Active remodeling of the chromatin landscape directs extravillous trophoblast cell lineage development

Kaela M. Varberg[1,2,*,†], Esteban M. Dominguez[1,2,†], Boryana Koseva[3], Ross P. McNally[1,2,¥], Ayelen Moreno-Irusta[1,2], Emily R. Wesley[3], Khursheed Iqbal[1,2], Warren A. Cheung[3], Hiroaki Okae[4], Takahiro Arima[4], Michael Lydic[5], Kristin Holoch[5], Courtney Marsh[1,5,] Michael J. Soares[1,2,5,6,*], and Elin Grundberg[1,2,3,*]

[1]Institute for Reproduction and Perinatal Research, University of Kansas Medical Center, Kansas City, Kansas 66160

[2]Department of Pathology & Laboratory Medicine, University of Kansas Medical Center, Kansas City, KS 66160

[3]Genomic Medicine Center, Children's Mercy Research Institute, Children's Mercy Kansas City, Kansas City, MO 64108

[4]Department of Informative Genetics, Environment and Genome Research Center, Tohoku University Graduate School of Medicine, Sendai 980-8575, Japan

[5]Department of Obstetrics and Gynecology, University of Kansas Medical Center, Kansas City, Kansas 66160

[6]Center for Perinatal Research, Children's Mercy Research Institute, Children's Mercy, Kansas City, MO 64108

Running title: Epigenomic Regulation of Extravillous Trophoblast Cells

[†]Co-first authors.

[¥]Present address: Department of Obstetrics and Gynecology, Northwestern University Feinberg School of Medicine, Chicago, IL 60611

***Correspondence:** Dr. Kaela M. Varberg or Michael J. Soares, The University of Kansas Medical Center, Department of Pathology and Laboratory Medicine, 3901 Rainbow Blvd, MS3050, Kansas City, KS 66160, Phone: 913-588-5774, Fax: 913-588-7180, Email: kvarberg@kumc.edu or msoares@kumc.edu; or Dr. Elin Grundberg, Genomic Medicine Center, Children's Mercy Research Institute, Children's Mercy, Kansas City, MO 64108, Phone: 816-731-7084, Email: egrundberg@cmh.edu

Keywords: trophoblast stem cells, extravillous trophoblast cells, differentiation, chromatin landscape, trophoblast gene regulatory networks

ABSTRACT

The extravillous trophoblast (**EVT**) cell lineage is a key feature of placentation and critical for spiral artery remodeling and successful pregnancy. Our knowledge of transcriptional regulation driving EVT cell development is limited. Here, we mapped the transcriptome and epigenome landscape as well as chromatin interactions of human trophoblast stem (**TS**) cells and their transition into the differentiated EVT cell lineage. Chromatin accessibility in intergenic regions was more extensive in EVT cells than in TS cells in the stem state, which is consistent with increased enhancer-driven gene regulation. Using reference epigenome annotation, we noted that 18% of the chromatin landscape in EVT cells was uncharted. We linked regulatory regions to their cognate target genes and characterized the three-dimensional organization of the TS cell functional genome by high throughput chromosome conformation capture. Integration of chromatin accessibility, long-range chromatin interactions, transcriptomic, and transcription factor binding motif enrichment enabled identification of transcription factors and regulatory mechanisms associated with EVT cell lineage development. Subsequent analyses elucidated functional roles for *TFAP2C*, *SNAI1,* and *EPAS1* in the regulation of EVT cell lineage development. *EPAS1* was identified as an upstream regulator of key EVT cell transcription factors, including *ASCL2* and *SNAI1* and together with its target genes, was linked to diseases of pregnancy, including idiopathic recurrent pregnancy loss and low birth weight. Collectively, we have revealed activation of a dynamic regulatory network that provides a framework for understanding EVT cell specification in trophoblast cell lineage development and human placentation.

INTRODUCTION

Pregnancy disorders such as recurrent pregnancy loss (**RPL**), preeclampsia, intrauterine growth restriction, and preterm birth can be the result of inadequate delivery of maternal blood to the growing fetus, a consequence of insufficient spiral artery remodeling[1,2]. Spiral artery remodeling is largely dependent on the function of a specialized population of trophoblast cells termed extravillous trophoblast (**EVT**) cells[3,4]. EVT cells are terminally differentiated trophoblast cells that invade into and restructure the uterine compartment, including transforming tightly coiled spiral arterioles into large, distended vessels capable of increasing blood delivery to the developing fetus[5,6]. EVT cells take two routes of invasion into the uterine parenchyma: interstitial and endovascular. Interstitial EVT cells migrate within the stromal compartment between the vasculature, whereas endovascular EVT cells invade into vessels where they replace endothelial cells and adopt a pseudo-endothelial phenotype[4,7]. In addition to increasing blood volume, EVT cell-remodeled spiral arteries decrease blood flow velocity, and consequently shear stress damage to uterine vessels[8,9]. The contributions of EVT cells to uterine transformation and successful pregnancy imparts importance to understanding the acquisition of an EVT cell fate. To date there is a limited knowledge of regulatory networks controlling EVT cell differentiation.

Comparisons of rodent embryonic stem (**ES**) and trophoblast stem (**TS**) cells using an assortment of approaches to characterize the transcriptome and epigenome have helped define the regulatory landscape critical for establishment of the trophoblast cell lineage[10–23]. These insights helped identify core transcription factors (**TFs**) essential for reprogramming mouse fibroblasts into TS cells[24,25]. Advancements have also been achieved in understanding higher-order regulatory networks controlling aspects of rodent trophoblast cell differentiation, including the identification of putative TF-gene regulated networks[14,19,26–32]. The execution of similar experimentation regarding human trophoblast cell development would be informative.

A range of *in vitro* model systems for investigating human trophoblast cell lineage development have been proffered but all have fallen short in adequately replicating establishment of the EVT cell lineage[33]. A fundamental problem has been in distinguishing features intrinsic to EVT cells from those associated with the immortalization or transformation processes used to establish the *in vitro* models. In 2018, culture conditions needed to maintain human TS cell renewal and to differentiate TS cells into EVT cells were established[34]. The human TS cell model is robust and has been utilized to provide new insights into the development of the human trophoblast cell lineage[32,35–42]. However, systematic profiling of open chromatin, chromatin contacts and associated transcriptional signatures in human TS cells and differentiated EVT cells have not been performed to date.

To this end, we applied multiple functional genomics approaches using next-generation sequencing (**NGS**) and contrasted the regulatory landscape of trophoblast cells in the stem state and following differentiation into EVT cells using human TS cells as a model[34]. The experimentation included systematic assessments of the transcriptome by RNA-sequencing (**RNA-Seq**), chromatin accessibility using Assay for Transposase-Accessible Chromatin-sequencing (**ATAC-Seq**), and high throughput chromosome conformation capture (**Hi-C)**, respectively. We integrated these analyses to identify higher-order regulation of human EVT cell differentiation, which was further validated by single-cell RNA sequencing (**scRNA-Seq**) of first trimester human placenta samples, to provide a rich resource for dissecting functional genomics of placental-related diseases of pregnancy. Using this resource, we identified candidate TFs and demonstrated their critical role in EVT cell lineage development with links to placental disorders.

RESULTS

An atlas of regulatory DNA and transcriptional profiles of human TS and EVT cells

We developed a comprehensive functional genomics resource of trophoblast cell development by utilizing human TS cells maintained in the stem state and following eight days of differentiation to EVT cells[34] (**Figure 1A**). We verified morphological changes as TS cells transition from discrete colonies into individual, elongated EVT cells (**Figure 1B**). We generated RNA-Seq and ATAC-Seq profiles for multiple replicates and donor TS cell lines across stem and EVT cell developmental states (**Supplementary Table 1**). Global gene regulatory patterns coincided with known invasive/EVT cell identifiers such as major histocompatibility complex, class I, G (***HLA-G***), matrix metallopeptidase 2 (***MMP2***), achaete-scute homolog 2 (***ASCL2***), and integrin subunit alpha 1 (***ITGA1***)[32] (**Figure 1C**). Overall, we identified 2267 transcripts that were significantly upregulated (log2 fold change >1, adjusted $p<0.05$) in EVT cells compared to trophoblast cells in the stem state and 2284 transcripts that were significantly downregulated (**Supplementary Table 2, Figure 1D**). The robustness and consistency of the TS cell model was confirmed by two methods. First, TS cell transcriptomes obtained from independent experiments were contrasted to previously reported transcriptomes for the same donor line[34] (Pearson correlation 0.64) (**Supplementary Figure 1A-B**). Second, transcriptomes from multiple donor lines, including the CT27 (46, XX) and CT29 (46, XY) TS cell lines, were compared across cell states (Pearson correlation 0.89) (**Supplementary Figure 1C**). Thus, the *in vitro* human TS cell culture model can be successfully leveraged to gain insight into EVT cell lineage development.

EVT differentiation drives global changes in chromatin accessibility

Although reference maps of regulatory DNA have been generated across hundreds of human cell and tissue types and states, trophoblast cell lineage representation within these reference maps is sparse and is restricted to either bulk placental samples or embryonic stem

cell-derived trophoblast cells without distinction of specific cell lineages or developmental states[43,44]. Using ATAC-Seq at high sequencing depth across multiple replicates, we have expanded these resources and provided maps of regulatory regions in the trophoblast cell stem state and following differentiation to EVT cells. We analyzed these regulatory regions in multiple ways. First, we called ATAC-Seq peaks, which are markers of open chromatin, jointly across replicates and identified, at 1% false discovery rate (**FDR**), 142,520 and 165,094 open chromatin regions in stem and EVT cell developmental states, respectively (**Supplementary Table 3**). We then conservatively assessed shared accessible regions by overlapping specific coordinates and included a minimum one base pair (**bp**) overlap. This conservative assessment identified 73% of the accessible regions as shared across cell states and 27% of the regions as unique to the EVT cell state. The accessible regions that were shared between stem and EVT cells were distributed similarly across the genome compared to other reported human tissues and cells, with the majority mapping to intronic (44%) or intergenic (32%) regions and a smaller subset mapping to promoter regions (14%) (**Figure 2A**). However, when contrasting stem versus EVT cell-specific regions, we noted that the genomic distribution of EVT cell-specific regions was shifted, with a larger proportion mapping to intergenic regions (40% versus 34%) and a smaller proportion mapping to promoter regions (4% versus 8%). This shift points toward increased enhancer-driven gene regulation in EVT versus stem state cells (**Figure 2A**). These observations are reinforced by contrasting sequence coverage in ATAC-Seq peaks, where the coverage is enriched near transcription start sites (**TSS**) for shared accessible regions, but not for EVT cell-specific regions (**Figure 2B**).

We expanded our annotation of stem and EVT cell open chromatin regions to include the atlas of regulatory DNA that was created from 16 different human tissue types or states[43]. This extensive atlas was used as a comparative reference standard. We found that as many as 97% of shared stem and EVT cell regions were previously annotated to regulatory DNA, including 11% that were specific to placental or embryonic/primitive origin. Interestingly, within

accessible chromatin regions that were unique to EVT cells, 17% were placental-or embryonic/primitive specific, and a similar proportion were unannotated. Specifically, 18% of the EVT cell-specific chromatin accessible regions identified have not been captured in stem state TS cells or in previous bulk placenta tissue genome-wide mapping (**Figure 2C**).

To identify consensus regulatory regions that display different transcription factor (**TF**) binding patterns in stem versus EVT cells, we next applied an alternative approach that utilized individual ATAC-Seq replicates in stem state cells (n=7) and EVT cells (n=7). The consensus regulatory regions we considered in this analysis were 1) covered by at least 10 counts per million reads, 2) at least 400 bp in size, 3) mapped to an ATAC-Seq peak, and 4) appeared in more than two of the samples. For the identified consensus regions (n=38,785), the individual replicates clustered together with the largest amount of variability attributed to differences between the two cell groups (**Supplementary Figure 2A-B**). Based on estimated binding affinity measured by differences in ATAC-Seq read densities, we identified a total of 9,334 regions that were differentially enriched (5% FDR) and considered more accessible in EVT cells (**Supplementary Table 4**). We then integrated gene expression data for regions in proximity (within 10 kb) of a gene measured by the RNA-Seq (n=1,576 genes) and found a striking enrichment of EVT cell-upregulated genes mapping near the differentially accessed regulatory regions (log2fold >1; adjusted $p<0.05$), which corresponded to 2.6-fold enrichment ($Chi^2=615$; Fisher $p<0.0001$) (**Supplementary Table 5**). We noted that many of the differentially bound ATAC-Seq regions mapped near genes implicated in the regulation of intrauterine trophoblast cell invasion (e.g., *MMP2*, *ITGA1*, *HLA-G*, and *ASCL2*).

Candidate transcriptional regulators of EVT cell differentiation identified in super-enhancers'

Regulatory DNA that is specific to EVT cells can provide insight into potential transcriptional regulators of EVT cell differentiation and its associated gene expression profile. Due to the significant enrichment of EVT cell-specific regulatory regions near genes with EVT

cell-specific expression patterns, we hypothesized that a subset of these regions would map to super-enhancers. A super-enhancer is operationally defined as a large cluster of transcriptional enhancers in close proximity that drive expression of a gene(s)[45]. Super-enhancers determine cell identity and can provide insight regarding key TFs underlying the differentiation process[19,46,47]. To test this, we used the ATAC-Seq data to identify super-enhancers based on established algorithms[48]. We identified a total of 1,296 super-enhancers, of which 309 mapped to 611 differentially bound ATAC-Seq peaks (**Supplementary Table 6**).

Four sets of regulatory regions were used to perform motif enrichment of known TF families compared to the genome background. The regulatory regions included: 1) all ATAC-Seq regions identified in EVT cells, 2) the EVT cell-unique ATAC-Seq regions identified as regions not shared with the stem state, 3) the differentially enriched EVT cell ATAC-Seq regions based on estimated TF binding affinity, and 4) regions in (3) that overlap a super-enhancer. In all four analyses (**Supplementary Table 7**, we found TEA Domain Transcription Factor 3 (***TEAD3)*** and TEA Domain Transcription Factor 1 (***TEAD1)*** as the top two ranked TF DNA binding motifs where both genes were also significantly upregulated in EVT cells (*TEAD3*: log2fold change=1.6, adjusted p=2E-10; *TEAD1*: log2 fold change=1.3, adjusted p=1E-3; **Supplementary Table 2**). In addition to *TEAD1* and *TEAD3*, another 20 TF DNA binding motifs were identified among the top 50 across all sets, including Activator Protein 1 (**AP-1**), Activating Protein 2 (**AP-2**), and GATA TF DNA binding motifs (**Supplementary Table 8**). *TEAD3* has been linked to trophoblast cell regulation[44], and a paralog, TEA Domain Transcription Factor 4 (***TEAD4)***, is a known determinant of the stem state[39,49]. *TEAD1*, *TEAD3*, and *TEAD4* each displayed unique regulatory landscapes (**Supplementary Figure 3A-C**). *TEAD1* and *TEAD4* were both proximal to super-enhancers (**Supplementary Figure 3A, Supplementary Figure 3C**). *TEAD1 and TEAD3* are more highly expressed in EVT cells compared to stem state cells, while *TEAD4* is expressed more highly in the stem state (**Supplementary Figure 3D**).

