Stem Cell Reports, Volume *12*

Supplemental Information

Single-Cell RNA Sequencing of Human Embryonic Stem Cell Differentiation Delineates Adverse Effects of Nicotine on Embryonic Development

Hongchao Guo, Lei Tian, Joe Z. Zhang, Tomoya Kitani, David T. Paik, Won Hee Lee, and Joseph C. Wu

Supplemental Information

Single-Cell RNA-Sequencing of Human Embryonic Stem Cell Differentiation Delineates Adverse Effects of Nicotine on Embryonic Development

Hongchao Guo[1,2,3,*], Lei Tian[1,2,3,*], Joe Z. Zhang[1,2,3], Tomoya Kitani[1,2,3], David T. Paik[1,2,3], Won Hee Lee[4], Joseph C. Wu[1,2,3]

[1]Stanford Cardiovascular Institute, Stanford, CA 94305, USA; [2]Institute for Stem Cell Biology and Regenerative Medicine, Stanford, CA 94305, USA; [3]Division of Cardiology, Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA; [4]Department of Basic Medical Sciences, University of Arizona College of Medicine, Phoenix, AZ 85004, USA

[*]Co-first author

Supplemental Methods

Reads mapping. FastQ files containing sequenced reads were mapped to the human reference genome (GRCh38) following "One Library, Multiple Flowcells" pipeline with Cellranger (version 2.1.1). Briefly, Cellranger firstly used STAR software to align sequenced reads to the reference genome. Then mapped reads were assigned to intergenic, intronic, and exonic regions based on the GTF annotation. A read was exonic if more than half of it intersects with an exon. Next, for reads that are not only aligned to a single exonic locus but also to 1 or more non-exonic loci, the exonic locus was prioritized and the read was considered to be confidently mapped to the exonic locus with a mapping quality score 255. Only reads that were confidently mapped to the transcriptome were used for UMI counting. Finally, we detected 6,847 and 5,646 cells for control and nicotine EBs, respectively. More details of mapping statistics were shown in **Table S1**.

Bioinformatics analysis. Cells with fewer than 200 or more than 6,000 expressed genes, and a high percentage of mitochondrial genes (>20%) were removed. Among them, 6,766 and 5,514 cells passed the filtering. We next normalized the gene expression values for each cell by the total expression, multiplied by 10,000, and made a log-transformation. Variable expressed genes across the single cells under cutoff with an average expression of more than 0.0125 and less than 3, and dispersion of more than 0.5, were detected for down-stream analysis. For dimensionality reduction and clustering, we used scaled z-scored residuals after regressing out mitochondrial percentage and the number of UMIs. Principal Component Analysis (PCA) was performed using variable expressed genes, and top 20 principal components (PCs) were used for cell clustering with a graph-based clustering approach at resolution 0.8. We ran t-SNE with the same number of PCs and default parameters to visualize the clustering results. Differentially expressed genes for each

cluster were detected by comparing cells within the cluster with other cells using the Wilcoxon rank sum test under cutoff with a *P* value less than 1%, log fold-change more than 0.25, and more than 25% cells expressing the gene. Cell cycle phase score for each cell was calculated based on the expression of G2M and S phase markers obtained from previous publication (Nestorowa et al., 2016). All these steps were performed using Seurat R package (Butler et al., 2018). We only detected the positive markers for each cluster. Functional enrichment analyses of differentially expressed genes were performed using the Bioconductor package "GeneAnswers" (Feng et al., 2018) in R (R Core Team, 2017). Functional annotations were from Gene Ontology (GO) (GO.db, Ashburner et al., 2000) and Reactome Pathway (Reactome.db, Fabregat et al., 2018) databases. Chord plots were generated with GOplot R package (Walter et al., 2015).

Merged data with combined raw data matrix from control and nicotine-exposed embryoid bodies (EBs) were used for integrative analysis. Steps for quality control, data normalization and transformation, dimensionality reduction, and clustering were the same as described above. Within each cluster, we detected differentially expressed genes (DEGs) between control and nicotine-exposed EBs with *P* value less than 1%, log fold-change more than log(1.2), and more than 10% cells expressing the gene. Gene expressions from published datasets were directly downloaded from GEO databases with GEO accession numbers GSE69844 (hepatic cell line, HepaRG) (De Abrew et al., 2016), GSE89923 (human gingival epithelium cell line, HGEC) (Gumus et al., 2008), and GSE56383 (human smooth muscle cell, HSMC) (Yoshiyama et al., 2014). We also included an unpublished dataset that contains human iPSC-derived endothelial cells (hiPSC-ECs) with nicotine exposure and controls. Fold-changes of DEGs between nicotine-exposed and control EBs for each dataset were calculated for comparison.

