Title

Single-cell multimodal analyses reveal epigenomic and transcriptomic basis for birth defects in maternal diabetes

Journal

Nature Cardiovascular Research, 2(12)

ISSN

2731-0590

Authors

Nishino, Tomohiro
Ranade, Sanjeev S
Pelonero, Angelo
et al.

Publication Date

2023-12-01

DOI

10.1038/s44161-023-00367-y

Peer reviewed

HHS Public Access

Author manuscript

Nat Cardiovasc Res. Author manuscript; available in PMC 2024 August 23.

Published in final edited form as:

Nat Cardiovasc Res. 2023 December ; 2(12): 1190–1203. doi:10.1038/s44161-023-00367-y.

Single Cell Multimodal Analyses Reveal Epigenomic and Transcriptomic Basis for Birth Defects in Maternal Diabetes

Tomohiro Nishino[1,2], **Sanjeev S. Ranade**[1,2], **Angelo Pelonero**[1,2], **Benjamin J. van Soldt**[1,2], **Lin Ye**[1,2], **Michael Alexanian**[1,2], **Frances Koback**[1,2], **Yu Huang**[1,2], **Nandhini Sadagopan**[1,2,3], **Adrienne Lam**[1,2], **Lyandysha V. Zholudeva**[1,2], **Feiya Li**[1,2], **Arun Padmanabhan**[1,2,3], **Reuben Thomas**[1], **Joke G. van Bemmel**[1,2], **Casey A. Gifford**[1,2], **Mauro W. Costa**[1,2], **Deepak Srivastava**[1,2,4,5,*]

[1]Gladstone Institutes; San Francisco, CA, USA.

[2]Roddenberry Center for Stem Cell Biology and Medicine at Gladstone; San Francisco, CA, USA.

[3]Division of Cardiology, Department of Medicine, University of California, San Francisco; San Francisco, CA, USA.

[4]Division of Cardiology, Department of Pediatrics, University of California, San Francisco; San Francisco, CA, USA.

[5]Department of Biochemistry and Biophysics, University of California, San Francisco; San Francisco, CA, USA.

Abstract

Maternal diabetes mellitus is among the most frequent environmental contributors to congenital birth defects, including heart defects and craniofacial anomalies, yet the cell types affected and mechanisms of disruption are largely unknown. Using multi-modal single cell analyses, here we show that maternal diabetes affects the epigenomic landscape of specific subsets of cardiac and craniofacial progenitors during embryogenesis. A previously unrecognized cardiac progenitor subpopulation expressing the homeodomain-containing protein ALX3 showed prominent chromatin accessibility changes and acquired a more posterior identity. Similarly, a subpopulation of neural crest-derived cells in the second pharyngeal arch, which contributes to craniofacial structures, displayed abnormalities in the epigenetic landscape and axial patterning defects. Chromatin accessibility changes in both populations were associated with increased retinoic acid signaling, known to establish anterior-posterior identity. This work highlights how

*Corresponding author.: deepak.srivastava@gladstone.ucsf.edu.

Author Contributions Statement:

T.N. and D.S. conceived and directed the study. T.N. and Y.H. performed animal work. T.N. and S.S.R. collected heart tissues and isolated single cells for subsequent scRNA-seq and scATAC-seq. T.N., S.S.R., A. Pelonero, B.v.S, and F.K. analyzed scRNA-seq and scATAT-seq and developed computational methods. T.N., B.v.S., L.V.Z., and F.L. performed RNA in situ hybridization and subsequent tissue clearing and imaging. T.N., L.Y., N.S., A.L., A. Padmanabhan, and M.W.C designed, performed and analyzed luciferase assay, mouse lineage trace experiment. T.N., S.S.R., M.A., A. Pelonero, J.V.B, C.A.G., M.W.C., and D.S. interpreted the data. R.T. reviewed statistical methods. T.N, M.W.C. and D.S. wrote the manuscript with contributions of M.A.

Code availability:

All codes are available on GitHub (https://github.com/SrivastavaLab-Gladstone/Nishino_DM_2022).

Competing Interests Statement:

D.S. is a scientific co-founder, shareholder and director of Tenaya Therapeutics. The remaining authors declare no competing interests.

an environmental insult can have highly selective epigenomic consequences on discrete cell types leading to developmental patterning defects.

Cardiac and craniofacial birth defects often occur in conjunction due to shared cardio-pharyngeal and neural crest-derived progenitors, and reciprocal signaling between the two populations [1,2]. Numerous genetic syndromes affect both regions, as do environmental influences. Among the latter, pregestational maternal diabetes (PGDM), and the associated hyperglycemia, is relatively common given the increasing prevalence of type II diabetes worldwide, and leads to a 4 to 5-fold increase in the incidence of birth defects [3]. Similarly, heart and craniofacial regions are also severely affected upon rare external exposure to retinoic acid [4]. Retinoic acid embryopathy is related to disruption of anterior-posterior patterning, particularly through dysregulation of the Hox code in pharyngeal arches and specification of the anterior and posterior regions of second heart field progenitors that contribute to the outflow and inflow tracts of the heart, respectively [5,6]. The mechanisms by which hyperglycemia increases the risk for birth defects are unclear, although there is some evidence that oxidative stress and increases in the metabolite beta hydroxybutyrate may lead to epigenomic changes [7–10]. To address this question, we analyzed the epigenomic landscape during mouse development at single cell resolution and deciphered cell-specific mechanisms that may result in malformations.

Cardiac malformations in mouse maternal diabetes mellitus

To investigate the cellular consequences of PGDM during embryonic cardio-pharyngeal development, we induced diabetes mellitus in female mice by intraperitoneal delivery of streptozotocin (STZ) for five consecutive days, resulting in loss of insulin-producing cells. Females with blood glucose levels greater than 250 mg/dL were mated with untreated males (Fig. 1A, Extended Data Fig. 1A). As expected, embryos from PGDM females had significantly higher frequency of cardiac malformations that included conotruncal ventricular septal defects, atrial septal defects, and outflow tract (OFT) anomalies when compared to embryos from vehicle (VEH) treated females (Extended Data Fig. 1B, Supplementary Table 1, Supplementary Video 1A–B). Neural tube defects and craniofacial defects were also observed in the STZ-induced PGDM model, as previously reported, mimicking the human condition [11–13].

Changes in cardio-pharyngeal development due to PGDM

To determine how PGDM affects the transcriptional and epigenomic state within individual cells during cardio-pharyngeal development, we conducted single cell RNA-seq (scRNA-seq) and single-cell sequencing assay for transposase-accessible chromatin (scATAC-seq) on the cardiac and surrounding pharyngeal region at E10.5 from VEH- or STZ-treated females (Fig. 1A, Extended Data Fig. 2A). At this time point, the precursors of the major areas affected in this model are present, including neural crest-derived cells. After dissociating the dissected cardio-pharyngeal region into single cells, each individual sample was split for parallel scRNA-seq and scATAC-seq experiments. Using unsupervised clustering with Seurat [14], single-cell transcriptomes of over 31,000 cells were clustered, identified and

labeled based on known marker genes. This initial analysis confirmed the presence of all expected cell types at E10.5 (Fig. 1B–C, Extended Data Fig. 2B, left, C).

In parallel, analysis of single-cell chromatin accessibility data by scATAC-seq of 41,610 high-quality cells with batch correction identified 25 clusters using ArchR [15]. Cluster identity was assigned based on gene activity score of known cell type specific marker genes, a measure of the chromatin accessibility of the promoter region and gene body of each gene (Fig. 1D–E, Extended Data Fig. 2B, right, and 2D). The specific peaks for each cluster were then identified, and transcription factor (TF) motifs enriched in those peaks were analyzed by HOMER [16], representing distinct cell types (Fig. 1F). Furthermore, integration of scRNA-seq and scATAC-seq data supported our cluster annotations based on chromatin accessibility, demonstrating that all cell types expected from our microdissection were captured in the scATAC-seq experiment (Extended Data Fig. 2E).

Annotation of scATAC-seq data suggested that PGDM affected the cellular population distribution, leading to a statistically significant increase in cardio-pharyngeal mesodermal progenitors, while decreasing the number of more differentiated cardiomyocytes and neural crest derivatives (Fig. 1G, left, Supplementary Table 2A). scRNA-seq data corroborated an increase in cardio-pharyngeal mesodermal progenitors with decreases in neural crest cells and cardiomyocytes (Fig. 1G, right, Supplementary Table 2B). Interestingly, all of the affected cell populations are directly involved in the types of cardiac and craniofacial defects observed in the setting of PGDM.

Cell-specific epigenomic and transcriptomic changes in PGDM

While numbers of cells may be important, dysregulation of gene networks within specific cell types may reveal the mechanistic basis for the morphogenetic defects associated with PGDM. We therefore performed peak calling for each scATAC-seq cluster, identifying a total of 492,330 sites, with about 76% being distal peaks (> 3 kb away from transcription start site (TSS)) (Extended Data Fig. 2F–G). Differentially accessible chromatin regions (DARs) were analyzed between the control (VEH) and hyperglycemic (STZ) conditions for each individual cluster. Surprisingly, of the total 4324 DARs detected, ~97% were concentrated in only two cell types: mesodermally-derived cardio-pharyngeal progenitors (48.8%) and neural crest-derived cells (48.5%). Upon deeper analysis of these two cell compartments, we found that DARs were mostly enriched in a subset of clusters within each cell type. Among the seven subclusters of cardio-pharyngeal mesodermal progenitors, over 90% of DARs were detected in a single cluster (C15). Similarly, of the five neural crest sub-clusters, only two had DARs, with 80% of those being in a single cluster (C19) (Fig. 1H). We did not observe any correlation between detected DAR numbers and cell number or cellular distribution between conditions in each cluster. These results indicate that maternal PGDM-induced changes in embryonic chromatin accessibility are exquisitely cell-specific despite the universal exposure to hyperglycemia, and the cell types most affected in our data are consistent with those most critical for PGDM-related congenital diseases.

Given the concentration of DARs in neural crest-derived cells and cardio-pharyngeal mesodermal cells, we re-clustered the scRNA-seq data, focusing on those populations.

Neural crest cell sub-clusters were annotated based on expression of *Hox* genes and other markers of neural crest-derivatives that distinguish cells that reside in different pharyngeal and aortic arches (Fig. 2A–C) [17,18]. Six distinct neural crest-related clusters were identified representing pharyngeal arch 2 (PA2), PA3, PA4/6, migrating neural crest progenitors (NC-prog), and the neural crest-derived smooth muscle cell progenitors (SMC-prog) and smooth muscle cells (SMC) of the aortic arch arteries and outflow tract. Analyses of the cellular distribution among each sample showed that PGDM led to an increase of undifferentiated NC-progenitors with a concomitant decrease in the *Rgs5*-positive SMC-progenitor population of the cardiac outflow tract (Extended Data Fig. 3A). Integration of the annotated neural crest scRNA-seq data with the neural crest portion of the scATAC-seq data revealed that the neural crest-related clusters with the largest number of DARs, C19 and C20, represented cells from PA2 and PA3/4/6, respectively (Fig. 2D–F, Extended Data Fig. 3B).

We performed TF motif analysis among the DARs in PA2 cranial neural crest cells and PA3/4/6 cardiac neural crest cells to determine how the gene regulatory regions were affected by PGDM and hyperglycemia. The AP2 motif, typical of undifferentiated neural crest cells [19,20], was the most enriched in DARs with increased accessibility upon STZ exposure in both clusters, suggesting an impairment of differentiation in these populations. Retinoic acid-related RXR motifs were also enriched in the more accessible DARs of PA2, while the TWIST1 motif was enriched in less accessible DARs, supporting impairment of differentiation and consistent with *Twist1*'s role in cranial neural crest cell differentiation [21,22] (Fig. 2G–H, Extended Data Fig. 3C–D). Similarly, in the PA3/4/6 cardiac neural crest, SIX, TBX, and SMAD2/4 motifs were enriched in DARs with reduced accessibility upon STZ treatment, suggesting the disruption of the core differentiation regulatory networks [23,24] (Extended Data Fig. 3D). Although unsupervised clustering could not differentiate PA3 and PA4/6 in the C20 population, we were able to distinguish between PA3 and PA4/6 cells within C20 based on gene scores of several genes enriched in PA3 or PA4/6 by scRNA-seq, including *Hox* genes, *Six2,* and *Tbx2* (Fig. 2C–D, Extended Data Fig. 3E–F). Subsequent DAR analysis between VEH and STZ in PA3 or PA4/6 showed an unbalanced difference in numbers of detected DARs, with changes detected almost exclusively in PA4/6, which is involved in pulmonary artery and aortic arch development (Extended Data Fig. 3G). TF motif enrichment analysis on DARs detected in PA4/6 reproduced the DAR findings in C20, including AP2 motif enrichment in more accessible loci in STZ and SMAD motif enrichment in less accessible loci in STZ, suggesting the chromatin accessibility change in C20 likely derived from PA4/6 changes (Extended Data Fig. 3D, H). In addition, the Sox TF motifs that form the core network in neural crest cell specification and migration [25] were recognized in the more accessible DAR loci in STZ (Extended Data Fig. 3H). In summary, our results indicate that in the setting of PDGM/hyperglycemia, cells in PA2 and *Six2*-high PA4/6 display transcriptomic and epigenomic changes suggesting a less differentiated status.

Within the PA2 cells, a subpopulation was highly enriched in the STZ group compared to VEH group (Fig. 2B, E) leading us to investigate the identity and transcriptional consequences of PGDM in these cells. scRNA-seq reclustering of PA2 identified three distinct clusters, with most of the changes under STZ conditions localized to cluster 0 (Fig. 2I). This cluster was marked with *Tfap2a* expression, typical of the less differentiated

neural crest cells, and was highly enriched upon STZ treatment, indicating that relatively undifferentiated cells within PA2 were most susceptible to PGDM and hyperglycemia, consistent with enrichment of the AP2 motif in more accessible DARs within PA2 (Fig. 2H–J, Extended Data Fig. 4A). In addition, *Nr2f1*, which forms a TF complex with *Tfap2a* in cranial neural crest cells to maintain a relatively undifferentiated state [26], was significantly upregulated in cluster 0 (Extended Data Fig. 4B).

Tfap2a is responsible for positional specification of neural crest cells along both the anterior-posterior (A-P) and dorsal-ventral (D-V) axis. The A-P axis is partly regulated through one of *Tfap2a's* direct downstream target genes, *Hoxa2* [27], which was also significantly upregulated in cluster 0 among STZ offspring (Extended Data Fig. 4C). *Hoxa2* directly downregulates *Six2* and dictates A-P patterning [28], which is consistent with the inverse correlation in expression levels of these two genes in PA2, PA3, and PA4/6, shown in Figure 2C. We also found *Dlx5* and *Dlx6*, which dictate D-V patterning in PA-related craniofacial development and are also downstream target genes of *Tfap2a* [29], were significantly downregulated in cluster 0 cells in STZ (Fig. 2K, Extended Data Fig. 4D).

In addition to pattern formation, the dysregulated genes are involved in cell differentiation and migration. For example, *Hoxa2* transiently activates *Meox1*, which is important for skeletal development from PA2 neural crest cells. While *Meox1* is expressed as early as E9.0, it is normally downregulated by E10.5 in PA2 [30]. Interestingly, *Meox1* transcripts persisted at a high level at E10.5 in cluster 0, where *Hoxa2* was up-regulated in STZ embryos. Furthermore, *Postn*, a direct target gene of *Meox1* [31] normally expressed in PA1 to dictate tooth development [32], was also upregulated in these cells (Extended Data Fig. 4C). Other than dysregulation of genes related to PA patterning and development, the top GO terms enriched in cluster 0 DEGs in STZ were associated with cell migration, and these results are consistent with DAR findings. (Fig. 2L, Extended Data Fig. 4E–F). Taken together, multi-modal analysis of neural crest cells at single cell resolution suggests that a subset of PA2 neural crest cells that are relatively undifferentiated are modified epigenetically by PGDM, resulting in transcriptional dysregulation of genes related to patterning, cell migration and cellular differentiation.

Alx3-expressing SHF cardiac progenitors affected by PGDM

We performed a similar characterization of mesodermally-derived cardio-pharyngeal progenitor cells and cardiomyocytes, as they also displayed a large number of DARs upon hyperglycemia. Upon re-clustering this population using scRNA-seq, we detected five cardiomyocyte sub-types and ten mesodermal progenitor sub-types (Fig. 3A–B, Extended Data Fig. 5A). The integration of scRNA-seq with scATAC-seq data (Fig. 3C, Extended Data Fig. 5B) revealed that the C15 cluster in scATAC-seq, which displayed the largest number of DARs, corresponded to one of two anterior heart field cell (AHF) clusters, which we labeled AHF2. AHF1 and AHF2 had very similar transcriptional profiles, but AHF2 displayed lower expression of AHF marker gene *Hand2* and near absence of *Rgs5* and *Armh4* (Fig. 3B, D, Extended Data Fig. 5C). This AHF2 population was clearly distinguishable from posterior second heart field (pSHF) cells, which are marked by *Foxf1*

and *Osr1* (Fig. 3A–B). The transcription factor *Alx3* was specifically expressed in the AHF2 population, uniquely marking this newly recognized subset of the AHF (Fig. 3D).

