffer and processed for quantification as described (Lee et al., 2011). Sulfated glycosaminoglycan (s-GAG) content was quantified using Dimethylmethylene blue (DMMB). Hydroxyproline (OH-Pro) content was quantified by acid hydrolysis of papain-digested samples followed by neutralization, oxidation, and addition of 4-dimethylaminobenaldehyde. Hoechst 33258 dye was used to quantify DNA content in the samples, and total s-GAG or OH-Pro content were normalized to amount of DNA. Proteoglycan production was assessed after 7 days by Alcian Blue staining. Micromass cultures were fixed in 4% PFA for 15 min, rinsed with 0.2N HCl, and stained with 0.1% Alcian Blue solution at pH1 (diluted with 0.2N HCl from 0.3% w/v Alcian blue in 70% ethanol solution; Sigma) overnight at room temperature. Cultures were rinsed thoroughly with distilled water and imaged using a Zeiss Axio Zoom V16.

Endothelial Progenitor Cell Differentiation and Analysis

Differentiation of iPSCs derived from late-EPCs (described in Chang et al., 2013) was performed using an adapted protocol from Tatsumi *et al* (Tatsumi et al., 2011). Briefly, cells were pretreated for 48 hours with E6 supplemented with 10 ng/ml of both BMP4 and Activin A. Subsequently, cells were cultured in DMEM/F12 (Life Technologies) supplemented with B27 (Life Technologies), N2 (Life Technologies) and BIO (Sigma) for 72 hours with daily media change performed. Cells were then cultured in StemPro-34 (Life Technologies) supplemented with 50 ng/ml of $VEGF_{165}$ (R&D Systems) for a further 48 hours prior to MAC-selection with CD144

microbeads (Miltenyi Biotech). CD144 enriched cells were then further expanded and routinely passaged in StemPro-34 media supplemented with 50 µg/ml $VEGF_{165}$.

Endothelial cell differentiation was assessed by flow cytometry against a panel of endothelial cell surface markers. Briefly, following six days of differentiation, cells were harvested and dissociated into single cell suspension using TrypLE. A total of 2 $x10^5$ cells in 200 µl of flow buffer was enumerated and aliquotted into each tube. Directly conjugated antibodies (all from BD Bioscience) to CD31 (Cat. No. 340297), CD34 (Cat. No. 345802), CD45 (Cat. No. 555482), CD144 (Cat. No. 560411) and VEGFR (Cat. No. 560494) were then added at the recommended manufacturer's dilution into each tube (i.e. 20 µl per test) and incubated on ice in the dark for 30 mins. To halt the staining process, 400 µl of flow buffer was added per tube and the entire volume of 600 µl was subjected to centrifugation at 400rpm for 5 mins. Supernatant was decanted and cells resuspended in 300 µl flow buffer prior to analysis on an Attune acoustic focus cytometer (Life Technologies). To gate samples for FACS analysis, cells (20,000 events) were initially gated by FSC-A vs SCS for the exclusion of debris. Data were analyzed using FlowJo v10.0.6 software.

Definitive Endoderm Differentiation and Analysis

Pre-differentiation of H1, H7 or H9 hESCs was performed by SCS in E8, E6 or BA as described in the main text. For endoderm induction, BMP4 and Activin A concentration were adjusted to 10 and 60 ng/mL, respectively. After 48 hours in pre-differentiation conditions (E8, E6, BA), media was replaced with E6 alone, E6 with high Activin A concentration (100 ng/mL) or E6 with Activin A (100 ng/mL) and LDN193189 (250 nM). Cells were allowed to differentiate for an additional 48 hours, with media refreshed daily. Analysis of definitive endoderm induction was performed by immunostaining and high-content imaging of SOX17 and FOXA2, as well as gene expression of *SOX17, FOXA2, AFP*, and *GATA6*.

Neural Induction

Single-cell seeding and 48-hour differentiation in E8, E6 or BA was performed as described above. After 48 hours, media was aspirated and replaced with either fresh E6 or E6 supplemented with 10µM SB431542 and 100 nM LDN193189. Media was replaced daily for 3 additional days of differentiation. At the 48-hour and 5-day time points, RNA samples were collected for gene expression analysis, and wells were fixed in 4% PFA for immunofluorescent imaging.

Teratoma Formation and Analysis

Cells pre-differentiated in E8, E6 or BA were dissociated by TrypLE, and $1.0x10^6$ cells were embedded in Matrigel and injected into the tibialis anterior muscles of *NOD/SCID* mice (Charles River Laboratory). These procedures were approved by the University of Ottawa Animal Care Veterinary Services (protocol #OHRIT-1666). Tumors were allowed to form for 9-18 weeks before teratomas were excised, fixed in 4% formaldehyde, and embedded in paraffin. Paraffin-embedded teratomas were sectioned and stained with hematoxylin and eosin (H&E) or by immunofluorescence (IF). For IF, sodium citrate/pressure cooker antigen retrieval was performed. Slides were blocked and permeabilized in 2% BSA, 0.01% Triton X-100 prior to overnight primary antibody incubation at 4°C. Details for SOX2, SOX17 and Desmin antibodies can be found in Supplemental Methods. Secondary antibodies were added for 1 hour at room temperature (AlexaFluor 680 or 488, 1:400). Nuclei were stained with Hoechst 33342 for 15 minutes prior to coverslipping. Imaging of H&E sections was performed using an Aperio CS2 scanscope (Leica Biosystems), and IF sections were imaged using the Cellomics ArrayScan. Quantification of cartilage regions in H&E sections and positive staining in IF was performed using custom scripts written for ImageJ.

Primary Antibodies

OCT4A (Cell Signaling Technology C52G3; 1:400), Brachyury T (R&D Systems AF2085, 5 µg/mL), SOX2 (R&D Systems MAB2018, 8 µg/mL; EMD Millipore AB5603, 1:200), Nestin (EMD Millipore ABD69, 1:500), Desmin (Abcam ab191181, 1:200) and SOX17 (R&D Systems AF1974, 1:100).

Supplemental References

Baddeley, A., and Turner, R. (2005). **spatstat** : An *R* Package for Analyzing Spatial Point Patterns. J. Stat. Softw.

Baddeley, A., Rubak, E., and Turner, R. (2015). Spatial Point Patterns: Methodology and Applications with {R} (London: Chapman and Hall/CRC).

Chang, W.Y., Lavoie, J.R., Kwon, S.Y., Chen, Z., Manias, J.L., Behbahani, J., Ling, V., Kandel, R.A., Stewart, D.J., and Stanford, W.L. (2013). Feeder-independent derivation of induced-pluripotent stem cells from peripheral blood endothelial progenitor cells. Stem Cell Res. *10*, 195–202.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.

Kuleshov, M. V, Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90-7.

Lee, W.D., Hurtig, M.B., Kandel, R.A., and Stanford, W.L. (2011). Membrane Culture of Bone Marrow Stromal Cells Yields Better Tissue Than Pellet Culture for Engineering Cartilage-Bone Substitute Biphasic Constructs in a Two-Step Process. Tissue Eng. Part C Methods *17*, 939–948.

Liao, Y., Smyth, G.K., and Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics *21*, 3448–3449.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. *102*, 15545–15550.

Tatsumi, R., Suzuki, Y., Sumi, T., Sone, M., Suemori, H., and Nakatsuji, N. (2011). Simple and highly efficient method for production of endothelial cells from human embryonic stem cells. Cell Transplant. *20*, 1423–1430.