Cell-cell communications. We downloaded human ligand-receptor pairs curated by Ramilowski et al. (Ramilowski et al., 2016). Ligands/receptors expressing in a cluster were defined if more than 25% cells in that cluster had an expression value larger than 0. The ribbons connecting cell clusters were colored according to the cluster broadcasting the ligand and connected to the cluster expressing the receptor. The size of a ribbon is proportional to the number of ligand-receptor pairs, thus demonstrating the activity of the communications. We used the igraph R package for visualization (Csardi and Nepusz, 2006).

Differentiation of hESCs to cardiomyocytes. For cardiac differentiation, a chemically defined monolayer differentiation protocol was used as previously described (Burridge et al., 2014). Briefly, hESCs at ~90% confluence were incubated with a differentiation basal medium comprising RPMI 1640 medium (GIBCO) and B27 supplement minus insulin (GIBCO). CHIR99021 (Selleck Chemicals), a selective glycogen synthase kinase 3β inhibitor, was added to the differentiation basal medium. On day 2, the medium was removed and replaced with differentiation basal medium minus CHIR99021. On day 3, the Wnt antagonist, IWR-1 (Selleck Chemicals), was added to the medium. After 48 hours, the medium was removed and replaced with the differentiation basal medium without any inhibitors. On day 7, the cells were incubated with the complete cardiomyocyte medium consisting of RPMI 1640 medium and B27 supplement plus insulin (GIBCO). The medium was changed every 2 days. Monolayers of hESC-derived cardiomyocytes were cultured for ~30 days and subsequently dissociated for experimental use with TrypLE Express (GIBCO). One μM nicotine was added during the cardiomyocyte differentiation process.

Immunofluorescence and confocal microscopy. hESC-derived cardiomyocytes were fixed in 4% paraformaldehyde in PBS. Following permeabilization with 0.3% Triton X-100, hESC-derived cardiomyocytes were stained with primary antibodies against cardiac troponin T type 2 (TNNT2; ab45932, Abcam) and alpha Actinin antibody (ACTN; ab68167, Abcam). After reaction with the primary antibodies, cells were incubated with the appropriate Alexa Fluor-conjugated secondary antibodies (Santa Cruz Biotechnology or Life Technologies). Images of the stained cells were obtained under a brightfield microscope (Leica). Confocal images were taken by using a 63 × Plan-Apochromat oil immersion objective (Carl Zeiss) and a LSM 510 Meta confocal microscope (Carl Zeiss). Images were analyzed by using ZEN software (Carl Zeiss) and ImageJ software (National Institutes of Health).

Quantitative RT-PCR. Total RNA was isolated using Direct-zol™ RNA MiniPrep (Zymo Research) following the manufacturer's instructions. For the reverse transcription, iScript™ cDNA Synthesis Kit (Bio-Rad) was used, and cDNA template was synthesized based on 0.5 μg of total RNA. After diluting 10 times of cDNA, 2 μL of the cDNA template, 0.5 μL of TaqMan® primer sets (Life Technology), 5 μL of TaqMan® Master Mix (Life Technology), and 2.5 μL ddH2O were mixed in the reaction system. Real-time quantitative PCR was performed with CFX Connect™ Real-Time PCR Detection System (Bio-Rad). Each reaction was run in replicates to minimize variation. Expression values were normalized to the average expression of GAPDH. Taqman assay IDs for GAPDH, HMGB1, and TLR4 are Hs02758991_g1, Hs01590761_g1, and Hs00152939_m1, respectively.

Western blot. Cultured cells were homogenized in RIPA buffer with proteinase inhibitors and

quantified by Pierce™ BCA Protein Assay Kit (Pierce Biotechnology Inc.). A total of 20 μg protein was separated by NuPAGE® Novex® 4-12% Bis-Tris Protein Gel (Invitrogen, Carlsbad, CA). After transfer to an Amersham™ Hybond™ Blotting Membranes (GE Healthcare), the protein lanes were analyzed by western blot using specific antibodies against GAPDH (MA5-15738-HRP, Thermo Fisher Scientific) and BNIP3 (ab10433, Abcam). Band intensity was analyzed and quantified by ImageJ Fuji program.