To determine if the AHF2 cells are truly of AHF origin, we examined overlap of *Alx3* expression with AHF-derived cells marked by Cre-recombinase under control of the Mef2c AHF-specific enhancer (Mef2c-AHF-Cre: Ai6 AHF lineage tracing mouse model) [33]. RNA *in situ* hybridization with light sheet microscopy revealed that *Alx3*-positive cells were distributed in a continuum from the distal outflow tract towards the posterior bilateral sides of the embryo. Of these, the more anterior *Alx3*-positive cells overlapped with the Mef2c-AHF-Cre lineage, demonstrating that *Alx3* marks a subset of the AHF cells (Fig. 3E, Supplementary Video 2). We further found that *Alx3*-expressing cells resided posterior to the *Armh4*-expressing AHF1 cells, abutting the AHF1 cells of the distal outflow tract (Extended Data Fig. 6A–B, Supplementary Video 3). Optical sections demonstrated that the *Alx3*-expressing cells and the *Armh4*-expressing cells share similar AHF regions at the distal outflow tract region without overlapping one another (Fig. 3F). Thus, PGDM affects a very specific subset of the SHF that appears to lie between the classically recognized AHF and pSHF.

To investigate the *Alx3*-enriched subpopulation of the SHF, we analyzed the mouse cardiac development single cell atlas data from E7.75 to E9.25 previously generated by our lab [34]. *Alx3*-positive cells ($Alx3^{Pos}$) were detectable within clusters annotated as AHF, with significant expression detected in a small number of cells by E9.25 (Extended Data Fig. 6C–E). DEG analysis between $Alx3^{Pos}$ and $Alx3^{Neg}$ AHF cells at E9.25 demonstrated that $Alx3^{Pos}$ AHF cells expressed higher levels of genes related to epithelial-to-mesenchymal transition and skeletal system morphogenesis, with lower levels of genes involved in cardiomyocyte development (Extended Data Fig. 6F). By E10.5, the $Alx3^{Pos}$ AHF cells expanded and although the number of DEGs between AHF2 and AHF1 increased considerably, they still displayed similar GO terms (Extended Data Fig. 6G–H).

To determine the cellular fate of *Alx3*-positive AHF2 cells, we introduced Cre recombinase into the endogenous mouse Alx3 locus using CRISPR-Cas9 mediated homologous recombination and crossed these mice with the Ai6 reporter mice (Supplementary Fig. 1A-B). Upon Cre exposure, loxP-stop-loxP present on the Ai6 mice is excised, leading to permanent ZsGreen1 labeling of cells including *Alx3*-positive cells and all their descendants. Outside of the heart, we found prominent ZsGreen1 expression in the frontonasal protrusion area, and forelimb and hindlimb regions, where Alx3-LacZ mice had previously been shown expression [35] (Supplementary Fig. 1C). In neonatal hearts, we identified ZsGreen1-positive cells in a subset of cells within the valve region of the OFT and in the dorsal wall and septal wall of the atrium, close to the entrance of the inferior or superior vena cava, which are also marked by *Alx3*-derived cells. ZsGreen1-positive cells were also found in the mesenchymal tissues around large vessels, such as the aorta and pulmonary artery (Fig. 3G, Supplementary Video 4). These findings suggest that the *Alx3*-positive AHF2 cells contribute to specific cells in the OFT region and atrium.

DARs within AHF2 cells with increased accessibility in STZ compared to VEH were enriched in GO terms related to anterior/posterior specification, mesenchymal development,

and mesonephros development. In contrast, DARs with reduced accessibility were enriched in terms related to Wnt signaling, maintenance of cell number, differentiation of cardiac muscle, and heart development (Extended Data Fig. 7A). TF motif analysis in DARs with increased accessibility in PGDM/hyperglycemia revealed enrichment for HOX gene groups (Hoxa2, Hoxb4, Hoxc9, Hoxa9), as well as COUPTFII and RXR, which are involved in establishing pSHF cells, and are not normally accessible in AHF cells [36,37]. In contrast, motifs enriched in DARs with decreased accessibility in PGDM/hyperglycemia included PBX1, MEIS1 and PKNOX1, which normally function together to activate more anterior Hox genes [37,38]. Of note, disruption of *Pbx* genes has been reported to cause outflow tract malformations both in mice and human [39]. In addition, motifs for TBX and TGIF, both involved in cardiac differentiation [40], were enriched in closed DARs (Fig. 4A). Thus, the AHF2 cells, which at baseline are more posterior in expression pattern to AHF1 cells and have a decreased cardiomyocyte signature, appeared to be even more “posteriorized” and less cardiac-like in offspring of STZ-treated mothers.

To investigate the effects of PGDM/hyperglycemia more precisely on *Alx3*^{Pos} AHF2 cells, we re-clustered the *Alx3*-expressing AHF2 cells and detected three distinct sub-clusters (Fig. 4B). Marker genes for Cluster 0 were enriched for genes related to cardiomyocyte development, including cardiac TFs like *Hand1/2*, *Isl1*, and *Tbx2/5*, while those for Cluster 1 and 2 were enriched for genes involved in cell migration, receptor protein tyrosine kinase signaling, and skeletal system morphogenesis (Extended Data Fig. 7B). Cluster 1 and 2 cells were increased in number under hyperglycemia, while Cluster 0 numbers were decreased compared to control (Fig. 4B, Supplementary Table 2C). These findings suggest that PGDM/hyperglycemic conditions led to a cellular state transition from Cluster 0 towards Cluster 1 and 2, characterized by further loss of cardiac TF expression and gain of a skeletal-like state. This was further confirmed by DEG analysis between the STZ and VEH groups in *Alx3*^{Pos} AHF2 cells (Fig. 4C). Notably, the pSHF Hox gene, *Hoxb1* [36,37], was expressed almost exclusively within clusters 1 and 2 where STZ cells increased, and was nearly absent in cluster 0, whereas STZ treatment led to both significant increases in the number of *Hoxb1* positive cells and its expression levels (Fig. 4D). This was corroborated by the STZ-mediated increase of the chromatin accessibility of the distal regulatory element of *Hoxb1*, calculated as a “peak-to-gene link” by ArchR. (Extended Data Fig. 7C).

Cluster 2 was notable not only for *Hoxb1* expression among the STZ group but also for decreased expression of *Crabp1* and increased *Crabp2* transcripts, competing modulators of retinoic acid (RA) signaling (Fig. 4E). *Crabp1*, normally highly expressed in the AHF, enhances RA catabolism, while *Crabp2*, which is expressed at higher levels in the pSHF, facilitates nuclear transport of RA and subsequent transcriptional activation, resulting in posteriorization of SHF progenitors [41]. This raised the possibility that, under PGDM/hyperglycemia, abnormal anterior-posterior patterning may occur in the *Alx3*^{Pos} AHF2 due to ectopic retinoic acid-induced posteriorization, including induction of *Hoxb1* expression.

Retinoic acid signaling dysregulation in PGDM-affected cells

Given the dysregulated *Crabp1* and *Crabp2* expression and anterior-posterior patterning defects observed in the STZ group, we specifically evaluated RA signaling in cells from

VEH or STZ treatment. Using ChromVAR analysis, which can assess enrichment of specific TF activity at the single cell level, we found that PGDM drove significantly increased enrichment for RA signaling-related TFs (RAR and RXR) in most cell types, including PA2 neural crest-derived cells and AHF2 cells (Extended Data Fig. 8A–B, Supplementary Table 3). This was in agreement with the TF motif enrichment analysis in DARs, which showed that the RXR motif and its downstream effectors' motifs, such as COUP-TFII and HOX, were highly enriched in more accessible DARs under PGDM/hyperglycemia in PA2 and PA4/6 neural crest and AHF2 cells (Fig. 2H, 4A, Extended Data Fig. 3H).

To investigate mechanisms by which altered RA signaling may contribute to transcriptional and epigenomic changes in hyperglycemia, we selected distal regulatory regions with differential accessibility that had peak-to-gene links with differentially expressed genes between VEH and STZ, and also had RA signaling-related motifs within the scATAC-seq peaks. Among these, we focused on several candidate enhancers for dysregulated genes in PA2 neural crest cells, with further experimentation related to *Tfap2a* and *Nr2f1* gene activation (Fig. 5A). *Tfap2a* is implicated in neural crest cell induction and migration, and together with *Nr2f1*, binds to active enhancers that are critical for maintaining cranial neural crest cells in an undifferentiated state [26]. We performed luciferase reporter assays using a validated RA-responsive *Cyp26a1* enhancer element as a positive control and found that high glucose itself did not enhance transcription, suggesting that heightened RA signaling is not a direct consequence of hyperglycemia (Fig. 5B). We further found PA2 neural crest candidate regulatory elements for *Tfap2a* and *Nr2f1* were able to activate transcription in the presence of RAR/RXR receptors and showed further modulation by addition of RA. RA-dependent activation was abolished upon deletion of putative RAR/RXR binding sites (Fig. 5C–D). The dysregulation of RA signaling present in STZ embryos was supported by analysis of the data using Weighted Correlation Network Analysis (WGCNA) [42]. This approach enables identification of co-variate transcriptional modules, gene regulatory networks (GRN), across individual cells without the influence of possible bias introduced through sub-clustering processes and manual annotation of the neural crest compartments. Among the total of 16 modules detected, WGCNA revealed a PGDM-induced abnormality in a GRN that was present only in PA2 of STZ embryos (Extended Data Fig. 9A, red box). This module contained genes involving in nervous system development and endochondral bone morphogenesis. In particular, *Tfap2a* downstream genes such as *Eya1*, *Dlx5/6*, and *Hoxa2* constituted a core network within this GRN, consistent with PA2-specific aberrant RA signaling leading to increased *Tfap2a*/*Nr2f1* and subsequent patterning defects in the setting of PGDM (Extended Data Fig. 9B–C).

Similarly, WGCNA using the mesodermal cardio-pharyngeal cells identified total of 15 modules, and we focused on a GRN with the most significant module activity score change between VEH and STZ samples in the AHF2 population (AFH2_mod1 highlighted in a red box) (Extended Data Fig. 9D–F). This module contained genes related to cardiac outflow tract and right ventricle development, including critical regulators such as *Hand1/2*, *Isl1*, *Bmp4*, and *Gata6*, which were downregulated in STZ as members of a core network of the GRN. This is consistent with the diminished "anterior" nature of the *Alx3*^{Pos} AHF2 in offspring of PGDM mothers. To experimentally determine if RA signaling activity, normally present in the pSHF, was ectopically enhanced in AHF2 cells within STZ-treated embryos,

we exposed a cohort of transgenic female mice containing a LacZ reporter driven by a RA response element (RARE-LacZ) to STZ to establish a PGDM model in this genetic background. As expected, direct detection of LacZ expression in VEH offspring using RNAscope indicated RA activity was evident in the pSHF region, but not in *Alx3*-expressing cells. In contrast, STZ-treated offspring demonstrated an enlarged LacZ-positive domain with marked anterior extension compared to the VEH control samples (Fig. 5E–F, Extended Data Fig. 10A–C, Supplementary Video 5A–B). This result provided *in vivo* evidence that "posteriorization" of the *Alx3*^{Pos} AHF2 population occurs with anterior expansion of RA signaling in offspring of mothers with PGDM, likely contributing to impaired outflow tract development.

Discussion

In this study, we used single cell multi-modal analyses to uncover mechanisms by which environmental influences such as PGDM can cause developmental anomalies. We demonstrated selective epigenomic vulnerability of a small subset of cells that help explain cardiac and craniofacial defects that are observed in PGDM despite all embryonic cells experiencing similar environmental exposure. Specifically, we found that a previously unrecognized subpopulation of the *Alx3*^{Pos} AHF is uniquely "posteriorized" in the setting of PGDM. Furthermore, a subset of relatively less differentiated neural crest cells within PA2 and PA4/6 also had evidence of abnormal cell differentiation and anterior-posterior patterning. Analyses of regulatory elements in these specific cell types demonstrated that RA signaling and downstream *Hox* genes were dysregulated in AHF2 and PA2, with ectopic RA activity resulting in disruption of proper patterning and loss of cell type specific transcription signatures (Fig. 5G). Mechanistically, it appears RA metabolism dysregulation by altered *Crabp1* and *Crabp2* expression contributes to the alterations in AHF2, while RXR-mediated transcriptional upregulation of *Tfap2a* and *Nr2f1* disrupts PA2 differentiation and patterning. In particular, the presence of putative binding sites of *Tfap2* and *Nr2f1* in the Rxr-dependent *Tfap2a* regulatory element suggests a positive feedback mechanism reinforcing RA signaling-mediated transcriptional dysregulation under PGDM/hyperglycemia in PA2 neural crest cells (Fig. 5G).

The recognition of selective epigenetic alteration of the Alx3-expressing cells of the AHF led to the identification of this unique sub-population of cardiac progenitors. Despite being adjacent to the typical AHF cells, our lineage-tracing experiments suggest that the Alx3-positive AHF cells contribute to distinct parts of the neonatal heart. Descendants of Alx3-expressing cells were found in the aortic valves and outflow tract mesenchyme, as well as the atrial wall and inflow vessels. The regions containing Alx3 descendants correlate with the defects observed, and indicate a higher sensitivity of these cell types to hyperglycemic conditions.

These findings begin to address how an environmental influence can affect specific embryonic cell types and lead to unique developmental phenotypes and congenital birth defects. In the future, it will be important to understand why certain cell types are more vulnerable than others and how environmental influences as occur in PGDM cause the very specific epigenomic shifts described here. While PGDM clearly causes shifts

in RA signaling, the precise mechanism by which regulators of RA metabolism and RXR-dependent transcription are altered will require further study. For example, increased levels of beta-hydroxybutyrate present in ketotic mothers with PGDM may contribute to specific epigenomic alterations through histone modification in the embryo [9,10]; future advances that enable interrogation of individual epigenomic marks in single cells will help explore this possibility. Ultimately, discovery of the consequences of environmental factors on embryonic development using multi-modal single cell genomics will not only lead to potential gene-environment combinations that may contribute to the complex and multifactorial diseases, but also preventive approaches to protect vulnerable cells.

Materials and Methods

Animal

Animal studies were conducted in compliance with all relevant ethical regulations in the animal use protocols, UCSF animal use guidelines and the NIH Guide for the Care and Use of Laboratory Animals. All protocols relating to animal use were approved by the Institutional Animal Care and Use Committee (IACUC; AN189140) at UCSF which is accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC). C57BL/6J wild type (JAX stock no. 000664), RARE-hsp68LacZ (JAX stock no. 008477), and B6.Cg-Gt(ROSA)26Sortm6(CAG-ZsGreen1)Hze/J (Ai6) (JAX stock no. 007906) mice were purchased from Jackson Laboratory (Bar Harbor, ME), Mef2c-AHF-Cre mice [33] were provided by Brian Black's lab. *Alx3*-Cre mice were produced by blastocyst injection of ribonucleoprotein complexes consisting of purified Cas9 protein (IDT), a guide RNA (crRNA targeting *Alx3* locus and universal 67mer tracrRNA; IDT), and a double-stranded DNA template for homology directed repair that led to the in-frame insertion of a Cre immediately downstream of the start codon in exon 1 of the endogenous *Alx3* locus (details of gRNAs and HDR are provided in the Tables section). Blastocysts were transferred to pseudopregnant females and pups were weaned at 4 weeks of age. Founder animals were screened by junctional region PCR amplifications and further confirmed by Sanger sequencing of those junctional regions and inserted Cre sequence. Positive founders were outcrossed to wild type C57BL/6J animals and pups were screened for germline transmission of the Alx3 Cre targeted allele by PCR genotyping. All animals were housed in a 12-hour light/dark cycle and given ad libitum access to standard chow and water unless otherwise documented.

Model of maternal diabetes mellitus

Diabetes induction was performed using streptozotocin (STZ) (Sigma-Aldrich, S0130) as previously described with minor modifications [43]. Diabetes mellitus was induced in ~10-week-old female C57BL/6 mice or RARE-hsp68LacZ mice by intraperitoneal injections of 60 mg/kg body weight STZ for 5 consecutive days. Two weeks after the last STZ injection, blood glucose levels were measured after 6 hr fasting using a Zoetis, AlphaTRAK2 glucometer. Mice were diagnosed as diabetic if blood glucose measurements were greater than or equal to 250 mg/dL and were subsequently bred to 12-week-old C57BL/6 male mice. Pregnant females were identified by echocardiography performed at E6.5 or E7.5 and euthanized to harvest embryos at E10.5 for single-cell RNA sequencing, single-cell

ATAC sequencing, and whole-mount *in situ* hybridization experiments, and at E18.5 for histological examination.

Heart morphology examination by micro-CT

E18.5 mouse embryos were dissected in cold PBS with 1% FBS and the heart samples were further microdissected, fixed with 4% PFA overnight at 4°C degrees with rocking, followed by washing with cold PBS twice for 5 min each. Then, samples were stained with 1% Phosphotungstic acid hydrate (Sigma-Aldrich, P4006) in 70% ethanol for three days at 4°C and were embedded in 0.5% agarose gel and imaged for micro-CT using μCT50 (SCANCO Medical AG). The images were processed with OsiriX MD (Pixmeo SARL) and were examined by two researchers independently.

