Guided Differentiation of Pluripotent Stem Cells for Cardiac Cell Diversity

Erik McIntire[1], Kenneth A Barr[2], Yoav Gilad[1,2]
1 Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America
2 Department of Medicine, University of Chicago, Chicago, Illinois, United States of America

We have developed a new guided induced pluripotent stem cell differentiation protocol that rapidly generates a temporally and spatially diverse range of cardiac-relevant cell types. In just 8-10 days, we consistently reproduce all cell types that exist in mature cardiac organoids. Using guided differentiation cultures one can rapidly characterize, at a population level, regulatory variation and gene by environment interactions in multiple cardiac cell types.

Induced pluripotent stem cells (iPSCs) are readily generated from somatic cells[1] and can be differentiated into specialized cell types, including cardiomyocytes, neurons, hepatocytes, and many others. iPSCs thus facilitate controlled *in vitro* experimentation in specific cell types, and allow for population-level studies of dynamic traits in different human cell types. For example, Knowles *et al.* used iPSC-derived cardiomyocytes (iPSC-CMs) to demonstrate that inter-individual gene regulatory variation can explain differences between individuals in sensitivity to the cardiotoxic side effects of the chemotherapeutic drug doxorubicin[2]. Knowles *et al.* is one of many examples of studies that used differentiated iPSCs to identify genetic variants that regulate gene expression levels (expression quantitative trait loci, or eQTLs) and other molecular phenotypes, which are useful for understanding the regulatory mechanisms that modulate disease risk. However, eQTL studies utilizing iPSCs have historically been limited in scope by standard bulk analysis of gene expression, which precludes characterizing multiple cell types simultaneously. The limitations of bulk analysis favor differentiation of iPSCs into homogenous monocultures. Consequently, many cell types remain understudied.

Unlike bulk RNA-sequencing, single-cell RNA-sequencing (scRNA-seq) can capture information about cellular heterogeneity and cell-type-specific gene expression, facilitating experimentation with more complex iPSC-derived cultures, such as organoids. Organoids comprise multiple cell types that self-organize and recapitulate major features of *in vivo* organs, including basic function, structural resemblance and constituent cell types. For instance, Lewis-Israeli *et al.* developed heart organoids that autonomously contract, exhibit a chamber-like morphology, and harbor various cardiac-relevant cell types such as cardiomyocytes, fibroblasts, and endothelial cells[3]. Such properties make organoids very useful as disease models. However, because organoid generation typically involves long-term (weeks or months) cell culture and differentiation procedures tailored to individual iPSC lines, organoids are not well-suited for population-level studies such as eQTL mapping efforts, which necessitate data collection from many (dozens or hundreds) individuals.

In order to study cell-type-specific gene regulation at the population level, we developed a 'guided' iPSC differentiation approach that offers high cellular diversity, fast experimental throughput, and reproducibility across cell lines. Guided differentiation, as opposed to 'directed' differentiation, gently biases iPSCs towards the cardiac mesoderm lineage, while maintaining the greatest degree of cellular diversity possible. Within guided cultures, individual cells differentiate at dissimilar rates, permitting maximum diversity with respect to function (cell type) and time (cell differentiation stage). The inclusion of early progenitor cells, in addition to differentiated cells, provides continuous developmental trajectories for detecting transient genetic effects that typically can only be captured by using time-course experimental design (e.g., Strober *et al.* (2019)[4]). In what follows, we demonstrate that guided cultures contain functionally diverse cell types, which represent multiple stages of cardiac differentiation, and show that the transcriptional profiles of single cells in our cultures are comparable to that of single cells from mature cardiac organoids. In the methods section, we provide a detailed protocol to establish guided differentiating cardiac cultures.

We performed guided differentiation (Fig. 1A) using three iPSC lines to generate three-dimensional cardiac cell cultures. First, we formed iPSCs into three-dimensional aggregates measuring ~300μM in diameter. We cultured iPSC aggregates in the formation plate using Essential 8™ medium (E8) with 10μM CEPT[5] for 24 hours and transferred the aggregates to ultra-low attachment plates for an additional 48 hours. We then biased iPSC aggregates towards cardiac lineage using temporal Wnt modulation[6,7]. We exchanged E8 for 'heart medium' (RPMI-1640 with a 2% v/v concentration of B-27 supplement, sans insulin) plus the Wnt activator CHIR99021 ("Chiron") at a final concentration of 4μM. After 24 hours, we exchanged Chiron+ heart medium for base heart medium. We refreshed the heart medium 48 hours later, this time adding IWP4, a Wnt

inhibitor, at a final concentration of 2μM. After 48 hours, we replaced IWP4+ heart medium with base heart medium for 48 hours before harvest. Contraction became visible 5 days after Wnt activation, or 8 days overall.

We collected cells at day 10 using the 10X Genomics platform for sequencing on an Illumina NovaSeq 6000. After filtering and normalizing the data, we performed principal component analysis with 5,000 highly variable features and used the top 50 principal components for graph-based unsupervised clustering. Guided differentiation cultures include a diversity of cell types from all three germ layers. We annotated cells using marker gene expression and differential expression analysis (Fig. 1B; see methods), and identified neuroectoderm, endothelial / endocardial cells, epicardial cells, fibroblasts, cardiac progenitor cells, and cardiomyocytes. We also annotated cellular populations of foregut endoderm and hepatic endoderm per marker gene expression (Fig. 1C). Subcluster analysis of foregut endoderm revealed posterior and anterior foregut populations (Extended Fig. 3), which have been reported previously[8].

As guided differentiation yields both terminal and progenitor cell types, it can potentially recapitulate data from a more standard differentiation time-course experiment. To demonstrate this, we performed Seurat integration[9] of our scRNA-seq data with data from a previous 16-day time-course study of iPSC-CM differentiation performed by Elorbany *et al.* (2022)[10]. Sub-setting cells by their collection day illustrates progression through iPSC-CM differentiation. While, as expected, cell states were rather specific to certain days in the time-course study, the 'single time-point' guided differentiation culture harbored cells throughout the entire UMAP space (Fig. 2), suggesting that guided differentiation produced cells that are transcriptionally similar to those collected during the entire time-course study.