Quantification and statistical analysis. Statistical analyses for each experiment are described in the figure legends or in the appropriate text. Multiple group comparisons were calculated using one-way ANOVA. Pairwise comparisons were carried out using the two-tailed unpaired Student's t test. $*P < 0.05$; $**P < 0.01$; $***P < 0.005$; $**** P < 0.001$. All error bars are defined as standard error mean of the mean (SEM) unless otherwise indicated.

Supplemental References

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: Tool for the unification of biology. Nat. Genet. *25*, 25–29.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411.

Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. InterJournal, Complex Systems 1695 (5): 1--9 (2006).

De Abrew, K.N., Kainkaryam, R.M., Shan, Y.Q.K., Overmann, G.J., Settivari, R.S., Wang, X.H., Xu, J., Adams, R.L., Tiesman, J.P., Carney, E.W., et al. (2016). Grouping 34 chemicals based on mode of action using connectivity mapping. Toxicol Sci 151, 447-461.

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., et al. (2018). The reactome pathway knowledgebase. *46*, 649–655.

Feng, G., Du, P., Kibbe, W.A., and Lin, S. (2018). GeneAnswers, integrated interpretation of genes overview of GeneAnswers installation of GeneAnswers package. *2*.

Gumus, Z.H., Du, B., Kacker, A., Boyle, J.O., Bocker, J.M., Mukherjee, P., Subbaramaiah, K., Dannenberg, A.J., and Weinstein, H. (2008). Effects of tobacco smoke on gene expression and cellular pathways in a cellular model of oral leukoplakia. Cancer Prev Res 1, 100-111.

Nestorowa, S., Hamey, F.K., Pijuan Sala, B., Diamanti, E., Shepherd, M., Laurenti, E., Wilson, N.K., Kent, D.G., and Göttgens, B. (2016). A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. Blood *128*, e20–e31.

Burridge, P.W., Matsa, E., Shukla, P., Lin, Z.C., Churko, J.M., Ebert, A.D., Lan, F., Diecke, S., Huber, B., Mordwinkin, N.M., et al. (2014). Chemically defined generation of human cardiomyocytes. Nat Methods 11, 855-860.

R Core Team (2018). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. Available online at https://www.R-project.org/.

Ramilowski, J.A., Goldberg, T., Harshbarger, J., Kloppmann, E., Lizio, M., Satagopam, V.P., Itoh, M., Kawaji, H., Carninci, P., Rost, B., et al. (2016). A draft network of ligand–receptor-mediated multicellular signalling in human. Nature Communications volume 6, Article number: 7866 (2015)

Walter, W., Sánchez-Cabo, F., and Ricote, M. (2015). GOplot: An R package for visually combining expression data with functional analysis. Bioinformatics *31*, 2912–29

Yoshiyama, S., Chen, Z.Y., Okagaki, T., Kohama, K., Nasu-Kawaharada, R., Izumi, T., Ohshima, N., Nagai, T., and Nakamura, A. (2014). Nicotine exposure alters human vascular smooth muscle cell phenotype from a contractile to a synthetic type. Atherosclerosis 237, 464-470.

Supplemental Figure Legends

Figure S1. Quality control of data set and cell type identification in control and nicotine-exposed embryoid bodies (EBs). (A) Scatter diagram showing the gene numbers (nGene) (left) and transcript numbers of single cells (nUMI) (right). **(B)** t-SNE plots of single cells from control and nicotine-exposed EBs at day 21. We identified six main types of progenitor cells in both control (top) and nicotine-exposed (bottom) EBs, including neural progenitor cells (Clusters 4, 6, and 8 in control EBs; Clusters 3, 5, 8, and 11 in nicotine-exposed EBs), stromal progenitor cells (Cluster 3 in control EBs; Cluster 6 in nicotine-exposed EBs), endothelial progenitor cells (Cluster 10 in control EBs; Cluster 10 in nicotine-exposed EBs), epithelial progenitor cells (Cluster 2 in control EBs; Cluster 2 in nicotine-exposedEBs), muscle progenitor cells (Clusters 5 and 11 in control EBs; Clusters 7 and 12 in nicotine-exposed EBs), and liver progenitor cells (Cluster 7 in control EBs; Cluster 4 in nicotine-exposed EBs). In addition, undifferentiated stem-like cells (USCs, Cluster 1) and undetermined cells (UDCs, Cluster 9) were also identified. **(C)** Heatmaps showing the expression pattern of the top 10 differentiated genes in each progenitor cell type in control (top) and nicotine-exposed (bottom) EBs. Representative differential genes for each cell type in control and nicotine-exposed EBs are listed on the right. The full list of differential genes for each cluster in control and nicotine-exposed EBs is shown in **Table S2**. **(D)** Significant gene markers for each cluster in combined EBs were selected to perform GO analysis. GO terms with $P < 0.05$ are shown. Gene number of each GO term is listed on the left. *P* value is shown as –log 10 (*P value*). Neural, muscle (Clusters 7 and 13), and epithelial (Clusters 2 and 12) progenitor cells consisted of several sub-clusters. EpiPCs were divided into two subsets: Clusters 2 and 12. Neural progenitor cells were divided into four subsets: Clusters 3, 4, 8, and 10. Muscle progenitor cells were divided into two subsets: Clusters 7 and 13. The full list of differential genes for each cluster