Embryo dissection, single-cell library preparation and sequencing

C57Bl6/J E10.5 mouse embryos were dissected in cold PBS (Life Technologies, 14190250) with 1%FBS (ThermoFisher Scientific, 10439016) and truncal region was further microdissected to include the heart tube region, the SHF and mesodermal regions behind the heart tube, as well as the second pharyngeal arches (Extended Data Fig. 2A). Three embryos with matched somite counts each were collected from females in the VEH or STZ treatments. Dissected cardiac tissue was incubated in 200 μl TrypLE (ThermoFisher Scientific, 12563029) for 5 min, triturated with a 200-μl pipette tip, and incubated for an additional 5 min. The TrypLE solution was quenched with 600 μl PBS with 1% FBS. Cells were filtered through a 70-μm cell strainer (BD Falcon, 08-771-2), centrifuged at 150g for 3 min, and resuspended in PBS with 1% FBS. The cell suspension was prepared by adjusting the cell number to recover 10K cells per sample according to the manufacturer's instructions (Chromium Single Cell 3' Reagent Kit v.3, 10X Genomics). GEM generation for scRNA-seq was performed on the Chromium controller. The remaining cells were then subjected to nuclear isolation according to the 10x nuclei isolation protocol (Nuclei Isolation for Single Cell ATAC Sequencing, CG000169, 10X Genomics). The nuclei solution was adjusted so that the target nuclei number was 10K per sample (Chromium Next GEM Single Cell ATAC Reagent Kits v1.1, 10X Genomics), followed by transposition reaction. GEM generation for scATAC-seq was performed using the Chromium controller (10X Genomics).

The subsequent scRNA-seq and scATAC-seq library preparations were performed according to the protocols described above using Chromium Single Cell 3′ Reagent Kits v3 (PN-1000075) and Chromium Next GEM Single Cell ATAC Reagent Kits v1.1 (PN-1000176), respectively. Additionally, Chromium Chip B Single Cell Kit (PN-1000153) and Chromium i7 Multiplex Kit (PN-120262) were used for scRNA-seq library preparation, and Chromium Next GEM Chip H Single Cell Kit v1.1 (PN-1000161) and Single Index Kit N Set A (PN-1000212) were used for scATAC-seq library preparation. The scRNA-seq libraries were sequenced using HiSeq4000 (Illumina) and NovaSeq (Illumina), and the scATAC-seq libraries were sequenced using NovaSeq, in accordance with the manufacturer's protocol.

scRNA-seq analysis

Raw sequencing data were preprocessed with the Cell Ranger v.5.0.0 pipeline (10X Genomics). Data from all samples was merged by cellranger-aggr in the pipeline described above and normalized to the same sequencing depth, resulting in a single gene-barcode matrix. Further analysis was performed using Seurat v4.0 R package [14] with reference to the Seurat web tutorials. Low quality cells were removed from analyses, keeping the cells with number of genes per cells between 4,000 and 9,500, with a UMI count per cell between 10,000 and 100,000, with mitochondrial gene percentage between 1 and 6%, and with ribosomal gene percentage between 10 and 30%. After the filtering step, for normalization, the data were transformed using SCTransform function, setting mitochondria gene percentages and ribosomal gene percentages as a variance to be regressed. Then, principal component analysis (PCA) was performed with RunPCA. Batch correction among each embryo was performed using Harmony [44]. Cells were clustered using top 32 principal components and visualized using a Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction (RunUMAP, FindNeighbors, and FindClusters). For clustering, a vector of resolution parameters was passed to the FindClusters function and the optimal resolution that established discernible clusters with distinct marker gene expression was selected. To identify marker genes, the clusters were compared pairwise for differential gene expression using the FindAllMarkers function (min.pct = 0.45, logfc.threshold=0.4, and return.thresh (p-value cut-off) = 1×10^{-10}). By cross-referencing the resulting cluster-specific marker genes with known cell type-specific maker genes, we classified them into clusters of seven cell types. The numbers of cells in each of those major cell type clusters in each condition (VEH vs STZ) were then calculated and plotted in a bar plot using the ratio of each cluster to the whole cells in each condition. Cardiomyocyte and Meso-CardiopharyngealProgenitors clusters, NeuralCrest cluster, PA2 neural crest cluster, and AHF2 cluster were re-clustered for the further analyses. Clustering was undertaken as previously described with minor modification. For the SCTransform step, in both subset analyses, we used mitochondria gene percentage and cell cycle scores as a variance to be regressed. In the mesoderm-cardiomyocyte subset analyses, top 30 PCAs were used for dimensional reduction. In the neural crest cell subset analyses, top 34 PCAs were used for dimensional reduction. In the PA2 neural crest cell subset analyses, top 22 PCAs were used for dimensional reduction. In the AHF2 cell subset analyses, top 30 PCAs were used for dimensional reduction. Differential gene expression analysis between conditions (VEH and STZ) in each sub cell types was performed using the FindAllMarkers function. For visualization of gene expression data, the following Seurat functions were used: FeaturePlot, VlnPlot, DotPlot, and DoHeatmap. Enrichr [45] was used for gene ontology (GO) enrichment analysis. All the indicated p-values for the GO term analyses are adjusted p-values calculated with Fisher exact test.

Gene regulatory network analysis

WGCNA (v1.70–3) [42] was used to identify gene regulatory networks which were dysregulated under hyperglycemic conditions. Transcriptional data for either mesoderm or neural-crest cells were retrieved via Seurat's GetAssayData function, returning a cell by gene expression matrix. This matrix was directly input to WGCNA's blockwiseModules function (minModuleSize = 10, mergeCutHeight = 0.15) so that individual cells are

considered as individual samples upon input. The resultant modules' constituent gene activity was subsequently scored per-cell with Seurat's AddModuleScore function in each source Seurat object. Statistical significance of the difference in the mean module scores between treatment conditions within previously-annotated mesoderm or neural-crest sub-clusters was assessed using a linear mixed-effects model (lmerTest v3.1–3). The mouse identification from which each cell is derived is modeled as a random effect, and Benjamini-Hochberg multiple-testing correction was applied to resultant p-values (stats v4.0.5). Modules with the lowest p-value per sub-cluster of interest were subsequently input to the StringDB web-interface (www.string-db.org) to compute, analyze, and visualize protein-protein interaction networks.

scATAC-seq analysis

Raw sequencing data were preprocessed with the Cell Ranger ATAC v.2.0 pipeline (10X Genomics). Subsequent analyses were performed using ArchR v1.0.1 R package [15] with reference to the ArchR web tutorials. Low quality cells were removed based on TSS enrichment score and the number of fragments per nuclei (createArrowFiles function with minTSS=7, minFrags=19952, maxFrags=1e+6), and nuclei doublets were removed through addDoubletScores and filterDoublets functions, resulting in making an ArchRproject to be analyzed. Through the genome-wide tiling with iterative Latent Semantic Indexing, the dimensionality reduction was performed using the addIterativeLSI function. Batch correction among each embryo was performed using the addHarmony function. Then, clustering and the 2D embedded visualization in UMAP space were performed using the addClusters function, and the addUMAP and plotEmbedding functions, respectively. Gene scores (GS), which were calculated by the accessibility of promoter and gene body regions of each gene and can be treated as a proxy of expression levels of a corresponding gene, were extracted to identify the cluster features using the getMarkerFeatures with useMatrix="GeneScoreMatrix". For peak calling per cluster, the addGroupCoverages and the addReproduciblePeakSet functions with peakMethod = "Macs2" were used. Furthermore, to identify the cluster specific feature peaks, the getMarkerFeature functions with useMatrix="PeakMatrix" was used, and transcription factor (TF) binding site motif enrichment analysis on the resultant peaks was performed using HOMER [16]. For differential accessible region (DAR) analysis, peak calling was performed for each condition (VEH or STZ) in each cluster and differential analyses were performed using the getMarkerFeatures function by setting a cluster from STZ as useGroups and a corresponding cluster from VEH as bgdGroups in the function. The statistically significant DARs were defined with FDR less than 0.05 and Log2 fold change greater than 1. Motif enrichment analyses on the detected DARs were performed using the peakAnnoEnrichment function. To integrate the scRNA-seq data from the same fetus heart samples, the aforementioned Seurat object with whole cell types were preprocessed with the FindVariableFeatures from the Seurat package to extract variable genes and transformed by the as. Single Cell Experiment function from Seurat packages. Then, the Seurat object was integrated using the addGeneIntegrationMatrix with useImputation=F. Considering the marker GS, marker peaks and the results of scRNA-seq data integration, the found clusters were annotated. Integration with scRNA-seq data was validated by calculating prediction scores, integration scores during integration, or Jaccard indices between the scATAC-seq annotation and the scRNA-seq annotation [46]. To

assess enrichment of TF activity per cell, ChromVAR analyses for retinoic acid signaling related receptor TFs were performed using the addDeviationsMatrix and getVarDeviations functions. The differentialDeviations function was used to determine whether there is a significant difference between the bias-corrected deviations for a TF motif between conditions per each cell type [47]. To identify an association of each peak and expression of a corresponding gene per cluster, the addPeak2GeneLinks function was used. The found associations between peak to gene were integrated with the differential expressed genes in the same cluster between conditions VEH and STZ to identify the candidate enhancers for those DEGs.

Single cell proportion test

For statistical tests addressing cell type population changes between conditions, a permutation test is used to calculate a p-value for each cluster implemented in the R package scProportionTest with “n_permutations = 10,000” (https://github.com/rpolicastro/scProportionTest/releases/tag/v1.0.0). scRNA populations were analyzed directly from Seurat as described in package documentation, and scATAC populations were analyzed by manually exporting ArchR metadata to scProportionTest.

Peak annotation

ChIPpeakAnno R library [48] was used for converting a bed file of detected peaks out from ArchR pipeline to a GRanges format. Then, the annotatePeak function from ChIPseeker R library [49] was used for annotating genome location with “tssResion” augment setting as 2 kb. The plotDistToTSS function was used for visualizing the location of each peak from TSS.

GREAT analyses

Inferring the functional significance of cis-regulatory regions found in scATAC-seq analyses was performed with GREAT v4.0.4 webtool [50] (http://great.stanford.edu/public/html/index.php) using default parameters. Detected regions were filtered with a significant threshold set at an adjusted p-value less than 0.05 and Log2 fold change greater than 1.

in situ hybridization experiments

Whole mount RNA *in situ* hybridization was performed in E10.5 embryos using Multiplex Fluorescent Reagent Kit v2 (Advanced Cell Diagnostics, 323100) as previously reported [34]. Each experiment was repeated using at least three biological replicates to determine the spatial expression of the genes. The *in situ* hybridization probes used in this study are as follows: Mm-Alx3-C2 (Advanced Cell Diagnostics, 1127551-C2), ZsGreen-C3 (Advanced Cell Diagnostics, 461251-C3), Mm-Armh4-C1 (Advanced Cell Diagnostics, 1085041-C1), and Ecoli-LacZ-C3 (Advanced Cell Diagnostics, 313451-C3). Whole-mount embryos were imaged in 0.1% PBST using the Leica MZ FLIII fluorescence stereomicroscope (acquisition software LAS v.3.7.4).

For light sheet microscopy imaging, the whole mount samples were microdissected to keep the area of interest and were mounted in 1.5% low melting temperature agarose (Fisher Scientific, Hampton, NH) in a glass capillary with matching piston rod (Sigma

Aldrich, Steinheim). Samples were then cleared in EasyIndex optical clearing solution with a refraction index of 1.465 (Lifecanvas technologies, Cambridge, MA) at 4 °C for two nights minimum before imaging, by extending the agarose-embedded sample into the solution. Samples were then imaged on a Zeiss Z.1 Lightsheet microscope. Images were processed with Fiji and Imaris.

Alx3$^{Cre/+}$*:Ai6* lineage tracing

Alx3$^{Cre/+}$ founder mice, regardless of their sex, were crossed with Ai6 mice to perform linage tracing of *Alx3* expressing cells during heart development. After birth, the hearts were dissected, fixed with 4% paraformaldehyde for 1 hr, and followed with tissue clearing steps with the modified QUBIC protocol; this included washing with PBS, delipidating samples with CUBIC-L at 37 ° C with agitation for 3 days, washing with PBS for 1 day, and RI matching with 50% CUBIC-R+(M)/H2O for 1 day followed with CUBIC-R+(M) for 3 days.

Images were taken by UltraMicroscope Blaze (Miltenyi) and processed with Imaris (ver 10.0.1).

Luciferase assay

Candidate enhancer sequences identified from the integrated analysis from scRNA-seq and scATAC-seq were PCR amplified (TAKARA, PrimeSTAR GXL) from mouse genomic DNA and PCR products cloned into pGL4.23 (Promega) pre-digested with XhoI and HindIII using Cold Fusion Cloning kit (System Biosciences). Deletion of putative TF binding sites from those enhancers were performed using QuickChange II XL site-directed mutagenesis kit (Agilent Technologies). All constructs were confirmed by Sanger Sequencing (Quintara Biosciences). The primers used for cloning enhancer sequences and establishing deletion mutants are listed in the Supplementary Table 4. For the luciferase assay, 7×10^4 HeLa cells were seeded per-well in 24 well plates 24 hours before transfection. All enhancer or mutant enhancer reporter constructs (200ng) were transfected with 20ng of pRL (Renilla luciferase control vector, Promega) and either empty vector (pcDNA3.1, Thermofisher Scientific) or hRARß-RXRa (Addgene, #135415) [51] with 2.4ul Fugene HD (Promega) in opti-MEM (Gibco). All transfectants were stimulated with either vehicle (DMSO) or 1 μM ATRA for 8 hours starting 24 hours after transfection. Each construct was transfected in quadruplicate and repeated at least three times. Luciferase assays were conducted 8 hours after stimulation using a dual luciferase kit (Dual-Luciferase® Reporter Assay System, Promega) according to manufacturer's protocol.

Luciferase activity was measured in SpectraMax MiniMax 300 imaging cytometer with SoftMax Pro6, Version 6.4 (Molecular Devices). pGL3-Basic-hCYP26A1-FL-luciferase (Addgene, #135566) [52] was used as a positive control in the assay system. The relative luciferase activity of each construct (arbitrary unit) was reported as the fold changed compared with empty vector.

X-gal staining

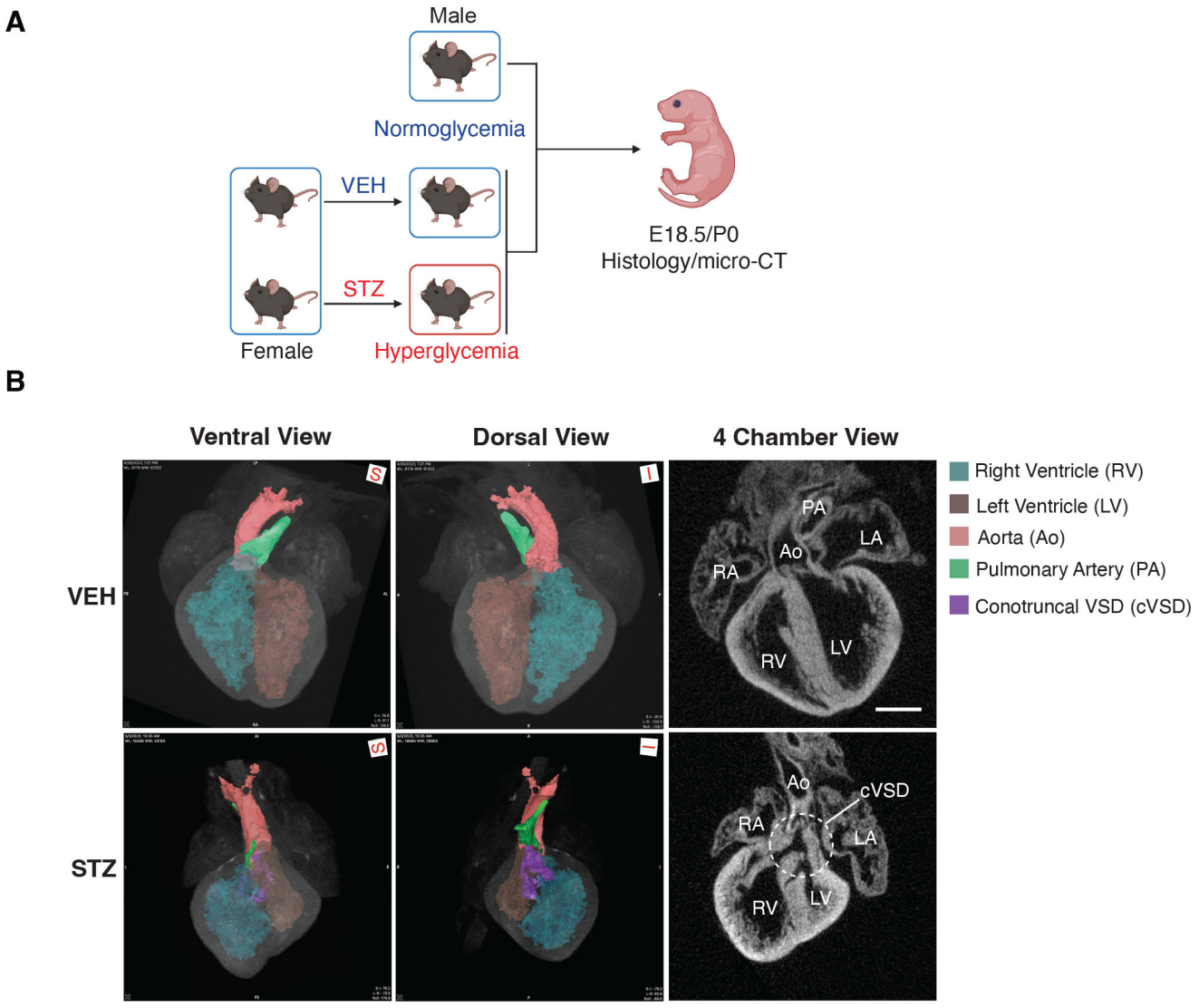
E10.5 mouse heart samples were microdissected in the same manner as above, were fixed with 4%PFA for 30 min on ice, and were washed three times for 10 min at room

temperature with X-gal buffer (5mM EGTA, 2mM MgCl2-6H2O, 0.02% IGEPAL NP40, 24 μM Deoxychalate with DPBS Ca/Mg). Then the samples were stained with X-gal staining solution (0.5 mg/ml X-gal, 5 mM K3Fe(CN)6, 5 mM K4Fe(CN)6.3H2O) for 1.5 hr at room temperature. Samples were washed three time in X-gal buffer for 10 min each time, followed by post-fixation with 4% paraformaldehyde overnight at 4 ° C. After washing with X-gal buffer three times, samples were transferred into 80% glycerol/PBS and were imaged using the Leica MZ 16F stereomicroscope (acquisition software LAS v.4.12).