AP-2 and GATA TFs are known regulators of trophoblast cell development[50–53]. Indeed, the most significant motif enrichment among all active chromatin regions in EVT cells and within EVT-specific super-enhancers following those for the TEAD family members was *TFAP2C* (also referred to as AP-2 gamma), which has previously been implicated in the regulation of early decision making in trophoblast cell lineage development[52,53] (**Supplementary Table 7A, Supplementary Table 7D**). However, the potential contributions of *TFAP2C* to TS cell maintenance and terminal trophoblast differentiation remain elusive. Consequently, we investigated the role of *TFAP2C* in the regulation of trophoblast cell development.

Functional assessment of TFAP2C actions in trophoblast cell development

To evaluate the functions of *TFAP2C* in trophoblast cell development, we first localized *TFAP2C* in placental specimens and then performed knockdown experiments in the TS cell model. Using *in situ* hybridization, we found *TFAP2C* to be expressed within the EVT cell column of first trimester human placental specimens **(Figure 3A).** *TFAP2C* co-localized primarily with *CDH1* (cytotrophoblast marker[54]) and less with *PLAC8* (EVT cell marker)[55,56] transcripts in EVT cell columns, consistent with its enrichment in the cytotrophoblast/trophoblast progenitor cell population at the base of the column (**Figure 3A**). To determine expression profiles in the TS cell model, we applied RT-qPCR and western blotting to measure *TFAP2C* transcript and protein levels, respectively. *TFAP2C* transcript is expressed in stem cells and EVT cells at comparable levels (**Figure 3B**)**,** with slightly higher expression of *TFAP2C* at the protein level in the stem state (**Figure 3C**). The regulatory landscape surrounding the *TFAP2C* locus shows evidence of active chromatin near TSS of the gene in both stem state and EVT cells. **(Supplementary Figure 4**). To elucidate the function of *TFAP2C* in TS cell differentiation we used a gene silencing approach. *TFAP2C*-specific short hairpin RNA (**shRNA**) stably delivered in the stem state interfered with *TFAP2C* transcript (**Figure 3D**) and protein expression (**Figure 3E**). *TFAP2C* disruption inhibited TS cell proliferation, altered TS cell

morphology (**Figure 3F**), and adversely affected cell survival in culture conditions promoting EVT cell differentiation. Disruption of *TFAP2C* also resulted in broad transcriptomic changes, as identified by RNA-Seq. Consistent with impaired cell proliferation, several known genes contributing to cell cycle progression were downregulated, including *AURKA*, *AURKB*, *PLK1*, *MYC*, *EGR1*, *CDK1*, *CCNA2*, *CCNB1*, and *MKI67* (**Figure 3G**). Transcript signatures indicative of the stem state (e.g., *TEAD4*, *PEG3*, and *PEG10*) and EVT cell state (e.g., *IGF2*, *ISM2*, *CDKN1C*, *FOXO4*, *GCM1*, *ASCL2*, *SNAI1*, *FSTL3*, *PLAC8*, and *NOTUM*) were also inhibited **(Figure 3G).** Notable upregulated genes following *TFAP2C* disruption included *CNN1* and *LAMB3*, which are associated with the stem state, as well as *MMP2*, *MCAM*, and *SPARC*, which are linked to the EVT cell state. Ingenuity pathway analysis of the top 500 differentially regulated genes following *TFAP2C* disruption highlighted kinetochore metaphase signaling, cell cycle control of chromosomal replication, as well as cyclins and cell cycle regulation as the top functional pathways (**Supplementary Figure 5**). Overall, we conclude that *TFAP2C* contributes to cell cycle regulation and is critical for maintenance of TS cells in a stem state and early EVT cell lineage development.

Long-range chromatin interactions define unique gene regulation machinery in EVT cells

To identify long-range chromatin interactions involving EVT cell super-enhancers and regulatory regions identified by ATAC-Seq we performed Hi-C chromatin capture. We obtained long-range intra-chromosomal contacts after aligning and filtering Hi-C read pairs across all Hi-C samples and called chromatin loops at 5 kb resolution. Using this threshold, we identified 11,999 stem state and 14,516 EVT cell chromatin loops with a minimum and median distance between anchors corresponding to 30 kb and 260-275 kb, respectively (**Supplementary Table 9**). These observations are consistent with efforts using other cell models[57,58]. We found that CCCTC-binding factor (**CTCF**) binding motifs were the most enriched TF binding motifs within the loop anchors (covering 60% of all EVT loop anchors). This observation confirms previous

efforts that indicate the important role of CTCF in shaping the three-dimensional structure of the genome[59,60], in addition to the function it serves as a transcriptional repressor and insulator[57].

To identify long-range chromatin interactions, or Hi-C loops, that are specific to either stem state or EVT cells (i.e., differential loops), Hi-C loops from the stem state were integrated using the same parameters described for EVT cells above. Using previously established algorithms[61], we identified 99 and 349 loops to be stem- and EVT-cell specific, respectively (**Supplementary Table 9**). These results point to two major properties of cell-specific, Hi-C loops: 1) they are rare due to strongly conserved chromatin conformation across cell states as shown previously[61] and 2) they are more common in EVT cells, which could be indicative of increased enhancer-driven gene regulation (**Figure 2A**).

To examine the enrichment of chromatin accessibility near the Hi-C loop anchors, we integrated reads from the ATAC-Seq peaks from the respective cell types. Among all chromatin loops, we found a clear enrichment of accessible chromatin near the center of the anchors (**Supplementary Figure 6A-B**). EVT cell-specific loops displayed a similar pattern; however, there was no enrichment of accessible chromatin near the center of the anchors for stem state-specific loops likely due to the low number of loops that were identified (**Supplementary Figure 6C-D**).

Using the intersected dataset of chromatin loop anchors that also show chromatin accessible (ATAC-Seq) regions, we focused on the 9,334 regulatory regions identified by ATAC-Seq that were differentially enriched and more accessible in EVT cells (**Supplementary Table 4**). From this integrated analysis we found that 1,151 EVT cell-specific regions defined by ATAC-Seq overlapped at least one Hi-C loop anchor across a total of 1,518 EVT cell Hi-C loops. Interestingly, of the 1,518 EVT cell chromatin loops identified, 119 contained differentially bound regulatory regions at both anchors and 1,094 had at least one anchor mapped within 10 kb of a known gene (**Supplementary Table 10**). Among the genes identified with long-range chromatin interactions of regulatory DNA, several have been implicated in EVT cell

differentiation and function including *ASCL2*, *MMP2*, *HLA-G*, and *ITGA1*[32,56,62–64] (**Figure 4**, **Supplementary Figure 7**). Specifically, we identified a super-enhancer 50 kb upstream of *ASCL2* interacting with its promoter region (**Figure 4**).

For orthogonal validation of the EVT 'regulome', the top 100 EVT cell-specific chromatin loop-gene associations were validated in an independent dataset. Specifically, we used publicly available single-cell gene expression data from EVT cells (n=693 cells) isolated from normal first trimester placental samples[65,66] and contrasted with syncytiotrophoblast (SCT) and villous cytotrophoblast (VCT) expression profiles (n= 7,834 cells) (**Supplementary Figure 8**). Approximately 70% of these genes that were found to possess EVT-cell specific gene regulation and significantly upregulated in TS cell-derived EVT showed increased expression in the primary EVT cells from the single-cell dataset in comparison to SCT and VCT cells (**Supplementary Table 11**). We classified these EVT cell-specific genes based on TF status and revealed five TF genes (*ASCL2*, *DLX6*, *SNAI1*, *MYCN*, and *EPAS1*). The regions surrounding each of the five TFs show evidence of EVT cell-specific regulation (**Figure 5, Supplementary Figure 9**). While *ASCL2* contributes to EVT cell differentiation[32], the contributions of the other identified TFs in trophoblast cell differentiation are less understood.

Functional investigation of transcriptional regulators of EVT cell differentiation

To evaluate the functions and potential contributions of *SNAI1* and *EPAS1* to EVT cell development, we localized each target in placental specimens by *in situ* hybridization and performed loss-of-function experiments in the TS cell model (**Figures 6-8**). First, we examined the spatial distribution of each TF in first trimester human placental tissue and showed that both TFs were more abundant in distal regions of the EVT column and co-localized with *PLAC8*, a marker of differentiated EVT cells[55,56] (**Figure 6A**). *SNAI1* and *EPAS1* (also called HIF2A) were significantly upregulated in EVT cells compared to stem state cells at the transcript (**Figure 6B**)

and protein levels (**Figure 6C**). Taken together, both TFs are expressed in EVT cells *in situ* and following their *in vitro* cell differentiation from TS cells.

To assess impacts on EVT cell development, we stably expressed shRNAs specific to each TF in the stem state and then attempted to differentiate the cells into EVT cells (**Figures 7-8**). shRNAs were identified that significantly inhibited TF expression at transcript (**Figure 7A, Figure 8A**) and protein levels **(Figure 7B, Figure 8B).** Knockdown of either *SNAI1*, or *EPAS1* were each compatible with maintenance of the TS cell stem state. However, deficits in each TF negatively affected aspects of EVT cell differentiation, with *EPAS1* disruption showing the most extensive morphologic impact (**Figure 7C, Figure 8C**). *EPAS1* knockdown cells consistently failed to exhibit the characteristic elongation normally associated with EVT cell differentiation.

Parameters of EVT cell differentiation in control and TF knockdown cells were further investigated using RNA-Seq which revealed that both TFs contributed to the regulation of several genes associated with EVT cell lineage development (**Figures 7-8**). *SNAI1* disruption led to downregulation of several genes associated with the EVT cell state (e.g., *FLT4, LAMA4, IGF2*, *ISM2*, *ASCL2*, *FSTL3*, *DLX6*, and *NOTUM*) and upregulation of EVT cell interferon-responsive transcripts (e.g., *IFIT1, IFIT2, IFIT3, IFITM1, IFI27, IFI35, IFIH1, OAS2, OAS3,* and *MX1)* as well as several stem state-associated transcripts (e.g., *ITGB6*, *GJA1*, *LAMB3*, *F3*, and *TAGLN*) (**Figure 7D**). IFN-responsive transcript expression appears to be a feature of human trophoblast cell differentiation[67] and may be associated with repetitive element expression and double stranded RNA stress[68]. Similar to morphological changes, disrupting *EPAS1* had the most extensive impact on the EVT cell transcriptome. Signature EVT cell-specific transcripts were prominently downregulated (e.g., *NOTUM*, *DIO2*, *HTRA1*, *MMP2*, *PDE6H*, *ADAM12*, *MCAM*, *HLA-G*, *ASCL2*, *DLX6*, *DLX5*, *ISM2*, *CCR1*, *SNAI1*, *ITGA1*, and *GCM1*), whereas several stem state markers (*ITGB6*, *LAMB*, *F3*, *JUN*, *EGLN3*, *CYR61*, and *NPPB*) were upregulated in *EPAS1* disrupted cells (**Figure 8D)**. In summary, disruption of *SNAI1* or *EPAS1* each resulted in downregulation of EVT cell-specific transcripts.

Differences among the activities of each TF as a transcriptional repressor versus activator (calculated based on Log2FC>3 or <-3) were evident. *SNAI1* and *EPAS1* had repression to activation ratios of 70:30 and 60:40, respectively. The known repressor actions of *SNAI1*[69] are consistent with the profile.

In addition to the common feature of the TFs impacting known EVT-cell markers, each TF also contributed to the regulation of unique transcript signatures. In fact, pathway analysis performed on the top 500 significantly altered (upregulated or downregulated) genes upon silencing of each TF revealed that the most significant biological function associated with the differentially regulated genes for both TFs was "Migration of Cells" (*SNAI1*: p=1.5E-39; *EPAS1*: p=8.3E-49; **Supplementary Table 12, Supplementary Table 13**). Additional pathway analysis performed on the subset of *EPAS1* gene targets mapping to the "Migration of Cell" function (n=209 genes) identified "Invasion of Cells" as another top function (p=5.3E-70). Many genes mapping to this "Invasion of Cells" cluster are known contributors to the regulation of EVT cell invasion into the uterus (**Figure 8E**).

EPAS1 is an upstream TF regulator and dysregulated in placenta failure

As noted in the functional assays, *EPAS1* disruption had the greatest impact on EVT cell morphology and resulted in the downregulation of multiple key genes involved in EVT cell differentiation and invasion. To test the potential of *EPAS1* to act as a direct transcriptional regulator, ATAC-Seq peaks containing *EPAS1* (HIF2A) motifs in EVT cells were intersected with RNA-Seq data from control and *EPAS1* knockdown cells (**Supplementary Table 14**). This analysis identified *EPAS1*-regulated genes, as detected by RNA-Seq analysis (log2 foldchange <-1; adjusted p <0.05), being significantly correlated with chromatin accessible regions containing *EPAS1* (HIF2A) binding motifs. This corresponded to a 1.5-fold enrichment (Chi^2=62; Fisher p<0.0001) when considering all EVT cell peaks and a 2.5-fold enrichment when restricting to those EVT cell peaks with differential binding affinity (Chi^2=71; Fisher p<0.0001).

Among these putative direct targets of *EPAS1* action were *PLAC8*, *DIO2*, *MMP2*, *MYCN*, *ADAM12*, *HLA-G*, and *DLX6.*

As failure of proper EVT cell invasion into the uterine compartment at the maternal fetal interface may lead to pregnancy complications, we next assessed whether there is evidence linking *EPAS1* to adverse pregnancy outcomes. To this end, single-cell gene expression data was generated from first trimester placental samples (4,787 cells) derived from individuals suffering from idiopathic RPL. Single EVT cells (n=194) were profiled, and expression profiles were compared to compatible single-cell expression profiles of EVT cells (n=693) from gestational age-matched normal control placental samples (**Supplementary Figure 10**). Of the 2,967 differentially expressed genes (absolute log2fold change >1, adjusted p<0.05) identified in EVT cells isolated from RPL specimens (**Supplementary Table 15**), *EPAS1* was ranked among the top 200. Further, *EPAS1* expression was significantly higher in placentas from RPL patients (log2fold change=1.6, adjusted p=7.2E-9, **Figure 8F**).