in combined EBs is shown in **Table S3** and differential genes-related GO terms for each cluster is shown in **Table S4**.

Figure S2. Cell type proportions and GO analysis for each individual in combined EBs. (A) Percentage of total cells in control and nicotine-exposed EBs is determined for each cluster. C, control EBs; N, nicotine-exposed EBs. **(B)** Proportion of cells in G2M, S, or G1 phase for each cell type. **(C)** Enriched differentially expressed genes (DEGs) related to pathways in each cell type upon nicotine exposure. The ribbons connect each gene to the assigned pathways. The color of a ribbon is consistent with the color at the pathway side to distinguish pathways. The gradient of red/blue is proportional to the fold-change of genes connected between the nicotine and control. The full list of DEGs is shown in **Table S5** and the list of DEG-related GO terms is shown in **Table S6**. **(D)** Nicotine reduced BNIP3 protein level. Upper: Representative western blot of BINP3 protein expression in control and nicotine-exposed EBs; Lower: Quantification of BINP3 expression from three independent experimetns, normalized to GAPDH. $^{**}P < 0.01$. **(E)** Genes that have consistent changes on RNA and protein expression level between found in available nicotine-associated proteomic dataset in PubMed and our scRNA-seq. **(F)** t-SNE plots of nicotine receptors, *CHRNA5*, and *CHRNB1*. Each cell on the t-SNE plot is colored according to gene expression. Gray: cells with no expression. Blue: cells with expression.

Figure S3. Generation and characterization hESC-derived cardiomyocytes. (A) Schematic for differentiation of hESC-derived cardiomyocytes (-insulin: B27 and RPMI without insulin, with glucose; + Insulin: B27 and RPMI with insulin and glucose; -Glucose: B27 and RPMI without insulin without Glucose). Scale bar, 100 μm. **(B)** Immunofluorescence analyses showing hESC-

derived cardiomyocytes expressing TNNT2 (green) and ACTN1 (red). Scale bar, 20 μm. **(C)** Representative FACS analysis of TNNT2 expression in hESC-derived cardiomyocytes with TNNT2 antibody (upper) and rabbit IgG antibody (lower). Statistical analysis data from C. (**$P < 0.01$; unpaired Student's t-test; and n = 3). **(D)** Cell viability assay based on quantitation of the ATP present in control and nicotine-exposed hESCs. Cell viability were measured from three independent experiments. * $P<0.05$.

