

Supplementary Information

Single-cell Transcriptomic Profiling Unveils Dysregulation of Cardiac Progenitor Cells and
Cardiomyocytes in a Mouse Model of Maternal Hyperglycemia

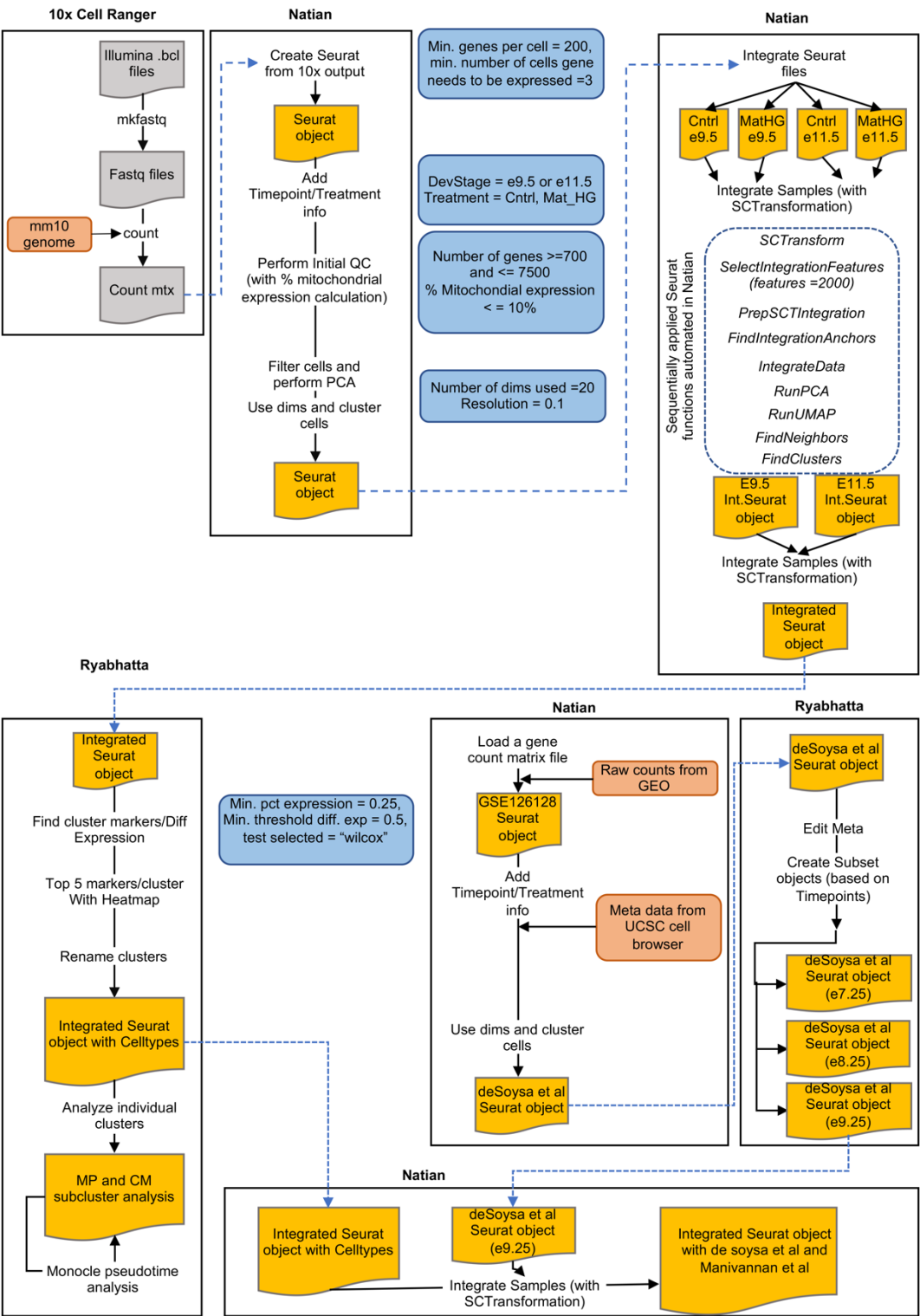
Authors: Sathiyarayanan Manivannan^{1,2#}, Corrin Mansfield^{1,2#}, Xinmin Zhang³,
Karthik. M. Kodigepalli⁴, Uddalak Majumdar^{1,2}, Vidu Garg^{1,2,5,6}, and Madhumita Basu^{1,2,5*}

Affiliations:

¹ Center for Cardiovascular Research, Abigail Wexner Research Institute at Nationwide
Children's Hospital, Columbus, Ohio, United States of America. ² Heart Center,
Nationwide Children's Hospital, Columbus, Ohio, United States of America. ³ BioInfoRx
Inc. Madison, Wisconsin, United States of America. ⁴ Department of Pediatrics, Medical
College of Wisconsin, Milwaukee, Wisconsin, United States of America. ⁵ Department of
Pediatrics, The Ohio State University College of Medicine, Columbus, Ohio, United States
of America. ⁶ Department of Molecular Genetics, The Ohio State University, Columbus,
Ohio, United States of America.