To assess the potential role of *EPAS1* in gene expression dysregulation in RPL, a comparison was performed between upregulated genes identified in EVT cells derived from RPL placentas (log2 fold change >1; adjusted p<0.05) and genes downregulated by *EPAS1* disruption in TS cell-derived EVT cells (log2 fold change <-1; adjusted p<0.05). This comparison identified a 1.7-fold enrichment (Chi^2=53; Fisher p<0.0001) indicative of common gene targets. Among these genes (i.e., upregulated in RPL and downregulated upon *EPAS1* disruption) were *EPAS1*, *NOTUM*, *DIO2*, *ADAM12*, *ITGA1*, *CSH1*, *PAPPA*, *PAPPA2*, *CDKN1C*, *LIFR*, *FN1,* and *FLT1*. Upregulation of *FLT1* has been shown to be linked to *EPAS1* upregulation during hypoxia-induced placental failure[70]. Failure of the placenta can also result in fetal growth restriction and low birth weight. Thus, *EPAS1* and its targets may also contribute to the development of these conditions. To this end, we accessed summary statistics from the largest genome-wide association study to date that links fetal genetic variation with birth weight[71] and identified loci linked to 273 genes (p<0.001)**.** Interestingly, a significant link (1.9-fold change,

Chi^2=32.7; Fisher p<0.0001) was identified between fetal genetic loci for birth weight and genes dysregulated by *EPAS1* disruption (absolute log2fold change >1, adjusted p<0.05). Genes included not only *EPAS1* itself (**Supplementary Figure 9C**), but also *FLT1*, PAPPA, PAPPA2, and *CDKN1C* (**Supplementary Table 16)**.

DISCUSSION

EVT cell invasion into the uterine compartment is a critical component of spiral artery remodeling and successful pregnancy[3,4]. Despite our understanding of the importance of EVT cell function, our knowledge of the underlying molecular mechanisms driving EVT cell development is limited. Here, we utilized the human TS model and generated an NGS-based functional genomics resource comprising atlases for transcriptomes by RNA-Seq, open (active) chromatin by ATAC-Seq, and chromatin interactions by Hi-C for TS cells cultured in the stem state and following differentiation to EVT cells. These datasets were used to identify higher order regulatory events driving EVT cell lineage specification including characterization of signature epigenomic landmarks characteristic of EVT cell lineage development. Global epigenomic events were reinforced by inspection at the level of individual genes known to drive EVT cell differentiation (e.g., *ASCL2*[32]) and downstream genes responsible for the phenotypic features of an EVT cell (e.g., *MMP2, ITGA1,* and *HLA-G*).

The architecture of the epigenome undergoes profound changes as stem cells differentiate into specialized cells[72,73]. The transformation is apparent at multiple levels. Some chromatin regions become more accessible and other regions are closed to the requisite machinery controlling gene transcription. This regulatory process is facilitated by formation of chromatin loops allowing for long-range interactions of regulatory elements including enhancers and promoters. The net effect is distinct transcript profiles associated with stem versus differentiated cells, resulting in cells possessing different morphologies and functional attributes.

These architectural and functional changes are exemplified in human TS cells differentiating into EVT cells. TS cells cultured in the stem state are tightly associated, and cellular efforts are directed to proliferation, whereas differentiation into EVT cells is characterized by cell colony dispersal, the manifestation of prominent cellular extensions, and cell migration. Trophoblast cell differentiation status is further defined by cell-specific patterns of gene expression, chromatin accessibility and long-range chromatin interactions – all occurring in a coordinated manner. This new map of trophoblast-lineage specific regulatory DNA generated here consists not only of distinct active chromatin regions specific to EVT cells but also extends current maps of open chromatin for human cell types and tissues[43]. Increased chromatin accessibility in EVT cells was enriched in intronic and intergenic regions and linked to increased RNA expression. These observations are in line with reference epigenome mapping efforts[43] and point towards pronounced activity of distal regulatory regions controlling cell lineage development through involvement of master TFs regulating gene expression in EVT cells.

Motif enrichment within chromatin accessible regions revealed potential 'core' TF-DNA interactions that define the EVT developmental state including *TEAD3* and *TFAP2C*. *TEAD3* was identified in a large integrative analysis of over 100 cell types to bind to enhancers that were specific to embryonic stem cell differentiated into trophoblast by treatment with bone morphogenetic protein 4[43]. *TFAP2C* is a known regulator of the mouse TS cell stem state[10,11,15,74]. Consistent with these observations we showed that *TFAP2C* is also critical for human TS cell expansion and the onset of EVT cell differentiation. In addition, *TFAP2C* represents a key factor in the derivation of TS cells from fibroblasts and pluripotent stem cells[24,25,52,53]. A recent CRISPR-Cas9 screen has identified other transcriptional regulators critical to maintaining the TS cell stem state (*ARID3A*, *GATA2*, *TEAD1*, and *GCM1*)[75].

Our genome-wide integrative analyses with multiple layers of epigenome information including active regulatory regions and their interactions led to the identification of an established TF regulator of EVT cell development, *ASCL2*[32]. The epigenomic landscape of *ASCL2* changed as trophoblast cells differentiated from the stem state to EVT cells. This transition was characterized by the appearance of a long-range interaction of a super-enhancer with the *ASCL2* promoter. We also identified compelling new TF candidates controlling EVT cell lineage development including *SNAI1, DLX6*, *MYCN*, and *EPAS1* – each possessing epigenomic signatures indicative of gene activation in both cultured EVT cells and primary EVT cells. Silencing of *SNAI1* resulted in disruptions in EVT cell development and affected the expression of a subset of transcripts characteristic of EVT cells. *SNAI1* is an established transcriptional repressor with known roles in promoting epithelial to mesenchymal transition[69], including a potential involvement in EVT cell development[76]. *EPAS1* encodes hypoxia inducible factor 2 alpha (**HIF2A**), a basic helix-loop-helix domain TF, coordinating cellular responses to low oxygen and developmental processes with a connection to trophoblast cell biology[77,78]. The actions of *EPAS1* on EVT cell development were much more extensive than observed for *SNAI1* and included transcriptional regulation of virtually all transcripts defining the EVT cell phenotype. Interestingly, these transcriptional *EPAS1* targets also included TFs contributing to the regulation of EVT cell differentiation (e.g., *ASCL2* and *SNAI1*). Thus, *EPAS1* can be placed upstream relative to other transcriptional regulators directing EVT cell differentiation (**Figure 8G**). *EPAS1* mRNA and protein accumulate as EVT cells differentiate and do so independent of low oxygen tensions. Mechanisms controlling EVT cell dependent *EPAS1* transcriptional activation and *EPAS1* (HIF2A) protein stabilization are unknown. Elucidation of factors controlling EVT cell *EPAS1* will add to understanding the regulation of EVT cell differentiation.

The biology of *EPAS1* in placentation is complicated. *EPAS1* expression is not restricted to EVT cells. Syncytiotrophoblast also express *EPAS1*[79,80]. Thus, additional layers of regulation

impart specificity for *EPAS1* actions in EVT cells versus syncytiotrophoblast. These disparate actions may be linked to contributions of cell specific co-regulators[81,82], canonical versus non-canonical *EPAS1* signaling[78,83], or the involvement of other unappreciated modulator(s) of *EPAS1* action. The relevance of *EPAS1* to placentation is evident in its association with placental disease. RPL is characterized by elevated EVT cell *EPAS1* expression. *EPAS1* is also dysregulated in preeclampsia[84]. Under such circumstances, *EPAS1* may be responding to a failed placenta and the ensuing hypoxia. Context is important, and inappropriate *EPAS1* may hinder rather than support placentation. Compromised placental function negatively impacts fetal growth, which is also connected to misexpression of *EPAS1* gene targets and is demonstrated by genetic linkage to low birth weight[71].

In summary, we have provided insights into a trophoblast regulatory network controlling EVT cell lineage development. Datasets generated from this research represent a resource for developing new hypotheses that can be tested *in vitro* with human TS cells and *in vivo* using appropriate animal models. They will also provide a valuable reference dataset for newly established TS cell lines derived from pluripotent stem cells[52,52,85–91]. *TFAP2C*, *SNAI1*, and *EPAS1* provide entry points and a framework for understanding EVT cell differentiation and human placentation. The challenge will be to harness this new knowledge to develop effective strategies for diagnosing and treating diseases of placentation.

MATERIALS AND METHODS

Human TS cell culture

Human TS cells[34] were cultured in 100 mm tissue culture dishes coated with 5 µg/mL collagen IV (CB40233, Thermo-Fisher, Waltham, MA). Human TS cells were maintained in Complete human TS cell medium [DMEM/F12 (11320033, Thermo-Fisher), 100 µm 2-mercaptoethanol, 0.2% (vol/vol) fetal bovine serum (**FBS**), 50 µM penicillin, 50 U/mL streptomycin, 0.3% bovine serum albumin (BP9704100, Thermo-Fisher), 1% Insulin-Transferrin-Selenium-Ethanolamine solution (vol/vol, Thermo-Fisher)], 1.5 µg/mL L-ascorbic acid (A8960, Sigma-Aldrich, St. Louis, MO), 50 ng/mL epidermal growth factor (**EGF**, E9644, Sigma-Aldrich), 2 µM CHIR99021 (04-0004, Reprocell, Beltsville, MD), 0.5 µM A83-01 (04-0014, Reprocell), 1 µM SB431542 (04-0010, Reprocell), 0.8 mM valproic acid (P4543, Sigma-Aldrich), and 5 µM Y27632 (04-0012-02, Reprocell). EVT cell differentiation was induced by plating human TS cells onto 6-well plates pre-coated with 1 µg/mL collagen IV at a density of 80,000 cells per well. Cells were cultured in EVT Differentiation Medium [DMEM/F12 (11320033, Thermo-Fisher), 100 µm 2-mercaptoethanol, 0.2% (vol/vol) FBS, 50 µM penicillin, 50 U/mL streptomycin, 0.3% bovine serum albumin (BP9704100, Thermo-Fisher), 1% Insulin-Transferrin-Selenium-Ethanolamine solution (vol/vol, Thermo-Fisher)], 100 ng/mL of neuregulin 1 (**NRG1**, 5218SC, Cell Signaling, Danvers, MA), 7.5 µM A83-01 (04-0014, Reprocell, Beltsville, MD), 2.5 µM Y27632 (04-0012, Reprocell), 4% KnockOut Serum Replacement (**KSR**, 10828028, Thermo-Fisher), and 2% Matrigel® (CB-40234, Thermo-Fisher). On day 3 of EVT cell differentiation, the medium was replaced with EVT Differentiation Medium excluding NRG1 and with a reduced Matrigel® concentration of 0.5%. On culture day 6 of EVT cell differentiation, the medium was replaced with EVT Differentiation Medium excluding NRG1 and KSR, and with a Matrigel concentration of 0.5%. Cells were analyzed on day 8 of EVT cell differentiation.

Cells were harvested from plates with trypsin, counted, and viability determined with a Countess II automated cell counter (Thermo-Fisher). Sample replicates with the highest viability

and an adequate cell count were selected to be used for ATAC-Seq. Cells used for RNA isolation were centrifuged at 300 g for 8 min, supernatants were removed, and cell pellets were resuspended in 1 mL of TRIzol (15596026, Thermo-Fisher).

Omni-ATAC library preparation and sequencing

The Omni-ATAC library preparation was performed on freshly isolated cells according to the established protocol[92] until the post-transposition cleanup step. After cleanup of the transposed DNA was complete, samples were stored at -20°C until library amplification. Samples were subsequently thawed at room temperature and library construction completed. The samples underwent 5 cycles of PCR before an aliquot was used in a quantitative PCR (**qPCR**) reaction to determine whether additional amplification was required. An Applied Biosciences Viia7 Real-Time PCR System (Applied Biosciences, Beverly Hills, CA) was used for qPCR and the number of additional PCR cycles needed was determined from the multicomponent plot, which plots fluorescence versus the number of PCR cycles. The maximum fluorescence for each sample was found from the plot, ¼ of the maximum fluorescence value was calculated, and the number of additional PCR cycles needed was calculated by determining the cycle number that reached ¼ maximum fluorescence on the graph. In cases where the ¼ maximum fluorescence value was halfway between two cycle numbers, the lower number of cycles was selected since slight under-amplification was more favorable than overamplification. The original samples were then PCR amplified with their determined number of additional cycles. Cleanup of the samples was performed after PCR amplification according to the published protocol[92]. Libraries were quantified using a Qubit dsDNA BR Assay Kit (Q32853, Thermo-Fisher) and the size was determined with a High Sensitivity DNA Bioanalyzer Kit (5067-4626, Agilent, Santa Clara, CA) and sequenced on an Illumina Nova6000Seq paired-end using Nextera Sequencing primers.

Basic quality control of raw FASTQ files was performed using HTStream (Version 1.3.2; https://github.com/s4hts/HTStream). PhiX reads were removed using hts_SeqScreener with default parameters. Duplicates were removed using hts_SuperDeduper (-e 250000). Adapter sequences were trimmed with hts_AdapterTrimmer (-p 4). Unknown nucleotides (N) were removed using hts_Ntrimmer. Base quality trimming was performed using a minimum average quality score of 20 (-q 20) in a 10-bp sliding window (-w10) with hts_QWindowTrim. Trimmed reads shorter than 50 bp as well as orphaned reads from a pair were removed using hts_LengthFilter (-n -m 50). Reads were aligned to the human genome (GRCh38.86) using BWA mem (Version 0.7.17-r1188). The alignments were shifted using alignmentSieve (deepTools; Version 3.5.1) using the --ATACshift option. Alignments in the ENCODE blacklisted regions (https://doi.org/10.5281/zenodo.1491733) were removed using Bedtools intersect (with option `-v`; Version 2.30.0). Peaks for each sample were called independently using the callpeak function in MACS3 (Version 3.0.0a6) with a minimum FDR of 0.01 (-q 0.01). Peak regions were annotated in regard to the genomic feature of their location using HOMER's annotatePeaks.pl script as well as overlap with DNAse I hypersensitive sites (min overlap of 1bp)[43]. Unique peaks (e.g., EVT cell) were determined based on complete lack of overlap with a comparing set of peak regions (e.g., stem state).

The webtool GREAT was used to retrieve genes with an overlapping regulatory domain, where the regulatory domain of a gene is the region 5 kb upstream and 1 kb downstream of the transcription start site (**TSS**). Additionally, distal gene elements were retrieved where the regulatory domain of the genes extends up to 1 Mb. Finally, each peak was annotated by mapping it to the most significant differentially expressed gene (EVT vs stem) with a TSS within 10 kb of the peak region.

ATAC-seq read counts per peak were generated using the Bioconductor package DiffBind (Version 3.2.7). Peaks with less than 10 counts per million (**CPM**) were excluded. The

differential binding affinity of a region between EVT and stem samples was assessed in the Bioconductor package limma (Version 3.48.3) with an FDR of 0.05.

Super-enhancers were called using HOMER (version 4.7.2) findpeaks (http://homer.ucsd.edu/homer/ngs/peaks.html) using in "-style super" mode.