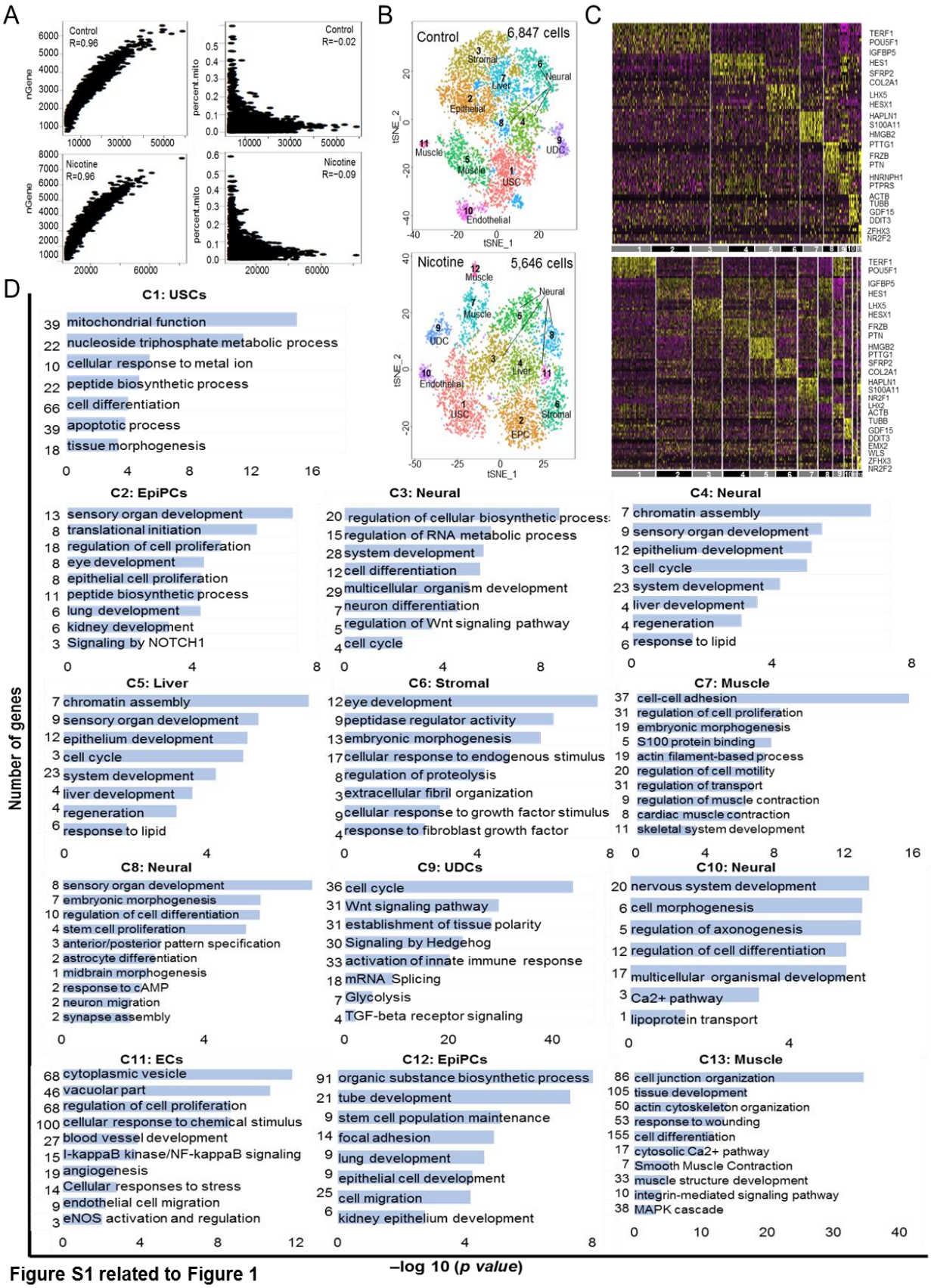

Figure S1 related to Figure 1

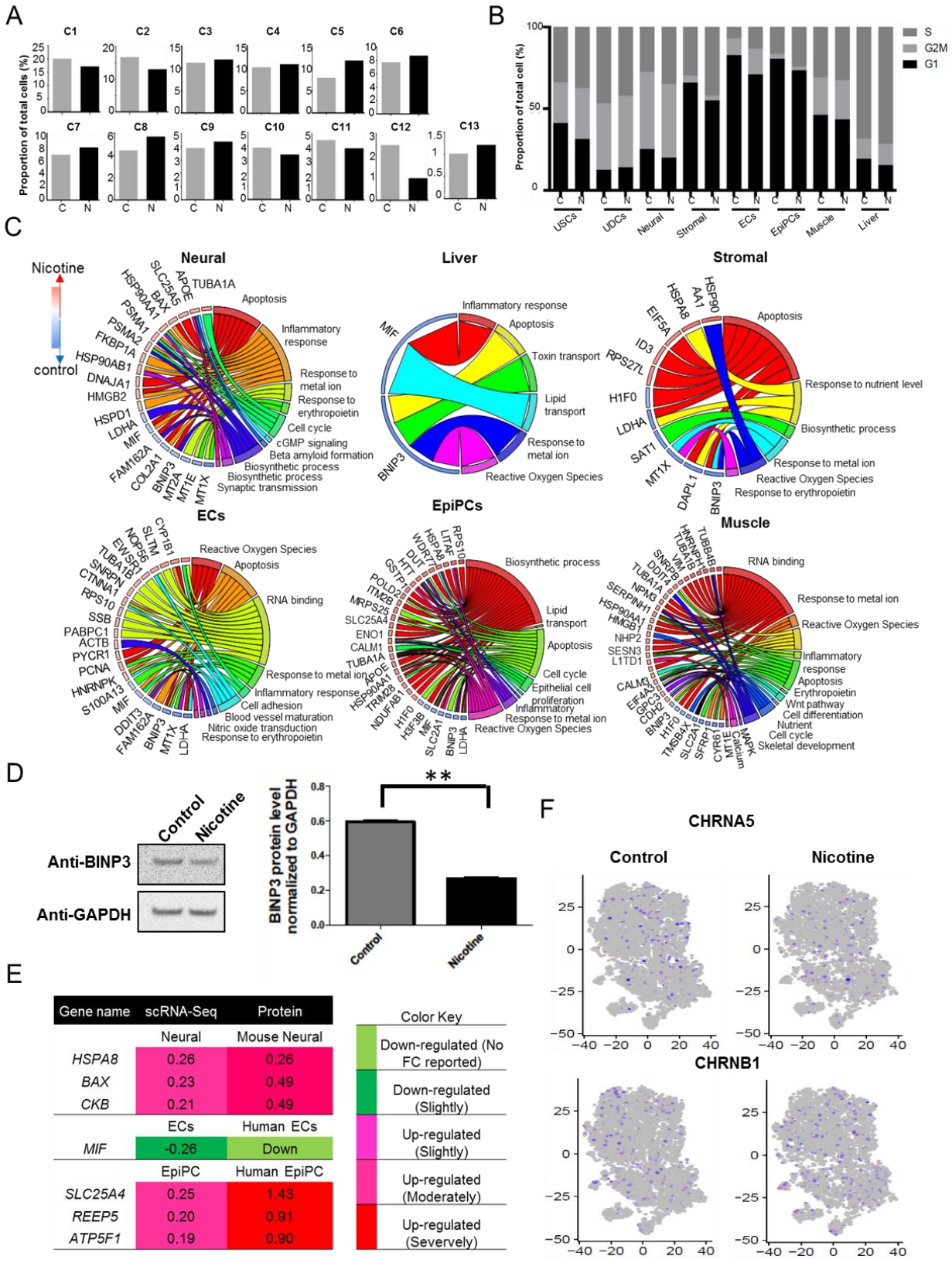

Figure S2 related to Figure 2

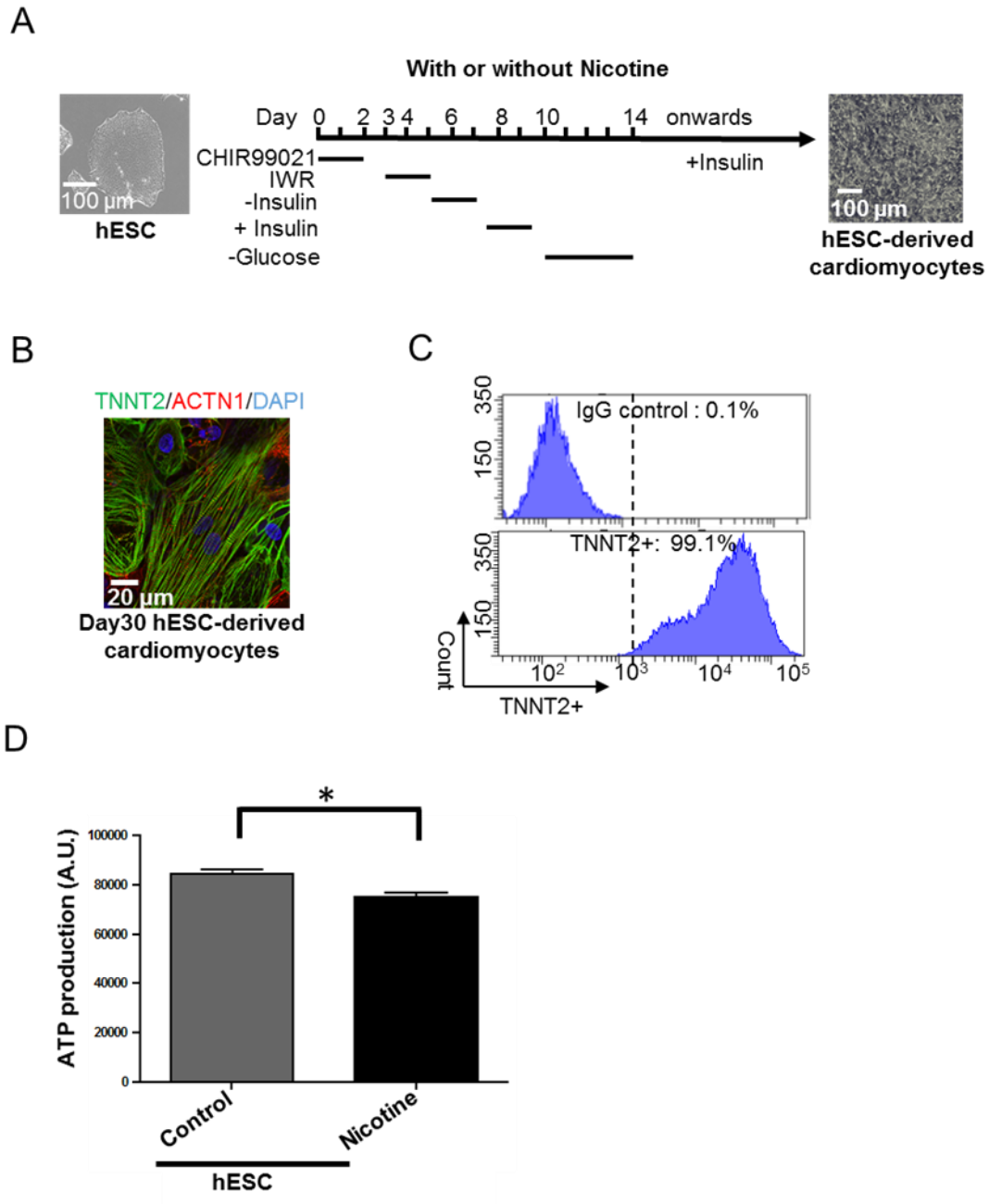

Figure S3 related to Figure 5

Supplemental Tables

Table S1. Mapping statistics, related to Figure 1.

Sample	Estimated number of cells	Mean reads per cell	Mean genes per cell	Read mapped confidently to genome	Read mapped confidently to transcriptome
Control EBs	6,847	35,989	2,495	95.3%	70.7%
Nicotine EBs	5,646	54,034	2,896	93.3%	72.3%

Table S2. Statistics for differentially expressed genes among clusters in control and nicotine-exposed EBs, related to Figure 1. "pct.1" is the proportion of cells that express the gene in the target cluster, as labeled in the "cluster" column. "pct.2" is the proportion of cells that express the gene in the other clusters. (Differentially expressed genes for control and nicotine EB.xlsx).