30 **Supplementary Figures and Captions**

31

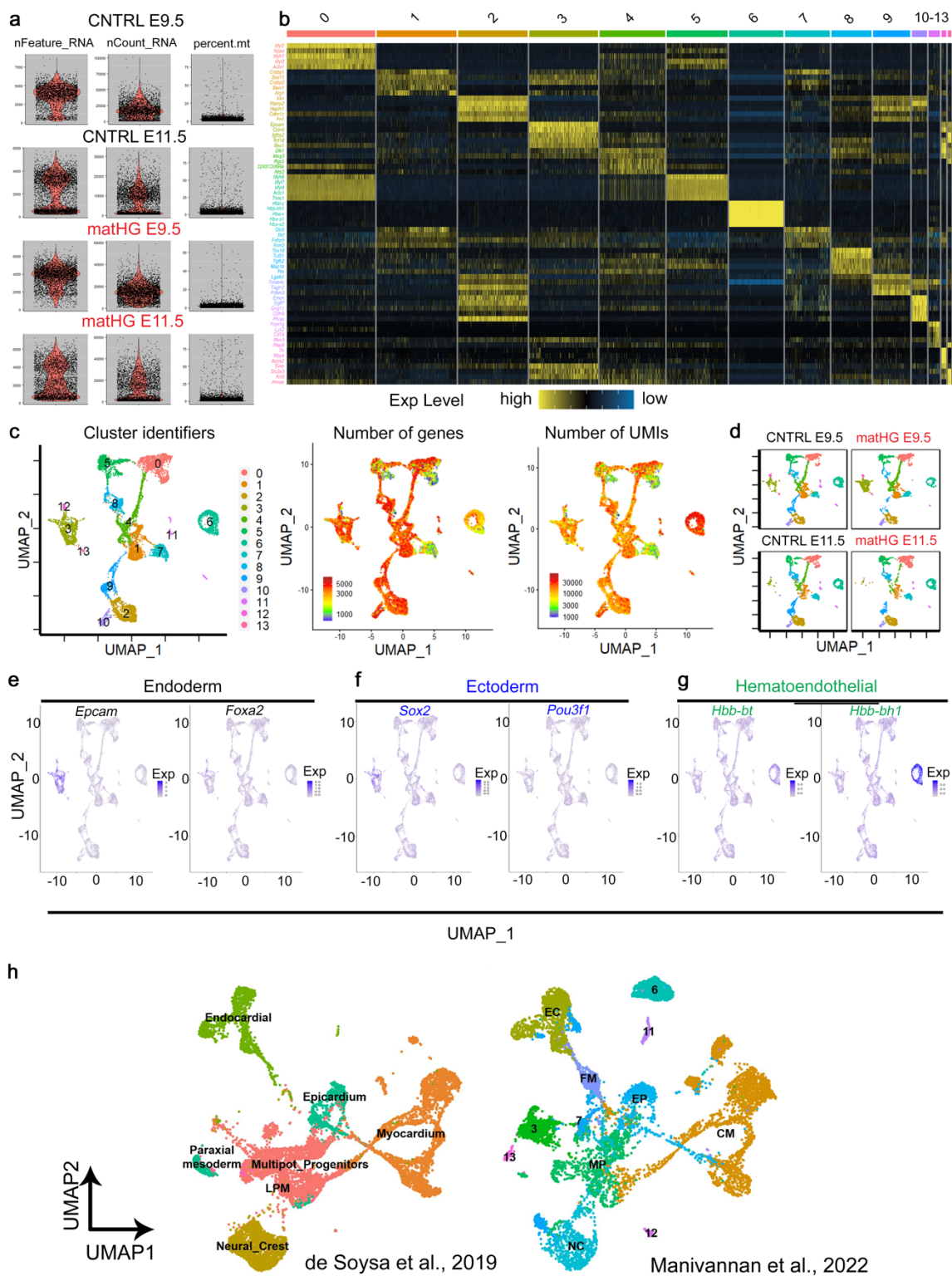


32

33

Supplementary Figure 1: Data processing pipelines and filtering cutoffs

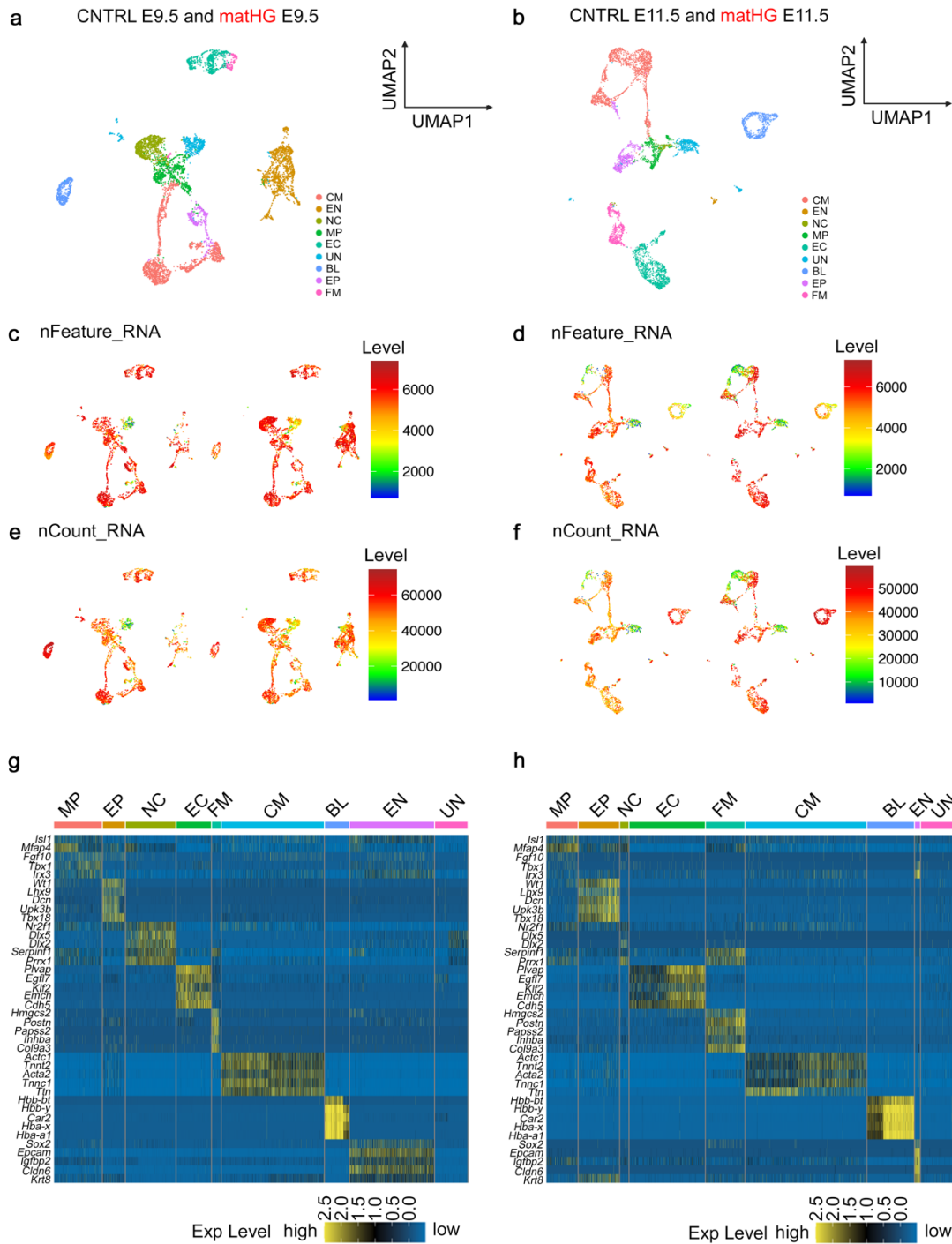
Flow diagram shows processing pipelines used to process and integrate individual samples. Primary processing steps were done using 10x Genomics Cell Ranger. The secondary processing and integration of individual samples were performed using Natian and Ryabhatta (Graphical user interfaces to run Seurat processing steps and analysis functions). Input data (references, GEO counts, metadata) to each step shown in salmon box, individual primary processing files shown in grey document shapes, secondary process files shown in orange document shapes. Parameters used for filtering and processing steps shown in blue boxes. Seurat functions used in integration in Natian (implicitly run) are shown in box with dotted outline. Data moved from one app to another are shown in blue dotted lines. A free (open access) manual is available through the following link: <https://natian-and-ryabhatta.web.app/>.



48
49
50
51

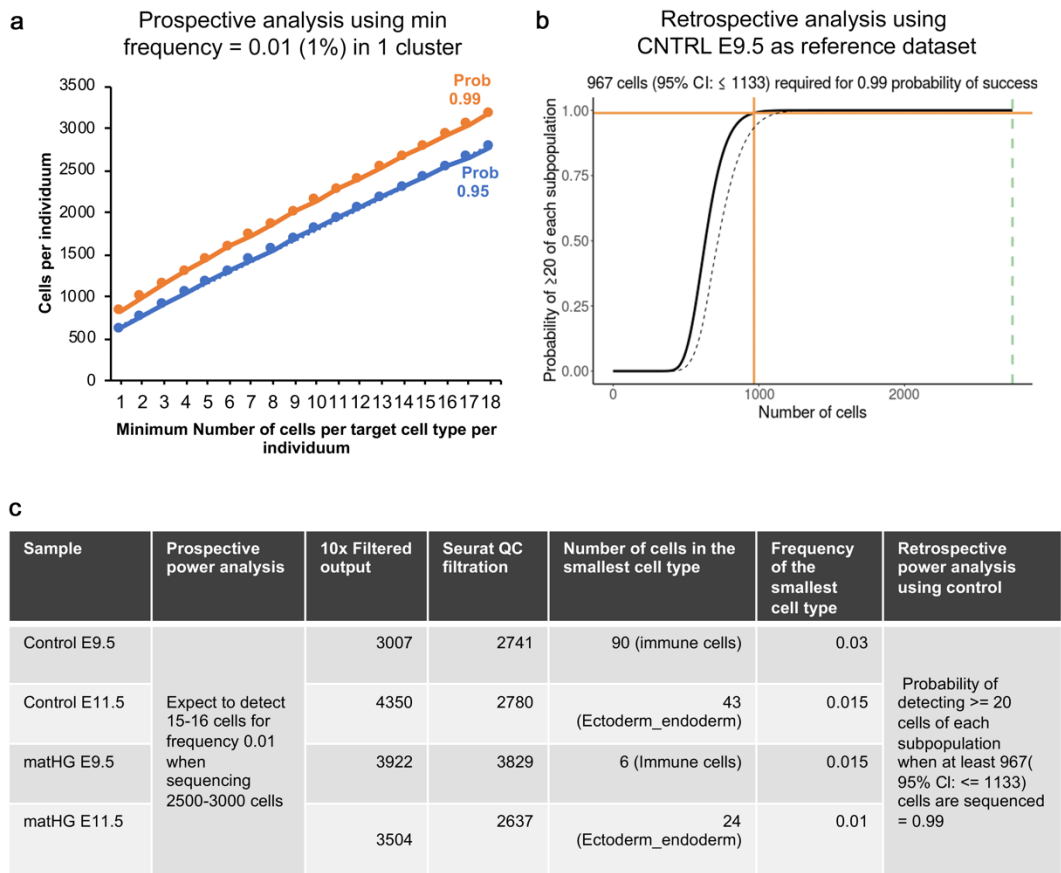
Supplementary Figure 2. Quality control and normalization of *in vivo* scRNA-seq data

(a) Quality control metrics used to assess the quality of 10xscRNA-seq libraries from CNTRL and matHG-exposed E9.5 and E11.5 hearts. Violin plots illustrating number of genes detected in each cell (nFeature-RNA), unique molecular identifiers (nCount-RNA) and less than 10% of the reads mapped to mitochondrial genes in each of the four cardiac tissue samples. **(b)** Unsupervised clustering shows a total of 14 clusters (0-13) and the top five marker genes per cluster in the heatmap. Normalized log expression levels are shown in yellow (high expression) and dark blue (low expression). **(c, d)** UMAP plots show the cluster identities and relationship between the number of genes and UMIs in merged datasets and in all four samples. **(e-g)** UMAP plots showing the expression of *Epcam*⁺, *Foxa2*⁺ endodermal, *Sox2*⁺, *Pou3f1*⁺ ectodermal and *Hbb-bt*⁺*Hbb-bh1*⁺ hematoendothelial clusters were discarded from subsequent analysis. **(h)** UMAP-plot (left panel) shows single-cell data from de Soysa et al., 2019, E9.25 *wt* embryonic heart cells (GSE126128) representing cell types labelled per metadata information from authors¹. UMAP-plot (right panel) shows single-cell data from current study, showing labels identified using marker genes in different clusters. UMAP plots generated using Ryabhatta from integrated data with both deSoysa et al;¹ and data from our study (see **Supplementary Fig. 1** for data processing details). CNTRL, control, matHG, maternal hyperglycemia, wt, wildtype, UMAP, Uniform manifold approximation and projection, UMI, unique molecular identifiers, MP, multipotent progenitors, EP, epicardial, NC, neural crest, EC, endocardial/endothelial, FM, fibromesenchymal, CM, cardiomyocytes, LPM, lateral plate mesoderm.



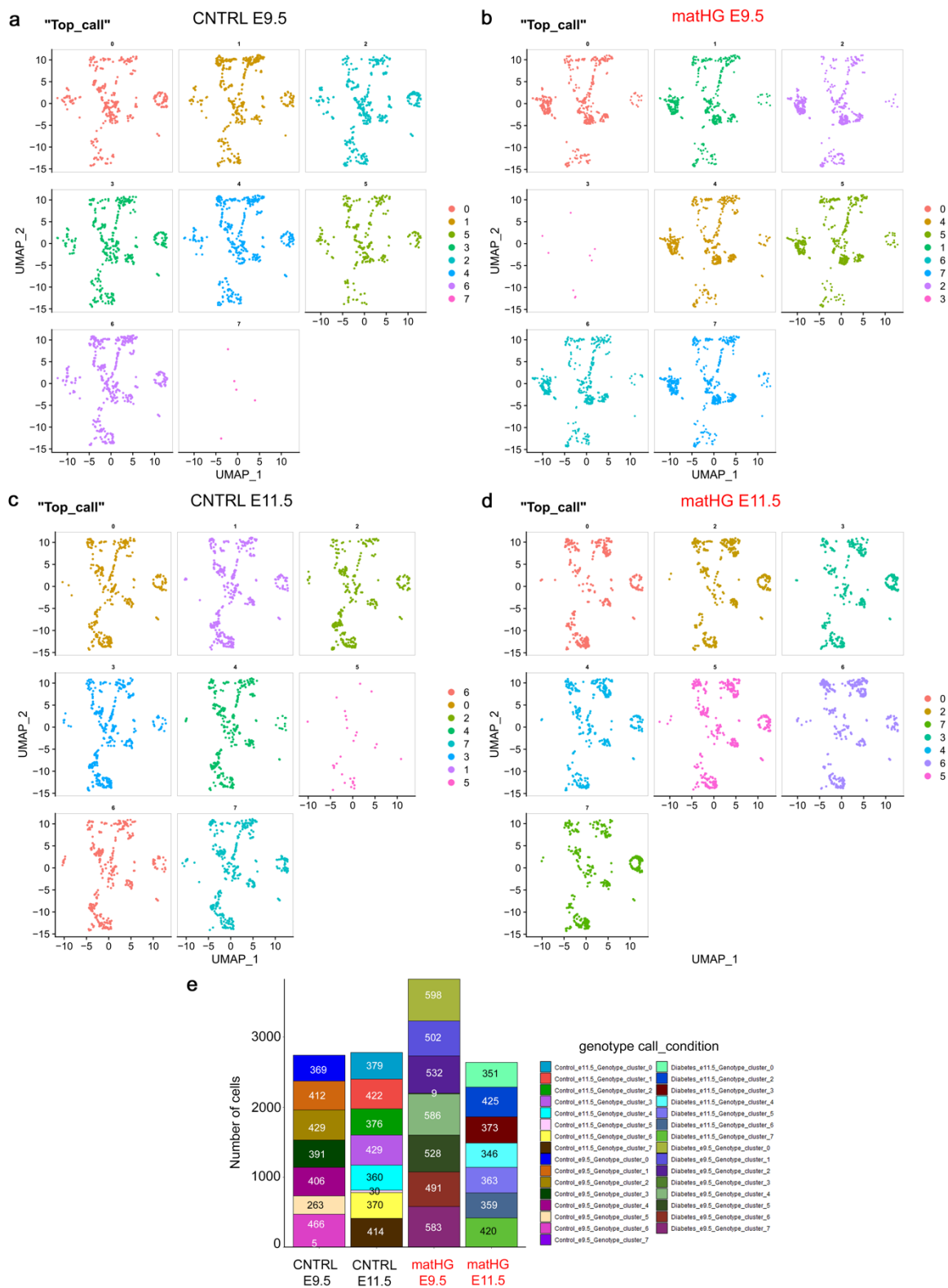
Supplementary Figure 3. Developmental stage specific unsupervised clustering of control and matHG-exposed scRNA-seq data

