Chapter 5 - Cell type-level analysis of genes implicated in pregnancy complications and fertility-related conditions

5.1 Overview

A number of studies have associated specific genes with increased incidence of pregnancy complications or other factors influencing fertility and reproductive success. However, the molecular and cellular mechanisms by which these genes contribute to the development of these conditions remain largely unknown. scRNA-seq data represents one means of furthering studies of candidate genes, as it facilitates identifying which specific cell types at the maternal-fetal interface preferentially express each implicated gene. Reciprocally, collective examination of the cell type-level expression patterns of all genes linked to a specific disease or condition enables us to map identified cell types to disease-specific or clinically relevant contexts. In this chapter, I describe our single-cell transcriptomics-level profiling of database- and literature-curated genes associated with fertility-related conditions or pregnancy outcome, and present the insights obtainable from these analyses with focus on genes linked to preeclampsia and abnormal fetal growth.

5.2 Curation of genes associated with pregnancy complications and fertility and intersection with genes upregulated in individual cell types at the maternal-fetal interface

We first curated a list of 306 genes previously associated with complications of pregnancy or fertility-related conditions from the NHGRI/EBI GWAS Catalog[146], OMIM database[147], and literature searches. A general summary of the conditions and genes studied is provided in **Table 1** and individual genes are described in Appendix 6.

Table 1. Number of genes and curation sources associated with each pregnancy complication or fertility-related condition of interest.

Disease or phenotype	# curated genes (total)	# maternally expressed	# fetally expressed	Sources
Abnormal birth weight/fetal growth	60	14	57	NHGRI/EBI, literature
Endometriosis/ovarian disease	53	53	2	NHGRI/EBI, literature
Gestational trophoblastic disorder/ hydatidiform mole	7	2	5	OMIM, literature
Age of menopause/menstrual onset	47	47	0	NHGRI/EBI, literature
Recurrent miscarriage	19	19	5	OMIM, literature
Placenta accreta	3	2	3	Literature
Placental abruption	21	17	9	OMIM, literature
Preterm birth	41	37	6	NHGRI/EBI, literature
Preeclampsia	95	57	50	NHGRI/EBI, OMIM, literature

We then intersected this gene list with the genes determined by droplet-based scRNA-seq to be significantly upregulated (adjusted $p < 0.05$) in each cell type relative to other populations at the

maternal-fetal interface. Finally, we visualized the cell type-averaged expression levels of all disease- and fertility-linked genes found to be significantly enriched in one or more of the cell populations (**Figure 1**). Collectively, each of the annotated cell types at the maternal-fetal interface exhibits upregulation of a subset of genes from our curated gene set. This underscores how the various cell populations in the decidua and placenta individually contribute to the proper functioning of the maternal-fetal microenvironment, or conversely, can be implicated in the processes leading to its dysregulation.