Statistics and Reproducibility

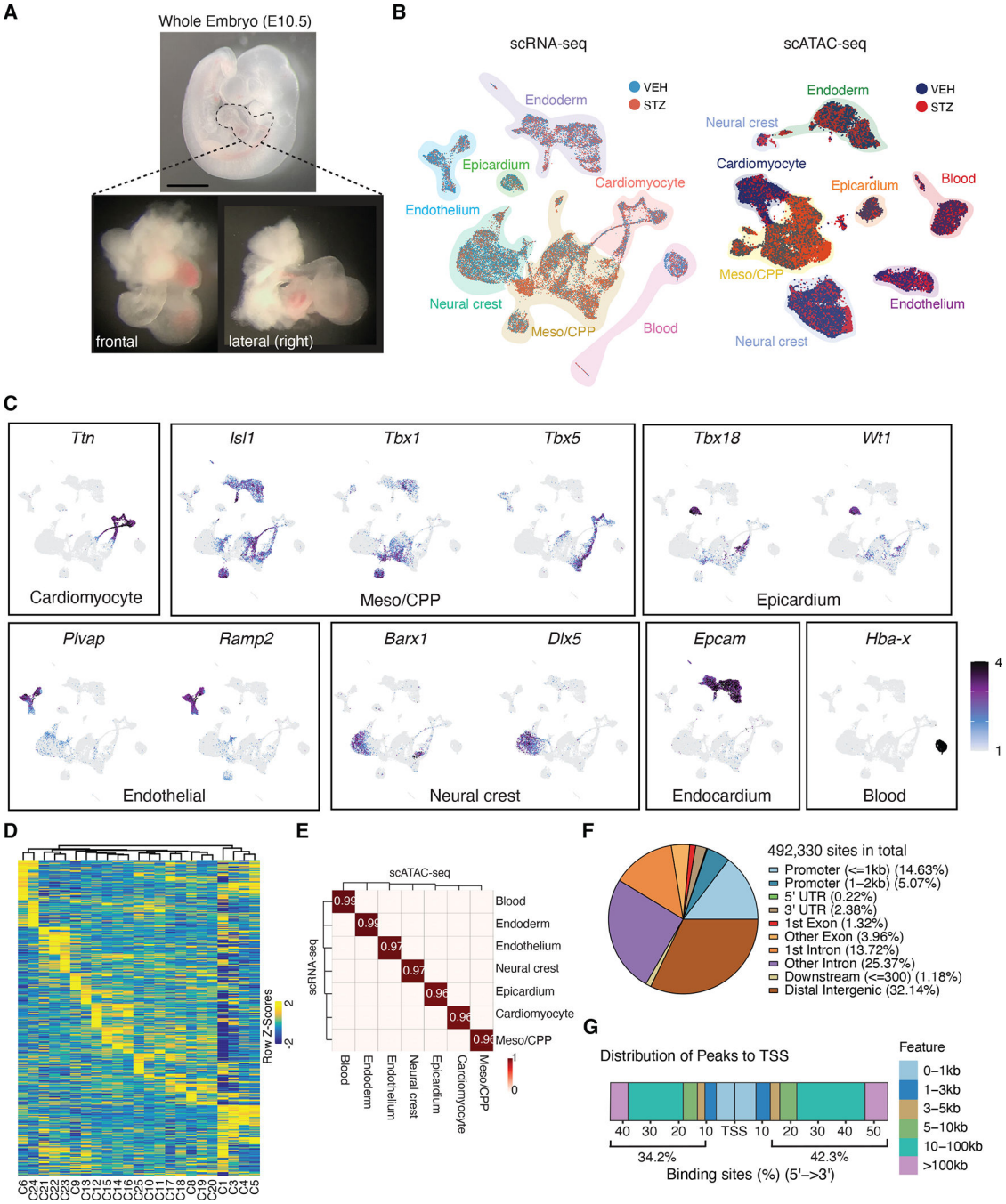
Statistical analyses were performed using GraphPad Prism 9. The number of replicates, statistical test used, test result, and the significance levels are described in each figure legend or Supplementary Table 1–3. For all the quantification, the mean ± s.e.m. is presented. No experimental sample was excluded from statistical analysis. When presenting representative results showing a gene expression pattern or a reporter gene expression pattern in an embryo, at least three independent embryos were analyzed.

Extended Data



Extended Data Fig. 1. Histological and micro-CT validation of the maternal diabetes model.

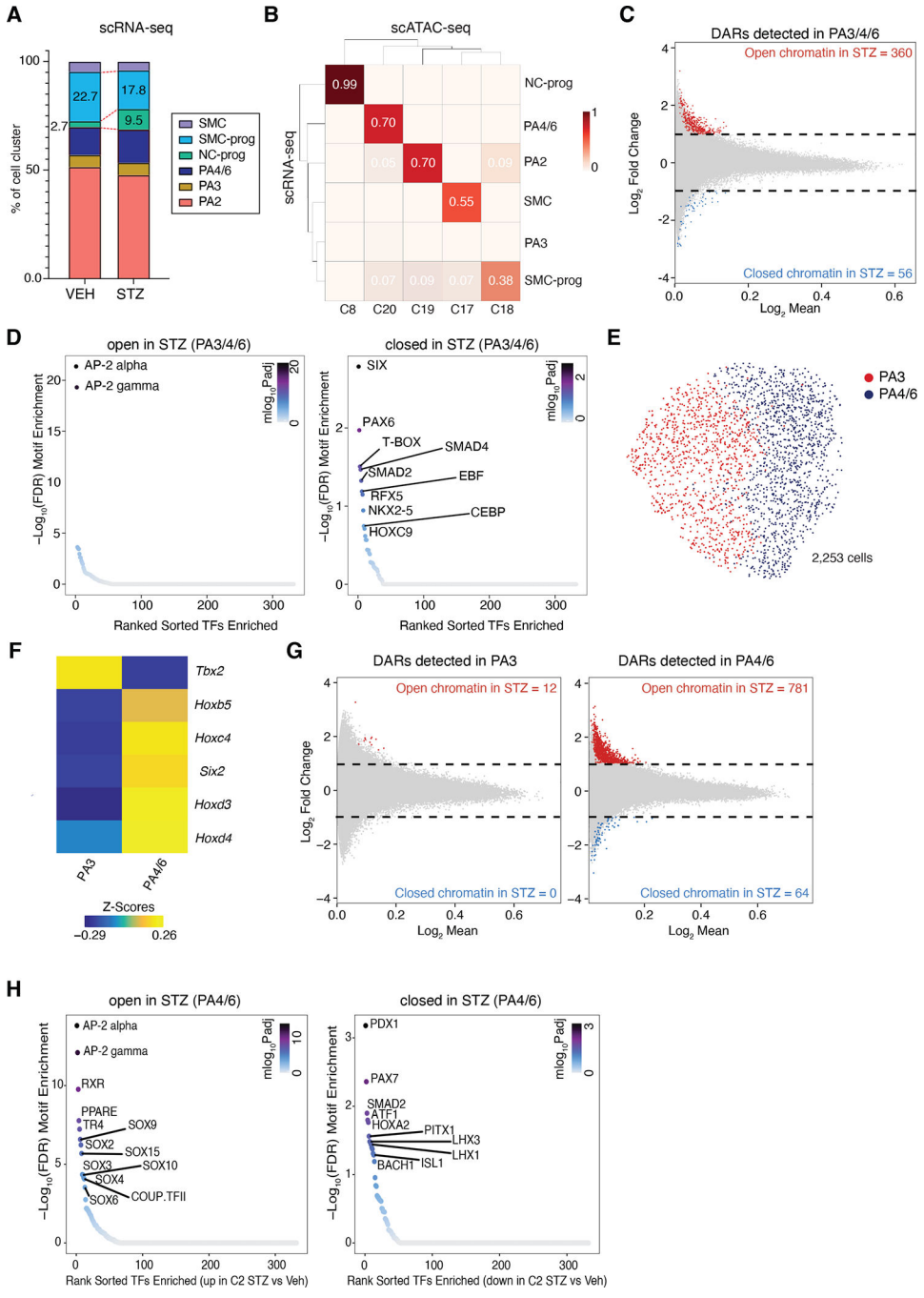
(A) The design of the *in vivo* maternal diabetic model experiment. After administration of either VEH or STZ, females in the STZ group with confirmed diabetes were mated with normoglycemic males, and heart samples at embryonic day 18.5 (E18.5) or postnatal day 0 (P0) were collected for histological examination. **(B)** Representative micro-CT images of the heart phenotypes detected in the diabetic model. The prevalence of each malformation is shown in Supplementary Table 1. The scale bar represents 500 μm.



Extended Data Fig. 2. Single cell multimodal analysis of cardio-pharyngeal region in maternal diabetes.

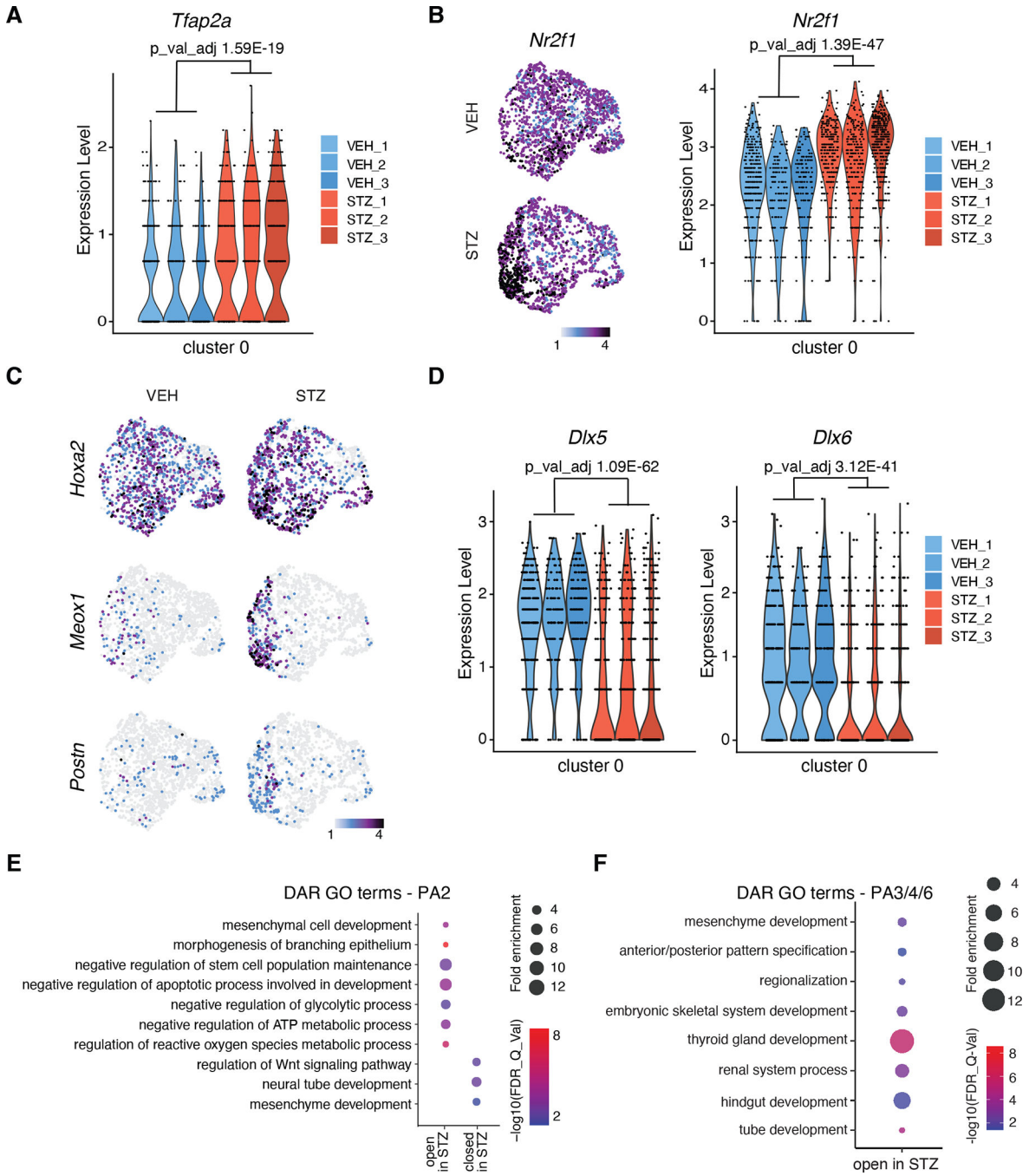
(A) Representative image of E10.5 embryo with detailed micro-dissected region used for scRNA/scATAC-seq experiment. Scale bar represents 1 mm. **(B)** scRNA-seq (left) or scATAC-seq (right) UMAP presentation colored by conditions. Different colors overlayed delineate cell type cluster annotations. **(C)** Expression patterns of representative cell type specific marker genes plotted on UMAP space shown in Fig. 1b. **(D)** Heatmap of marker gene scores per cluster of scATAC-seq. **(E)** Heatmap of Jaccard indices calculated

between scRNA-seq and scATAC-seq after integration. Values range from 0 to 1 (higher value represents closer annotation matching between the two modalities). **(F)** Genomic distribution of all the called peaks color coded by the genomic location as shown. Total called peaks = 492,330. **(G)** Distribution of all called peaks based on the distance from transcription start sites.



Extended Data Fig. 3. Maternal diabetes dysregulates epigenomic landscape of neural crest cells in pharyngeal arches 4 and 6.

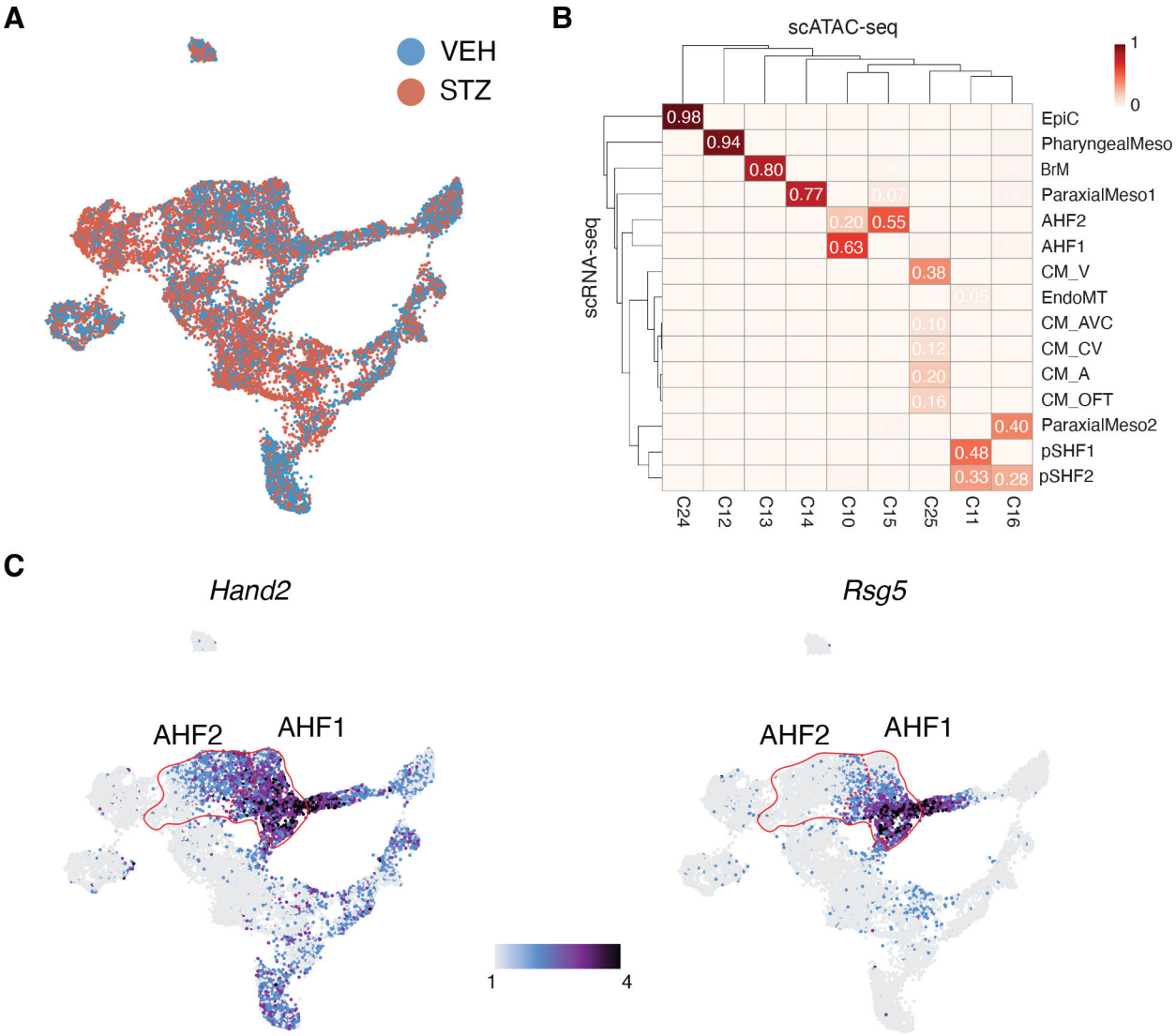
(A) Population distribution by sub-cell-type normalized to total number of cells per sample in neural crest cell subset data of scRNA-seq. Numbers inside the barplot represent the percentage of cell types of the total cell number. Statistics performed by permutation test in scRNA-seq data, comparing STZ vs. VEH, for NC-prog, FDR < 0.001, Log2FD = 1.88; for SMC-prog, FDR < 0.001, Log2FD = −0.33. **(B)** Heatmap of Jaccard indices calculated between neural crest cell scRNA-seq and scATAC-seq cell annotations after integration. Values range from 0 to 1 (the higher value represents closer annotation matching between those two modalities). **(C)** MA plot of DARs in PA3/4/6 population between VEH and STZ. Red dots represent the more accessible (open) (FDR < = 0.05 & Log2FC > = 1) and blue dots represent less accessible (closed) DARs in STZ (FDR < = 0.05 & Log2FC < = −1). **(D)** Enriched TF binding motifs in more accessible (left) or less accessible (right) DARs in STZ vs. VEH within the PA3/4/6 population. **(E)** scATAC-seq UMAP representation of neural crest cell C20 subset population colored by clusters (PA3 – dark red; PA4/6 – dark blue). **(F)** Heatmap of Gene Scores (GS) of curated marker genes based on scRNA-seq data for PA3 and PA4/6 neural crest. Scale indicates z-scored GS values. **(G)** MA plot of DARs between VEH and STZ in PA3 population (left) and PA4/6 population (right). Red dots represent the more accessible (open) (FDR < = 0.05 & Log2FC > = 1) and blue dots represent less accessible (closed) DARs in STZ (FDR < = 0.05 & Log2FC < = −1). **(H)** Enriched TF binding motifs in more accessible (left) and less accessible (right) DARs in STZ in PA4/6 population. NC-prog, neural crest cell progenitors; PA2, pharyngeal arch 2; PA3, pharyngeal arch 3; PA4/6, pharyngeal arch 4/6; SMC, smooth muscle cells; SMC-prog, smooth muscle cell progenitors.