(a, b) UMAP plots show distribution of nine clusters (CM, EN, NC, MP, EC, UN, BL, EP, FM) separated based on developmental stages (E9.5 and E11.5), when exposed to maternal CNTRL and HG environment. Colors indicate cluster identities. **(c-f)** Individual UMAP plots show the relationship between the number of genes (nFeature-RNA) and UMIs (nCount-RNA) in E9.5 and E11.5 samples subjected to intrauterine CNTRL and matHG environment. **(g, h)** Unsupervised clustering of CNTRL and matHG-exposed E9.5 and E11.5 single cell transcriptomes show nine clusters and the top five marker genes per cluster in the heatmap. Normalized log expression levels are shown in yellow (high expression) and dark blue (low expression). CNTRL, control, matHG, maternal hyperglycemia, CM, cardiomyocytes, EN, endoderm, NC, neural crest, MP, multipotent progenitor, EC, endocardial/endothelial, UN, unknown/unidentified, BL, blood, EP, epicardial, FM, fibromesenchymal, UMI, unique molecular identifiers.



Supplementary Figure 4: Power analysis on Single-cell sample size

(a) Statistical power analysis performed using SCOPIT using both prospective and retrospective mode. Graph shows total number of cells in each individual sample (Cells per individual) required to the minimum number of target cells in a rare cell type (population frequency 0.01 or 1%) in 1 cluster with either 0.99 or 0.95 probability of capture. (b) Graph shows the probability of capturing greater than or equal to 20 cells of a rare population with probability of 0.99 with dotted lines showing 95% confidence interval. (c) Table listing the results of prospective and retrospective power analysis, and the number of cells identified using the Cell Ranger non-droplets, and filtered cell through tertiary process steps.

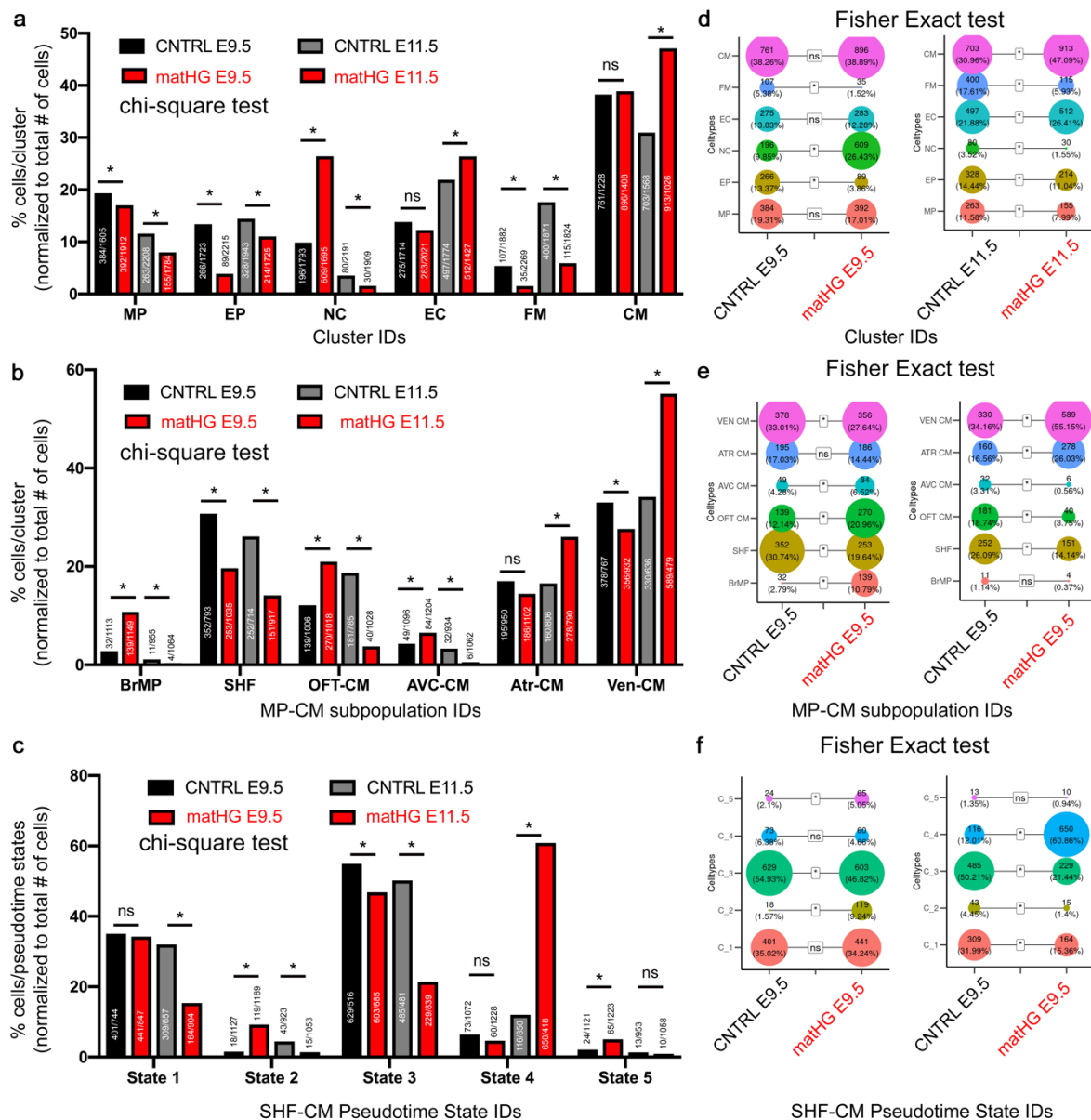


Supplementary Figure 5: Post-hoc sequencing genotyping of single-cell data

(a-d) UMAP plots shown for each sample, generated by adding the predicted genotype calls (0-7) as meta data for each cell in the Seurat object of individual samples. The UMAP plots were split according to the genotype calls (indicated by colors) using split on meta button in Ryabhatta. **(e)** Stacked-bar plot showing the number of cells called for each genotype in CNTRL and matHG-exposed E9.5 and E11.5 hearts using Souporecell. Colors in the bar indicate genotype calls per timepoint per maternal condition. CNTRL, control, matHG, maternal hyperglycemia, UMAP, Uniform manifold approximation and projection.

Supplementary Figure 6. Cluster identification based on discrete gene expression profile

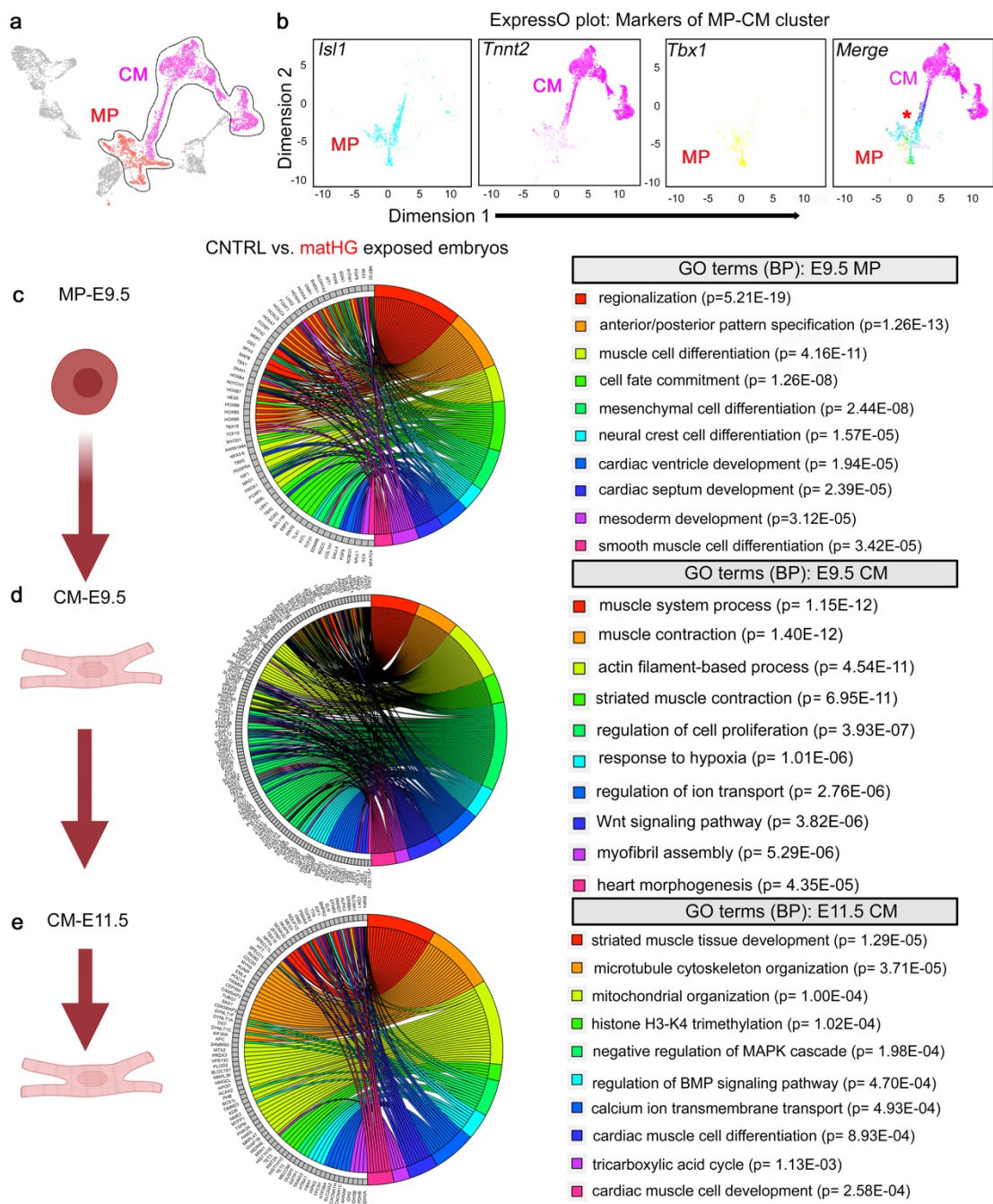
(a) UMAP plots illustrate cluster-specific expression of highly expressed marker genes applied to classify six broadly defined cardiac cell populations. The scale indicates Z-scored expression values (red = high expression, grey = no expression). Statistical tests for differential gene-expression applied to 8503 cells. **(b)** Dot plots represent the expression of lineage specific marker genes across MP, EP, NC, EC, FM, and CM clusters. Each dot is sized to represent the percentage of cells of each type expressing the marker gene and colored to represent the average expression of each marker gene across all cells, as shown in the key (dark blue = high expression and light blue = low expression). MP, multipotent progenitor, EP, epicardial, NC, neural crest EC, endocardial/endothelial, FM, fibromesenchymal and CM, cardiomyocytes, UMAP, Uniform manifold approximation and projection.