Table S3. Statistics for differentially expressed genes among clusters in the combined EBs, related to Figure 1. The format is the same as Table S2. (Differentially expressed genes for combined EBs.xlsx).

Table S4. Pathway enrichment of differentially expressed genes among clusters in the combined EBs, related to Figure 1. The format is the same as Table S3. (Pathway Enrichment for combined EBs.xlsx).

Table S5. Statistics for differentially expressed genes between nicotine-exposed and control EBs in each cluster, related to Figure 2. "pct.1" is the proportion of cells that express the gene in the target cluster, which is labeled in the "cluster" column, where "D" is control and "N" is

nicotine-exposed. “pct.2” is the proportion of cells that express the gene in the other clusters. (Nicotine.vs.Control.DEGs.xlsx).

Table S6. Pathway enrichment of differentially expressed genes between nicotine-exposed and control EBs in each cell type, related to Figure 2. The *P*-value was calculated under the hypergeometric test. nGenes is the number of DEGs involved in the pathway. Z-score was calculated using the following formula: z-score=(up-down)/$\sqrt{}$(up+down), where up/down is the number of genes up-regulated/down-regulated in the nicotine cells. (Nicotine.vs.Control.PathwayEnrichment.xlsx).

Table S7. Cross-talk among cell types in control and nicotine-exposed EBs, related to Figure 4. “Ligand Clusters” column lists the cell types with more than 25% cells expressing the ligand. “Receptor Clusters” column lists the cell types with more than 25% cells expressing the receptor. (Control and nicotine EBs.crossTalks.xlsx).