Extended Data Fig. 4. Maternal diabetes dysregulates epigenomic and transcriptional landscape associated with cell differentiation and patterning in pharyngeal arch 2 neural crest.

(A) Violin plot of *Tfap2a* expression levels in cluster 0 of UMAP in Fig. 2i across 3 VEH and 3 STZ embryos (Wilcoxon Rank Sum test). **(B)** Expression of *Nr2f1* mRNA on UMAP space for PA2 neural crest cells (VEH – left top; STZ – left bottom). Scale bar indicates z-scored expression values. Violin plot of *Nr2f1* expression levels in cluster 0 of UMAP in Fig. 2i (right) (Wilcoxon Rank Sum test). **(C)** Expression of indicated genes on UMAP space for PA2 neural crest cells. Scale bar indicates z-scored expression values. **(D)** Violin

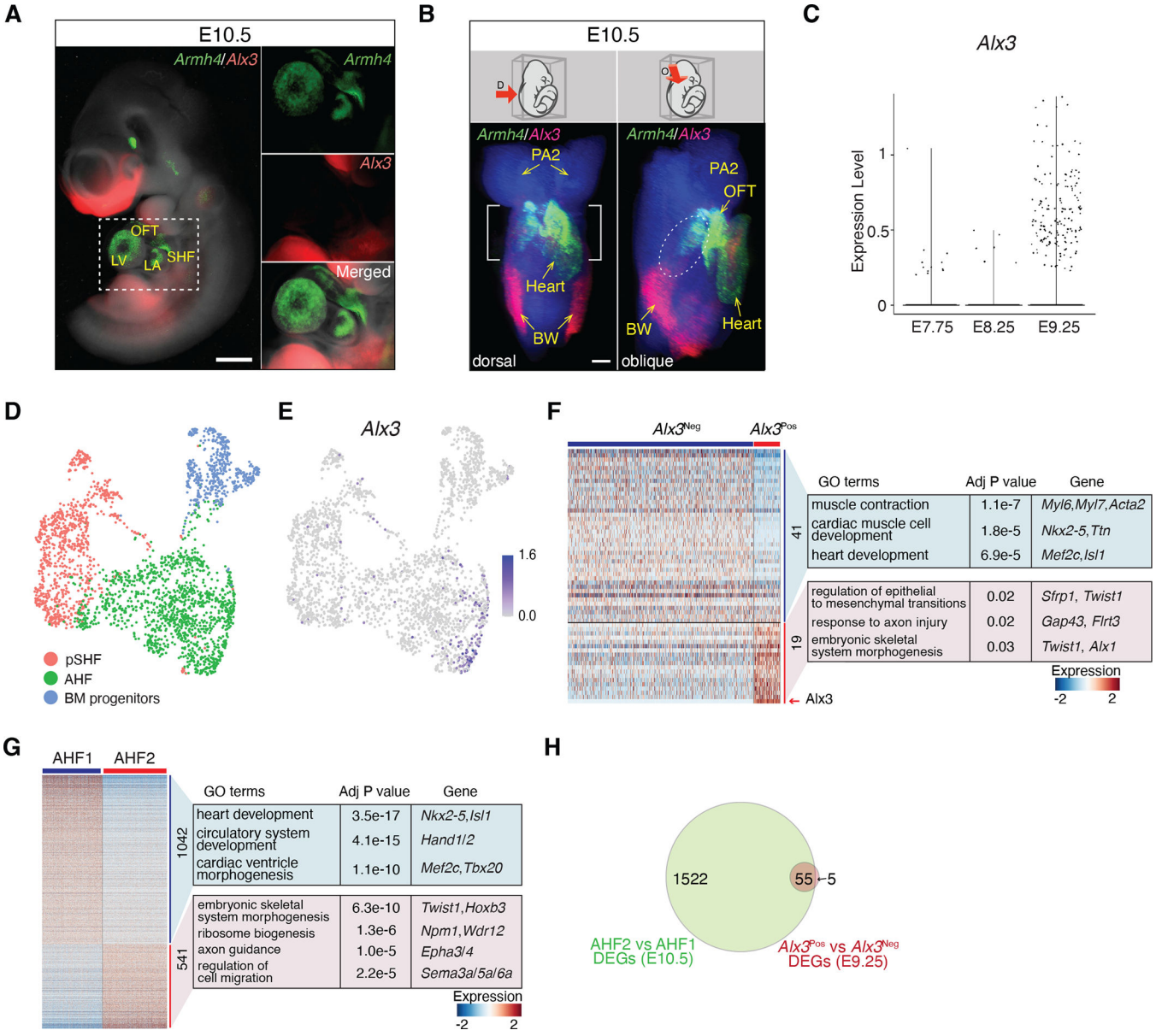
plots of *Dlx5* (left) and *Dlx6* (right) expression levels in cluster 0 of UMAP in Fig. 2i (Wilcoxon Rank Sum test). **(E)** Enriched GO terms in detected DARs in PA2 population using GREAT analysis. **(F)** Enriched GO terms in detected DARs in PA3/4/6 population using GREAT analysis.



Extended Data Fig. 5. Identification of distinct subsets of AHF progenitors.

(A) scRNA-seq UMAP representation of mesodermal population ('Meso/CPP', 'Cardiomyocyte', or 'Epicardium' in Fig. 1b) colored by conditions (VEH – blue; STZ – light red). **(B)** Heatmap of Jaccard indices between mesoderm cell scRNA-seq and scATAC-seq annotations after integration. Values range from 0 to 1 (the higher value represents closer annotation matching between those two modalities). CM_V, ventricular cardiomyocyte; CM_AVC, atrioventricular canal cardiomyocyte; CM_A, atrial cardiomyocyte; CM_SV, sinus venosus cardiomyocyte; CM_OFT, outflow tract cardiomyocyte; pSHF1/2, posterior second heart field 1/2; EndoMT, endothelial mesenchymal transition; EpiC, Epicardium; AHF1/2, anterior heart field 1/2; PharyngealMeso, pharyngeal mesoderm; ParaxialMeso1/2,

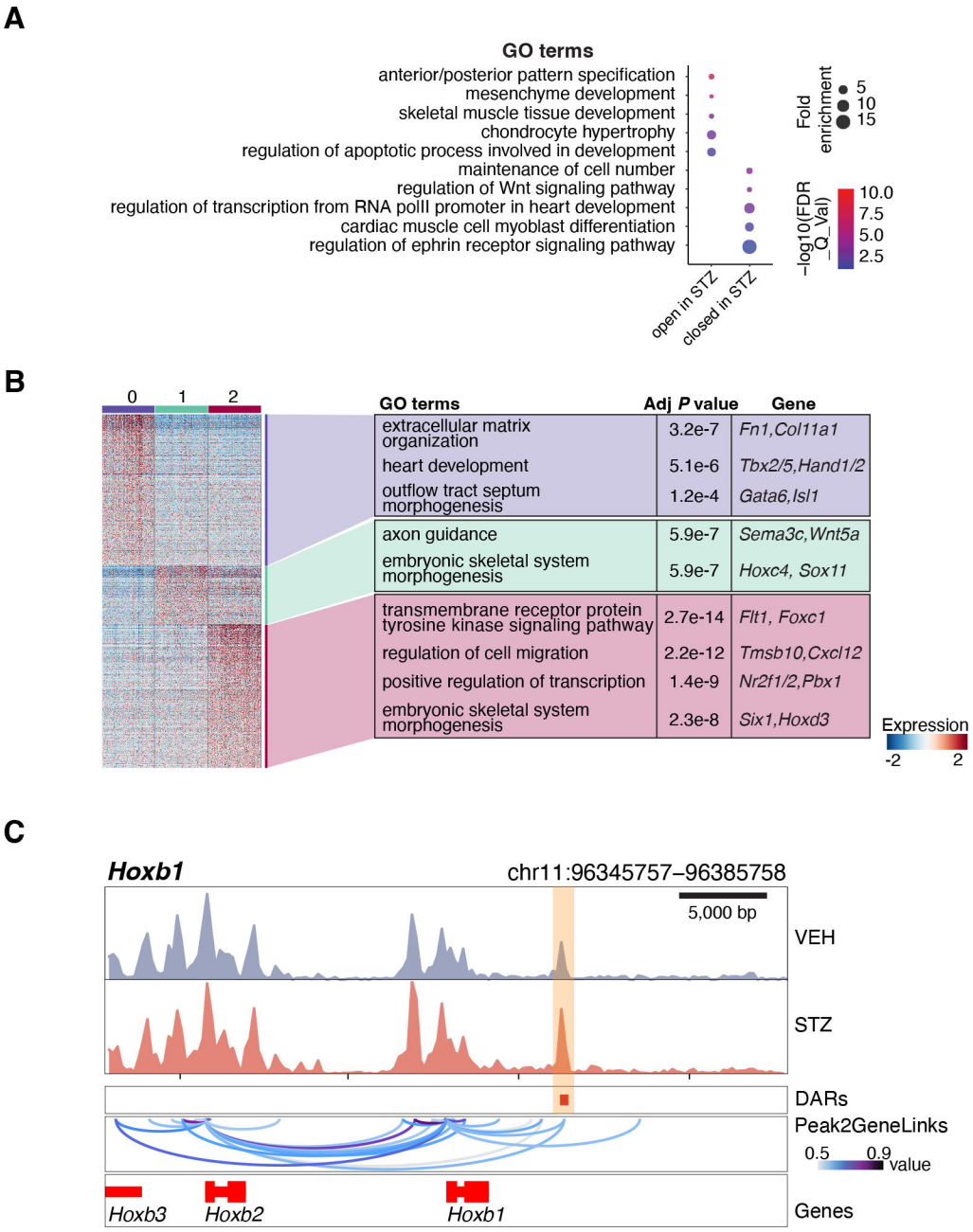
paraxial mesoderm 1/2; BrM, branchiomeric muscle. **(C)** Expression pattern of *Hand2* (left) and *Rgs5 (right)* on UMAP space. AHF1 and 2 are circled in red. Scale bar indicates z-scored expression values.



Extended Data Fig. 6. *Alx3*^{Pos} cells are a distinct subset of the AHF population.

(A) Representative images from RNA *in situ* hybridization for *Armh4* (green) and *Alx3* (red) in an E10.5 embryo from VEH treated female. The scale bar represents 500 μm. **(B)** Representative images from whole mount RNA *in situ* hybridization of E10.5 embryos using light sheet microscopy. *Armh4* (green) and *Alx3* (red) expression is shown from the dorsal view (D – left) and the right oblique view (O – right). A white bracket (left) highlights the anterior part of *Alx3*^{Pos} cells. A white dotted oval (right) highlights the *Alx3*^{Pos} cell streak on left side of the embryo from outflow tract (OFT) towards the posterolateral region.

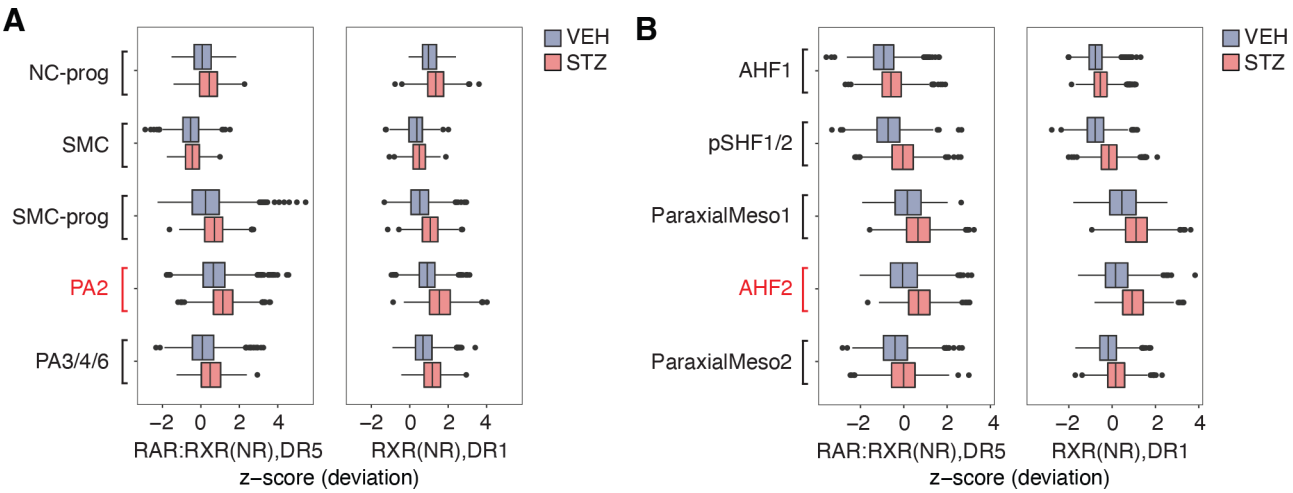
Still images were extracted from Supplementary video 2. Scale bar represents 100 μm. PA2, pharyngeal arch 2; BW, body wall. **(C)** The distribution of *Alx3* positive cells by scRNA-seq between E7.75 and E9.25. **(D)** scRNA-seq UMAP of cardiac progenitor cells at E9.25 from the same data as (C) color coded by cell type annotation. AHF, anterior heart field; BM progenitors, branchiomeric muscle progenitors; pSHF, posterior second heart field. **(E)** Expression of *Alx3* on the same UMAP as (D). **(F)** Heatmap of differentially expressed genes (DEGs) between $Alx3^{Neg}$ AHF and $Alx3^{Pos}$ AHF at E9.25. All detected DEGs that attained adjusted p-val < 0.05 and $Log_2FC > 0.25$ are shown. Top GO terms enriched in upregulated or downregulated DEGs are shown with representative genes composing each GO (Fisher's exact test, corrected for multiple testing using the Benjamini-Hochberg method). Scale bar indicates z-scored expression values. **(G)** Heatmap presentation of DEGs between AHF1 and AHF2 at E10.5 using only VEH cells in the scRNA-seq data. All detected DEGs that attained adjusted p-val < 0.05 and $Log_2FC > 0.25$ are shown. Top GO terms enriched in upregulated or downregulated DEGs are shown with representative genes composing each GO (Fisher's exact test, corrected for multiple testing using the Benjamini-Hochberg method). Scale bar indicates z-scored expression values. **(H)** Venn diagram representing the intersect between DEGs shown in (F-G).



Extended Data Fig. 7. PGDM disrupts anterior-posterior patterning in AHF2.