Supplementary Figure 7. Maternal hyperglycemia alters the proportion of cells in each cluster and pseudotime states in developing embryonic hearts

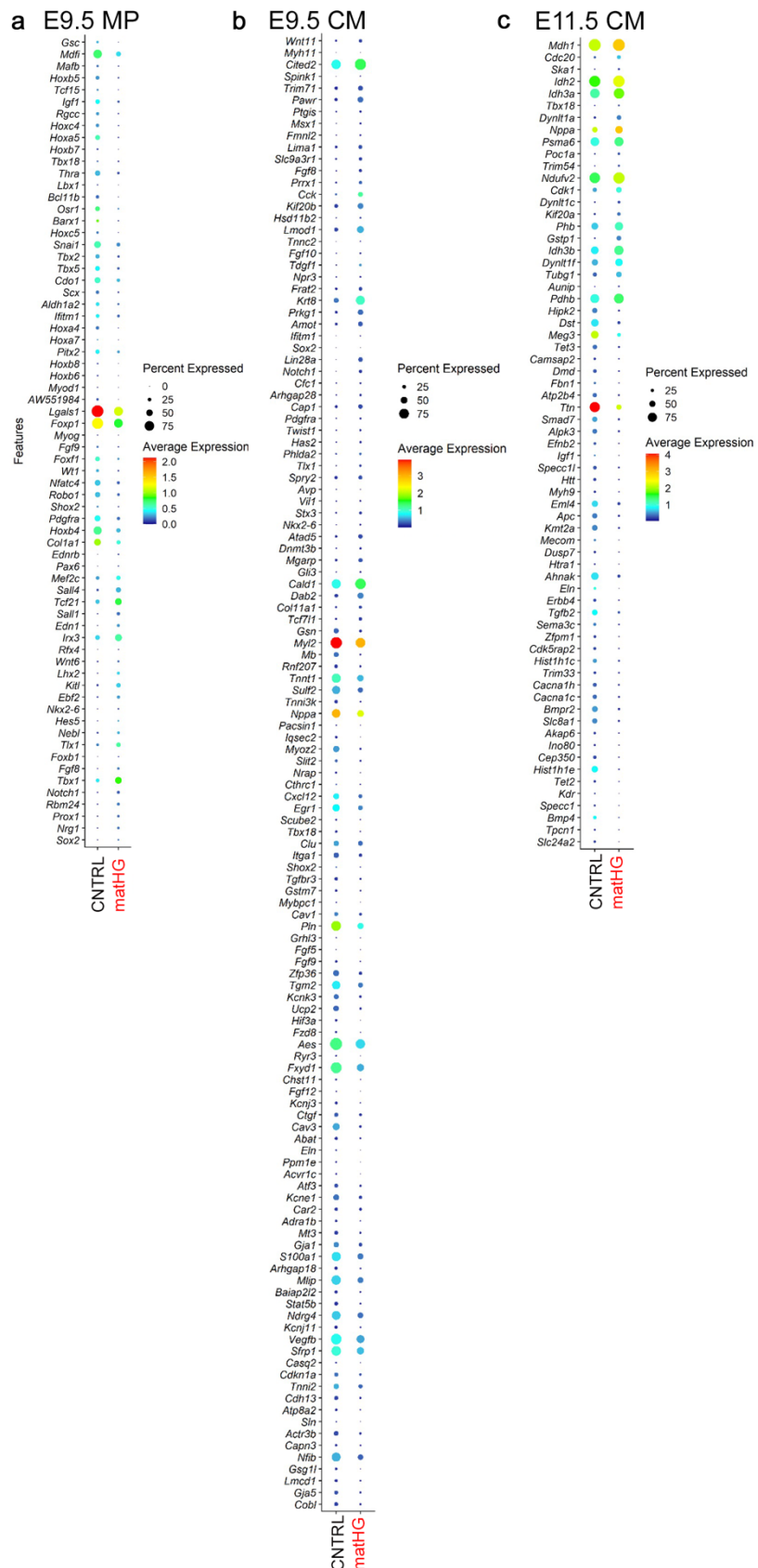
(a-c) Bar graphs show the distribution of cells in six clusters (a), in six MP-CM subclusters (b) and in five pseudotime states (c) from CNTRL and matHG exposed E9.5 and E11.5 hearts. For cell number comparisons, cells were normalized to the total number of cells analyzed, embedded in the individual bar graphs. Statistical significance between groups

determined by Chi-square test. **(d-f)** Bubble plots show the distribution of cells in six clusters **(d)**, in six MP-CM subclusters **(e)** and in five pseudotime states **(f)** from CNTRL and matHG exposed E9.5 and E11.5 hearts. Statistical significance between groups determined by Fisher Exact test. ns= non-significant and * indicates two-tailed p-value $\leq$ 0.05. CNTRL, control, matHG, maternal hyperglycemia, MP, multipotent progenitor, CM, cardiomyocytes, EP, epicardial, EC, endocardial/endothelial, FM, fibromesenchymal, NC, neural crest, BrMP, branchiomic muscle progenitors, SHF, second heart field, OFT, outflow tract, AVC, atrioventricular canal, Atr, atrial, Ven, ventricular.



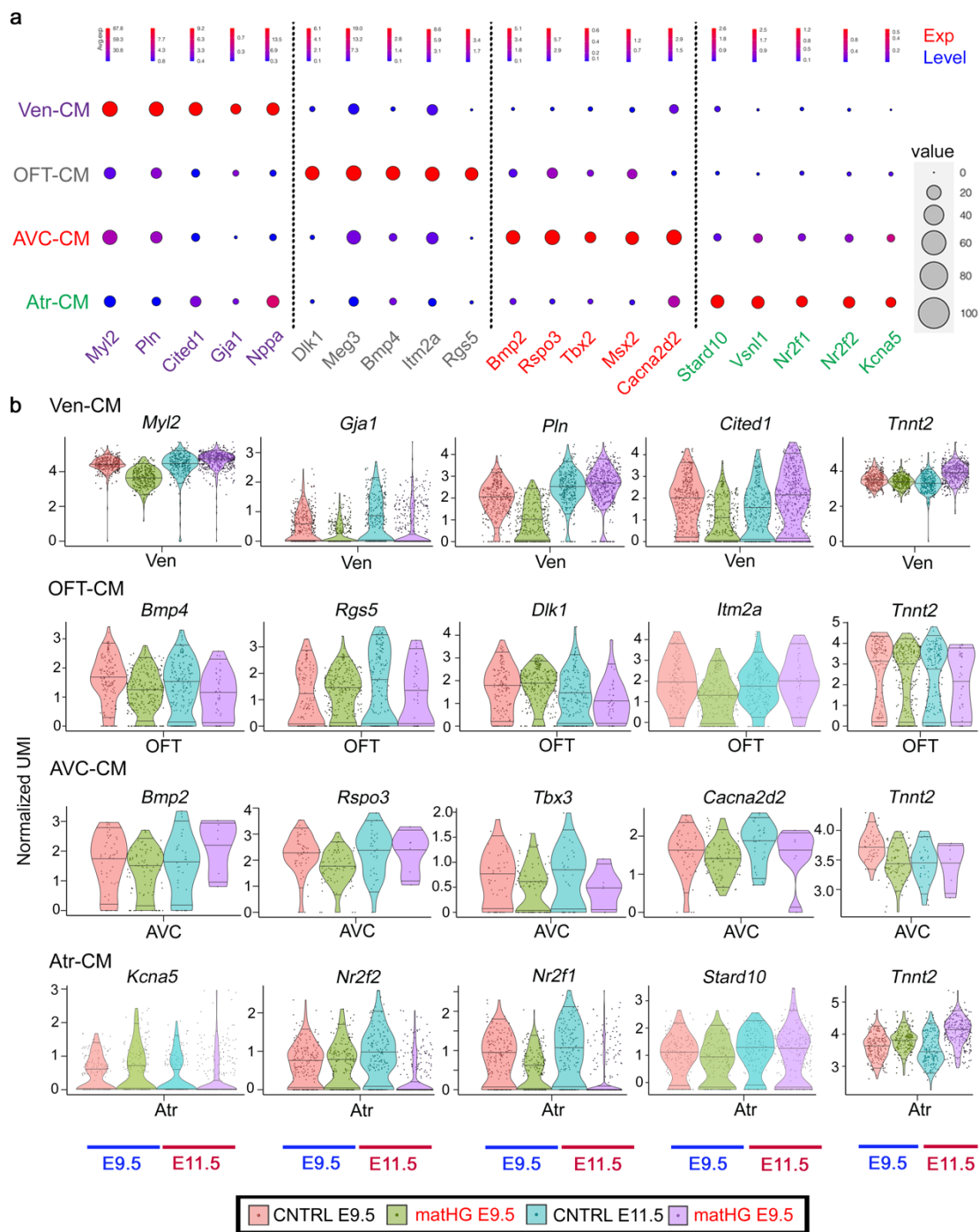
Supplementary Figure 8. Transcriptomic analysis of matHG-exposed hearts reveals gene expression differences in cardiac progenitor population and in cardiomyocytes.