To demonstrate similarity between the transcriptional profiles of cells from guided cultures and their *in vivo* counterparts, we used automated cell classification. For a cell reference, we leveraged a dataset of annotated fetal heart cells published by Miao *et al.* (2020)[11], which consisted of 11 different cell types (Fig. 3A). We utilized this reference to train the *scPred*[12] prediction model for automated cell annotation (Fig. 3B). Individual cells from our guided differentiation culture were assigned probabilities of belonging to each of the 11 reference cell types. As the guided differentiation culture harbors cell types that are not present in fetal heart tissue (e.g., foregut endoderm), we set the probability threshold to 0.9; cells below the threshold were classified as unassigned and not included in the UMAP plots. Based on this analysis, our guided differentiation culture putatively harbored all cell types except red blood cells (Fig. 3C). These included cardiomyocytes, fibroblasts, epicardial cells, endothelial / endocardial cells, and nervous system cells, consistent with the manual cell annotation. Additionally, *scPred* cell type predictions included immune cells, conduction system cells, and pericytes, which were not detected by manual annotation. *ScPred* classifies cells individually, without clustering, and is capable of identifying small cell populations that would otherwise be masked by clustering.

Finally, to benchmark our guided differentiation culture against more established *in vitro* models, we applied the *scPred* prediction model to scRNA-seq data from mature multi-lineage organoids (100-day culture) published by Silva, *et al.* (2021)[13] (Fig. 3D). Both iPSC-based cultures showed comparable levels of cell diversity and transcriptional similarity relative to the fetal heart tissue, demonstrating that our guided differentiation culture method can generate diverse, transcriptionally relevant cell types in just 10 days. The reduced time and labor needed for guided differentiation (relative to standard organoids) will enable dynamic population-level studies at scale, including eQTL mapping studies at cell type resolution. Moreover, the transcriptional similarity between guided cardiac cell types and their *in vivo* counterparts suggests that dynamic functional patterns identified using guided cultures will be functionally relevant to human tissues.

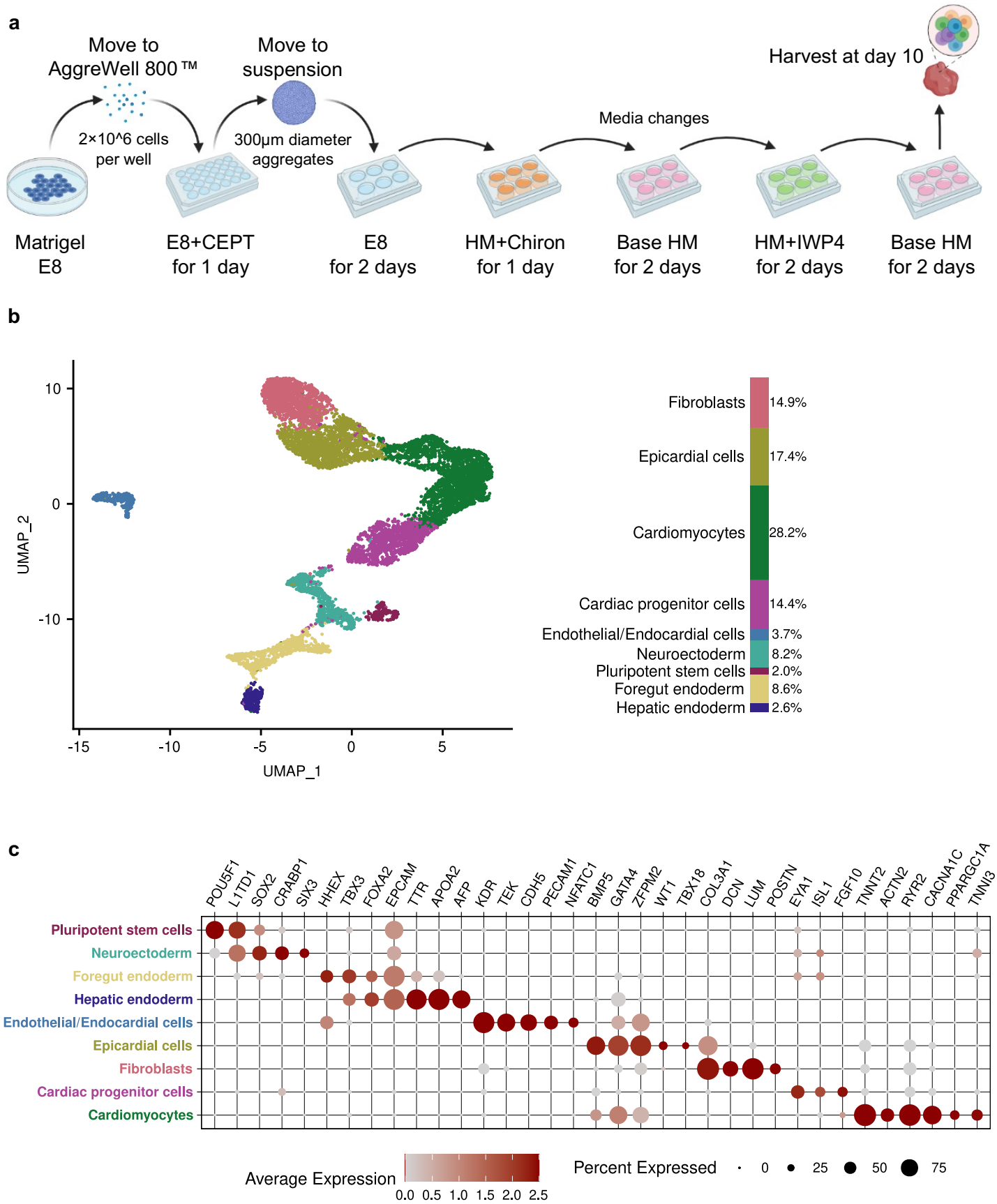


Fig 1 | Guided differentiation and characterization. a, Overview of the 10-day guided differentiation protocol. CEPT: Chroman 1, emricasan, polyamines, and trans-ISRIB. HM: heart medium. Chiron: CHIR99021. **b,** UMAP of 8,936 sequenced cells, manually annotated, alongside cell type proportions. **c,** Expression of canonical marker genes for each cell cluster.