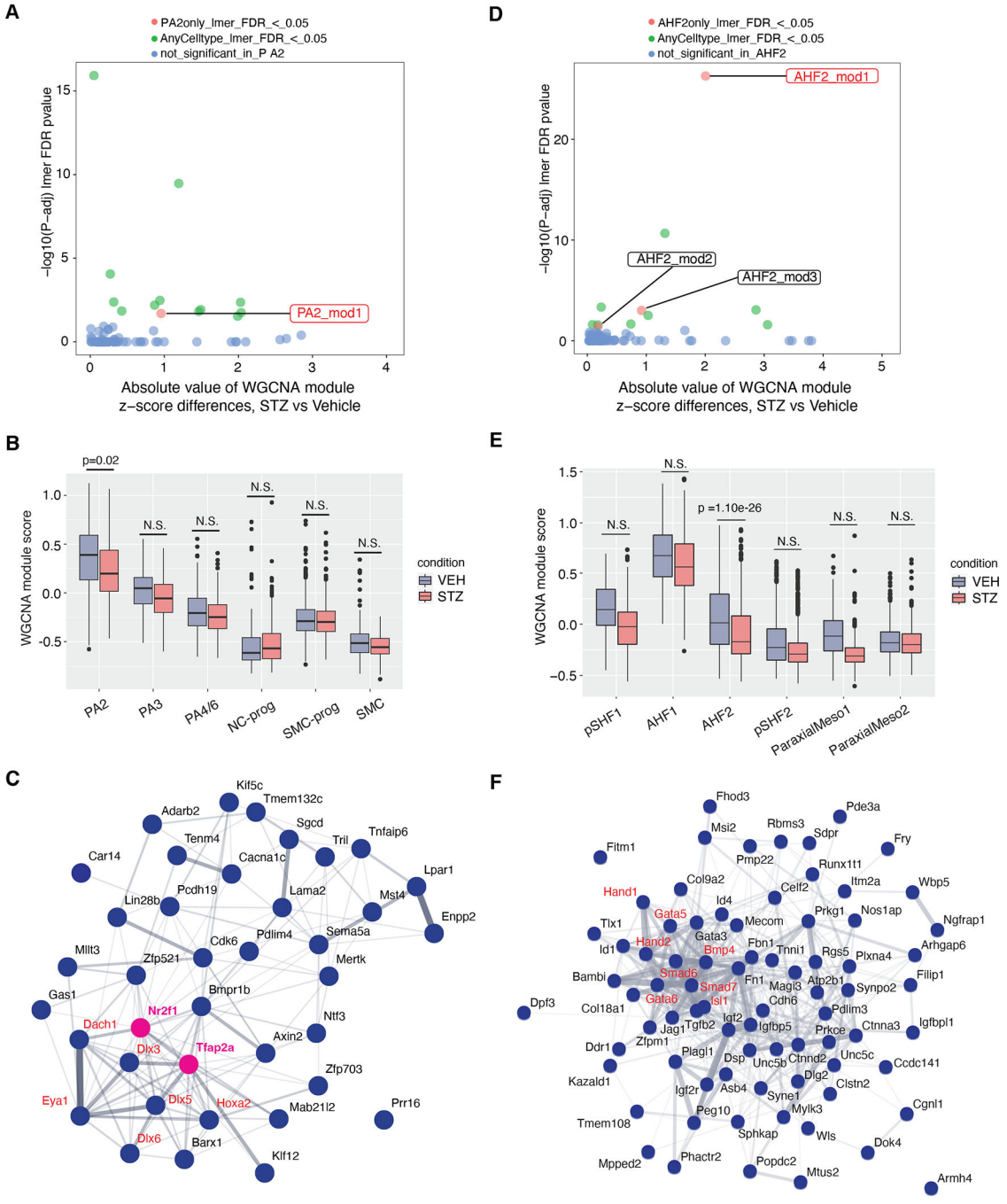
(A) Enriched GO terms in VEH vs. STZ DARs in AHF2 population using GREAT analysis. **(B)** Heatmap of marker genes of each of three subclusters found in *Alx3*^{Pos} AHF2. These marker genes were detected using only VEH-treated *Alx3*^{Pos} AHF2 cells (left). All marker genes that attained an adjusted p-val < 0.05 and $Log_2FC > 0.25$ are shown. Scale bar indicates z-scored expression values. Top GO terms enriched in marker genes for each sub cluster with statistical information and representative maker genes to corresponding GO term are shown (Fisher's exact test, corrected for multiple testing using the Benjamini-Hochberg method) (right). **(C)** Genome browser plots for *Hoxb1* locus. The top two rows represent the chromatin accessibility in VEH and in STZ within AHF2. The third track

from the top shows the genomic location of the DAR with more accessibility in STZ (red rectangles, highlighted by yellow box). The second track from the bottom represent the links between peaks and gene ('Peak2GeneLinks'), calculated by ArchR. Darker lines represent stronger links. The bottom track shows the gene location and transcriptional direction (red – positive strand; blue – negative strand).



Extended Data Fig. 8. Enhanced retinoic acid signaling in pharyngeal arch 2 and AHF2 in response to hyperglycemia.

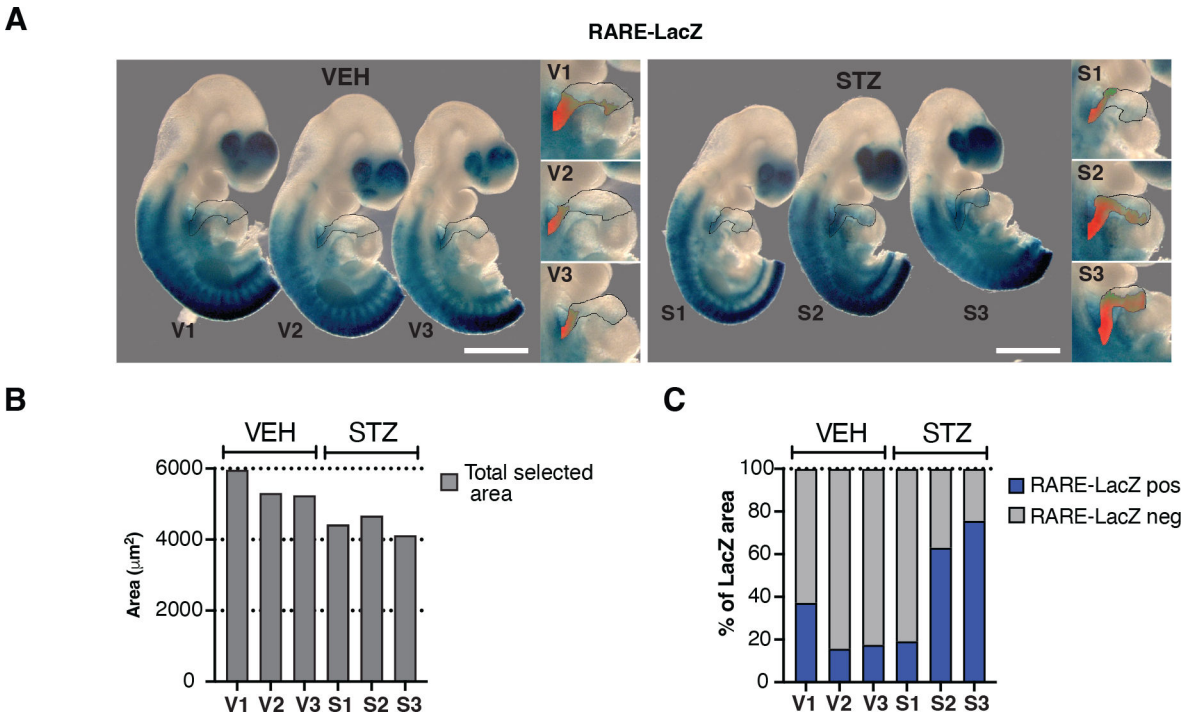
(A) Box plots of the distribution of ChromVAR deviation score for RAR and RXR transcription factor motifs for each cluster in the neural crest cell population. PA2 neural crest cells are highlighted in red. The X-axis shows the distribution of the Z-score. (STZ – red; VEH – blue) (n = 6 biological samples. n = 3 VEH replicates and n = 3 STZ replicates). **(B)** Box plots of the distribution of ChromVAR deviation score for RAR and RXR transcription factor motifs for each cluster in the mesoderm population. AHF2 cells are highlighted in red. The X-axis shows the distribution of the Z-score. (STZ – red; VEH – blue) (n = 6 biological samples. n = 3 VEH replicates and n = 3 STZ replicates). In the box plots, the central line indicates the median, box bounds represent the 25th and 75th percentiles, whiskers extend to values within 1.5 times the interquartile range, and outliers lie beyond this range.



Extended Data Fig. 9. Disrupted retinoic acid signaling is associated with dysregulation of gene regulatory networks in pharyngeal arch 2 and AHF2.

(A) Dot plot demonstrating the distribution of the WGCNA modules per cluster. X axis shows the Z-score differences of WGCNA module score per cluster between STZ and VEH and Y axis shows the statistical significance of the differences. Modules that are not statistically significant are shown in blue, and those that are statistically significant are shown in green or pink. Red label highlights selected module used for subsequent analysis. **(B)** Module scores for a gene module detected in the WGCNA analysis that

showed statistically significant variation between VEH and STZ only in PA2. Linear mixed effects models with mouse id as the random effect was used to test the significance of the mean difference in the module score between VEH and STZ (n = 6 biological samples. n = 3 VEH replicates and n = 3 STZ replicates) (linear mixed-effects model with Benjamini-Hochberg multiple-testing correction). **(C)** Map of functional protein-protein interactions (PPI) of genes composing the module described in (B), depicted using STRING. Genes composing a core of the PPI network and being downstream of *Tfap2* are highlighted in red and bold. (D) Dot plot demonstrating the distribution of the WGCNA modules per cluster. X axis shows the Z-score differences of WGCNA module score per cluster between STZ and VEH and Y axis shows the statistical significance of the differences. Modules that are not statistically significant are shown in blue, and those that are statistically significant are shown in green or pink. Red label highlights selected module used for subsequent analysis. **(E)** Module scores for a cardiac gene regulatory module detected in the WGCNA analysis that showed statistically significant variation between VEH and STZ only in AHF2. The same statistical test as (B) was used (n = 6 biological samples. n = 3 VEH replicates and n = 3 STZ replicates) (linear mixed-effects model with Benjamini-Hochberg multiple-testing correction). **(F)** Map of PPI of genes composing the module described in (E), depicted using STRING. Genes composing a core of the PPI network and being critical cardiac TFs or signaling genes are highlighted in red and bold. NC-prog, neural crest cell progenitors; PA2, pharyngeal arch 2; PA3, pharyngeal arch 3; PA4/6, pharyngeal arch 4/6; SMCs, smooth muscle cells; SMC-Prog, smooth muscle cell progenitors; pSHF1/2, posterior second heart field 1/2; AHF1/2, anterior heart field 1/2; ParaxialMeso1/2, paraxial mesoderm 1/2. In the box plots, the central line indicates the median, box bounds represent the 25th and 75th percentiles, whiskers extend to values within 1.5 times the interquartile range, and outliers lie beyond this range.



Extended Data Fig. 10. Anterior extension of retinoic acid signaling activity in STZ in vivo (complementary to Fig. 5e, f).

(A) X-gal staining of RARE-LacZ mouse fetuses at E10.5 from VEH and STZ groups (N = 3 each). Second heart field area and outflow tract area are circled with black lines. Highlighted area in magnified panels to the right show LacZ positive areas detected by threshold analysis as described in methods. The scale bar represents 1 mm. **(B)** Measurements of area of second heart field and outflow tract circled with black line in (A) (N = 3 each). **(C)** Percentage of LacZ positive area with in the second heart field area and outflow tract area circled with black line in (A) (N = 3 each).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We thank members of the Srivastava laboratory for discussion and feedback; Bethany Taylor from Gladstone Institutes for editorial and graphics assistance; Giovanni Maki from Gladstone Institutes for graphics assistance; and Kathryn Claiborn from Gladstone Institutes for editorial review. We acknowledge the center for advanced technology (CAT) for sequencing; the Gladstone Histology and Light Microscopy Core for their technical support and the Gladstone Animal Facility for support with mouse colonies. Fig. 1A, Fig. 5G, Extended Data Fig. 1A, and Supplementary Fig.1A were created with BioRender.com. National Institutes of Health/NHLBI grant P01 HL146366, R01 HL057181, R01 HL015100, R01 HL127240, Roddenberry Foundation, L.K. Whittier Foundation, Dario and Irina Sattui, Younger Family Fund, and Additional Ventures to D.S. The Japan Society for the Promotion of Science Overseas Research Fellowship to T.N. Additional Ventures to S.S.R. American Heart Association Postdoctoral Fellowship (#899270) to B.v.S. National Institutes of Health grant K08 HL157700, Sarnoff Cardiovascular Research Foundation, and Michael Antonov Charitable Foundation to A. Padmanabhan.

Data availability:

Data is available in the main text and the supplementary materials. All the sequencing data have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO series accession number GSE198905.

References and Notes

1. Hutson MR & Kirby ML Neural crest and cardiovascular development: A 20-year perspective. Birth Defects Res. C Embryo Today 69, 2–13 (2003). [PubMed: 12768653]
2. Srivastava D Making or breaking the heart: From lineage determination to morphogenesis. Cell 126, 1037–1048 (2006). [PubMed: 16990131]
3. Ogurtsova K et al. IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes Res. Clin. Pract. 128, 40–50, doi:10.1016/j.diabres.2017.03.024 (2017). [PubMed: 28437734]
4. Lammer EJ et al. Retinoic acid embryopathy. N. Engl. J. Med. 313, 837–841 (1985). [PubMed: 3162101]
5. Langston AW & Gudas LJ Retinoic acid and homeobox gene regulation. Curr. Opin. Genet. Dev. 4, 550–555, doi:10.1016/0959-437x(94)90071-a (1994). [PubMed: 7950323]
6. Waxman JS & Yelon D Increased Hox activity mimics the teratogenic effects of excess retinoic acid signaling. Dev. Dyn. 238, 1207–1213, doi:10.1002/dvdy.21951 (2009). [PubMed: 19384962]
7. Basu M & Garg V Maternal hyperglycemia and fetal cardiac development: Clinical impact and underlying mechanisms. Birth Defects Res. Part A: Clin. Mol. Teratol. 110, 1504–1516, doi:10.1002/bdr2.1435 (2018).
8. Helle E & Priest JR Maternal obesity and diabetes mellitus as risk factors for congenital heart disease in the offspring. Journal of the American Heart Association 9, e011541, doi:10.1161/JAHA.119.011541 (2020). [PubMed: 32308111]
9. Jovanovic L et al. The Diabetes in Early Pregnancy Study: beta-hydroxybutyrate levels in type 1 diabetic pregnancy compared with normal pregnancy. NICHD-Diabetes in Early Pregnancy Study Group (DIEP). National Institute of Child Health and Development. Diabetes Care 21, 1978–1984, doi:10.2337/diacare.21.11.1978 (1998). [PubMed: 9802754]
10. Xie Z et al. Metabolic regulation of gene expression by histone lysine beta-hydroxybutyrylation. Mol. Cell 62, 194–206, doi:10.1016/j.molcel.2016.03.036 (2016). [PubMed: 27105115]
11. Moazzen H et al. N-Acetylcysteine prevents congenital heart defects induced by pregestational diabetes. Cardiovasc. Diabetol. 13, 46, doi:10.1186/1475-2840-13-46 (2014). [PubMed: 24533448]
12. Yang P et al. Maternal hyperglycemia activates an ASK1-FoxO3a-caspase 8 pathway that leads to embryonic neural tube defects. Sci. Signal. 6, ra74, doi:10.1126/scisignal.2004020 (2013). [PubMed: 23982205]
13. Hrubec TC, Prater MR, Toops KA & Holladay SD Reduction in diabetes-induced craniofacial defects by maternal immune stimulation. Birth Defects Res. B Dev. Reprod. Toxicol. 77, 1–9, doi:10.1002/bdrb.20062 (2006). [PubMed: 16342200]
14. Hao Y et al. Integrated analysis of multimodal single-cell data. Cell 184, 3573–3587 e3529, doi:10.1016/j.cell.2021.04.048 (2021). [PubMed: 34062119]
15. Granja JM et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat. Genet. 53, 403–411, doi:10.1038/s41588-021-00790-6 (2021). [PubMed: 33633365]
16. Heinz S et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589, doi:10.1016/j.molcel.2010.05.004 (2010). [PubMed: 20513432]
17. Simoes-Costa M & Bronner ME Establishing neural crest identity: A gene regulatory recipe. Development 142, 242–257, doi:10.1242/dev.105445 (2015). [PubMed: 25564621]

18. Parker HJ, Pushel I & Krumlauf R Coupling the roles of Hox genes to regulatory networks patterning cranial neural crest. Developmental biology 444 Suppl 1, S67–S78, doi:10.1016/j.ydbio.2018.03.016 (2018). [PubMed: 29571614]
19. Dooley CM et al. The gene regulatory basis of genetic compensation during neural crest induction. PLoS Genet. 15, e1008213, doi:10.1371/journal.pgen.1008213 (2019). [PubMed: 31199790]
20. Rothstein M & Simoes-Costa M Heterodimerization of TFAP2 pioneer factors drives epigenomic remodeling during neural crest specification. Genome Res. 30, 35–48, doi:10.1101/gr.249680.119 (2020). [PubMed: 31848212]
21. Fan X et al. TWIST1 and chromatin regulatory proteins interact to guide neural crest cell differentiation. Elife 10, doi:10.7554/eLife.62873 (2021).
22. Soldatov R et al. Spatiotemporal structure of cell fate decisions in murine neural crest. Science (New York, N.Y 364, doi:10.1126/science.aas9536 (2019).
23. Akerberg BN et al. A reference map of murine cardiac transcription factor chromatin occupancy identifies dynamic and conserved enhancers. Nat. Commun. 10, 4907, doi:10.1038/s41467-019-12812-3 (2019). [PubMed: 31659164]
24. Prall OW et al. An Nkx2–5/Bmp2/Smad1 negative feedback loop controls heart progenitor specification and proliferation. Cell 128, 947–959 (2007). [PubMed: 17350578]
25. Schussler O et al. Cardiac neural crest cells: Their rhombomeric specification, migration, and association with heart and great vessel anomalies. Cell Mol. Neurobiol. 41, 403–429, doi:10.1007/s10571-020-00863-w (2021). [PubMed: 32405705]
26. Rada-Iglesias A et al. Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. Cell Stem Cell 11, 633–648, doi:10.1016/j.stem.2012.07.006 (2012). [PubMed: 22981823]
27. Maconochie M et al. Regulation of Hoxa2 in cranial neural crest cells involves members of the AP-2 family. Development 126, 1483–1494, doi:10.1242/dev.126.7.1483 (1999). [PubMed: 10068641]
28. Kutejova E, Engist B, Mallo M, Kanzler B & Bobola N Hoxa2 downregulates Six2 in the neural crest-derived mesenchyme. Development 132, 469–478, doi:10.1242/dev.01536 (2005). [PubMed: 15634706]
29. Van Otterloo E, Li H, Jones KL & Williams T AP-2alpha and AP-2beta cooperatively orchestrate homeobox gene expression during branchial arch patterning. Development 145, doi:10.1242/dev.157438 (2018).
30. Kirilenko P et al. Transient activation of meox1 is an early component of the gene regulatory network downstream of hoxa2. Mol. Cell Biol. 31, 1301–1308, doi:10.1128/MCB.00705-10 (2011). [PubMed: 21245383]
31. Alexanian M et al. A transcriptional switch governs fibroblast activation in heart disease. Nature 595, 438–443, doi:10.1038/s41586-021-03674-1 (2021). [PubMed: 34163071]
32. Kruzynska-Frejtag A et al. Periostin is expressed within the developing teeth at the sites of epithelial-mesenchymal interaction. Dev. Dyn. 229, 857–868, doi:10.1002/dvdy.10453 (2004). [PubMed: 15042709]
33. Verzi MP, McCulley DJ, De Val S, Dodou E & Black BL The right ventricle, outflow tract, and ventricular septum comprise a restricted expression domain within the secondary/anterior heart field. Developmental biology 287, 134–145 (2005). [PubMed: 16188249]
34. de Soysa TY 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). [PubMed: 31341279]
35. Beverdam A, Brouwer A, Reijnen M, Korving J & Meijlink F Severe nasal clefting and abnormal embryonic apoptosis in Alx3/Alx4 double mutant mice. Development 128, 3975–3986, doi:10.1242/dev.128.20.3975 (2001). [PubMed: 11641221]
36. Basson MA Signaling in cell differentiation and morphogenesis. Cold Spring Harb. Perspect. Biol. 4, doi:10.1101/cshperspect.a008151 (2012).
37. Stefanovic S et al. Hox-dependent coordination of mouse cardiac progenitor cell patterning and differentiation. Elife 9, doi:10.7554/eLife.55124 (2020).