(a) Merged UMAP plot demonstrate MP and CM clusters in CNTRL and matHG-exposed E9.5 and E11.5 hearts. Statistical tests for differential gene-expression applied to 4467 cells. **(b)** ExpressO plot show expression of well-known markers of MP (*Isl1* and *Tbx1*) and CM (*Tnnt2*) clusters. * Indicates presence of *Isl1⁺Tnnt2⁺* cells. **(c-e)** GOplots represent the analysis of the gene ontology (GO) terms enriched among the 262 DEGs in E9.5 MP, 357 DEGs in E9.5 CM and 326 DEGs in E11.5 CM clusters (DEG cutoff: $\text{Log2Foldchange} \geq +1$ or ≤ -1 and $p_{\text{adjusted}} \leq 0.05$). The left side of the circle displays the gene, and the right side shows the GO term associated biological processes. The assorted colors represent different GO terms, and the color of each GO terms is annotated, p-values represent enrichment of GO term. CNTRL, control, matHG, maternal hyperglycemia, MP, multipotent progenitors, CM, cardiomyocytes, DEG, differentially expressed gene, GO, gene ontology, BP, biological processes.



Supplementary Figure 9. Differentially expressed gene analysis in cardiac progenitor population and in cardiomyocytes.

(a-c) Bubble plots show differentially expressed genes in E9.5 MP, E9.5 and E11.5 CM using DESeq2. The statistical test applied for DESeq2 (pseudobulk) is the Wald test, with the P-value adjusted for multiple testing using the Benjamini and Hochberg method by default. Y-axis label represents genes, and X-axis label maternal diabetic status. Size of the bubble represent percentage of differentially expressed genes enriched in CNTRL vs. matHG-exposed MP and CM. The color of the bubble indicates expression level, where blue and red represents low and high gene expression levels, respectively. Statistical tests for differential gene-expression applied to 4467 cells. CNTRL, control, HG, hyperglycemia, MP, multipotent progenitors, CM, cardiomyocytes.

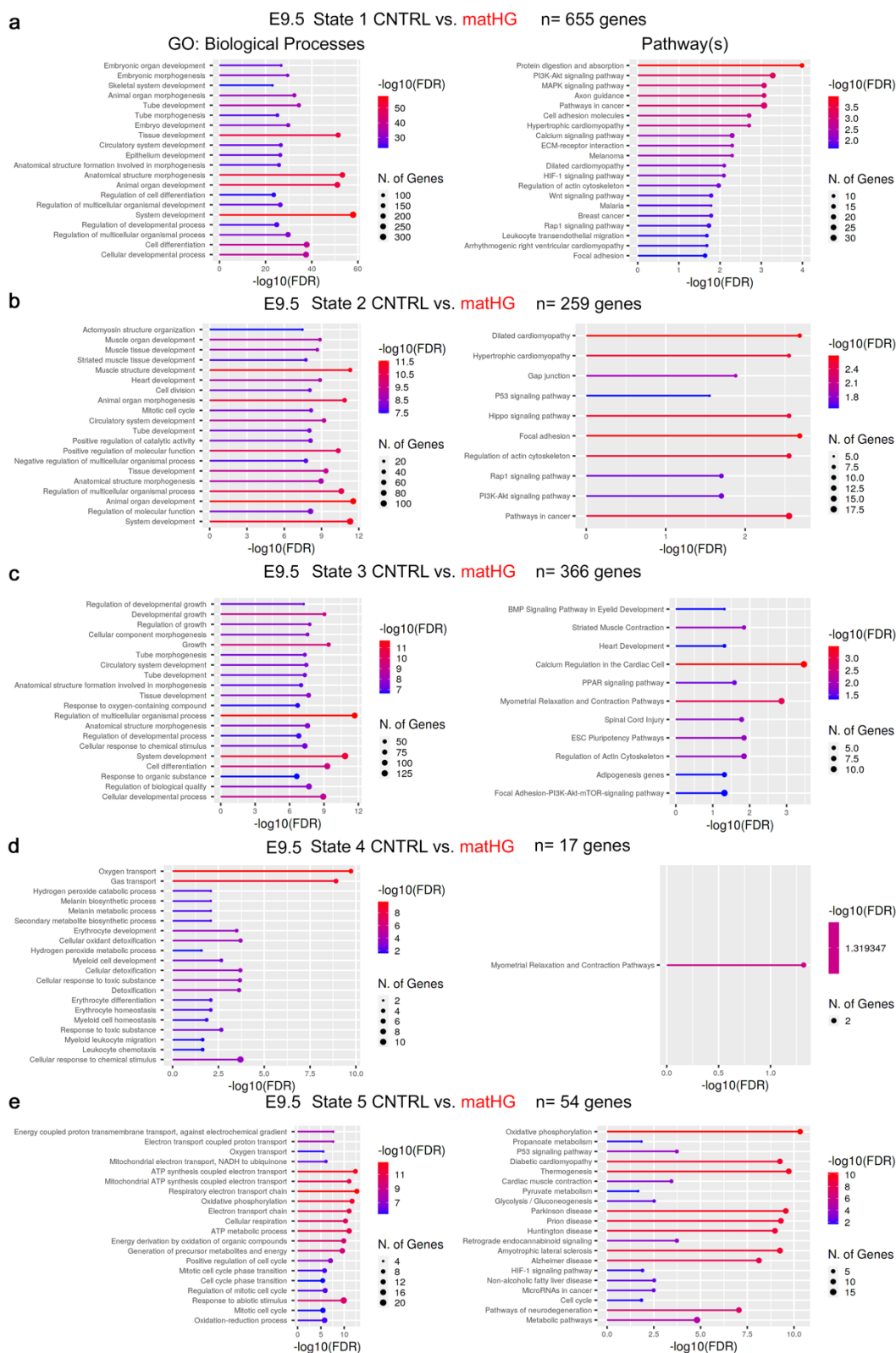


187
188
189
190

Supplementary Figure 10. Impact of maternal hyperglycemia on cardiomyocyte subtypes.

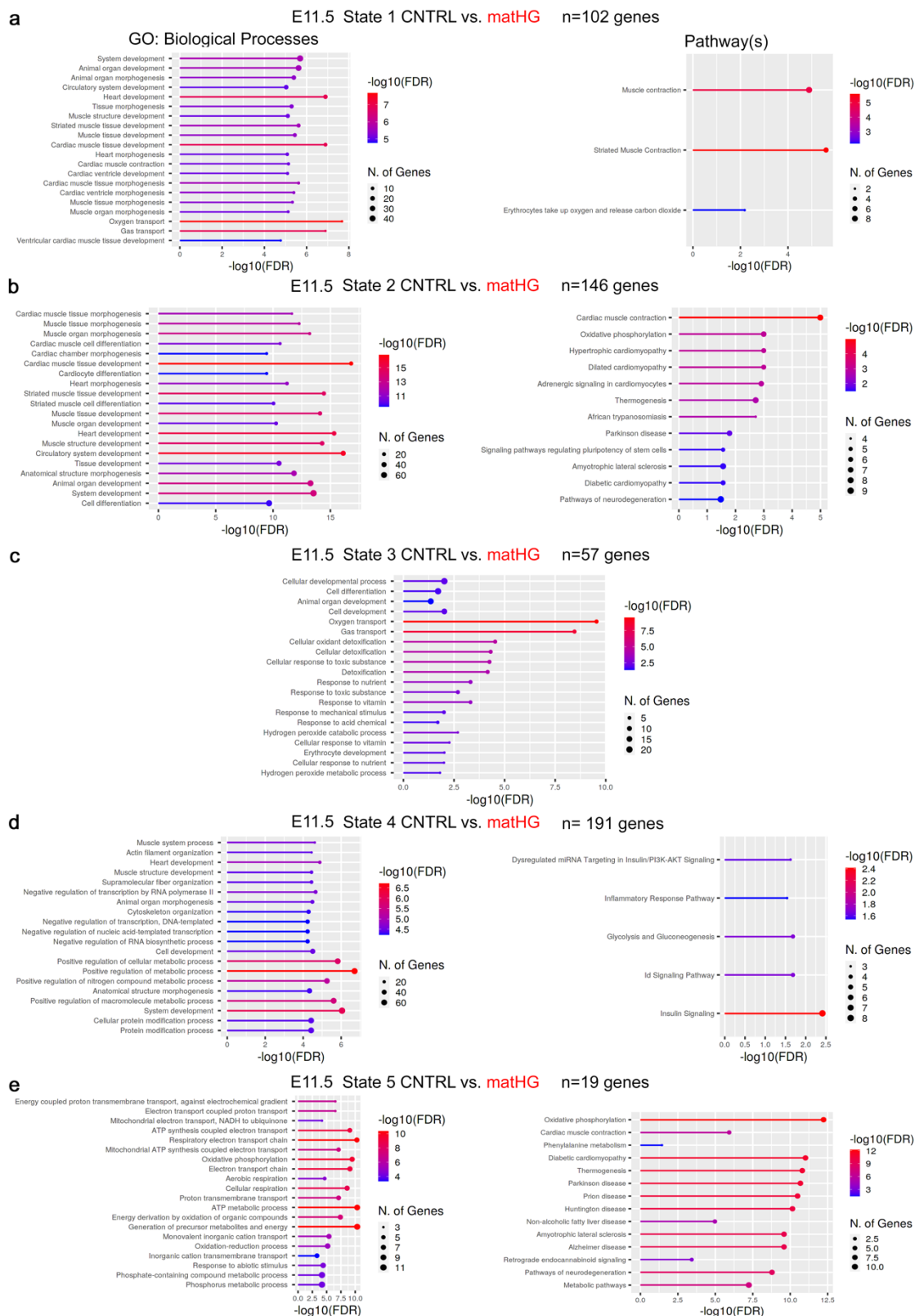
(a) Dot plots represent the expression of known marker genes across CM subtypes (Ven, OFT, AVC, and Atr-CM). Statistical tests for differential gene-expression applied to 3273 cells. Each dot is sized to represent the proportion of cells of each type expressing the marker gene and colored to represent the average expression of each marker gene across all as shown in the key (red = high expression and dark blue = low expression).