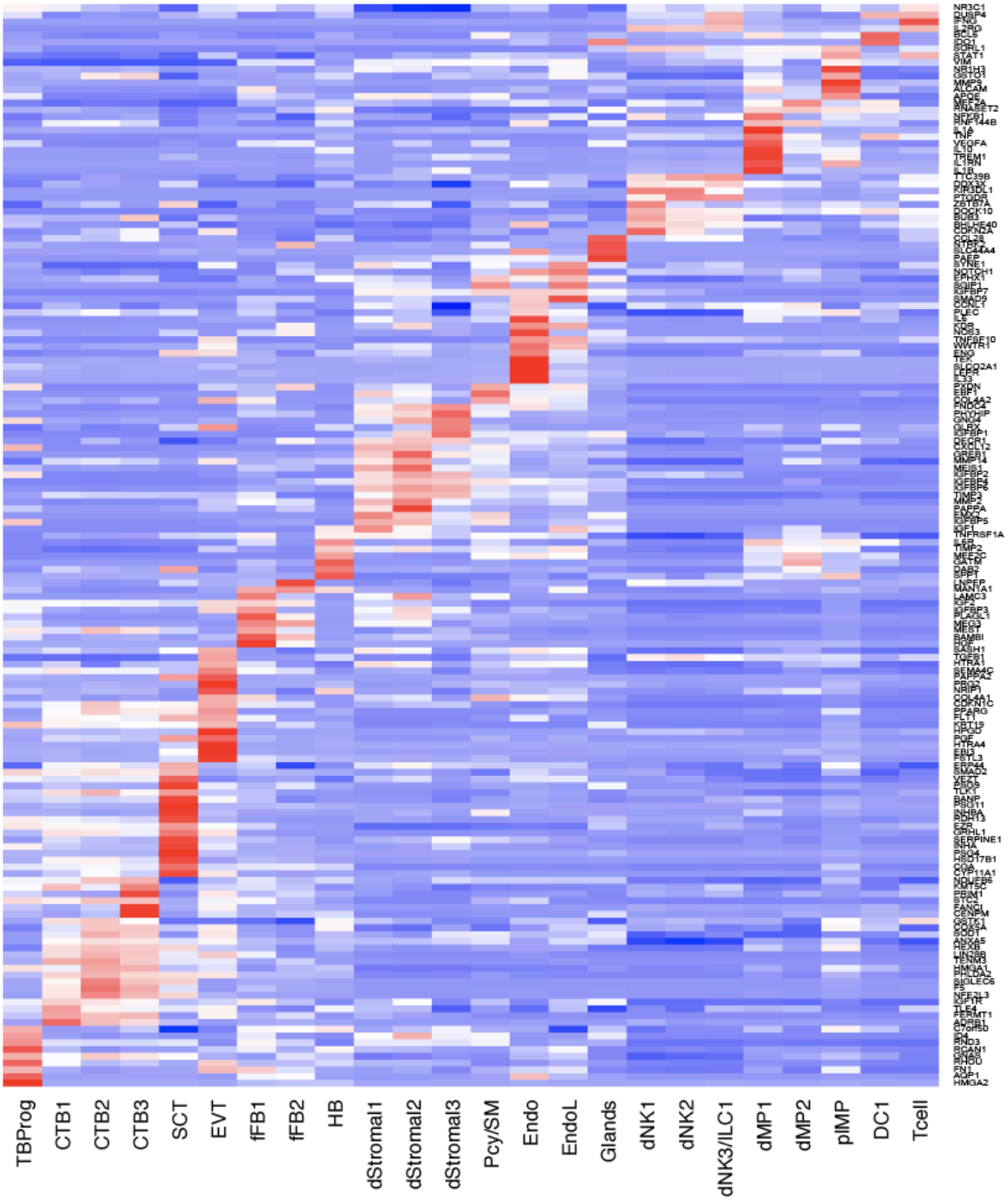

Figure 1. Genes associated with pregnancy complications or fertility are upregulated within all annotated cell types at the maternal-fetal interface. Heatmap plotting the disease- and fertility-linked genes which are

preferentially expressed in a particular cell type relative to all other populations identified at the maternal-fetal interface.

Many of the disease- and fertility-associated genes have been previously linked to maternal- or fetal-specific expression and function. After annotating each gene as being maternally and/or fetally associated based on literature curation efforts (Appendix 6), we sought to specifically intersect maternally expressed genes with the marker genes distinguishing maternal cell types, and fetally expressed genes with the marker genes distinguishing fetal cell types, in order to obtain a more functionally relevant understanding of the genes in their proper maternal or fetal context (**Figure 2a-b**). Results among maternal cells show that many maternally-associated genes are most highly expressed in the stromal cell and mononuclear phagocyte subsets, as well as in vascular smooth muscle and endothelial cells. Among the stromal cells and mononuclear phagocytes, relatively large numbers of genes are enriched in dStromal2, dMP1, and plMP (**Figure 2a**). Meanwhile, a majority of fetally-associated genes are specifically overexpressed among trophoblast cells, particularly in the differentiated SCT and invading EVT, although fibroblasts and Hofbauer cells were observed to upregulate a number of genes as well (**Figure 2b**).