RNA library preparation and sequencing

RNA was isolated with a TRIzol/chloroform precipitation followed by a cleanup with a RNeasy Mini Kit (74104, Qiagen, Germantown, MD). Frozen cell isolates in TRIzol were thawed on ice prior to starting the isolation. Once thawed, the samples were placed at room temperature for 5 min to promote dissociation of nucleoprotein complexes. Next, 200 µL of chloroform was added to each sample with vigorous shaking for 15 sec. Samples were placed at room temperature for 3 min before centrifugation at 12,000 g for 15 min at 4°C. After centrifugation, the upper, colorless, aqueous phase was removed and transferred to a new 1.5 mL tube and 1.5 volumes of 100% ethanol added. Isolates containing ethanol were mixed thoroughly by pipetting and 650 µl transferred to a RNeasy spin column. The protocol for the RNeasy Mini Kit was then followed to clean the sample post-TRIzol/chloroform extraction. RNA was then quantified using a Qubit RNA BR Assay Kit and RNA integrity assessed using RNA ScreenTape (5067-5577 and 5067-5576, Agilent) on the Agilent TapeStation platform (Agilent) prior to library preparation.

RNA libraries were prepared using a TruSeq Stranded RNA HT Sample Prep Kit (RS-122-2303, Illumina on a Caliper Sciclone G3 platform (PerkinElmer, Waltham, MA). Manual bead cleanup was performed using AMPure XP Beads (A63881, Beckman Coulter, Brea, CA) after library preparation to remove primer-dimers. 1.0x volume of beads was added to the entire library volume. Samples were mixed by pipetting and incubated at room temperature for 5 min before being placed on a magnetic stand until the supernatant cleared. Once clear, the

supernatant was removed and discarded, and the beads were washed twice with 80% ethanol while still on the magnet. Any residual ethanol was removed, and the beads were air-dried on the magnet for 5 min at room temperature. Once dry, 30 µL of Resuspension Buffer was added to each sample, the beads were resuspended by pipetting, and the solution incubated for 2 min at room temperature. Samples were then placed on a magnet until the supernatant was clear. The supernatant was then transferred to a new plate. A Qubit dsDNA BR Assay kit was used to determine the concentration of the library and a Fragment Analyzer Standard Sensitivity NGS Fragment Kit (DNF-473-10000, Agilent) was used to detect the size of the library and to verify removal of excess primer-dimers. Standard Illumina Free-Adapter Blocking was performed on the libraries. Cleaned, adapter-blocked libraries were loaded on a NovaSeq6000 (Illumina, San Diego, CA) with a run configuration of 151x8x8x151 and an average depth of 70 M paired-end reads per library.

Raw FASTQ files were trimmed using default parameters (-r 0.1 -d 0.03) in Skewer (Version 0.2.2) and reads shorter than 18 bp were discarded. Transcripts were quantified using Kallisto (Version 0.46.2). Differentially expressed genes at FDR of 0.05 were discovered using the Bioconductor package DESeq2 in R (Version 1.32.0). CT27 and CT29 RNA-sequencing data were processed individually following standard processing.

Chromatin capture by Hi-C

Cells were rinsed with 5 mL of phosphate buffered saline (**PBS**) and then 5 mL PBS containing 2% formaldehyde were added and incubated for 10 min. Then, 550 µL of Glycine 1.25 M (final conc 0.125M) were added and incubated for 5 min at room temperature before an additional 15 min incubation on ice. Liquid was removed and cells were rinsed 2 times with 5 mL ice-cold PBS. Cells were scraped and transferred into 2 mL low bind tubes to have the equivalent of

1*10^6 cells (1 Petri Dish). Nuclei were pelleted at 2,500 rpm for 10 min at 4˚C (repeat if necessary) and all supernatant was removed and flash frozen on dry ice and stored at -80˚C until ready for use. Crosslinked samples were thawed on ice and prepared for Hi-C using the Arima-HiC Kit (Arima Genomics, San Diego, CA). Cross-linking was performed according to the Arima-HiC protocol with the following modifications. At step 5, the incubation at 37°C was increased to a duration of 60 min and the samples were maintained overnight at 4°C after step 11. Samples were then cleaned with AMPure XP Beads as described in the protocol and all quantification and quality control steps that required the use of a Qubit were performed using a Qubit dsDNA HS Assay Kit. Proximally ligated DNA (1000 ng/sample) was used as an input into the library preparation protocol. Samples were sheared to 400 bp using a Covaris LE220-plus system using the manufacturer's recommended settings. Sizing was confirmed using a High Sensitivity DNA Bioanalyzer Kit. Size-selection as described in the Arima-HiC protocol was followed and the entire size-selected sample was used as input into biotin enrichment and library preparation with a KAPA HyperPrep Kit and Roche SeqCap Adapters (7141530001 or 7141548001, Roche). Instead of performing the Arima-HiC quality control step 2, the number of cycles needed for library amplification was determined as follows: perform 10 PCR cycles for 125-200 ng of input, perform 9 PCR cycles for 200-400 ng of input, perform 8 PCR cycles for 400-600 ng of input, and perform 7 cycles of PCR for inputs greater than 600 ng. The completed libraries were then quantified using a Qubit dsDNA BR Assay Kit and the library size was determined with a High Sensitivity DNA Bioanalyzer Kit. Standard Illumina Free-Adapter Blocking was performed on the libraries. Cleaned, adapter-blocked libraries were loaded on a NovaSeq6000 with a run configuration of 151x8x8x151 and an average depth of 900M PE reads per library.

To discover chromatin loops the standard Juicer pipeline[61] was used where contact maps for each sample were generated using the Arima Genomics fragment map file specific to

GRCh38 and default parameters. After merging contact maps for each cell state, stem and EVT, chromatin loops were discovered with HiCCUPS (Juicer Tools Version 1.22.01;[61]) using a matrix size of 1000 (-m 1000) at 5 Kb, 10 Kb, and 25 Kb resolution (-r 25000,10000,5000). Differential loops between EVT and stem were identified using HiCCUPSDiff (Juicer Tools Version 1.22.01;[61]) with a matrix size of 1000 (-m 1000). The chromatin loops as well as the differential loops were filtered to exclude contacts in the ENCODE blacklisted regions ([93]; https://doi.org/10.5281/zenodo.1491733) using Bedtools intersect (with option `-v`; Version 2.30.0;[94]). Furthermore, genes with a TSS within 10 Kb of a loop region were linked to each loop.

The bedtools toolkit (https://github.com/arq5x/bedtools2, version 2.30.0) was used to compute the depth of coverage of the ATAC read data genome-wide in 100 bp bins, normalizing number of ATAC reads in each bin for every 10 M total reads for a given sample. All bins within 20 kb of either end of each of the EVT and stem loops were then extracted. To generate control sets, 1000 randomly selected intervals (chose from chr1-22, X, Y) that do not intersect any of the loops and do not intersect our list of blocked regions were generated. Three different control sets with intervals of size 5 kb, 10 kb and 25 kb were generated. All ATAC bins within 20 kb of the intervals in each of these control sets were extracted.

Transcription factor motif analysis

Motif analysis for ATAC-Seq data was performed using the findMotifGenome.pl function in HOMER (option '-size-given200'; Version 4.11). Motifs were selected if (provide specific cutoff/selection criteria). Motif analysis for Hi-C data in the chromatin contact regions were discovered using HOMER findMotifGenome.pl function in HOMER (option '-size given'; Version 4.11;[95]).

Single-Cell RNA Analysis

Sample Collection and Processing

Study participants were screened at the University of Kansas Health System Advanced Reproductive Medicine Clinic. Inclusion criteria was age >18 years and <42 years at time of conception, BMI >18 kg/m^2 and <30 kg/m^2, and history of unexplained recurrent pregnancy loss (RPL) with recent diagnosis of miscarriage. Diagnosis of RPL was made using American Society for Reproductive Medicine Practice Bulletin (vol 113, no 3, March 2020) with RPL defined as "spontaneous loss of two or more pregnancies." For all patients known etiologies for miscarriage were ruled out before allowing participation in study including: Mullerian anomaly, polycystic ovary disease, thyroid disease, diabetes, coagulopathy, balanced translocation or structural rearrangement from sperm or oocyte contributor, and obesity. Miscarriage was diagnosed using American College of Obstetrician and Gynecologists Early Pregnancy Loss Bulletin (Number 200, November 2018) which defines miscarriage as "a nonviable, intrauterine pregnancy with either an empty gestational sac or a gestational sac containing an embryo or fetus without fetal cardiac activity within the first 12 6/7 weeks of gestation." All study participants had transvaginal ultrasound to confirm diagnosis of miscarriage.

Product of conception (**POC**) including placental and decidual tissue were collected from the study participants where all samples had normal karyotype based on clinical cytogenetics testing including chromosomal analysis (GTG banded chromosomes analyzed at the 450-550 band levels) and array comparative genomic hybridization performed in accordance with current International Standing Committee on Human Cytogenetic Nomenclature (ISCN 2009) (**Supplementary Table 17**).

Each sample was rinsed in PBS before the addition of 5 mL of digestion medium [DMEM/F12 (11320033, Thermo Fisher) with 10% FBS, and 100 units/mL of penicillin and 100

µg of streptomycin (15140122, Thermo Fisher) with 2.0 mg/mL collagenase type 1A] and incubated for 50 min at 37ºC. The sample was strained through a pre-wetted 100 µm filter in a 50 mL conical tube and rinsed twice with 5 mL of DMEM/F12 with 10% FBS and 100 units/mL penicillin and 100 µg/mL streptomycin and centrifuged at 300 xg for 8 min. Cell pellets were resuspended in 1 mL of cold Recovery Cell Culture Freezing Medium (12648010, Thermo Fisher), and the cell suspension was transferred to a cryogenic storage vial. The cryogenic storage vial was placed in a Corning CoolCell FTS30, which was then placed in a -80°C freezer overnight. Samples were stored at -80°C before being thawed and processed for single-cell RNA-Seq.

10X Genomics single-cell capture and sequencing

Prior to single-cell capture, samples were thawed in 10 mL of Thawing Medium consisting of DMEM/F-12 (11320033, Thermo-Fisher) supplemented with 10% FBS and 100 units/mL of penicillin and 100 µg/mL of streptomycin (15140122, Thermo-Fisher) that was prewarmed in a 37°C bead bath and centrifuged at 300 xg for 8 min. The supernatant was carefully removed without disturbing the cell pellets. The cell pellets were each resuspended in 0.5 mL of Thawing Medium, and the cell suspensions were placed on ice and passed through a prewetted 40 µm nylon mesh cell strainer. Cell suspensions were centrifuged at 300 xg for 8 min at 4°C, and the supernatant was carefully aspirated without disturbing the cell pellets. The cell pellets were resuspended in 100 µL of cold Thawing Medium, and cell count and viability were assessed using 0.4% Trypan Blue and a Countess II automated cell counter. For each sample, Chromium Chip B (100153, 10x Genomics) were loaded (**Supplementary Table 17**). Following cell loading, cDNA and library preparation was performed identically for all samples using the Chromium Single Cell 3' Library & Gel Bead Kit v3 (1000075, 10x Genomics) according to the manufacturer's protocol. Libraries were sequenced using an Illumina NovaSeq 6000 using 2x94 cycle paired-end. Sequenced reads were initially processed by the CellRanger pipeline which

includes FASTQ creation, read alignment, gene counting, and cell calling. The Cell Ranger GRCh38(v2020-A) genome was used as the reference for alignment.

Post-sequencing analysis of single-cell RNA datasets

The Cell Ranger matrix file from the single-cell RNA sequencing of RPL samples was processed along with the Cell Ranger matrices of two publicly available datasets[65,66] of gestational age-matched normal control POC samples profiled using the same platform (10X Genomics, **Supplementary Table 17**) using Seurat's standard workflow[96]. Each sample was individually processed with the SCTransform function[97] using the glmGamPoi package[98] with capture as a batch variable when applicable. The three datasets were integrated using the standard integration workflow for data normalized with SCTransform[97] and 3,000 integration features were selected for downstream analysis. After integration, linear dimensional reduction, nonlinear dimensional reduction, nearest neighbor finding, and unsupervised clustering were completed according to the standard workflow[96]. Cell type label transfer was performed by referencing the Vento-Tormo dataset[66]. The predicted cell IDs were generated using the FindTransferAnchors and TransferData functions of Seurat[96]. If necessary, cell type labels were manually adjusted based on marker gene expression to exclude ambiguous cell clusters. Cells with greater than 50% mitochondrial reads and less than 2,000 molecules were removed from all three datasets. The Vento-Tormo dataset[66] had an additional filtering where cells with a mitochondrial ratio of greater than 20% and less than 3500 UMIs and 1250 unique genes were removed.

Differential Expression Analysis

For differential expression analysis, genes were filtered according to the parameters used for Vento-Tormo dataset[66]. The raw UMI counts of samples with multiple captures were summed prior to differential expression analysis.

Genes that were expressed in three or more cells per dataset were kept for pseudobulk differential expression analysis. A matrix was aggregated to find the UMI counts by gene per sample ID within a specific cell type. Using the raw counts and metadata of a specific cell type, a DESeq dataset[99] was created, and the counts were transformed by the vst function for principal component analysis. Differential expression testing was performed by the DESeq function with default parameters on the generated DESeq dataset[99]. Pairwise differential expression contrasts were performed according to the inferred cell type labels (EVT or SCT/VCT) where Vento-Tormo[66] and Suryawanshi[65] datasets were combined as the control group. The contrast was performed with the results function from DESEq2 with default parameters (Benjamini-Hochberg p-value adjustment) and alpha equal to 0.05[99]. Log2 fold changes were shrunk with the lfcShrink function with the apeglm package[99,100]. For a gene to be significantly differentially expressed, the Benjamini-Hochberg adjusted p-value was less than 0.05.

Data visualization

Data from the single-cell RNA assays were normalized and scaled with the NormalizeData and ScaleDate functions of Seurat[96]. Violin plots and feature plots were then generated using default parameters.

Ethics Declaration

The study was approved by the Children's Mercy Institutional Review Board (study # 11120514). Informed written consent was obtained from the RPL participants before study inclusion.

Human placental specimens

Sections of paraffin-embedded first trimester placenta tissue were obtained from the Lunenfeld-Tanenbaum Research Institute (Mount Sinai Hospital, Toronto, Canada) with written informed consent. Prior approval was granted by the respective local human research ethics review committees at the Mount Sinai Hospital and the University of Kansas Medical Center.

In situ hybridization

Transcripts for *TFAP2C, EPAS1, SNAI1, PLAC8*, and *CDH1* were localized within human first trimester placental tissue sections. RNAscope Multiplex Fluorescent Reagent Kit version 2 (Advanced Cell Diagnostics) was used for in situ hybridization analysis. Probes were prepared to detect *TFAP2C* (NM_003222.3, 515921, target region: 664-1596), *EPAS1* (NM_001430.4, 410591, target region: 1332-2354), *SNAI1* (NM_005985.3, 560421, target region: 17-1233), *CDH1* (NM_004360.3, 311091-C2, target region: 263-1255), and *PLAC8* (NM_016619.3, 858491-C2, target region: 5-1448). *CDH1* and *PLAC8* probes were used to identify cytotrophoblast/trophoblast progenitor cells and differentiated EVT cells of the EVT cell column, respectively[54,55]. Fluorescence images were captured on a Nikon 90i upright microscope (Nikon) with a Photometrics CoolSNAP-ES monochrome camera (Roper).

shRNA-mediated gene silencing

Lentivirus-mediated shRNA delivery was used as a tool to test the biological roles for *TFAP2C*, *EPAS1*, and *SNAI1* in human TS cells. Several shRNAs designed for each target were subcloned into pLKO.1 and their efficacy tested by RT-qPCR and western blotting. An shRNA which does not recognize any known mammalian gene, pLKO.1-shSCR (Plasmid 1864) was used as a control (Addgene, Cambridge, MA). Sequences for shRNAs used in the analysis are provided in **Supplementary Table 18**. Construct design, lentivirus production, and lentiviral transduction were performed as previously described[32]. Statistical analyses were performed

using GraphPad Prism 8 software. Welch's *t* tests, Brown-Forsythe and Welch ANOVA tests, or Two-way ANOVA tests were applied when appropriate. Data is represented as mean ± standard deviation with the statistical significance level set at $P<0.05$.