(b) Violin plots demonstrate normalized UMI or gene-expression levels across CM subtypes in CNTRL and matHG-exposed E9.5 and E11.5 hearts. Each dot in the violin plot represents individual cells. CNTRL, control, matHG, maternal hyperglycemia, UMI, unique molecular identifier, Ven, ventricular, OFT, outflow tract, AVC, atrioventricular canal, and Atr, atrial, CM, cardiomyocytes.



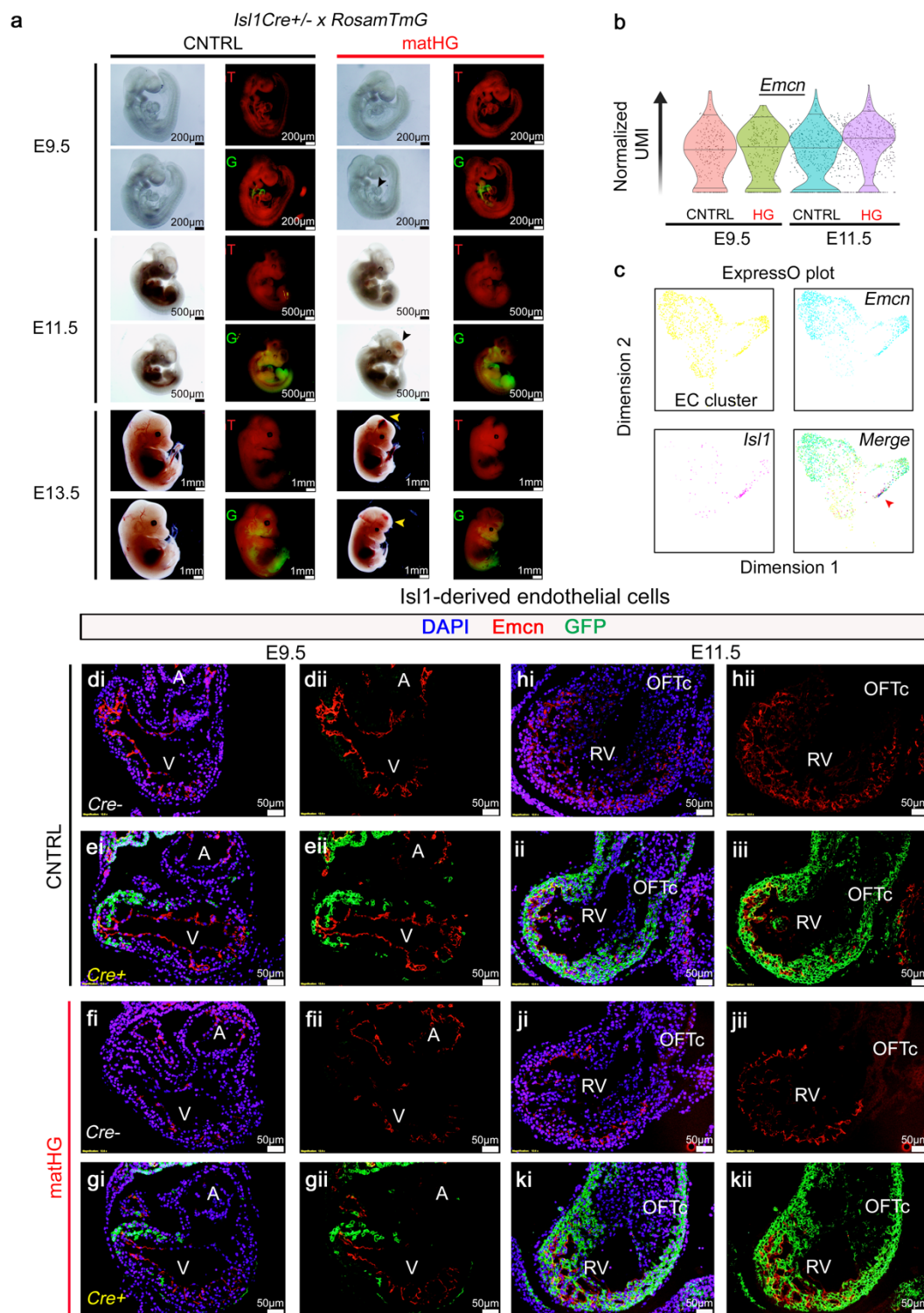
Supplementary Figure 11. Transcriptional changes in MP-CM subpopulations at E9.5 across the pseudotime trajectories

(a-e) Lollipop charts illustrate the top 20 GO-terms for Biological Processes and Pathways, sorted as descending negative logarithmic adjusted p-value of enrichment analysis ($-\log_{10}(\text{FDR})$) shown in the key. Numbers in circles represent the number of DEGs matched to a specific GO term. Five pseudotime states were compared between CNTRL and mathHG exposed E9.5 MP-CM subclusters. GO, gene ontology, FDR, false discovery rate, DEGs, differentially expressed genes, CNTRL, control, mathHG, maternal hyperglycemia.



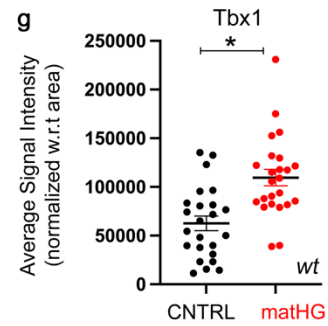
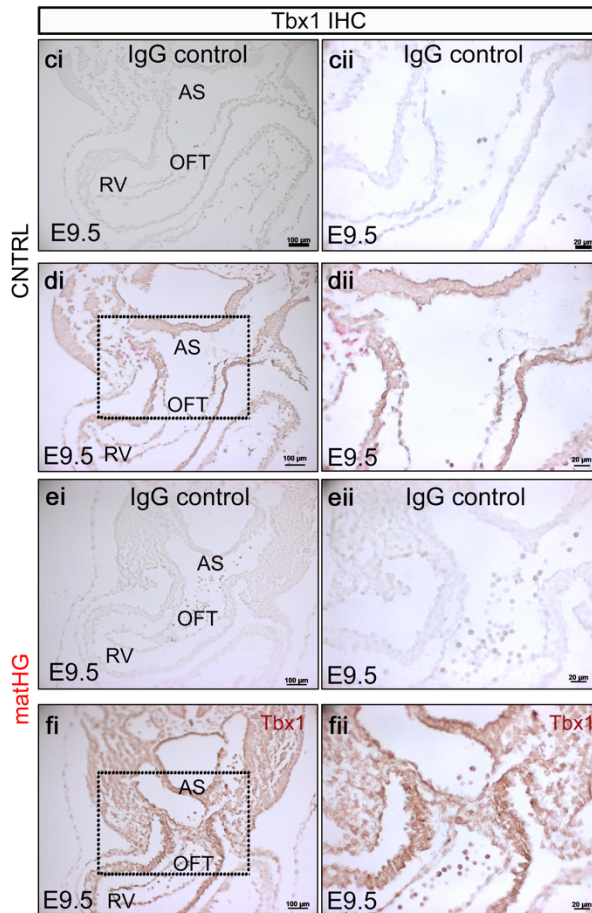
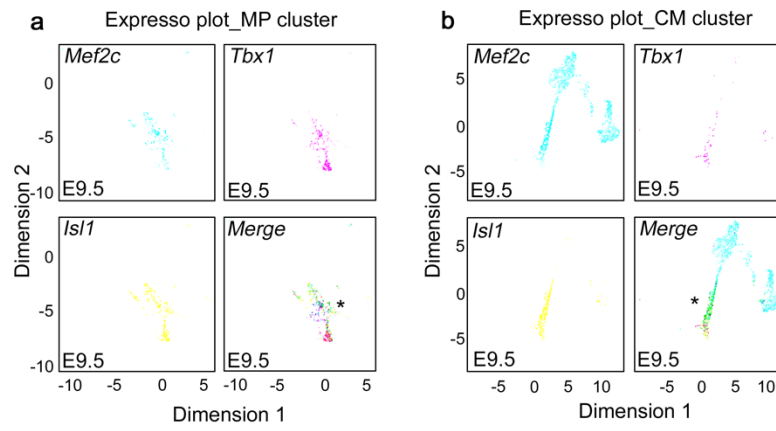
Supplementary Figure 12. Transcriptomic changes in MP-CM subpopulations at E11.5 across the pseudotime trajectories

(a-e) Lollipop plots represent the top 20 GO-terms for Biological Processes and Pathways, sorted as descending negative logarithmic adjusted p-value of enrichment analysis ($-\log_{10}(\text{FDR})$) shown in the key. Numbers in circles represent the number of DEGs matched to a specific GO term. Five pseudotime states were compared between CNTRL and mathHG exposed E11.5 MP-CM subclusters. No significantly enriched pathways were noted in cells present in State 3 at E11.5. GO, gene ontology, FDR, false discovery rate, DEGs, differentially expressed genes, CNTRL, control, mathHG, maternal hyperglycemia.



Supplementary Figure 13. *Isl1*-cell fate mapping studies under maternal hyperglycemic exposure.

(a) Representative whole-mount brightfield and fluorescence images of *Isl1**Cre*⁻; *Rosa*^{mTmG/+} (T; red, tdTomato⁺) and *Isl1**Cre*⁺; *Rosa*^{mTmG/+} (G; green, GFP⁺) littermate control hearts at E9.5, E11.5 and E13.5 exposed to CNTRL and matHG environment. GFP⁺ expression demonstrate the endogenous pattern of *Isl1* expression, and those genetically labelled descendants derived from *Isl1* progenitors. MatHG-exposed *Isl1**Cre*⁺; *Rosa*^{mTmG/+} embryos show intracerebral hemorrhage and exencephaly at E11.5 and E13.5 (shown in black and yellow arrowheads). **(b)** Violin plots show normalized UMI or expression of *Emcn*, an endothelial cell marker, in E9.5 and E11.5 scRNA-seq data. **(c)** ExpressO plots display *Isl1*⁺*Emcn*⁺ cells in EC cluster (red arrowheads). **(d-k)** Representative immunofluorescent images show GFP (green) and *Emcn* (red) protein expression in CNTRL and matHG-exposed Cre⁻ and Cre⁺ littermate controls (n≥2 embryos/timepoint/ maternal condition). Nuclei stained with DAPI shown in blue. *Isl1*-derived *Emcn*⁺GFP⁺ ECs shown in yellow. CNTRL, control, matHG, maternal hyperglycemia, EC, endocardial/endothelial. A, atria, V, ventricle, OFTc, outflow tract cushion, RV, right ventricle. Scale bars: 200μm (E9.5), 500μm (E11.5) and 1mm (E13.5) **(a)** and 50μm **(d-k)**.