38. Pineault N, Helgason CD, Lawrence HJ & Humphries RK Differential expression of Hox, Meis1, and Pbx1 genes in primitive cells throughout murine hematopoietic ontogeny. Exp. Hematol. 30, 49–57, doi:10.1016/s0301-472x(01)00757-3 (2002). [PubMed: 11823037]
39. Arrington CB, Dowse BR, Bleyl SB & Bowles NE Non-synonymous variants in pre-B cell leukemia homeobox (PBX) genes are associated with congenital heart defects. Eur. J. Med. Genet. 55, 235–237, doi:10.1016/j.ejmg.2012.02.002 (2012). [PubMed: 22426282]
40. Stennard FA et al. Murine T-box transcription factor Tbx20 acts as a repressor during heart development, and is essential for adult heart integrity, function and adaptation. Development 132, 2451–2462, doi:10.1242/dev.01799 (2005). [PubMed: 15843414]
41. Xavier-Neto J et al. A retinoic acid-inducible transgenic marker of sino-atrial development in the mouse heart. Development 126, 2677–2687 (1999). [PubMed: 10331979]
42. Langfelder P & Horvath S WGCNA: An R package for weighted correlation network analysis. BMC bioinformatics 9, 559, doi:10.1186/1471-2105-9-559 (2008). [PubMed: 19114008]
43. Furman BL Streptozotocin-induced diabetic models in mice and rats. Curr. Protoc. 1, e78, doi:10.1002/cpz1.78 (2021). [PubMed: 33905609]
44. Korsunsky I et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods 16, 1289–1296, doi:10.1038/s41592-019-0619-0 (2019). [PubMed: 31740819]
45. Chen EY et al. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. BMC bioinformatics 14, 128, doi:10.1186/1471-2105-14-128 (2013). [PubMed: 23586463]
46. Sarropoulos I et al. Developmental and evolutionary dynamics of cis-regulatory elements in mouse cerebellar cells. Science (New York, N.Y 373, doi:10.1126/science.abg4696 (2021).
47. Schep AN, Wu B, Buenrostro JD & Greenleaf WJ chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat. Methods 14, 975–978, doi:10.1038/nmeth.4401 (2017). [PubMed: 28825706]
48. Zhu LJ et al. ChIPpeakAnno: A Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC bioinformatics 11, 237, doi:10.1186/1471-2105-11-237 (2010). [PubMed: 20459804]
49. Yu G, Wang LG & He QY ChIPseeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31, 2382–2383, doi:10.1093/bioinformatics/btv145 (2015). [PubMed: 25765347]
50. McLean CY et al. GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. 28, 495–501, doi:10.1038/nbt.1630 (2010). [PubMed: 20436461]
51. Zolfaghari R et al. CYP26A1 gene promoter is a useful tool for reporting RAR-mediated retinoid activity. Anal. Biochem. 577, 98–109, doi:10.1016/j.ab.2019.04.022 (2019). [PubMed: 31039331]
52. Zolfaghari R & Ross AC Hepatocyte nuclear factor 4alpha (HNF4alpha) in coordination with retinoic acid receptors increases all-trans-retinoic acid-dependent CYP26A1 gene expression in HepG2 human hepatocytes. J. Cell. Biochem. 115, 1740–1751, doi:10.1002/jcb.24839 (2014). [PubMed: 24819304]

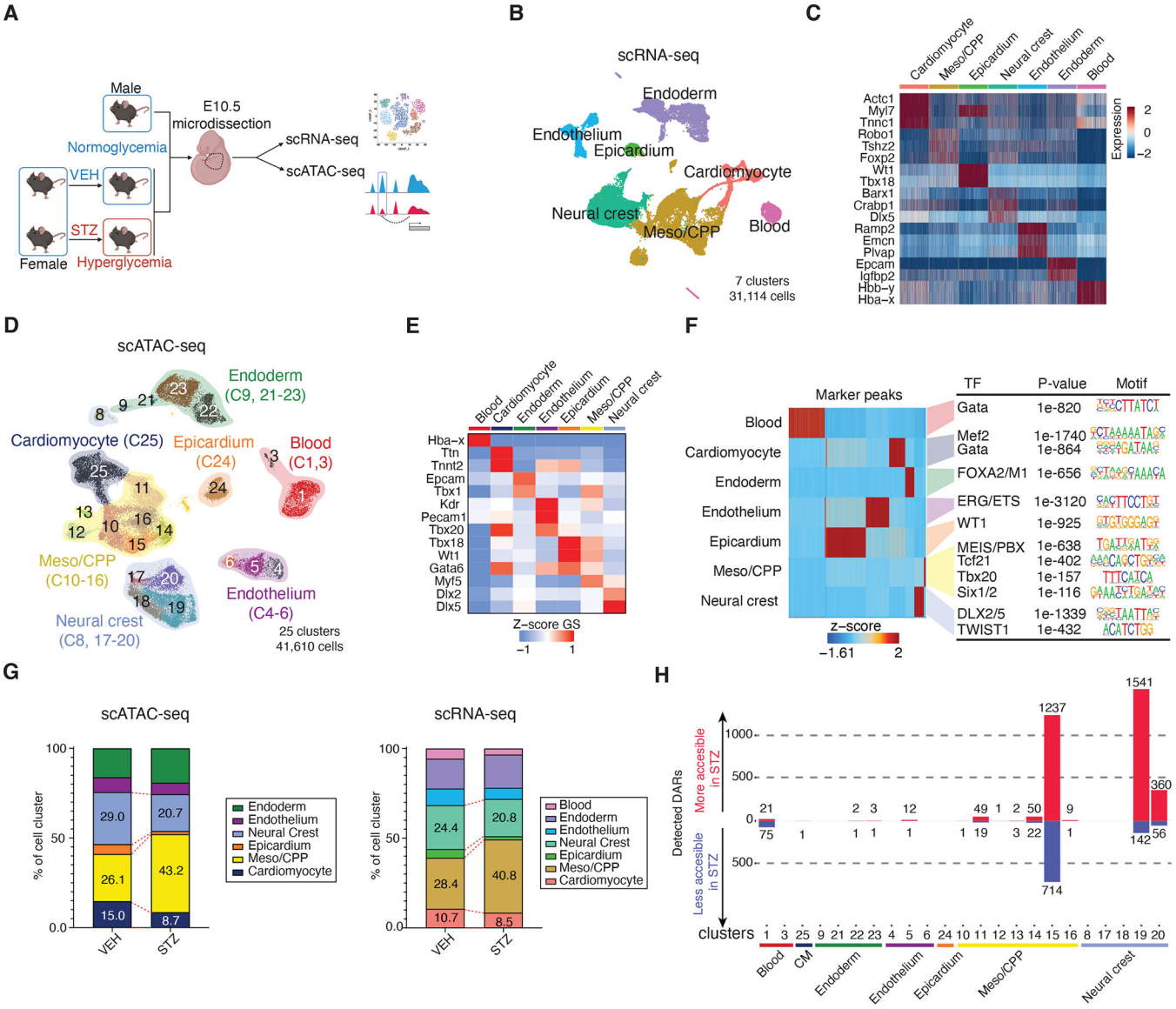


Fig. 1. Integrated Analysis of scRNA-seq and scATAC-seq Reveals Highly Selective Chromatin Accessibility Alterations in PGDM.

(A) Strategy for the *in vivo* PGDM model after treatment of either vehicle (VEH) or streptozotocin (STZ). Samples were collected from area indicated with dotted line at E10.5 and split to perform scRNA-seq and scATAC-seq using the 10x Genomics Chromium system. **(B)** UMAP representation of scRNA-seq of all captured cells from the cardio-pharyngeal area at E10.5 colored by cluster identity, including VEH and STZ samples (n=3 each). **(C)** Expression heat map of marker genes of each defined population. Scale indicates z-scored expression values. All genes represented have an adjusted p-value less than 1×10^{-10} (Wilcoxon rank-sum test). **(D)** UMAP representation of all captured cells in scATAC-seq colored by cluster identity and grouped based on cell types. Cells were collected from the same embryos as those used for scRNA-seq shown in (B). **(E)** Heatmap of Gene Scores (GS) from curated marker genes based on existing knowledge and scRNA-seq data for each cell type. Scale indicates z-scored GS values. **(F)** Top enriched TF motifs in

detected marker peaks (FDR < 1.00e-08 and Log2FC > 2) for each cell type analyzed using HOMER de novo analysis (hypergeometric test). **(G)** Population distribution of each cell type normalized to total number of cells per sample in scATAC-seq (left) and scRNA-seq (right). **(H)** Distribution of detected differentially accessible regions (DARs) between VEH and STZ groups in each sub cluster across all cell type. (FDR < 0.05 and |LogFC| > 1). Meso/CPP; Mesodermal cardio-pharyngeal progenitors.

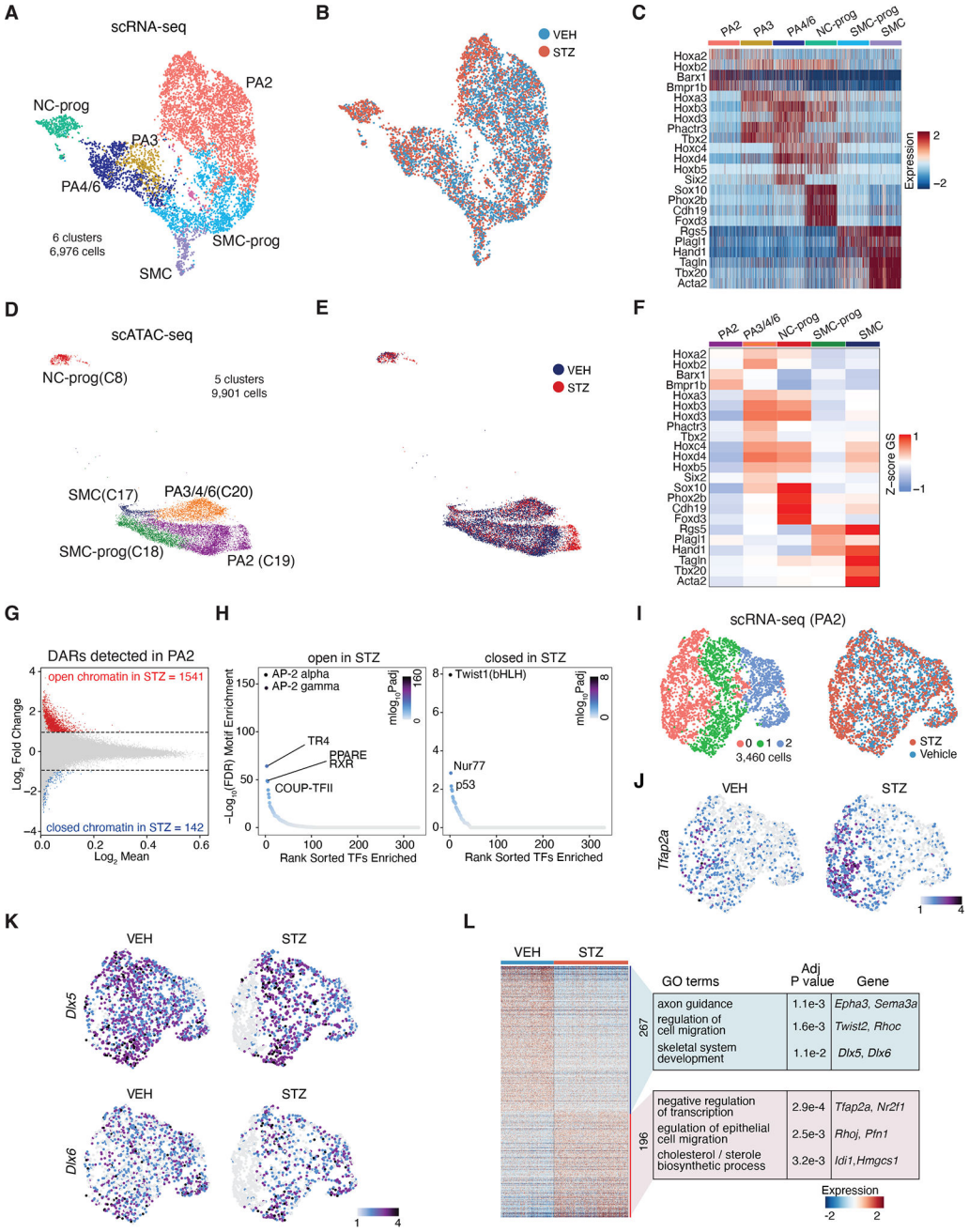


Fig. 2. Maternal Diabetes Disrupts the Epigenomic Landscape of Craniofacial Neural Crest Cells in the Second Pharyngeal Arch.

(A) scRNA-seq UMAP representation of the neural crest cell population from Figure 1B, including VEH (n=3) and STZ cells (n=3), colored by cluster identity. **(B)** scRNA-seq UMAP representation of neural crest cells colored by conditions. **(C)** Heatmap for marker gene expression in the neural crest cell sub-populations. The Hox genes delineate spatial information. Scale indicates z-scored expression values. **(D)** scATAC-seq UMAP representation of the neural crest cell population colored by cell type identity with original cluster numbers in parenthesis. **(E)** scATAC-seq UMAP representation of the neural crest cell population colored by conditions. **(F)** Heatmap of GS of curated marker genes based

on scRNA-seq data for each neural crest sub-type. Scale indicates z-scored GS values. **(G)** MA plot of differentially accessible regions (DARs) in PA2 population between VEH and STZ. Red dots represent the more accessible (open) chromatin region in STZ (FDR <= 0.05 & Log2FC >= 1) and blue dots represent less accessible (closed) chromatin region in STZ (FDR <= 0.05 & Log2FC <= −1). **(H)** Enriched TF binding motifs in more accessible (left) or less accessible (right) DARs in STZ within PA2. **(I)** scRNA-seq UMAP representation of PA2 neural crest cells. UMAP colored by cluster identity (left) or by conditions (right). **(J)** Expression of *Tfap2a* on UMAP of PA2 neural crest cells in VEH or STZ embryos. Scale indicates z-scored expression values. **(K)** Expression of *Dlx5* and *Dlx6* on UMAP of PA2 neural crest cells in VEH or STZ embryos. Scale indicates z-scored expression values. **(L)** Heatmap representation of differentially expressed genes (DEGs) in PA2 subset cluster 0 between VEH and STZ (adjusted p-val < 0.05 and Log2FC > 0.25). Top GO terms enriched in upregulated or downregulated genes are shown with representative genes for each GO term (Fisher's exact test, corrected for multiple testing using the Benjamini-Hochberg method). Scale indicates z-scored expression values. NC-prog, neural crest cell progenitors; PA2, pharyngeal arch 2; PA3, pharyngeal arch 3; PA4/6, pharyngeal arch 4/6; SMC, smooth muscle cells; SMC-prog, smooth muscle cell progenitors.