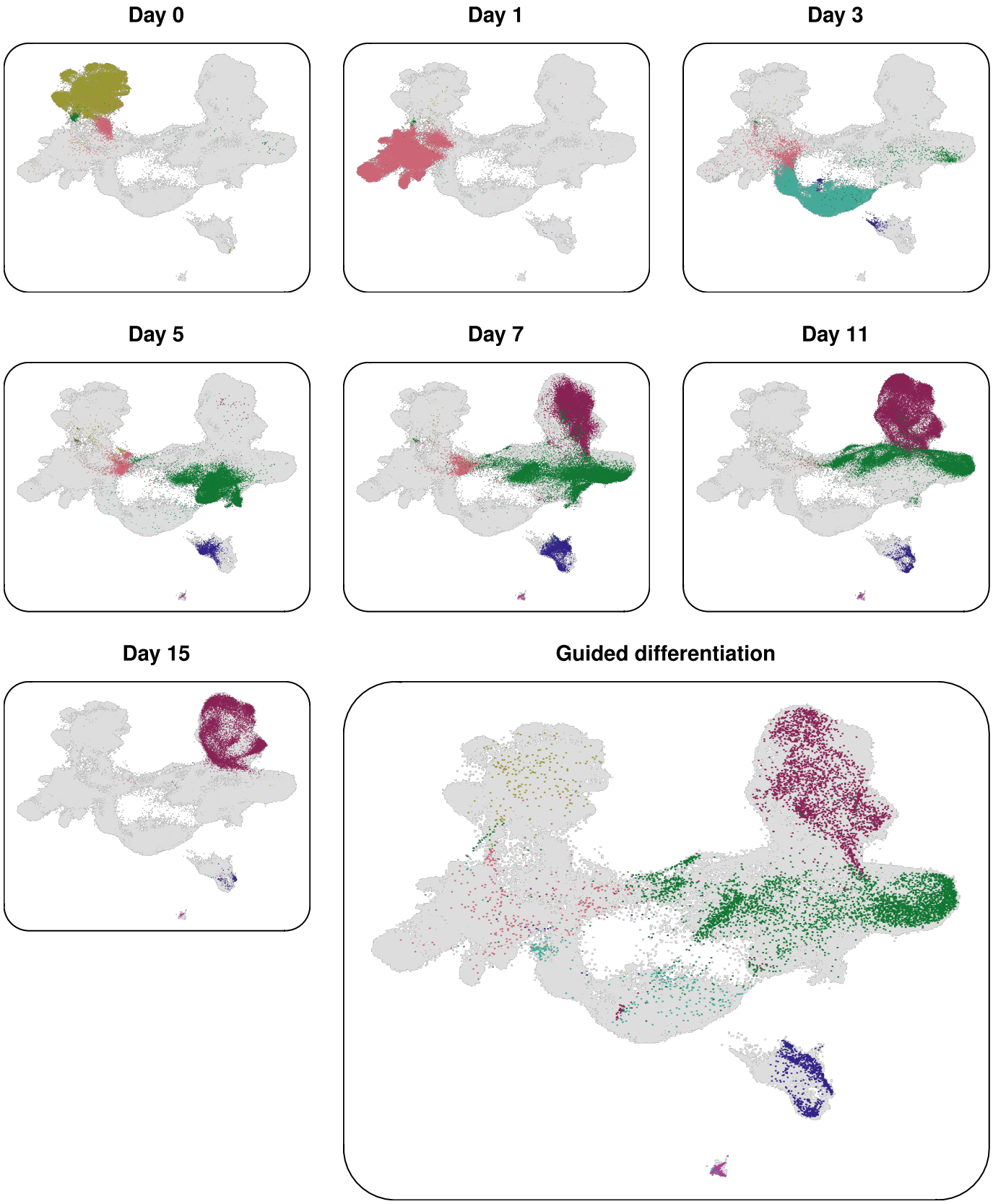


Fig 2 | UMAP plots of cells from Seurat integration of a 16-day iPSC-CM differentiation time-course dataset and the guided differentiation dataset. Cells are colored according to cluster and dataset; grey cells are those present in the entire integrated dataset but not present in the selected dataset. The small plots represent cells from the differentiation time-course (n = 230,849), plotted according to their day of collection. The large plot represents cells from a single collection of guided differentiation culture cells (n = 8,936).