Transcript quantification by RT-qPCR

RNA was isolated with TRIzol/chloroform precipitation (Thermo-Fisher, 15596018) as previously described[32]. cDNA was synthesized from 1 mg of total RNA using the High-Capacity cDNA Reverse Transcription Kit (Thermo-Fisher, 4368813) and diluted 10 times with ultra-pure distilled water. qPCR was performed using PowerSYBR Green PCR Master Mix (Thermo-Fisher, 4367659) and primers (250 nM each). PCR primer sequences are listed in **Supplementary Table 19**. Amplification and fluorescence detection were measured with a QuantStudio 5 Flex Real-Time PCR System (Thermo-Fisher). An initial step (95°C, 10 min) preceded 40 cycles of a two-step PCR (92°C, 15 s; 60°C, 1 min) and was followed by a dissociation step (95°C, 15 s; 60°C, 15 s; 95°C 15 s). The comparative cycle threshold method was used for relative quantification of the amount of mRNA for each sample normalized to housekeeping genes *B2M* or *POLR2A*.

Western blotting

Western blotting was performed as previously described[32]. Briefly, cell lysates were sonicated in radioimmunoprecipitation assay (**RIPA**) lysis buffer (sc-24948A, Santa Cruz Biotechnology, Dallas, TX) that was supplemented with Halt ™ Protease and phosphatase inhibitor cocktail (78443, Thermo-Fisher). Protein concentrations were measured with the DC Protein Assay (5000113-115, Bio-Rad) and RIPA buffer was used the blank standard. Proteins were separated by SDS-PAGE and transferred onto polyvinylidene difluoride (**PVDF**) membranes (10600023, GE Healthcare, Chicago, IL). Following the transfer, PVDF membranes were blocked with 5%

non-fat milk in Tris-buffered saline with 0.1% Tween 20 (**TBST**) and probed with primary antibodies to SNAI1 (3879S, Cell Signaling Technology; 1:1000), EPAS1 (66731-1 -Ig, Proteintech, 1:1000), TFAP2C (AP-2γ, 6E4/4, sc-12762, Santa Cruz Biotechnology, 1:750), and glyeraldehyde-3-phosphate dehydrogenase (**GAPDH**, ab9485, Abcam, 1:5000) overnight at 4°C. PVDF membranes were washed three times for 5 min each with TBST and incubated with secondary antibodies (goat anti-rabbit IgG HRP, A0545, Sigma Aldrich and goat anti-mouse IgG HRP, 7076, Cell Signaling Technology) for 1 h at room temperature. Immunoreactive proteins were visualized using the Luminata™ Crescendo Western HRP Substrate (WBLUR0500, Millipore) according to the manufacturer's instructions. Crescendo Western HRP Substrate (WBLUR0500, Millipore, Billerica, MA) using the manufacturer's instructions. Crescendo Western HRP Substrate (WBLUR0500, Millipore, Billerica, MA) using the manufacturer's instructions. Western HRP Substrate (WBLUR0500, Millipore, Billerica, MA) using the manufacturer's instructions.

RNA-seq analysis for silencing experiment

Stranded mRNA-Seq was performed using the Illumina NovaSeq 6000 Sequencing System at the University of Kansas Medical Center – Genomics Core (Kansas City, KS). Quality control was completed using the Agilent TapeStation 4200 with the RNA ScreenTape Assay kit (Agilent Technologies 5067-5576). Total RNA (1 µg) was used to initiate the library preparation protocol. The total RNA fraction was processed by oligo dT bead capture of mRNA, fragmentation, reverse transcription into cDNA, end repair of cDNA, ligation with the appropriate Unique Dual Index (UDI) adaptors, strand selection, and library amplification by PCR using the Universal Plus mRNA-Seq with NuQuant library preparation kit (0520-A01, Tecan Genomics). Library validation was performed using the D1000 ScreenTape Assay kit (5067-5582, Agilent) on the Agilent TapeStation 4200. The concentration of each library was determined with the NuQuant module of the library prep kit using a Qubit 4 Fluorometer (Thermo-Fisher/Invitrogen). Libraries were pooled based on equal molar amounts and the multiplexed pool was quantified, in

triplicate, using the Roche Lightcycler96 with FastStart Essential DNA Green Master (06402712001, Roche 0) and KAPA Library Quant (Illumina) DNA Standards 1-6 (KK4903, KAPA Biosystems). Using the qPCR results, the RNA-Seq library pool was adjusted to 2.125 nM for multiplexed sequencing. Pooled libraries were denatured with 0.2 N NaOH (0.04N final concentration) and neutralized with 400 mM Tris-HCl, pH 8.0. A dilution of the pooled libraries to 425 pM was performed in the sample tube on the instrument and followed by onboard clonal clustering of the patterned flow cell using the NovaSeq 6000 S1 Reagent Kit (200 cycle) (20012864, Illumina). A 2x101 cycle sequencing profile with dual index reads was completed using the following sequence profile: Read 1 – 101 cycles x Index Read 1 – 8 cycles x Index Read 2 – 8 cycles x Read 2 – 101 cycles. Following collection, sequence data were converted from .bcl file format to fastq file format using bcl2fastq software and de-multiplexed into individual sequences for data distribution using a secure FTP site for further downstream analysis.

Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; (https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE). Processed data can be visualized in the UCSC Genome Browser using the following link: https://genome.ucsc.edu/s/cmri_gmc_bioinformatics/EVT_STEM_Epigenome

COMPETING INTEREST STATEMENT

There is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

ACKNOWLEDGEMENTS

We thank Stacy Oxley, Brandi Miller, and Nhu Bui for their administrative assistance, Dan Louiselle and Rebecca Biswell for their work in sample processing, Adam Walters and Margaret

Gibson for their work in library preparation and NGS and Bradley Belden and Mackenzie Stevens for their work in clinical coordination.

AUTHOR CONTRIBUTIONS

K.M.V., M.J.S., and E.G. conceived and designed the research; K.M.V., E.D, A.M., and R.M. performed experiments; H.O., T.A., M.L., K.H., and C.M. provided reagents. K.M.V, E.D., B.K., A.M., E.W., R.P.M., K.I., W.C., C.M., M.J.S., and E.G. analyzed the data and interpreted results of experiments; K.M.V., B.K., M.J.S., and E.G. prepared figures and drafted manuscript; K.M.V., M.J.S., and E.G. edited and revised manuscript; All authors approved the final version of manuscript.

FUNDING

The research was supported by an NRSA postdoctoral fellowship to K.M.V. from the National Institutes of Health (F32HD096809), grants from the National Institutes of Health (HD020676, HD079363, HD099638, HD105734), and the Sosland Foundation. E.G holds the Roberta D. Harding & William F. Bradley, Jr. Endowed Chair in Genomic Research.

REFERENCES

1. Brosens, I., Pijnenborg, R., Vercruysse, L. & Romero, R. The 'Great Obstetrical Syndromes' are associated with disorders of deep placentation. *Am. J. Obstet. Gynecol.* **204**, 193–201 (2011).
2. Brosens, I., Puttemans, P. & Benagiano, G. Placental bed research: I. The placental bed: from spiral arteries remodeling to the great obstetrical syndromes. *Am. J. Obstet. Gynecol.* **221**, 437–456 (2019).
3. Pijnenborg, R., Vercruysse, L. & Hanssens, M. The uterine spiral arteries in human pregnancy: facts and controversies. *Placenta* **27**, 939–958 (2006).
4. Velicky, P., Knöfler, M. & Pollheimer, J. Function and control of human invasive trophoblast subtypes: Intrinsic vs. maternal control. *Cell Adhes. Migr.* **10**, 154–162 (2016).
5. Knöfler, M. *et al.* Human placenta and trophoblast development: key molecular mechanisms and model systems. *Cell. Mol. Life Sci. CMLS* **76**, 3479–3496 (2019).
6. Turco, M. Y. & Moffett, A. Development of the human placenta. *Development* **146**, dev163428 (2019).
7. Damsky, C. H. & Fisher, S. J. Trophoblast pseudo-vasculogenesis: faking it with endothelial adhesion receptors. *Curr. Opin. Cell Biol.* **10**, 660–666 (1998).
8. Bahlmann, F., Fittschen, M., Reinhard, I., Wellek, S. & Steiner, E. Reference values for blood flow velocity in the uterine artery in normal pregnancies from 18 weeks to 42 weeks of gestation calculated by automatic Doppler waveform analysis. *Ultraschall Med. Stuttg. Ger. 1980* **33**, 258–264 (2012).
9. Burton, G. J., Woods, A. W., Jauniaux, E. & Kingdom, J. C. P. Rheological and physiological consequences of conversion of the maternal spiral arteries for uteroplacental blood flow during human pregnancy. *Placenta* **30**, 473–482 (2009).

10. Kidder, B. L. & Palmer, S. Examination of transcriptional networks reveals an important role for TCFAP2C, SMARCA4, and EOMES in trophoblast stem cell maintenance. *Genome Res.* **20**, 458–472 (2010).
11. Home, P. *et al.* Altered subcellular localization of transcription factor TEAD4 regulates first mammalian cell lineage commitment. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 7362–7367 (2012).
12. Rugg-Gunn, P. J., Cox, B. J., Ralston, A. & Rossant, J. Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 10783–10790 (2010).
13. Kidder, B. L. & Palmer, S. HDAC1 regulates pluripotency and lineage specific transcriptional networks in embryonic and trophoblast stem cells. *Nucleic Acids Res.* **40**, 2925–2939 (2012).
14. Chuong, E. B., Rumi, M. A. K., Soares, M. J. & Baker, J. C. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat. Genet.* **45**, 325–329 (2013).
15. Latos, P. A. *et al.* Elf5-centered transcription factor hub controls trophoblast stem cell self-renewal and differentiation through stoichiometry-sensitive shifts in target gene networks. *Genes Dev.* **29**, 2435–2448 (2015).
16. Latos, P. A. *et al.* Fgf and Esrrb integrate epigenetic and transcriptional networks that regulate self-renewal of trophoblast stem cells. *Nat. Commun.* **6**, 7776 (2015).
17. Branco, M. R. *et al.* Maternal DNA methylation regulates early trophoblast development. *Dev. Cell* **36**, 152–163 (2016).
18. Rhee, C. *et al.* Mechanisms of transcription factor-mediated direct reprogramming of mouse embryonic stem cells to trophoblast stem-like cells. *Nucleic Acids Res.* **45**, 10103–10114 (2017).

19. Lee, B.-K. *et al.* Super-enhancer-guided mapping of regulatory networks controlling mouse trophoblast stem cells. *Nat. Commun.* **10**, 4749 (2019).
20. Schoenfelder, S. *et al.* Divergent wiring of repressive and active chromatin interactions between mouse embryonic and trophoblast lineages. *Nat. Commun.* **9**, 4189 (2018).
21. Tomikawa, J. *et al.* Exploring trophoblast-specific Tead4 enhancers through chromatin conformation capture assays followed by functional screening. *Nucleic Acids Res.* **48**, 278–289 (2020).
22. Todd, C. D., Deniz, Ö., Taylor, D. & Branco, M. R. Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells. *eLife* **8**, (2019).
23. Roberts, R. M. & Fisher, S. J. Trophoblast stem cells. *Biol. Reprod.* **84**, 412–421 (2011).
24. Benchetrit, H. *et al.* Extensive nuclear reprogramming underlies lineage conversion into functional trophoblast stem-like cells. *Cell Stem Cell* **17**, 543–556 (2015).
25. Kubaczka, C. *et al.* Direct induction of trophoblast stem cells from murine fibroblasts. *Cell Stem Cell* **17**, 557–568 (2015).
26. Abell, A. N. *et al.* MAP3K4/CBP-regulated H2B acetylation controls epithelial-mesenchymal transition in trophoblast stem cells. *Cell Stem Cell* **8**, 525–537 (2011).
27. Tuteja, G., Chung, T. & Bejerano, G. Changes in the enhancer landscape during early placental development uncover a trophoblast invasion gene-enhancer network. *Placenta* **37**, 45–55 (2016).
28. Home, P. *et al.* Genetic redundancy of GATA factors in the extraembryonic trophoblast lineage ensures the progression of preimplantation and postimplantation mammalian development. *Dev. Camb. Engl.* **144**, 876–888 (2017).
29. Mobley, R. J. *et al.* MAP3K4 controls the chromatin modifier HDAC6 during trophoblast stem cell epithelial-to-mesenchymal transition. *Cell Rep.* **18**, 2387–2400 (2017).
30. Xu, J. & Kidder, B. L. KDM5B decommissions the H3K4 methylation landscape of self-renewal genes during trophoblast stem cell differentiation. *Biol. Open* **7**, bio031245 (2018).

31. Kubota, K., Kent, L. N., Rumi, M. A. K., Roby, K. F. & Soares, M. J. Dynamic regulation of AP-1 transcriptional complexes directs trophoblast differentiation. *Mol. Cell. Biol.* **35**, 3163–3177 (2015).
32. Varberg, K. M. *et al.* ASCL2 reciprocally controls key trophoblast lineage decisions during hemochorial placenta development. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
33. Lee, C. Q. E. *et al.* What is trophoblast? A combination of criteria define human first-trimester trophoblast. *Stem Cell Rep.* **6**, 257–272 (2016).
34. Okae, H. *et al.* Derivation of human trophoblast stem cells. *Cell Stem Cell* **22**, 50-63.e6 (2018).
35. Ishiuchi, T. *et al.* Zfp281 shapes the transcriptome of trophoblast stem cells and Is essential for placental development. *Cell Rep.* **27**, 1742-1754.e6 (2019).
36. Takahashi, S. *et al.* Loss of p57KIP2 expression confers resistance to contact inhibition in human androgenetic trophoblast stem cells. *Proc. Natl. Acad. Sci. U. S. A.* (2019) doi:10.1073/pnas.1916019116.
37. Bhattacharya, B. *et al.* Atypical protein kinase C iota (PKCλ/ι) ensures mammalian development by establishing the maternal-fetal exchange interface. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 14280–14291 (2020).
38. Jaju Bhattad, G. *et al.* Histone deacetylase 1 and 2 drive differentiation and fusion of progenitor cells in human placental trophoblasts. *Cell Death Dis.* **11**, 311 (2020).
39. Saha, B. *et al.* TEAD4 ensures postimplantation development by promoting trophoblast self-renewal: An implication in early human pregnancy loss. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 17864–17875 (2020).
40. Perez-Garcia, V. *et al.* Placentation defects are highly prevalent in embryonic lethal mouse mutants. *Nature* **555**, 463–468 (2018).
41. Muto, M. *et al.* Intersection of regulatory pathways controlling hemostasis and hemochorial placentation. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2111267118 (2021).