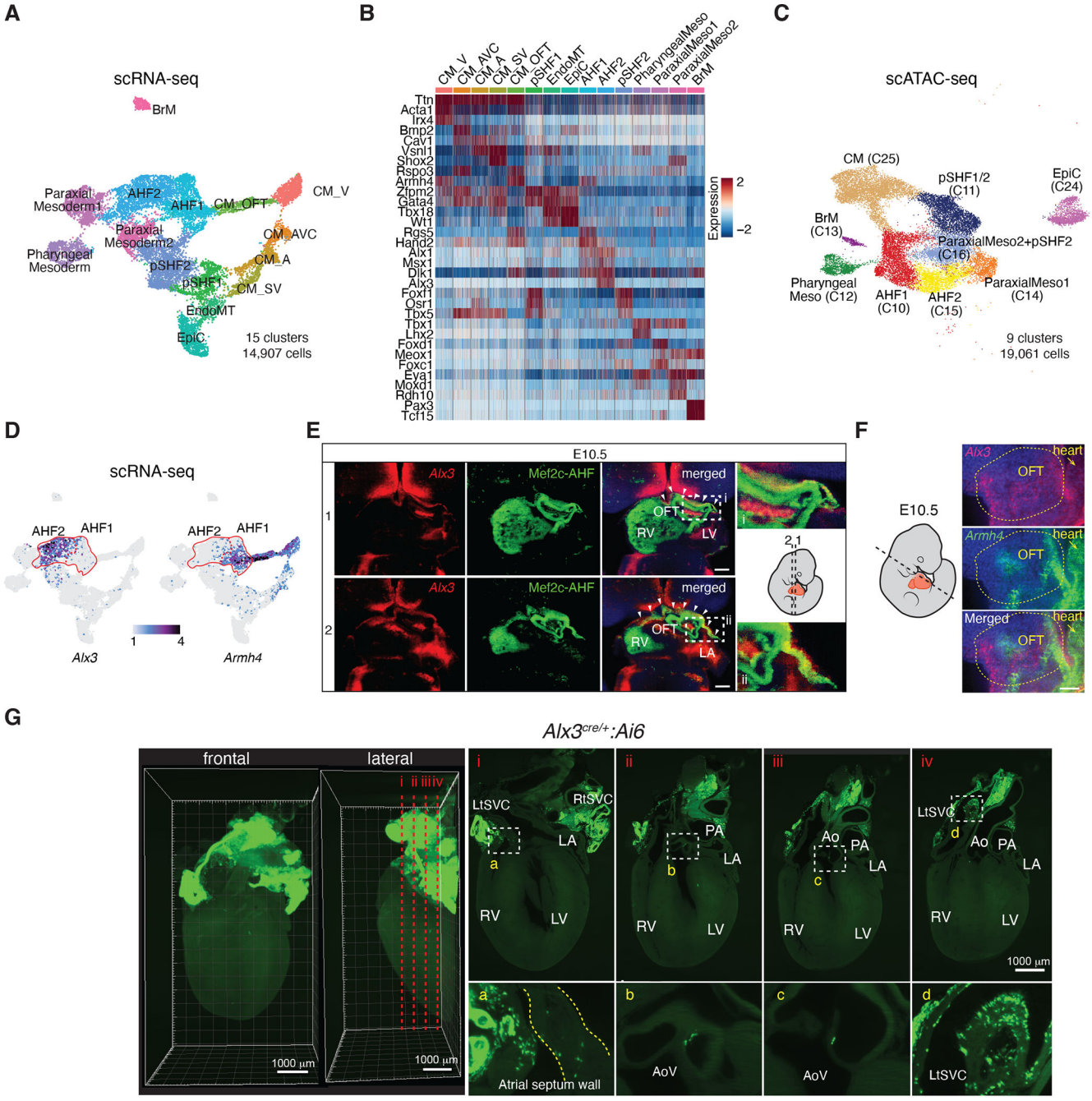


Fig. 3. Maternal Diabetes Epigenetically Alters *Alx3*-Positive Anterior Heart Field Cells, Which Contribute to the Aortic Valve and Atrial Wall.

(A) scRNA-seq UMAP representation of mesodermal cell subset population ("Meso/CPP", "Cardiomyocyte", or "Epicardium" in Figure 1B), including both VEH and STZ, colored by cluster identity. **(B)** Heatmap of expression levels of top marker genes for each sub-cell-type in the mesodermal cell population. Scale indicates z-scored expression values. All genes represented have an adjusted p-value smaller than 1e-10 (Wilcoxon rank-sum test). **(C)** scATAC-seq UMAP representation of mesodermal cell population, including both VEH and STZ, colored by cluster identity. **(D)** Expression pattern of *Alx3* (left) and *Armh4* (right).

AHF1 and 2 are circled in red. Scale bar indicates z-scored expression values. **(E)** Optical sections from the same data as Supplementary Video 1, with angle of section indicated in cartoon (section 1, top panels from a more ventral slice; section 2, bottom panels from a more dorsal slice). *Alx3* (red), Mef2c-AHF (green) and DAPI (blue) are shown. The scale bar represents 100 μm. **(F)** Optical sections of outflow tract region (circled by yellow dotted line) from the same data as Supplementary Video 2, with angle of section indicated in cartoon. *Armh4* (green, top), *Alx3* (red, middle), and DAPI (blue) are shown. The scale bar represents 50 μm. OFT, outflow tract. **(G)** 3D reconstructed images of a neonatal heart from an $Alx3^{Cre/+}$:Ai6 mouse using light sheet microscopy from frontal and lateral views. Optical sections were made at serial levels from (i) to (iv), as shown in the lateral view. In each optical section, areas of ZsGreen positive cells were magnified. (a) atrial wall, (b) and (c) aortic valve area, and (d) right superior vena cava. Scale bar 500 μm for two images on the left and 1000 μm for the images (i)-(iv).

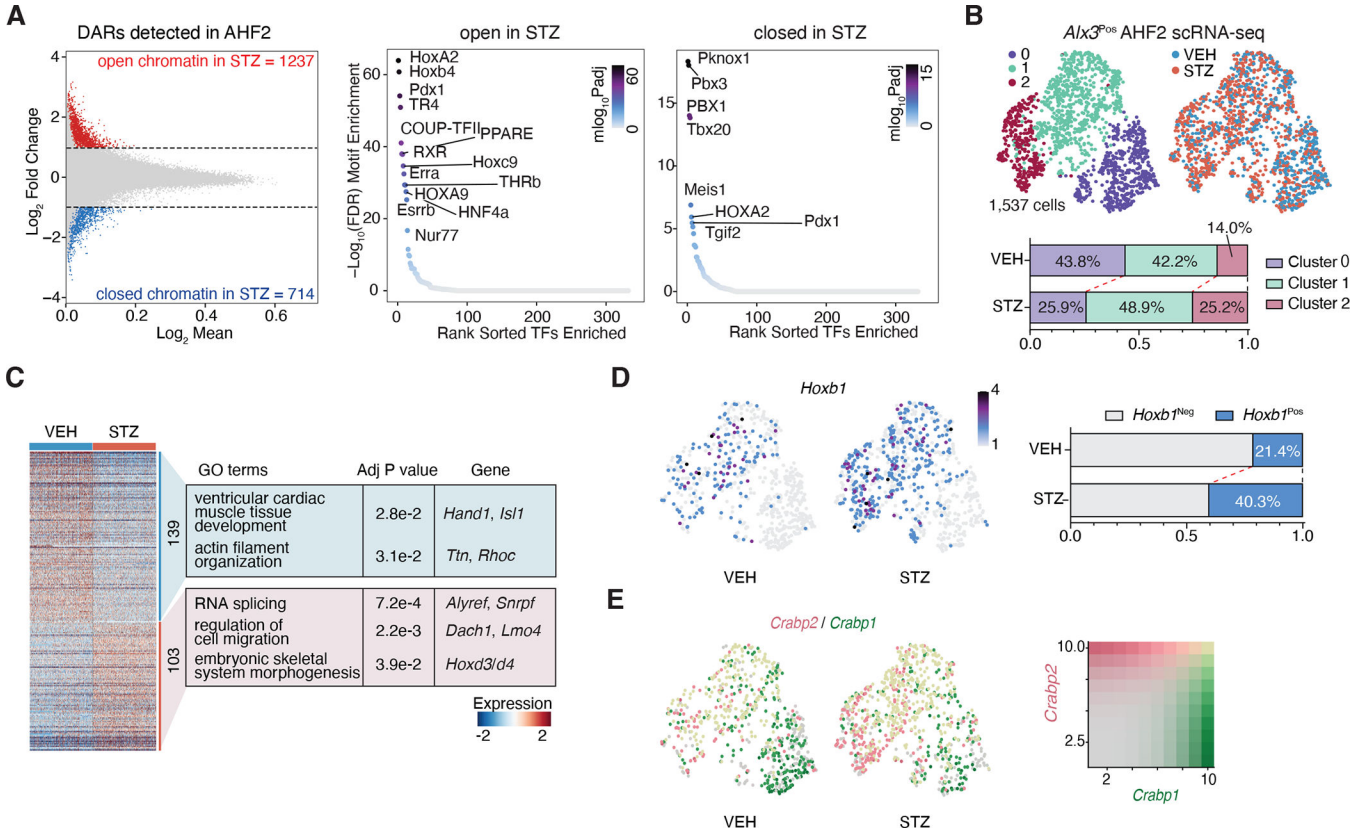


Fig. 4. Maternal Diabetes Epigenetically Alters *Alx3*-Positive Anterior Heart Field Cells with Disruption of Anterior-Posterior Patterning.

(A) MA plot of DARs between VEH and STZ (left) in AHF2 population. Red and blue dots represent more or less accessible chromatin regions, respectively in STZ (FDR <= 0.05 & Log2FC >= 1). Enriched TF binding motifs in more accessible open (middle) or less accessible closed (right) DARs in AHF2 population from STZ embryos. **(B)** scRNA-seq UMAP representation of *Alx3*^{Pos} AHF2 cells colored by cluster identity (upper left) or by conditions (upper right). Population distribution normalized to total number of cells per each sample in *Alx3*^{Pos} AHF2 cells. Each number inside the bar plot (bottom) represents the percentage of the corresponding cell type among the total cell number of each condition. **(C)** Heatmap of differentially expressed genes (DEGs) between VEH or STZ *Alx3*^{Pos} AHF2 cells. All the detected DEGs with adjusted p-val < 0.05 and Log2FC > 0.25 are shown. Top GO terms enriched in upregulated or downregulated DEGs are shown with representative genes (Fisher's exact test, corrected for multiple testing using the Benjamini-Hochberg method). Scale bar indicates z-scored expression values. **(D)** Expression of *Hoxb1* on UMAP of *Alx3*^{Pos} AHF2 cells. Scale bar indicates z-scored expression values. Percentage of *Hoxb1* positive cells to the total cell number in each condition (right). Statistics of STZ vs VEH performed by permutation test in scRNA-seq data for *Hoxb1* positive cells, p=0.0001, Log2FD=0.82. **(E)** Expression of *Crabp1* (green) or *Crabp2* (red) on UMAP of *Alx3*^{Pos} AHF2 cells. Scale grid indicates z-scored expression values for each gene. CM_V, ventricular cardiomyocyte; CM_AVC, atrioventricular canal cardiomyocyte; CM_A, atrial cardiomyocyte; CM_SV, sinus venosus cardiomyocyte;

CM_OFT, outflow tract cardiomyocyte; pSHF1/2, posterior second heart field 1/2; EndoMT, endothelial mesenchymal transition; EpiC, Epicardium; AHF1/2, anterior heart field 1/2; PharyngealMeso, pharyngeal mesoderm; ParaxialMeso1/2, paraxial mesoderm 1/2; BrM, branchiomeric muscle.

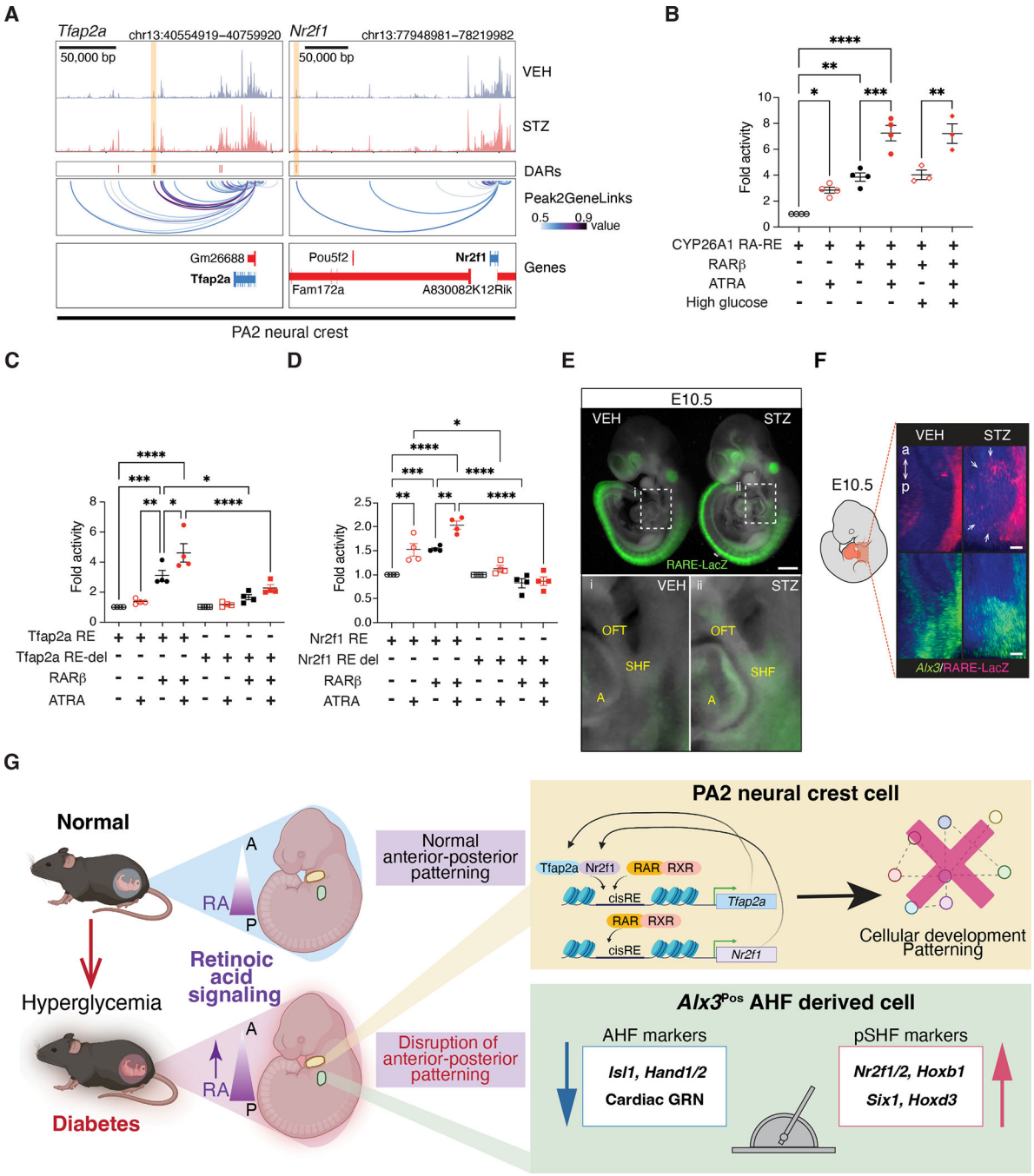


Fig. 5. Disruption of Retinoic Acid Signaling Associated with Pharyngeal Arch and *Alx3*^{pos} Anterior Heart Field Anterior-Posterior Patterning Defect.

(A) Genome browser plots for the *Tfap2a* or *Nr2f1* locus. The top two rows represent the chromatin accessibility in VEH or STZ. The third track shows genomic locations of detected DARs with more accessibility in STZ (red rectangles). The DARs selected for further analyses are highlighted in yellow boxes. The second track from the bottom represents the links between peaks and a gene ("Peak2GeneLinks"), calculated by ArchR. Darker lines represent stronger links. The bottom track shows the gene location and

transcriptional direction (red – positive strand; blue – negative strand). **(B)** Relative luciferase activity driven by the known RA response elements (RA-RE) for hCYP26A1 with exogenous RARß transduction, addition of 1μM ATRA and high glucose condition (25 mM). (n=3~4 independent experiments and n=3~4 technical samples with each; Tukey's multiple comparisons test on one-way ANOVA) **(C-D)** Relative luciferase activity driven by the candidate regulatory region for *Tfap2a* and *Nr2f1* with (RE)/without deletion (RE-del) of the putative RAR/RXR binding site under conditions indicated. (n=4 independent experiments and n=3~4 technical samples with each; Tukey's multiple comparisons test on one-way ANOVA) **(E)** Whole mount RNA *in situ* hybridization for LacZ transgene driven *in vivo* by a RA response element (RARE-LacZ) in E10.5 VEH or STZ embryos. Regions highlighted by the white dashed squares are shown in (i) and (ii). Scale bar represents 500 μm. OFT, outflow tract; SHF, second heart field; A, atrium. **(F)** Representative images of whole mount RNA in situ hybridization of the SHF area of E10.5 embryos in lateral view using light sheet microscopy. RARE-LacZ (magenta, top) and *Alx3* (green, bottom) are shown for VEH or STZ embryos. Still images were extracted from Supplementary Video 3. White arrows indicate ectopic LacZ signals in *Alx3* positive area. The scale bars represent 50 μm. The cartoon illustrates the approximate location shown in the images. **(G)** Model of PA2 neural crest and *Alx3*^{pos} AHF-specific epigenomic and transcriptional alterations and their consequences through aberrant RA signaling in maternal diabetes. In **(B-D)**: Data are presented as mean values +/– SEM. *p<0.05, **p<0.005, ***p<0.001, and ****p<0.0001 by Tukey's multiple comparisons test after ANOVA.