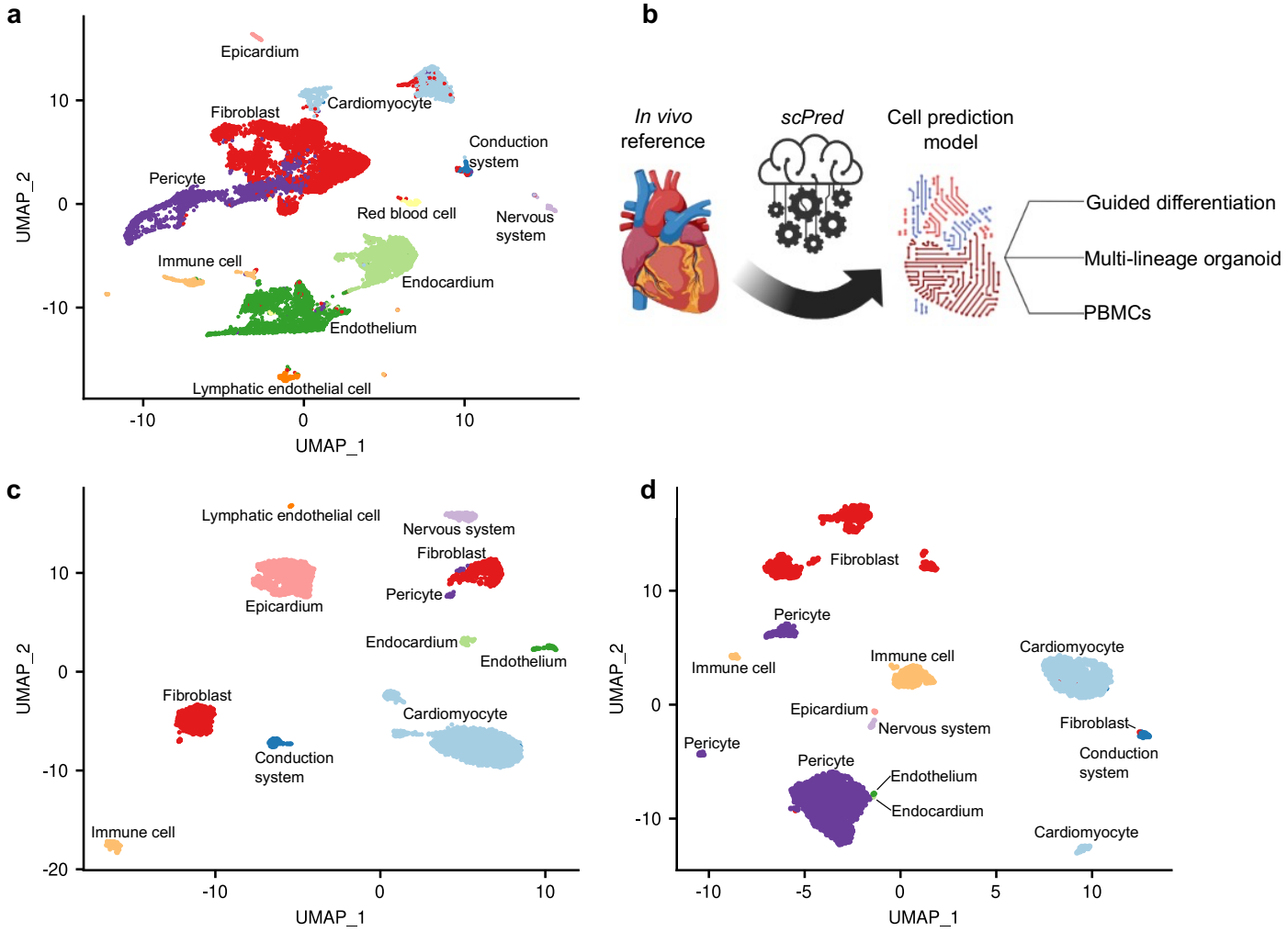


Fig 3 | Automated annotation using *scPred* trained on a fetal heart reference. a, UMAP of fetal heart cells (n = 13,569) with 11 cell types annotated by Miao *et al.* (2020)[11]. **b,** Overview of automated annotation process using *scPred*. Model was trained using the cell reference and applied to three datasets, including peripheral blood mononuclear cells (PBMCs) used here as a non-cardiac control (Extended Fig. 5). **c,** UMAP of classified cell types from the guided differentiation culture (n = 5,391). **d,** UMAP of classified cell types from multi-lineage organoid published by Silva *et al.* (2021)[13] (n = 3,009).

References

1. Takahashi, K. *et al. Cell* **131**, 861–872 (2007).
2. Knowles, D. A. *et al. eLife* **7**, e33480 (2018).
3. Lewis-Israeli, Y.R., Wasserman, A.H., Gabalski, M.A. *et al. Nat Commun* **12**, 5142 (2021).
4. Strober, B. J. *et al. Science* **364**, 1287–1290 (2019).
5. Chen, Y., Tristan, C.A., Chen, L. *et al. Nat Methods* **18**, 528–541 (2021).
6. Lian, X. *et al. Nat. Protoc.* **8**, 162–175 (2013).
7. Burridge, P. W. *et al. Nat. Methods* **11**, 855–860 (2014).
8. Drakhlis, L., Biswanath, S., Farr, CM. *et al. Nat Biotechnol* 39, 737–746 (2021).
9. Stuart, T. *et al. Cell* **177**, 1888-1902.e21 (2019).
10. Elorbany, R. *et al. PLOS Genet.* **18**, e1009666 (2022).
11. Miao, Y. *et al. Cell Stem Cell* **27**, 574-589.e8 (2020).
12. Alquicira-Hernandez, J., Sathe, A., Ji, H. P., Nguyen, Q. & Powell, J. E. *Genome Biol.* **20**, 264 (2019).
13. Silva, A. C. *et al. Cell Stem Cell* **28**, 2137-2152.e6 (2021).

Methods

Samples. We used 3 iPSC lines from unrelated Yoruba individuals from Ibadan, Nigeria (YRI): 19114 (female), 19130 (male), and 19152 (female). These iPSC lines were reprogrammed from lymphoblastoid cell lines (LCLs) and were characterized and validated previously[14]. We confirmed cell line identities using genotype data generated by the HapMap project from the original LCL lines[15].

IPSC maintenance. We maintained the 3 iPSC lines using Matrigel Growth Factor Reduced (GFR) Basement Membrane Matrix (Corning, 354230) with Essential 8™ (E8) medium (Thermo Fisher Scientific, A1517001) and Penicillin-Streptomycin (Lonza, 17-602F) in an incubator at 37°C and 5% CO_2. At roughly 80% confluency (approximately every 3-5 days), we passaged cell cultures using a dissociation reagent (0.5 mM EDTA, 300 mM NaCl in PBS) and seeded iPSCs with 10μM ROCK inhibitor Y-27632 (ab120129, Abcam).

Guided differentiation. Similar to directed differentiation, guided differentiation primarily generates cardiac-relevant cell types. However, it also yields cell types representative of all three germ layers and includes cell stages throughout cardiomyogenic lineage, resulting in a greater diversity of cell types. We performed guided differentiation of the 3 iPSC lines by forming three-dimensional aggregates and performing temporal Wnt modulation[6,7]. We formed iPSC aggregates using an AggreWell™800 24-well plate (STEMCELL Technologies, 34811). We coated each well with Anti-Adherence Rinsing Solution (STEMCELL Technologies, 07010). On day 0, we disassociated iPSCs seeded them onto the plate using E8 medium with 10μM CEPT[5], i.e., chroman 1 (Torcis, 7163), emricasan (MedKoo Biosciences, 510230), polyamines (Sigma-Aldrich, P8483), and trans-ISRIB (Torcis, 5284). Cell aggregates were formed at approximately 300μm diameter using 2 million cells per well (i.e., 1 million cells per mL) of the AggreWell™800 plate, as described previously by Branco *et al.* (2019)[16]. We centrifuged the plate at 100*g* for 3 min to aggregate cells together. Cells remained in the plate for 24h. After 24h, we transferred fully-formed iPSC aggregates to a suspension culture using an ultra-low adherent plate (STEMCELL Technologies, 100-0083) with E8 medium for 48h. On day 3, we changed the cell medium to “heart medium” plus 4μM CHIR99021 (STEMCELL Technologies, 72052). Heart medium is comprised of RPMI 1640 Medium, GlutaMAX™ Supplement, HEPES (Thermo Fisher Scientific, 72400047), 2% v/v B-27 Supplement, minus insulin (Thermo Fisher Scientific, A1895601), and Penicillin-Streptomycin. After 24h, on day 4, we changed medium to base heart medium for 48h. Contraction became visible on day 5 for all three lines. On day 6, we changed medium to heart medium plus 2μM IWP4 (STEMCELL Technologies, 72552) for 48h. Lastly, on day 8, we changed medium to base heart medium for 48h and collected cells on day 10.

Aggregate disassociation. We collected day 10 aggregates from the 3 iPSC lines and disassociated them by treating with room temperature AccuMax™ (STEMCELL Technologies, 07921) followed by incubation at 37°C for 10 min. Following incubation, we pipetted aggregates up-and-down for 30 sec with a p1000 wide bore pipette tip (Thermo Fisher Scientific, 2079GPK). Aggregates were then incubated for an additional 5 min at 37°C. We repeated the pipetting every 5 minutes until aggregates were disassociated, at which point we added 5mL of 4°C Bovine Serum Albumin (BSA) solution (Sigma-Aldrich, A8412) and centrifuged cells for 100*g* for 3 min. We resuspended cells in 1mL 4°C BSA solution and strained cells through a 40μm cell strainer (Bel-Art™, H136800040). We combined cells from each iPSC line in even proportions (500,000 cells per line) and centrifuged cells at 100*g* for 3 min. We then resuspended cells in 4°C BSA solution at a concentration of approximately 2 million cells per mL. Finally, we strained the cell suspension using a 40μm cell strainer.