42. Karvas, R. M. *et al.* Stem-cell-derived trophoblast organoids model human placental development and susceptibility to emerging pathogens. *Cell Stem Cell* **29**, 810-825.e8 (2022).

43. Meuleman, W. *et al.* Index and biological spectrum of human DNase I hypersensitive sites. *Nature* **584**, 244–251 (2020).

44. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

45. Pott, S. & Lieb, J. D. What are super-enhancers? *Nat. Genet.* **47**, 8–12 (2015).

46. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).

47. Moorthy, S. D. *et al.* Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. *Genome Res.* **27**, 246–258 (2017).

48. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).

49. Meinhardt, G. *et al.* Pivotal role of the transcriptional co-activator YAP in trophoblast stemness of the developing human placenta. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 13562–13570 (2020).

50. Biadasiewicz, K. *et al.* Transcription factor AP-2α promotes EGF-dependent invasion of human trophoblast. *Endocrinology* **152**, 1458–1469 (2011).

51. Krendl, C. *et al.* GATA2/3-TFAP2A/C transcription factor network couples human pluripotent stem cell differentiation to trophectoderm with repression of pluripotency. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E9579–E9588 (2017).

52. Liu, X. *et al.* Reprogramming roadmap reveals route to human induced trophoblast stem cells. *Nature* **586**, 101–107 (2020).

53. Zhu, M. *et al.* Developmental clock and mechanism of de novo polarization of the mouse embryo. *Science* **370**, eabd2703 (2020).

54. Renaud, S. J. *et al.* OVO-like 1 regulates progenitor cell fate in human trophoblast development. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6175-6184 (2015).

55. Chang, W.-L. *et al.* PLAC8, a new marker for human interstitial extravillous trophoblast cells, promotes their invasion and migration. *Dev. Camb. Engl.* **145**, dev148932 (2018).

56. Morey, R. *et al.* Transcriptomic drivers of differentiation, maturation, and polyploidy in human extravillous trophoblast. *Front. Cell Dev. Biol.* **9**, 702046 (2021).

57. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).

58. Greenwald, W. W. *et al.* Subtle changes in chromatin loop contact propensity are associated with differential gene regulation and expression. *Nat. Commun.* **10**, 1054 (2019).

59. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).

60. Kubo, N. *et al.* Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nat. Struct. Mol. Biol.* **28**, 152–161 (2021).

61. Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst.* **3**, 95–98 (2016).

62. Liu, Y. *et al.* Single-cell RNA-seq reveals the diversity of trophoblast subtypes and patterns of differentiation in the human placenta. *Cell Res.* **28**, 819–832 (2018).

63. Meinhardt, G. *et al.* Wnt-dependent T-cell factor-4 controls human extravillous trophoblast motility. *Endocrinology* **155**, 1908–1920 (2014).

64. Damsky, C. H. *et al.* Integrin switching regulates normal trophoblast invasion. *Dev. Camb. Engl.* **120**, 3657–3666 (1994).

65. Suryawanshi, H. *et al.* A single-cell survey of the human first-trimester placenta and decidua. *Sci. Adv.* **4**, eaau4788 (2018).

66. Vento-Tormo, R. *et al.* Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563**, 347–353 (2018).

67. West, R. C. *et al.* Dynamics of trophoblast differentiation in peri-implantation-stage human embryos. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 22635–22644 (2019).

68. Sheng, W. *et al.* LSD1 ablation stimulates anti-tumor immunity and enables checkpoint blockade. *Cell* **174**, 549-563.e19 (2018).

69. Nieto, M. A. The snail superfamily of zinc-finger transcription factors. *Nat. Rev. Mol. Cell Biol.* **3**, 155–166 (2002).

70. Sasagawa, T. *et al.* HIF-2α, but not HIF-1α, mediates hypoxia-induced up-regulation of Flt-1 gene expression in placental trophoblasts. *Sci. Rep.* **8**, 17375 (2018).

71. Warrington, N. M. *et al.* Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat. Genet.* **51**, 804–814 (2019).

72. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).

73. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).

74. Adachi, K. *et al.* Context-dependent wiring of Sox2 regulatory networks for self-renewal of embryonic and trophoblast stem cells. *Mol. Cell* **52**, 380–392 (2013).

75. Dong, C. *et al.* A genome-wide CRISPR-Cas9 knockout screen identifies essential and growth-restricting genes in human trophoblast stem cells. *Nat. Commun.* **13**, 2548 (2022).

76. E Davies, J. *et al.* Epithelial-mesenchymal transition during extravillous trophoblast differentiation. *Cell Adhes. Migr.* **10**, 310–321 (2016).

77. Fryer, B. H. & Simon, M. C. Hypoxia, HIF and the placenta. *Cell Cycle Georget. Tex* **5**, 495–498 (2006).

78. Patel, S. A. & Simon, M. C. Biology of hypoxia-inducible factor-2alpha in development and disease. *Cell Death Differ.* **15**, 628–634 (2008).

79. Rajakumar, A. & Conrad, K. P. Expression, ontogeny, and regulation of hypoxia-inducible transcription factors in the human placenta. *Biol. Reprod.* **63**, 559–569 (2000).

80. Fujii, T. *et al.* Enhanced HIF2α expression during human trophoblast differentiation into syncytiotrophoblast suppresses transcription of placental growth factor. *Sci. Rep.* **7**, 12455 (2017).

81. Loboda, A., Jozkowicz, A. & Dulak, J. HIF-1 and HIF-2 transcription factors--similar but not identical. *Mol. Cells* **29**, 435–442 (2010).

82. Pugh, C. W. Modulation of the hypoxic response. *Adv. Exp. Med. Biol.* **903**, 259–271 (2016).

83. Johnson, E. A. HIF takes it up a notch. *Sci. Signal.* **4**, pe33 (2011).

84. Rajakumar, A. *et al.* Selective overexpression of the hypoxia-inducible transcription factor, HIF-2alpha, in placentas from women with preeclampsia. *Biol. Reprod.* **64**, 499–506 (2001).

85. Castel, G. *et al.* Induction of human trophoblast stem cells from somatic cells and pluripotent stem cells. *Cell Rep.* **33**, 108419 (2020).

86. Cinkornpumin, J. K. *et al.* Naive human embryonic stem cells can give rise to cells with a trophoblast-like transcriptome and methylome. *Stem Cell Rep.* **15**, 198–213 (2020).

87. Guo, G. *et al.* Human naive epiblast cells possess unrestricted lineage potential. *Cell Stem Cell* **28**, 1040-1056.e6 (2021).

88. Io, S. *et al.* Capturing human trophoblast development with naive pluripotent stem cells in vitro. *Cell Stem Cell* **28**, 1023-1039.e13 (2021).

89. Wei, Y. *et al.* Efficient derivation of human trophoblast stem cells from primed pluripotent stem cells. *Sci. Adv.* **7**, eabf4416 (2021).

90. Yanagida, A. *et al.* Naive stem cell blastocyst model captures human embryo lineage segregation. *Cell Stem Cell* **28**, 1016-1022.e4 (2021).

91. Mischler, A. *et al.* Two distinct trophectoderm lineage stem cells from human pluripotent stem cells. *J. Biol. Chem.* **296**, 100386 (2021).

92. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).

93. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* **9**, 9354 (2019).

94. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma. Oxf. Engl.* **26**, 841–842 (2010).

95. 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 (2010).

96. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

97. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).

98. Ahlmann-Eltze, C. & Huber, W. glmGamPoi: fitting Gamma-Poisson generalized linear models on single cell count data. *Bioinforma. Oxf. Engl.* **36**, 5701–5702 (2021).

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

100. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinforma. Oxf. Engl.* **35**, 2084–2092 (2019).

FIGURES

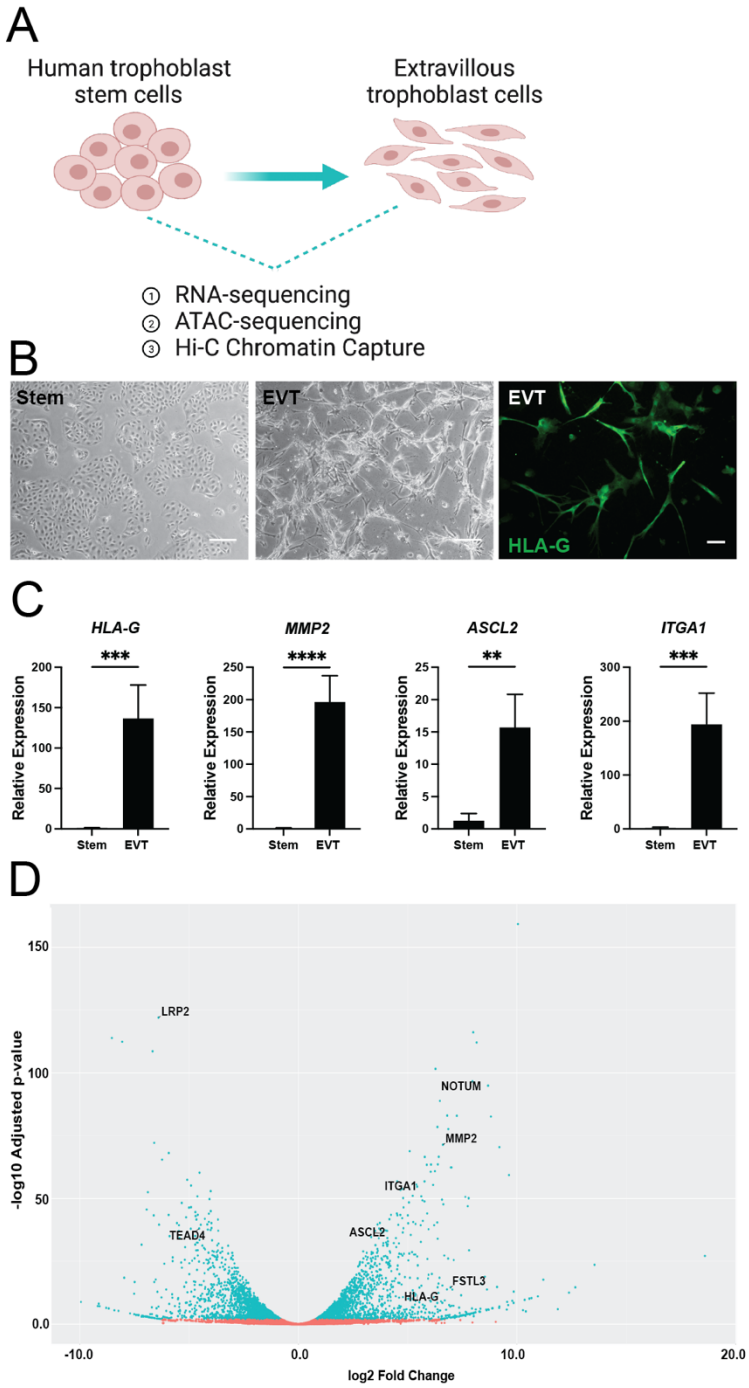

Figure 1. Morphologic and transcriptomic changes in EVT cells derived from human trophoblast stem (TS) cells. A) Simplified schematic depicting the three primary analyses (RNA-Seq, ATAC-Seq, and Hi-C) performed on human TS cell models in the stem state and

following eight days of extravillous trophoblast (**EVT**) cell differentiation. **B)** Phase contrast images of TS cells cultured in the stem and EVT cell states. Stem state cells maintain a cobble-stone morphology and proliferate to form discrete cell colonies. EVT cell differentiation is accompanied by cell elongation and upregulation of HLA-G (green, immunofluorescence panel). Scale bars represent 200 μm (phase contrast panels) and 100 μm (immunofluorescence panel). **C)** The relative expression of *HLA-G*, *MMP2*, *ASCL2*, and *ITGA1* in stem state and EVT cells (n=4 per group) measured by RT-qPCR (**$p<0.01$, ***$p<0.001$, ****$p<0.0001$). **D)** Volcano plot depicting transcriptomic changes in EVT cells compared to stem state cells measured by RNA-Seq. Differentially expressed transcripts (absolute log2 fold change >1, adjusted p value <0.05) are labeled in turquoise. Key stem state (*LRP2* and *TEAD4*) and EVT cell (*NOTUM*, *MMP2*, *ITGA1*, *ASCL2*, *HLA-G*, and *FSTL3*) transcripts are annotated.