Supplementary Figure 14. Perturbation in disease associated cardiac progenitor marker, *Tbx1* in response to maternal hyperglycemia.

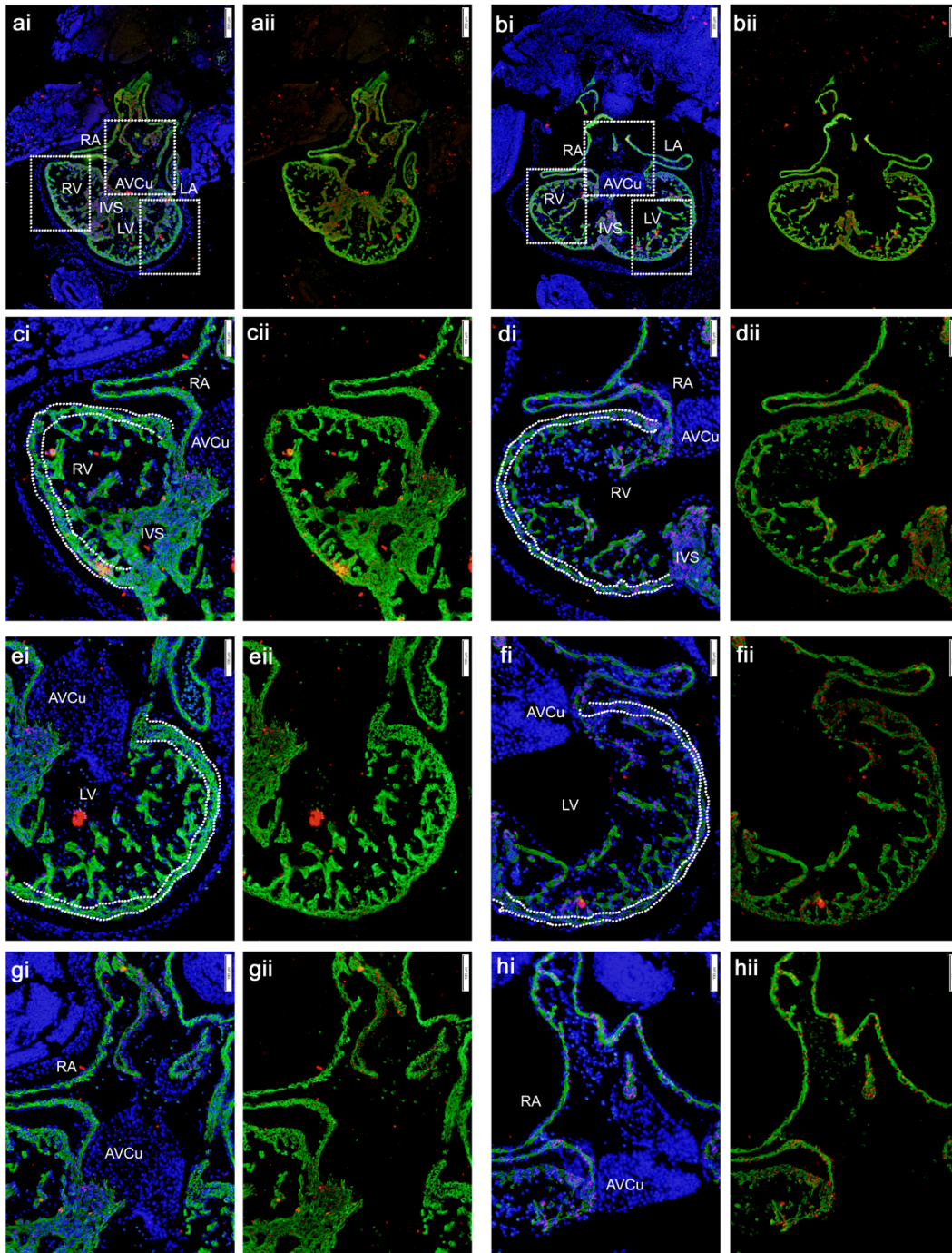
(a, b) ExpressO plots show transcript expression of *Mef2c*, *Tbx1*, and *Is/1* in MP and CM clusters at E9.5. Black asterisks indicate co-expression of these SHF progenitor markers. **(c-f)** Immunohistochemical staining of transverse cardiac sections with *Tbx1* in CNTRL vs. mathHG exposed E9.5 *wt* hearts (n=3 independent embryos/maternal diabetes condition). **cii-fii** show higher magnification images of **ci-fi** (indicated by black rectangular boxes). **ci** and **ei** represent rabbit IgG (negative) control. **(g)** Quantification of *Tbx1* protein expression in CNTRL and mathHG-exposed E9.5 hearts. Statistical comparisons made between CNTRL and mathHG groups by unpaired *t*-test with Welch's correction using GraphPad Prism 9. Data presented as mean $\pm$ SEM. * Indicates two-tailed p-value < 0.05. MP, multipotent progenitors, CM, cardiomyocytes, CNTRL, control, mathHG, maternal hyperglycemia, *wt*, wildtype, OFT, outflow tract, AS, aortic sac, IHC, immunohistochemistry. Scale bars: 100 μ m (**ci-fi**) and 20 μ m (**cii-fii**).