ScRNA-seq analysis. We collected cells from the 3 iPSC lines for scRNA-seq using the 10X Genomics Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (Dual Index). We targeted 10,000 cells total in one lane of a 10X chip and sequenced the library using an Illumina NovaSeq 6000 at the University of Chicago Functional Genomics Core Facility. We obtained 20,755 mean reads per cell. We aligned samples to the human genome (GRCh38) using Cell Ranger[17] v6.1.2 and assigned cells to individuals using Vireo v0.5.6[18]. We then analyzed count data in R (v4.2.0)[19] / RStudio (v2022.02.3+492)[20] using Seurat v4.3.0[9] with tidyverse (v1.3.1)[21]. For cell filtering, we removed cells classified as doublets or unassigned by Vireo. Additionally, we filtered out cells expressing less than 1,500 unique genes. We performed sctransform-based normalization[22] (v0.3.5) (sctransform function) using 5,000 variable features, performed dimensionality reduction (RunPCA function) and used 50 dimensions for uniform manifold approximation and projection (UMAP) embedding (RunUMAP function). We computed nearest neighbors using 50 dimensions (FindNeighbors function) and performed unsupervised clustering at a resolution of 0.2 (FindClusters function) to yield 9 cell clusters for

differential expression analysis using the Wilcoxon Rank Sum test. We annotated cell clusters based on canonical marker gene expression. We shaded cell clusters using color schemes adapted from Paul Tol[23] for all figures unless otherwise noted.

Foregut endoderm subcluster analysis. We subset the foregut endoderm cluster cells as their own Seurat object and re-normalized using the same approach described above. We performed unsupervised clustering at a resolution of 0.1 to yield 2 clusters for differential expression analysis. We annotated cell clusters based on canonical marker gene expression.

Seurat integration. In order to compare cells from our guided differentiation and cells from a 16-day iPSC-CM differentiation time-course published by Elorbany *et al.* (2022), we performed Seurat integration of the two scRNA-seq datasets. First, we filtered the iPSC-CM time-course dataset using the same parameters as in the original study; that is, 1) genes must be detected in at least 10 cells, 2) cells must contain at least 300 unique genes, 3) cells must have no more than 25% mitochondrial reads, 4) cells must have a doublet probability of 0.3 or less, 5) cell assignment must be unambiguous, 6) cells with feature or read counts more than 4 standard deviations away from the median are excluded. We filtered cells from the guided differentiation dataset using the criteria described in the scRNA-seq analysis section above. Using sctransform, we normalized each dataset individually using 5,000 variable features and selected 5,000 anchor features for integration (SelectIntegrationFeatures function). We then prepared the datasets for integration (PrepSCTIntegration function), determined a set of anchor features (FindIntegrationAnchors function), and integrated the datasets (IntegrateData function). Following integration, we performed dimensionality reduction, UMAP embedding, and computed nearest neighbors as described above. We performed unsupervised clustering at a resolution of 0.1 to yield 7 cell clusters. We subset cells by dataset (and for the time-course dataset, by day of collection) for visualization with UMAP.

Automated cell classification. We used *scPred*[12] v1.9.2 to create a prediction classifier model trained on annotated fetal heart cells published by Miao *et al.* (2020)[11]. The fetal heart dataset comprised 11 annotated cell types: endocardium, endothelium, lymphatic endothelial cell, cardiomyocyte, epicardium, smooth muscle cell, fibroblast, immune cell, nervous system, conduction system, and red blood cell. We normalized the fetal heart dataset using the same approach described above and then trained the classifier (getFeatureSpace and trainModel functions) using default parameters. For the multi-lineage organoid dataset, we filtered out cells expressing less than 1,500 unique genes, per the original study[13]. The multi-lineage organoid dataset, guided differentiation dataset, and peripheral blood mononuclear cells (PBMC) dataset[24] were all normalized using the same approach described above. We classified cells from each dataset (scPredict function) using a threshold of 0.9. We shaded cell clusters in figure 3 using RColorBrewer (v1.1-3)[25].

References

14. Banovich, N. E. *et al. Genome Res.* **28**, 122–131 (2018).
15. The International HapMap Consortium. *Nature* **426**, 789–796 (2003).
16. Branco, M. A. *et al. Sci. Rep.* **9**, 9229 (2019).
17. Zheng, G. X. Y. *et al*. *Nat. Commun.* **8**, 14049 (2017).
18. Huang, Y., McCarthy, D. J. & Stegle, O. *Genome Biol.* **20**, 273 (2019).
19. R Development Core Team. R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, 2019).
20. RStudio Team. RStudio: Integrated Development for R (RStudio, 2019).
21. Wickham, H. et al. Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
22. Hafemeister, C. & Satija, R. *Genome Biol.* **20**, 296 (2019).
23. Tol, Paul. "Colour schemes." SRON Technical Note 3.2 (2021).
24. 20k Human PBMCs, 3' HT v3.1, Chromium X Single Cell Gene Expression Dataset by Cell Ranger (2021).
25. Neuwirth, E. ColorBrewer Palettes [R package RColorBrewer version 1.1-3]. (2022).

Acknowledgements. We thank all members of the Gilad lab, especially Natalia Gonzales for providing feedback on the manuscript. This work was completed in part with resources provided by the University of Chicago's Research Computing Center. We thank the University of Chicago Functional Genomics Core Facility for their assistance with sequencing the libraries. Figures 1A and 3B were created using BioRender.com.

Data Availability. Code for data analysis available at: https://github.com/erikmcintire/GuidedDifferentiation
The fastq and scRNA-seq count files will be made public concurrent with manuscript publication.