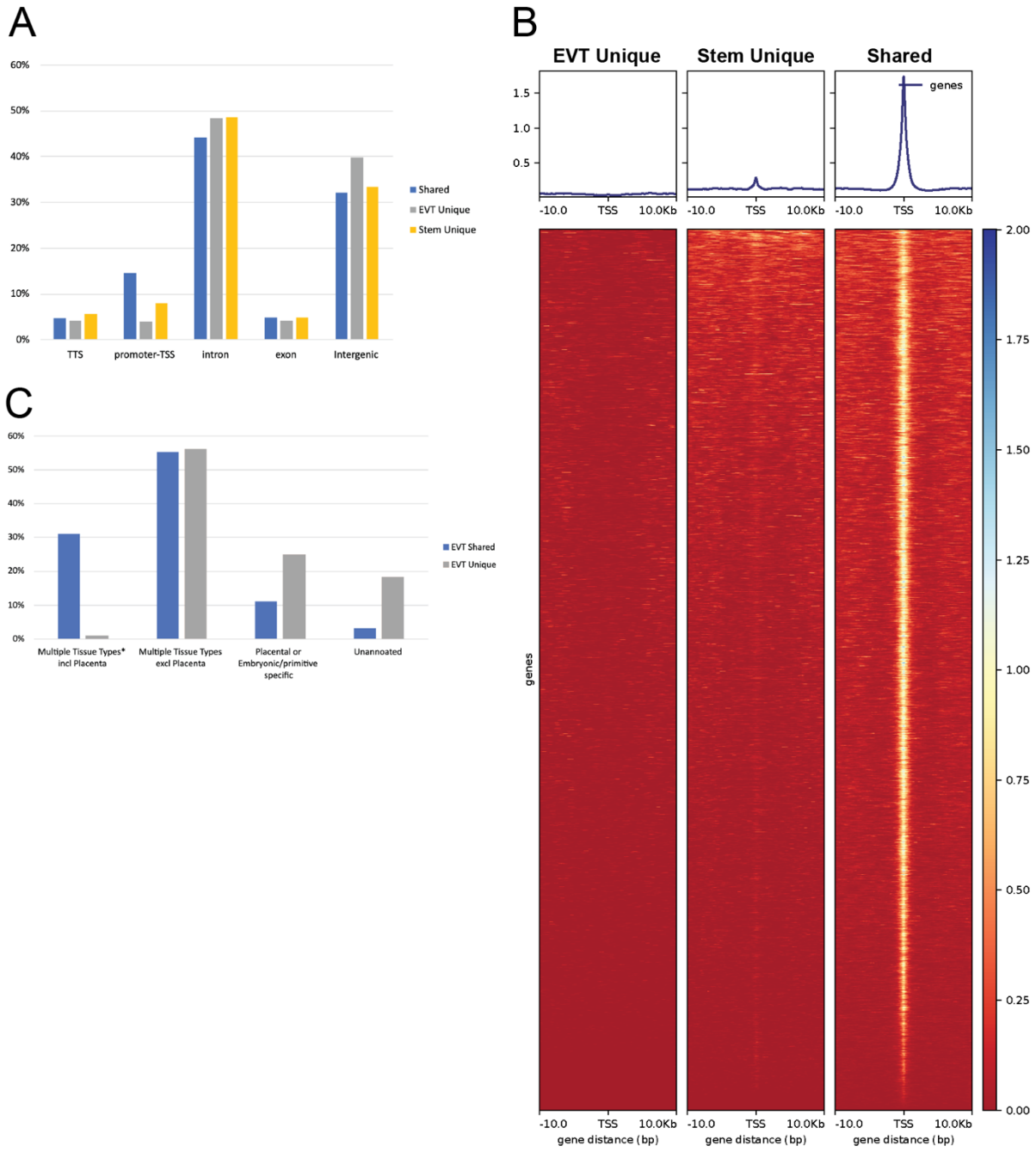

Figure 2. EVT cell differentiation drives global changes in chromatin accessibility. A) Percentage (%) distribution of chromatin accessible regions across five defined genomic regions (e.g., transcription termination site (**TTS**), promoter-transcription start site (**TSS**), intron, exon, and intergenic regions). Chromatin accessible regions were identified by ATAC-Seq and are classified into four primary groups, 1) regions that are shared between stem and EVT cells

(blue) and 2) regions that are unique to EVT cells (gray) or 3) unique to stem state cells (yellow). **B)** Heat maps showing the density of chromatin accessible regions within 10Kb of TSS that are unique to EVT cells, unique to stem state cells, or shared between the two cell states. Colorimetric scale (red to blue) represents peak number intensity. **C)** Percentage (%) distribution of EVT cell chromatin accessible regions shared with the stem state (blue) or unique to EVT cells (gray) that were annotated using an atlas of regulatory DNA[43] or unannotated.

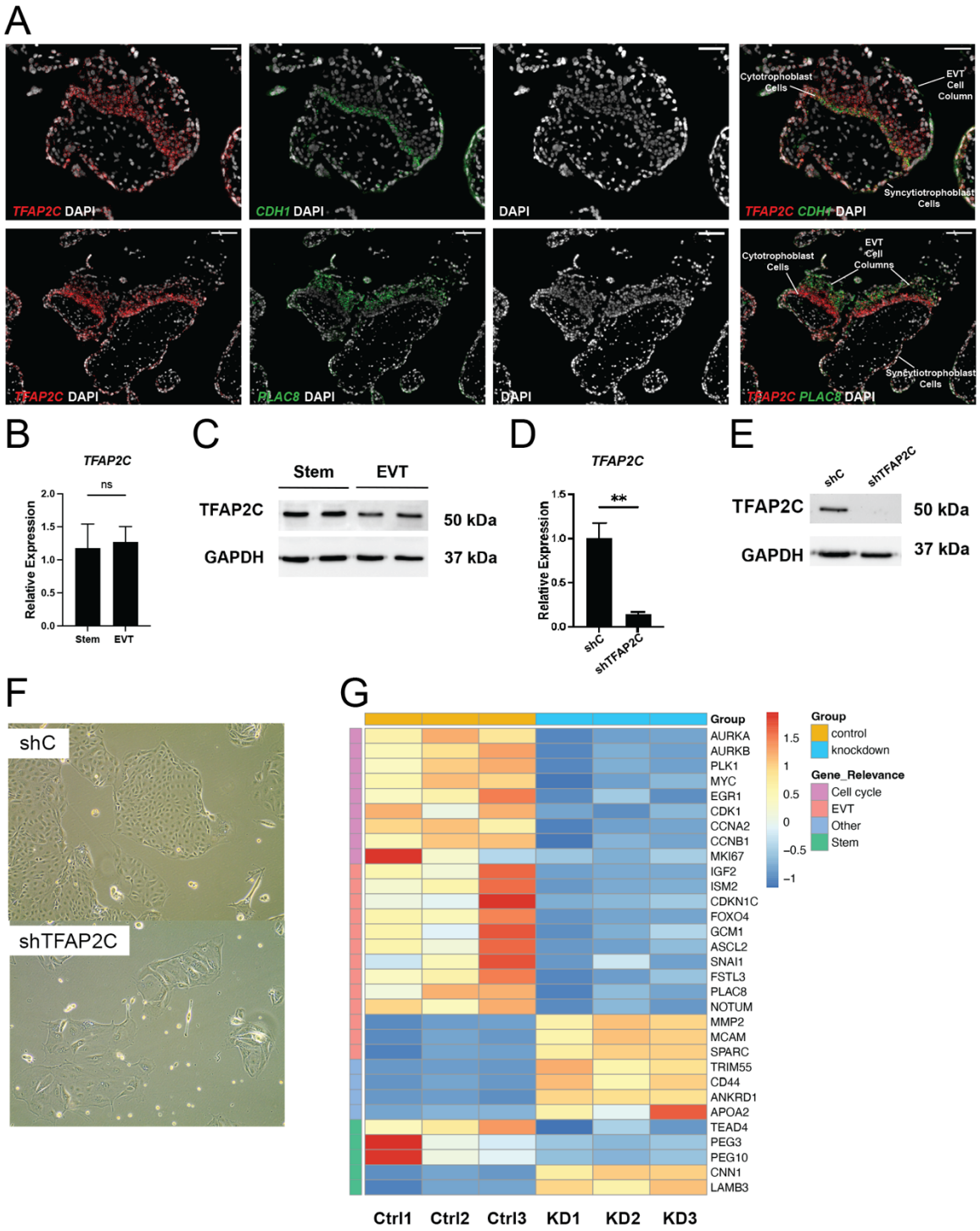

Figure 3. Functional assessment of *TFAP2C* actions in trophoblast cell development.

A) Human placental tissue specimen (12 weeks of gestation) probed for *TFAP2C*, *CDH1*, and *PLAC8* transcripts using *in situ* hybridization. DAPI labels cell nuclei. Overlay of merged immunofluorescence images: *TFAP2C* (red), DAPI (gray), and either *CDH1* (green) or *PLAC8*

(green), respectively. Scale bars represent 50 μm (*CDH1* image panels) and 100 μm (*PLAC8* image panels). **B)** Relative expression of *TFAP2C* normalized to *POLR2A* in stem and EVT cells measured by RT-qPCR ($p>0.05$; n=4 per group). **C)** TFAP2C (50 kDa) and GAPDH (37 kDa) proteins in stem and EVT cells assessed by western blot analysis. **D)** Relative expression of *TFAP2C* normalized to *POLR2A* in stem state cells transduced with lentivirus containing a control shRNA (**shC**) or a *TFAP2C*-specific shRNA (**shTFAP2C**; n=4 per group; $^{**}p<0.01$). **E)** TFAP2C (50 kDa) and GAPDH (37 kDa) proteins in stem state cells transduced with shC or shTFAP2C and assessed by western blot. **F)** Phase contrast images of stem state cells transduced with shC or shTFAP2C. **G)** Heat map based on scaled, normalized counts of selected differentially expressed transcripts generated from RNA-Seq in shC (control; orange) or shTFAP2C (knockdown; blue) transduced cells. Transcripts are clustered into four groups including cell cycle (purple), EVT-specific (pink), stem state-specific (green), or other (blue). Graphs in panels B and D depict mean ± standard deviation (SD).

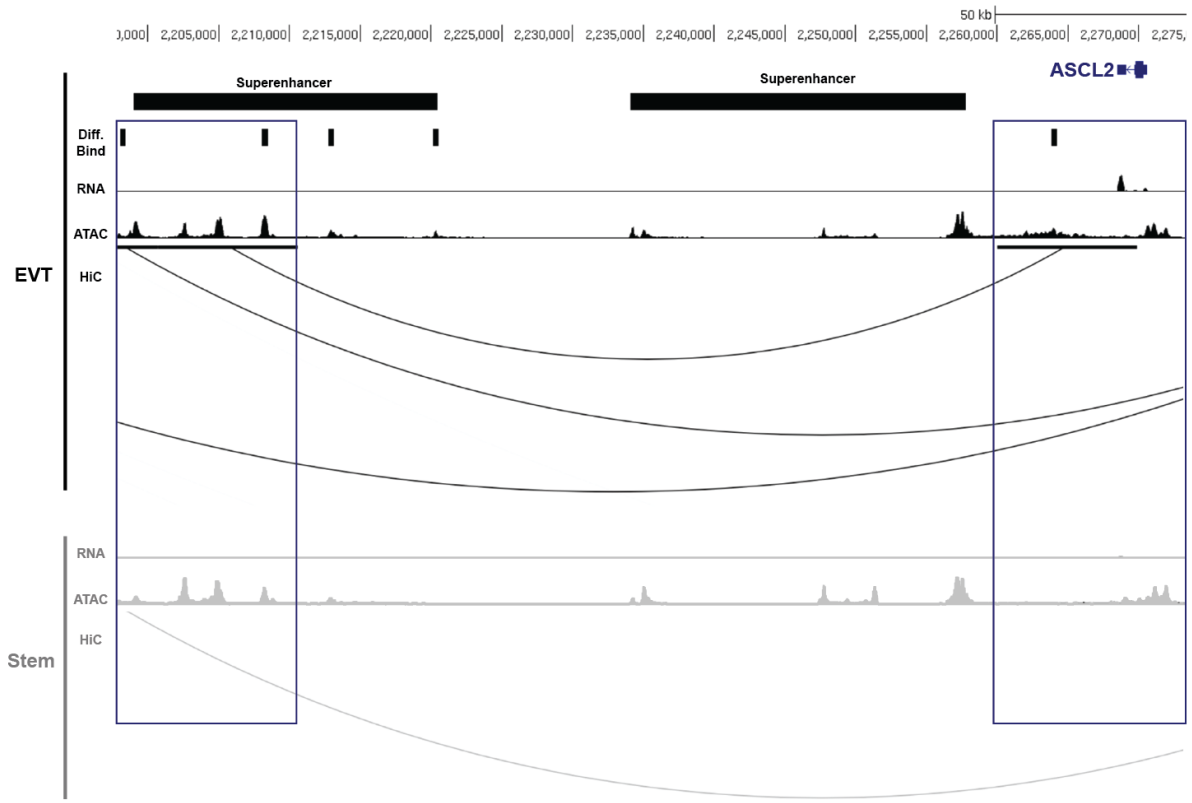

Figure 4. Long-range chromatin interactions define a unique regulatory landscape for the *ASCL2* gene in EVT cells. ATAC-Seq, RNA-Seq, and Hi-C assessments performed in EVT cells (top panel, black) and stem state cells (bottom panel, grey) and shown for regions surrounding *ASCL2* (blue). ATAC-Seq (Counts per million mapped reads, CPM, y-axis) and RNA-Seq data (Transcripts per million, TPM, y-axis) are shown in individual tracks, Hi-C loops, are shown below the ATAC-Seq tracks [EVT (black); Stem state (gray)]. Super-enhancers and differentially bound regions are specific to the EVT cell state and are depicted as black bars below the *ASCL2* track.

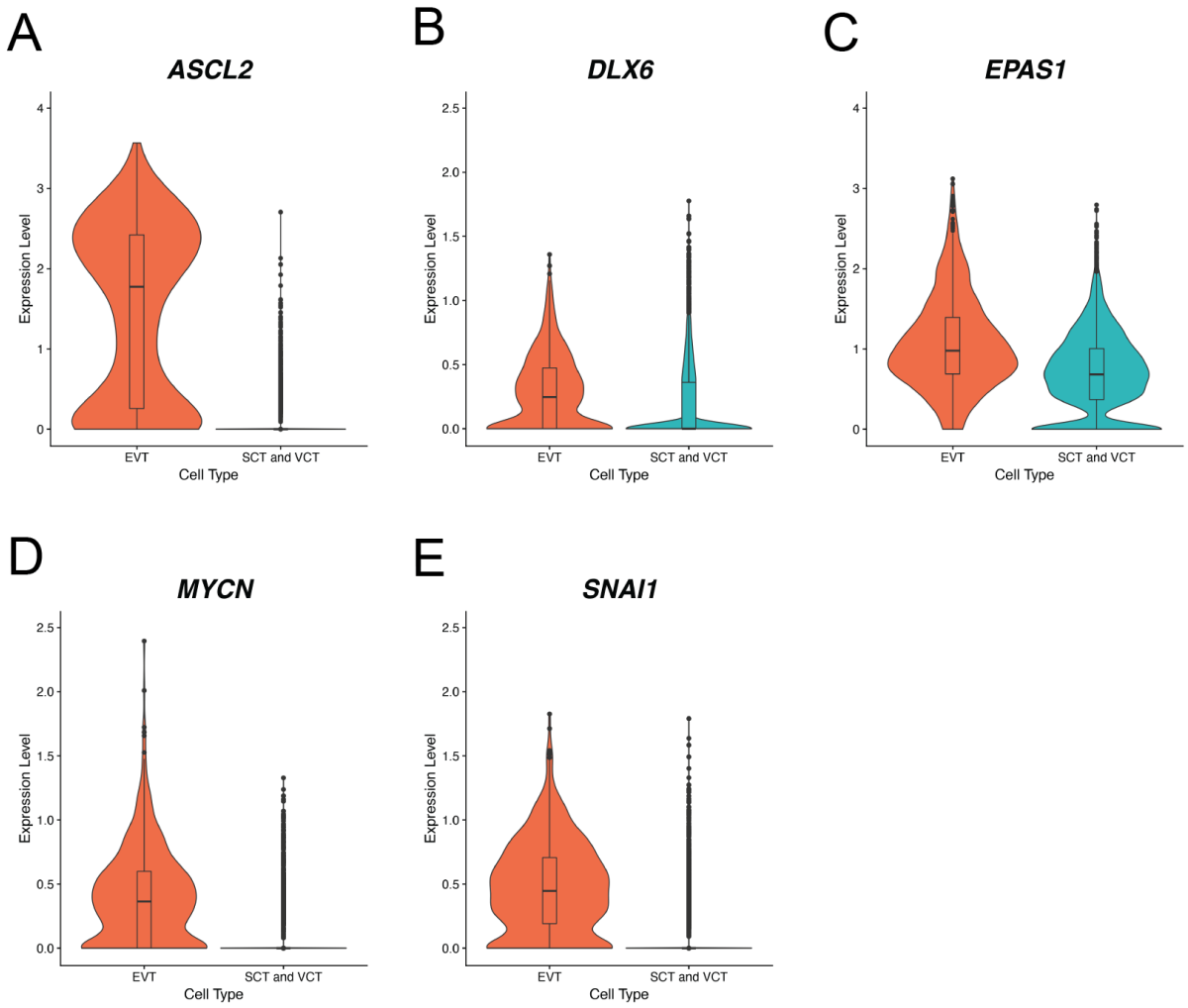

Figure 5. Expression of transcription factors in single EVT cells derived from human placenta. Violin plots generated from single-cell RNA-Seq analysis of human placental specimens (first trimester of gestation) for five transcription factors [*ASCL2* (**A**); *DLX6* (**B**); *EPAS1* (**C**); *MYCN* (**D**), and *SNAI1* (**E**)] in EVT cells (orange) compared with syncytiotrophoblast (**SCT**) and villous cytotrophoblast cells (**VCT**) combined (turquoise). Normalized and natural log-transformed single cell expression values are plotted on the y-axis.