E11.5

DAPI Tropomyosin Nkx2-5

CNTRL

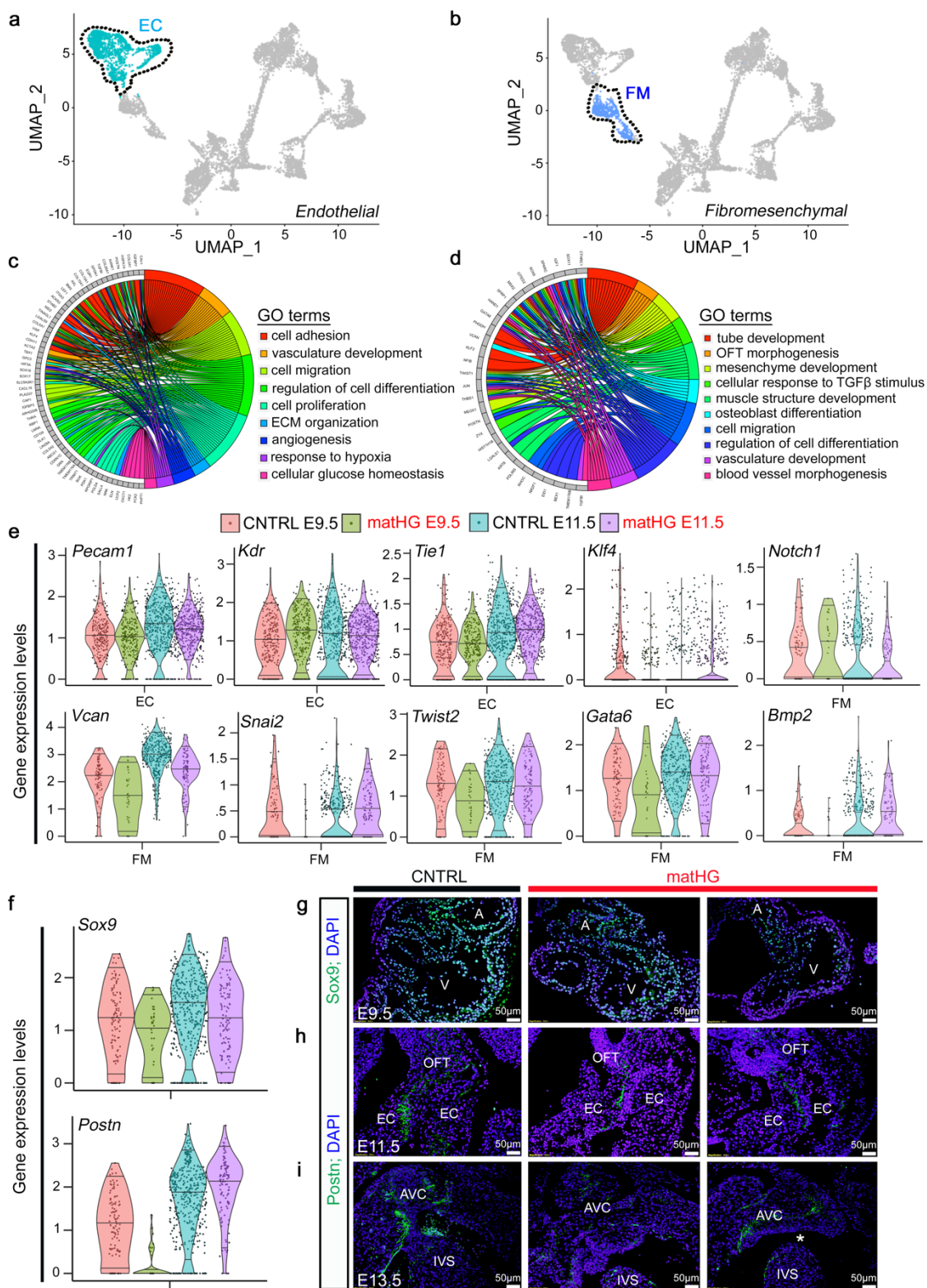
matHG



276
277

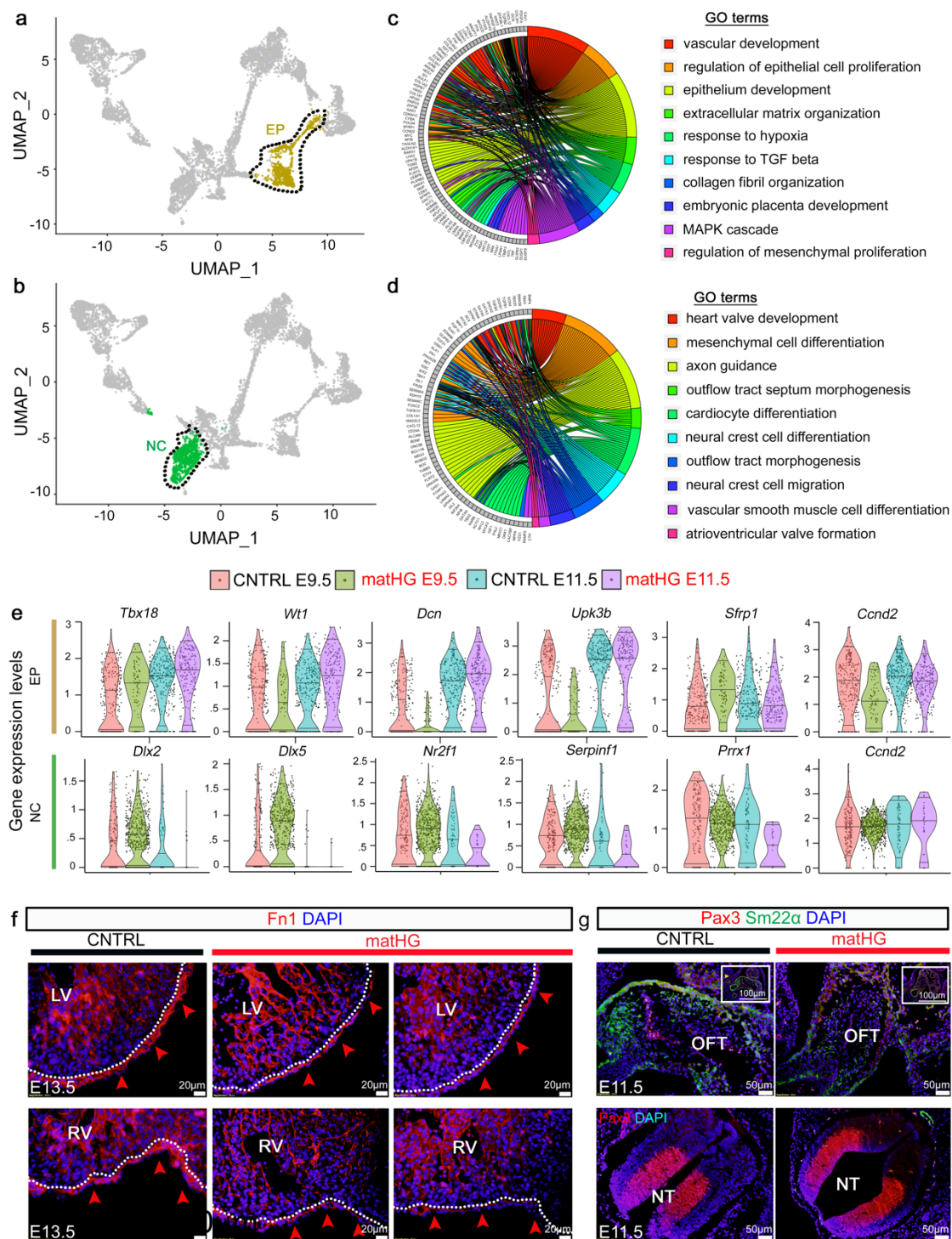
Supplementary Figure 15. Maternal hyperglycemic exposure affects the sarcomeric protein expression in E11.5 cardiomyocytes.

(a-h) Representative immunofluorescence images of CNTRL and matHG-exposed E11.5 hearts (n=2 independent embryos per group) labeled with sarcomeric protein, Tropomyosin and CM marker, Nkx2-5. TPM1+Nkx2-5+ expression found in both the atria and ventricles at the timepoint of analysis and demonstrate a decrease upon matHG exposure. C-H, higher magnification images of A and B, indicated by white boxes. Dotted white lines in C-F show thin myocardial wall and impaired trabeculation in matHG-exposed hearts. CM, cardiomyocytes, CNTRL, control, matHG, maternal hyperglycemia, wt, wildtype, RV, right ventricle, LV, left ventricle, RA, right atria, LA, left atria, IVS, interventricular septum, AVCu, atrioventricular cushion. Scale bars: 200µm (**ai, aii, bi, bii**) and 100 µm (**ci-hi, and cii-hii**).



Supplementary Figure 16. Single-cell RNA-seq reveals transcriptional changes in endocardial/endothelial and mesenchymal cells exposed to maternal hyperglycemia

(a, b) UMAP plots represent the EC and FM clusters from E9.5 and E11.5 embryonic hearts subjected to CNTRL and matHG. **(c, d)** GOplots represent the analysis of the GO terms enriched among the DEGs in E9.5 and E11.5 EC and FM clusters (DEG cutoff: $\text{Log2Foldchange} \geq 1$ or ≤ -1 and $P_{\text{adjusted}} \leq 0.05$). The left side of the circle displays the gene, and the right side shows the GO-term associated biological processes. The assorted colors represent different GO terms. **(e)** Violin plots show the normalized expression levels of highly variable genes in EC (*Pecam1*, *Kdr*, *Tie1*, and *Klf4*) and FM (*Vcan*, *Snai2*, *Twist2*, *Gata6*, *Bmp2*, and *Notch1*) clusters from CNTRL and matHG-exposed E9.5 and E11.5 embryonic hearts. **(f)** Violin plots illustrating *Sox9* and *Postn* gene-expression in E9.5 and E11.5 FM populations exposed to CNTRL and matHG environment. **(g-i)** Panels of immunofluorescent images show protein expression of Sox9 (green) and Postn (green) at E9.5, E11.5 and E13.5 hearts exposed to CNTRL and matHG. Nuclei stained with DAPI shown in blue. White asterisk denotes presence of VSD in matHG-exposed E13.5 embryo. Nuclei stained with DAPI in blue. CNTRL, control, HG, hyperglycemia, GO, gene ontology, FM, fibromesenchymal, A, atria; V, ventricle; OFT, outflow tract; EC, endocardial/endothelial cushion; AVC, atrioventricular canal; IVS, interventricular septum, VSD, ventricular septal defect. Scale bars: **g-i**: 50 μm .



313
314

Supplementary Figure 17. Single-cell RNA-seq reveals transcriptional changes in epicardial and cardiac neural crest cell populations exposed to maternal hyperglycemia

(a, b) UMAP plots represent the EP and NC clusters from E9.5 and E11.5 embryonic hearts subjected to CNTRL and matHG. **(c, d)** GOplots represent the GO terms enriched among the DEGs in E9.5 and E11.5 EP and NC clusters (DEG cutoff: $\text{Log2Foldchange} \geq 1$ or ≤ -1 and $P_{\text{adjusted}} \leq 0.05$). The left-side of the circle displays the gene, and the right-side shows GO-term associated biological processes. The assorted colors represent GO terms. **(e)** Violin plots show the normalized expression levels of highly variable genes in EP (*Tbx18*, *Wt1*, *Dcn*, *Upk3b*, *Sfrp1*, and *Ccnd2*) and NC (*Dlx2*, *Dlx5*, *Nr2f1*, *Serpinf1*, *Prrx1*, and *Ccnd2*) clusters from CNTRL and matHG-exposed E9.5 and E11.5 embryonic hearts. **(f)** Panels of immunofluorescent images show protein expression of *Fn1* (red arrowheads) in the EP from low to high-matHG exposed E13.5 LV and RV compared to CNTRL (indicated by dotted white line). **(g)** Panels of immunofluorescent images illustrate the protein expression of NC-markers, Pax3 (red) and Sm22 α (green) in E11.5 OFT exposed to matHG vs. CNTRL. Insets (white square boxes) show positive Pax3 expression in the NT. Nuclei stained with DAPI in blue. CNTRL, control, HG, hyperglycemia, GO, gene ontology, EP, epicardial, NC, neural crest, LV, left ventricle, RV, right ventricle, OFT, outflow tract, NT, neural tube. Scale bars: **f**: 20 μm and **g**:50 μm .

Supplementary Tables:

Supplementary Table 1: Distribution of *Isl1-Cre^{+/-}*; *Rosa^{mT/mG}* embryos in the setting of control and mathHG environment

Maternal Status	Maternal B.G $\pm$ SD (litters) [#]	Embryonic timepoint	Number of embryos	<i>Rosa^{mT/mG}</i> ; <i>Cre⁻</i>	<i>Rosa^{mT/mG}</i> ; <i>Cre⁺</i>	2-tailed P value (χ^2 test) [*]
CNTRL	220.7 $\pm$ 22.4 (3)	E9.5	28	11	17	0.5908
mathHG	335.0 $\pm$ 58.7 (3)		28	14	14	
CNTRL	205.3 $\pm$ 37.8 (3)	E11.5	22	14	8	0.2949
mathHG	420.8 $\pm$ 191.8 (4)		33	15	18	
CNTRL	244.7 $\pm$ 25.0 (3)	E13.5	26	13	13	0.7930
mathHG	624.0 $\pm$ 66.7 (4)		28	15	13	

[#] Unpaired t test indicates statistically significant differences in maternal B.G. at each developmental stage.

^{*}Chi-square test with Yates's correction indicates differences between number of embryos of each genotype recovered from *Isl1-Cre^{+/-}* and homozygous *Rosa^{mT/mG}* crossings.

Supplementary Table 2. Oligonucleotide sequences

Primer Name	Forward Primer	Reverse Primer
<i>GFP</i>	CTGCTGCCCCGACAACCAC	TGTGATCGCGCTTCTCGTT
<i>Isl1</i>	TCAGGTTGTACGGGATCAAAT GTAGAGGTGCAAAGTTACCAGCC	GTTCCGGCTGCCATTTGCAGA TTAGAGCCTGGTCCTCCTTCTG
<i>Tbx1</i>	CTGTGGGACGAGTTCAATCAG	TTGTCATCTACGGGCACAAAG
<i>Fgf10</i>	TTTGGTGTCTTCGTTCCCTGT	TAGCTCCGCACATGCCTTC
<i>Mef2c</i>	ATCCCGATGCAGACGATTCAG	AACAGCACACAATCTTTGCCT

<i>Tbx20</i>	CGAGCAGCTCCTCAAACAGA	CCGTGGCTGGTACTTATGCA
<i>Hand2</i>	GAGAACCCCTACTTCCACGG	GACAGGGCCATACTGTAGTCG
<i>Nkx2.5</i>	GCTTCAAGCAACAGCGGTAC	ACTTGTAGCGACGGTTCTGG
<i>Myl2</i>	ATCGACAAGAATGACCTAAGGGA	ATTTTTCACGTTCACCTCGTCCT
<i>Cited2</i>	TTCAAGTTGGCTGTCCCC	ACTGACGACATTCCACACCC

Supplementary References

- 1 de Soysa, T. Y. *et al.* Single-cell analysis of cardiogenesis reveals basis for organ-level developmental defects. *Nature* **572**, 120-124, doi:10.1038/s41586-019-1414-x (2019).