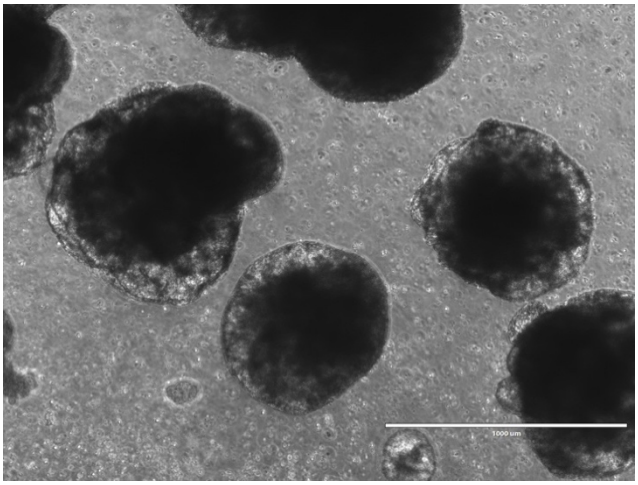
Author contributions. Y.G. conceived the study and supervised the project. E.M. developed the differentiation method and performed the differentiation experiments. E.M. and K.A.B performed single cell RNA-seq library preparation. E.M. analyzed and interpreted the data with support from K.A.B. E.M. drafted the manuscript with review and editing provided by Y.G.

Competing interests. Y.G. and E.M. are named as inventors with the University of Chicago on a patent application related to guided differentiation.

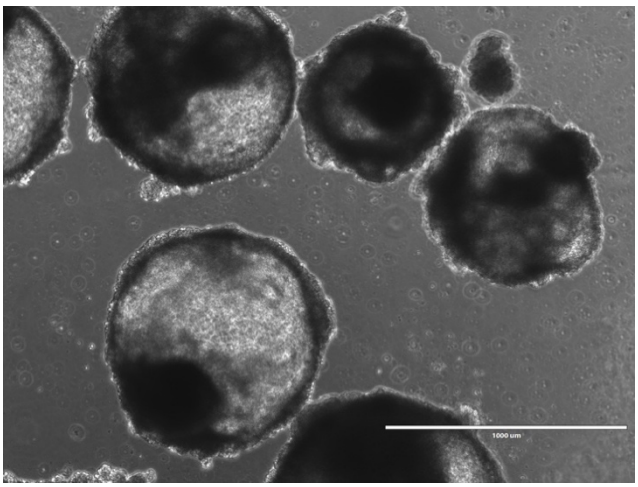
Funding. This work was funded by a MIRA award (R35GM131726) to Y.G. E.M. was supported by grants from the NIH (T32GM139782 and T32HL007381).