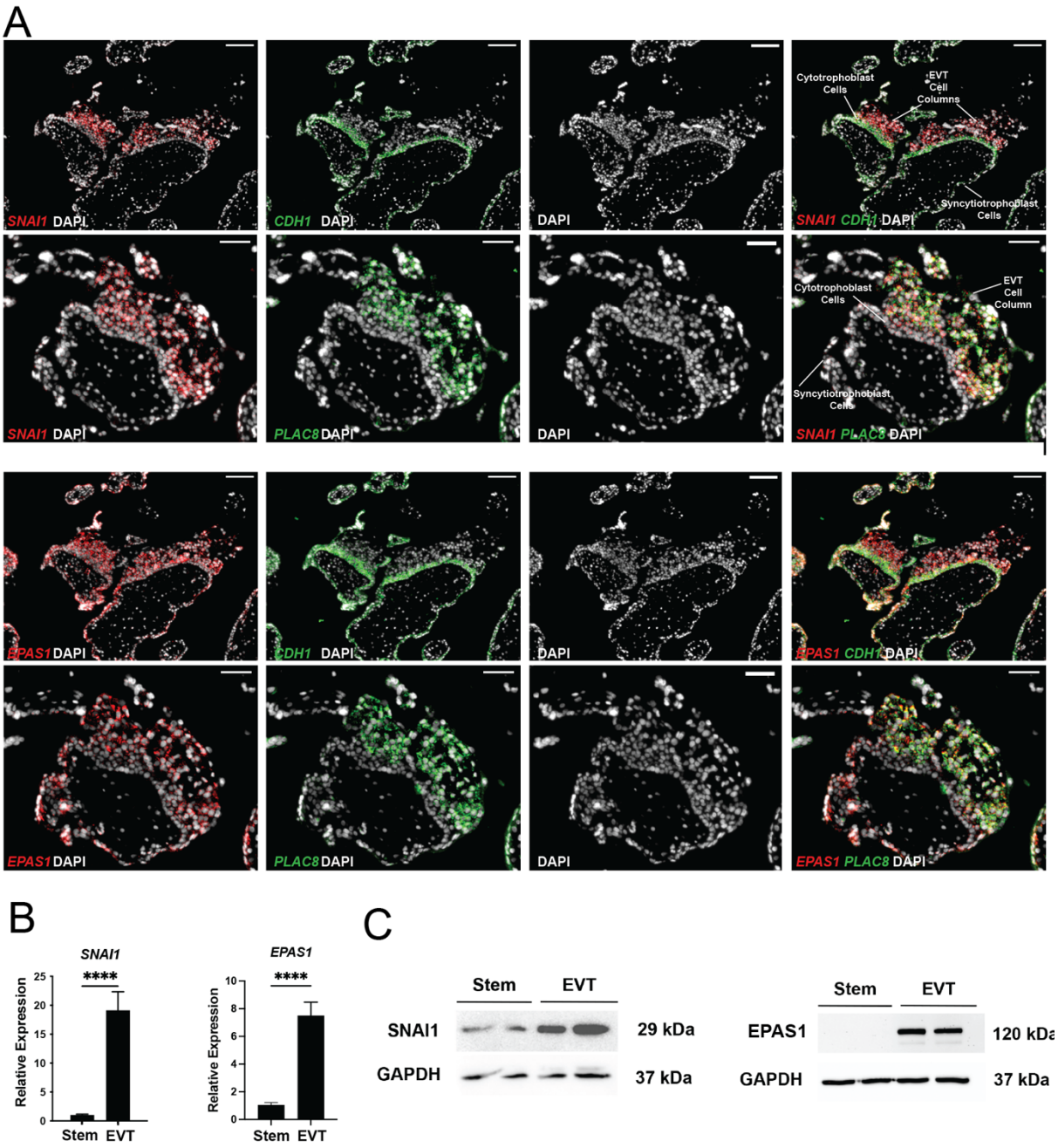

Figure 6. Expression patterns of *SNAI1*, and *EPAS1*. A) Human placental tissue specimen (12 weeks of gestation) probed for *SNAI1, EPAS1, CDH1* and *PLAC8* transcripts using *in situ* hybridization. DAPI labels cell nuclei. Overlay of merged immunofluorescence images: *SNAI1* (red), or *EPAS1* (red), with DAPI (gray), and *CDH1* (green) or *PLAC8* (green). Scale bars represent 100 µm in *CDH1* image panels and 50 µm in *PLAC8* image panels. **B)** Relative expression of *SNAI1* (Stem n=6; EVT n=4), and *EPAS1* (n=6 per group) normalized to *B2M* in

stem and EVT cells measured by RT-qPCR (*$p<0.05$ and ****$p<0.0001$). Graphs depict mean ± standard deviation (SD). **C)** SNAI1 (29 kDa), EPAS1 (120 kDa), and GAPDH (37 kDa) protein in stem and EVT cells assessed by western blot analysis.

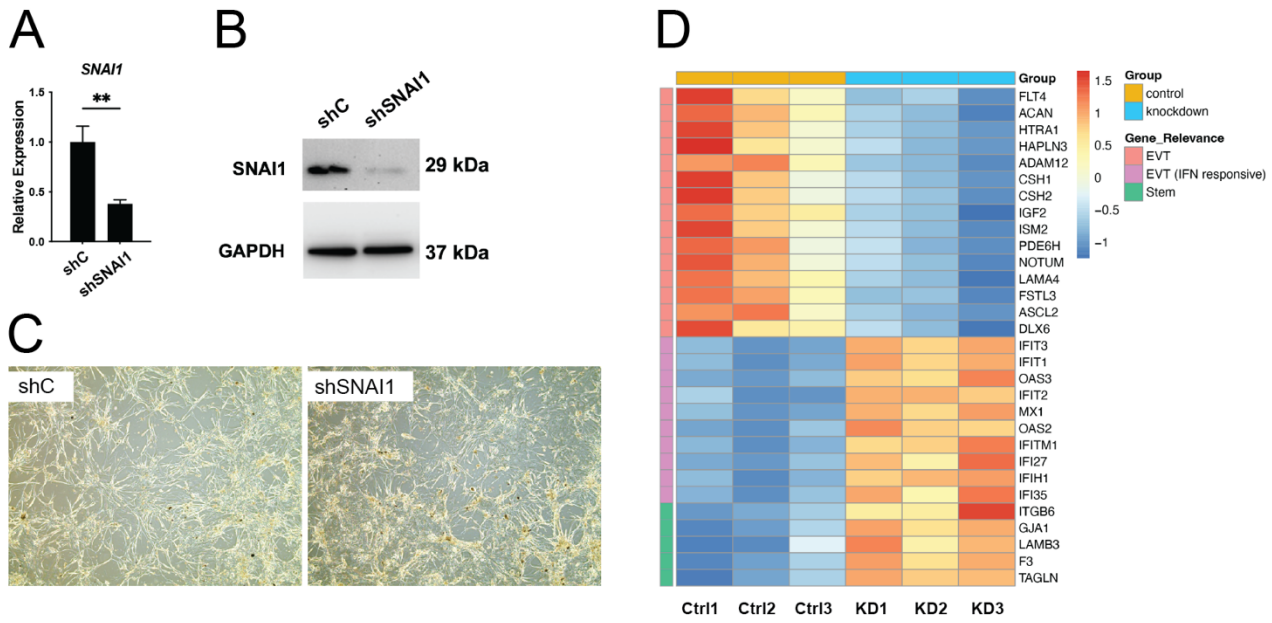

Figure 7. Functional investigation of *SNAI1* on EVT cell differentiation.

A) Relative expression of *SNAI1* (n=5 per group) normalized to *B2M* in EVT cells differentiated from stem state cells transduced with lentivirus containing a control shRNA (**shC**) or SNAI1-specific shRNA (**shSNAI1**). **B)** SNAI1 (29 kDa) and GAPDH (37 kDa) protein in EVT cells from shC control or shSNAI1 treated cells measured by western blot. **C)** Phase contrast images of shC control or shSNAI1 EVT cells following eight days of differentiation. **D)** Heat map based on scaled, normalized counts of selected transcripts generated from RNA-Seq in shC (control; orange) or shSNAI1 (knockdown; blue) transduced cells (n=3 per group). Transcripts are clustered into three groups including EVT-specific (pink), EVT interferon (IFN) responsive (purple), or stem state-specific (green).

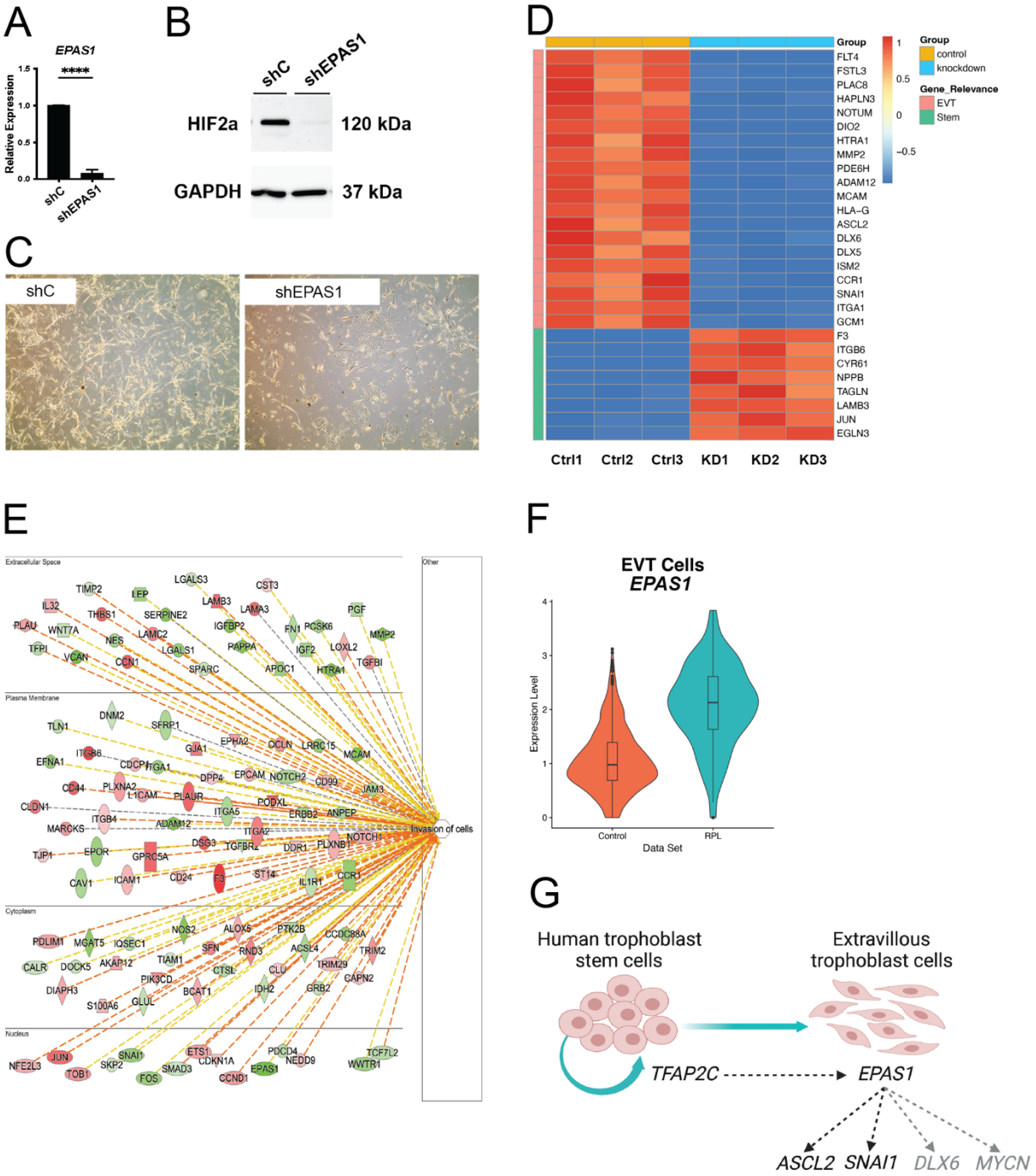

Figure 8. Functional investigation of *EPAS1* on EVT cell differentiation. A) Relative expression of *EPAS1* (n=3 per group) normalized to *B2M* in EVT cells differentiated from stem state cells transduced with lentivirus containing a control shRNA (**shC**) or a EPAS1-specific shRNA (**shEPAS1**). **B)** EPAS1 (120 kDa) and GAPDH (37 kDa) protein in EVT cells from shC

control or shEPAS1 treated cells measured by western blot. **C)** Phase contrast images of shC control or shEPAS1 EVT cells following eight days of differentiation. Heat map based on scaled, normalized counts of selected transcripts generated from RNA-Seq in shC control (orange) or shEPAS1 (blue) transduced cells (n=3 per group). Transcripts are clustered into two groups including EVT-specific (pink) or stem state-specific (green). **E)** Network schematic depicting results from Ingenuity Pathway Analysis performed on 209 genes among top 500 differentially expressed genes regulated (upregulated or downregulated) by *EPAS1* that was annotated to the biological function “Migration of Cells” (p=8.3E-49). The most significant biological function associated with this subset was “Invasion of Cells” (p=5.3E-70). The biological function is displayed as a network with subcellular layout with gene targets as nodes and the node shapes indicate the gene’s primary function. Node colors indicate increased (red) or decreased (green) expression of the genes following *EPAS1* disruption. **F)** Violin plot of *EPAS1* expression in EVT cells generated from single-cell RNA-Seq analysis of human first trimester placental samples derived from control patients (orange) and patients experiencing recurrent pregnancy loss (**RPL**, turquoise). The y-axis represents normalized and natural log transformed *EPAS1* expression levels. **G)** *TFAP2C* is important for maintenance of human TS cells in the stem state. *EPAS1* is induced upon EVT cell differentiation and acts as an upstream regulator of several EVT-specific transcription factors including *ASCL2*, *SNAI1*, *DLX6*, and *MYCN*.