Extended figures

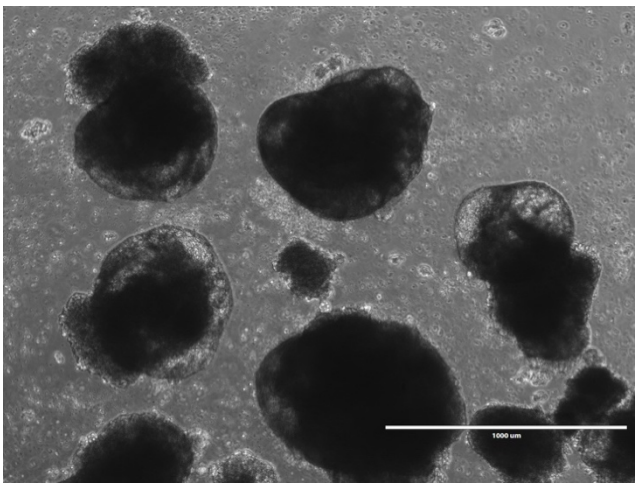
a



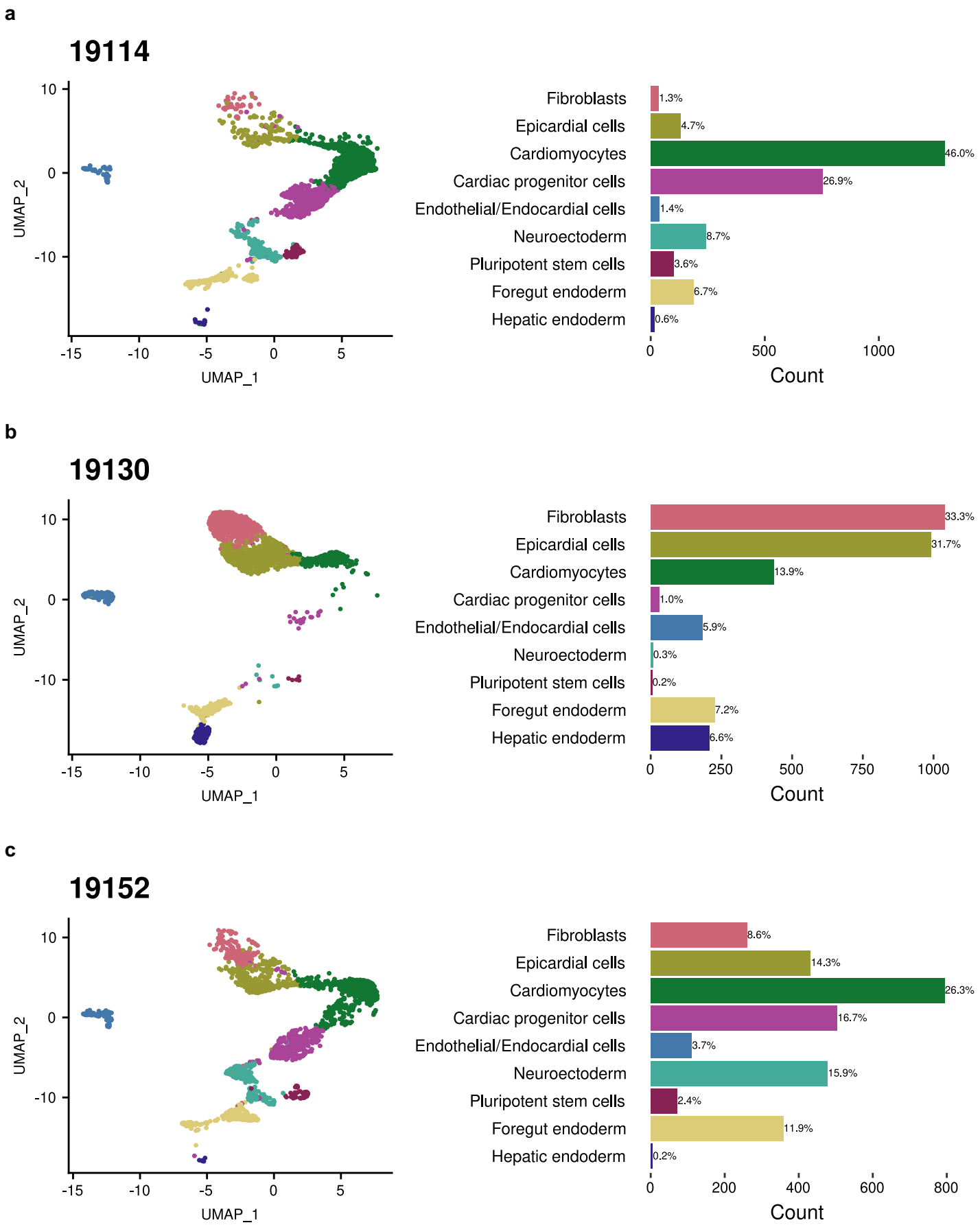
b



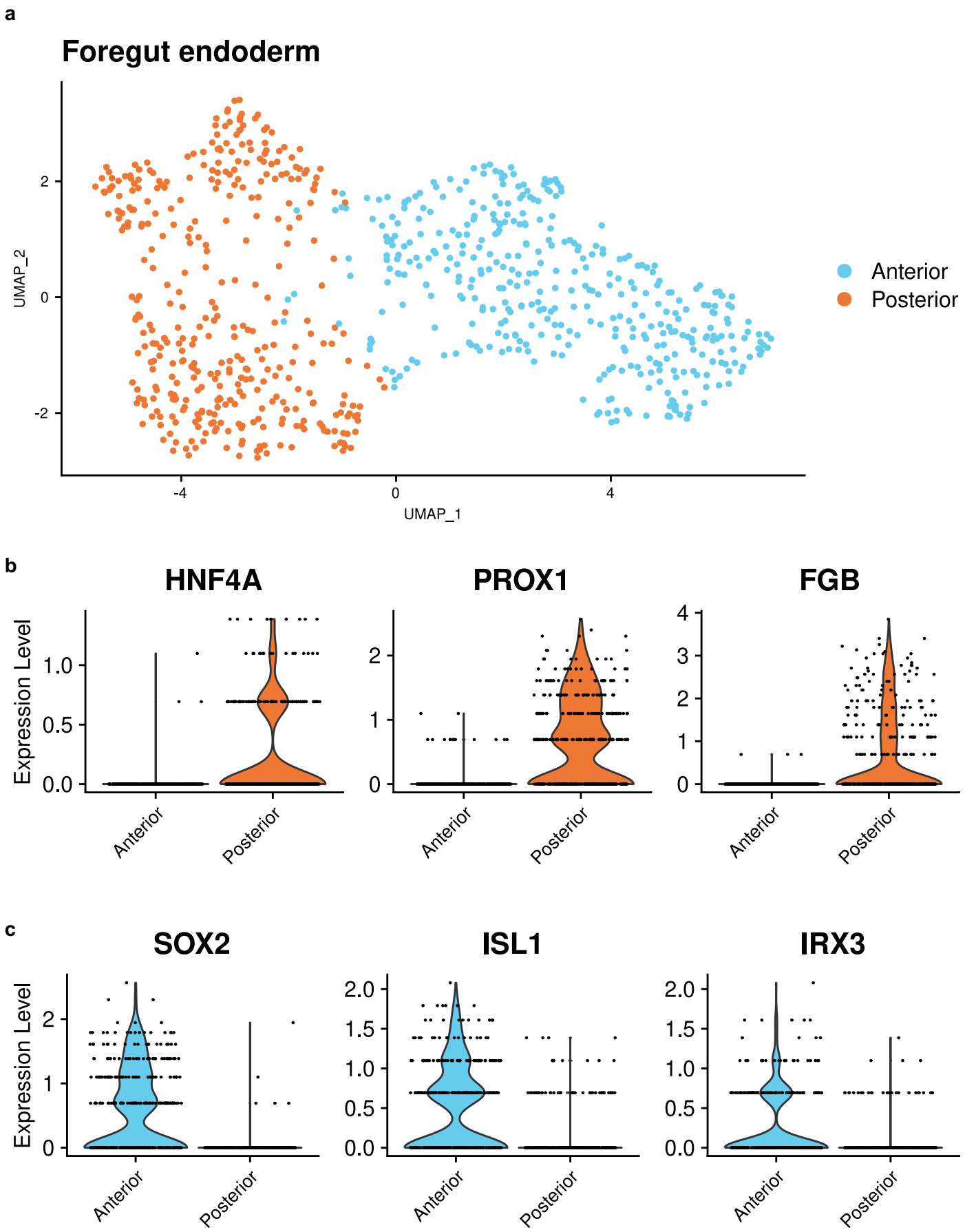
c



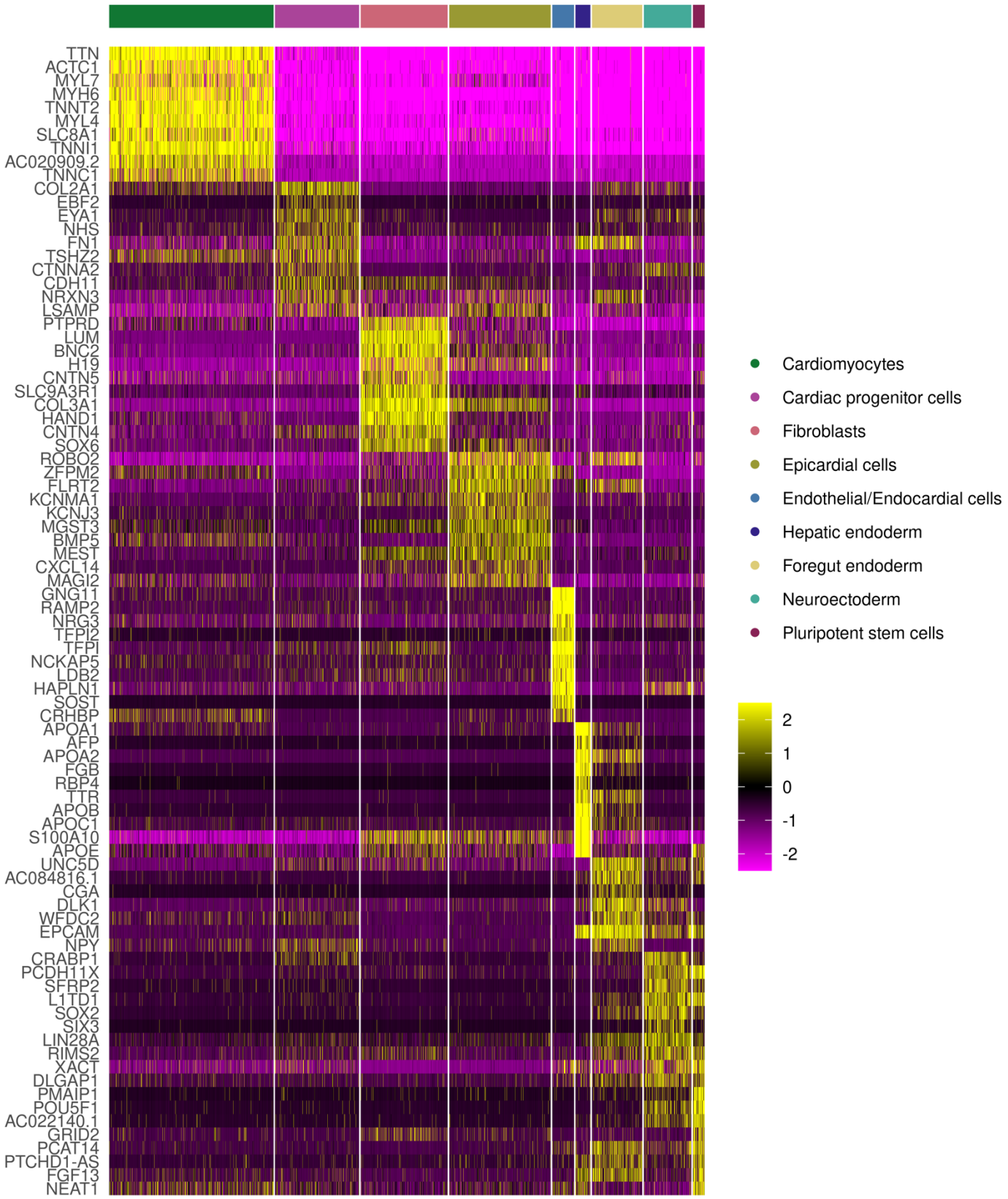
Extended Figure 1 | **a-c,** Representative images of guided differentiation cultures (day 10) for cell lines 19114 (a), 19130 (b), and 19152 (c).



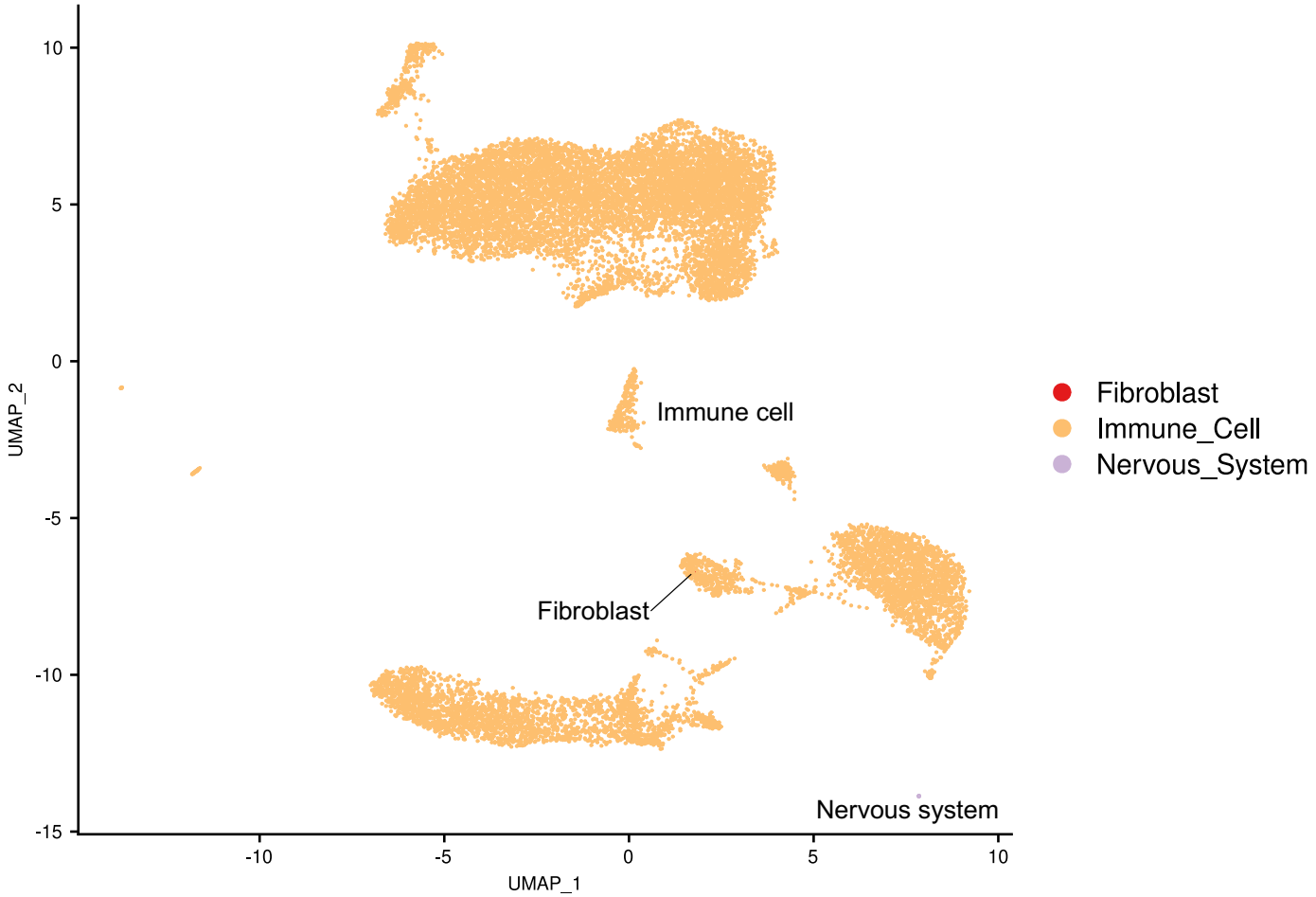
Extended Figure 2 | a-c, UMAP plots of individual cell lines**,** manually annotated, alongside overall proportion of each cell type. Cell line 19114, n = 2,798. Cell line 19130, n = 3,124. Cell line 19152, n = 3,014.



Extended Figure 3 | **a,** UMAP plot of subclustered foregut endoderm cells (n = 772). **b,** Violin plots comparing each cell subcluster for expression of posterior foregut marker genes. **c,** Violin plots comparing each cell subcluster for expression of anterior foregut marker genes.



Extended Figure 4 | Heatmap of the top 10 genes (per cell cluster) according to differential expression (log fold change >2).



Extended Figure 5 | UMAP of classified cell types from peripheral blood mononuclear cells (PBMCs), used here as a non-cardiac control (n = 15,297).