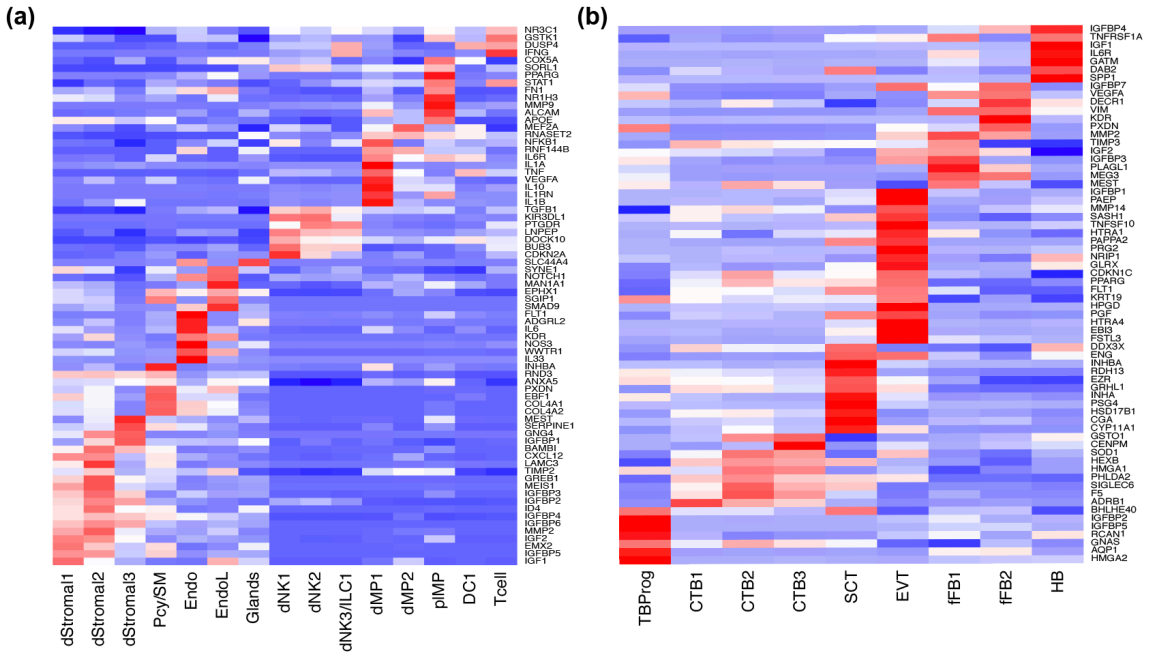

Figure 2. Specific maternal and fetal cell populations upregulate a number of the maternally-expressed and fetally-expressed genes associated with pregnancy complications and fertility. (a) Heatmap plotting the maternally expressed genes upregulated in identified maternal cell types; decidual stromal cells (dStromal1-3), vascular smooth muscle (Pcy/SM) and endothelial cells (Endo), and mononuclear phagocytes (dMP1/2 and plMP) exhibit upregulation of many of these genes. **(b)** Heatmap plotting the fetally expressed genes upregulated in

identified fetal cell types; the trophoblast cell subsets (TBProg, CTB1-3, SCT, and EVT), exhibit upregulation of many of these genes.

5.3 Analyses of maternally and fetally expressed genes implicated in preeclampsia and aberrant fetal growth

Next, we focused on individual pregnancy complications or fertility-related conditions, visualizing the cell type-level expression patterns of maternally and fetally expressed genes specifically associated with these conditions. We sought to gain a preliminary understanding of the relevance of each cell type to disease pathology, and reciprocally, determine in which cell type or types the gene set linked with each disease is most highly enriched. In addition to the results discussed below, heatmaps for other diseases and conditions in which relevant genes are highly expressed by specific cell populations are shown in Appendix 11.

Examining the genes associated with preeclampsia, a disorder characterized by hypertension and poor placentation, one notable observation is that trophoblast cells collectively express the largest proportion of fetal preeclampsia-linked genes (**Figure 3a**). Among these are genes involved in regulating cell growth and differentiation or apoptosis, such as *SOD1*, *HTRA1*, and *INHBA*; angiogenic factors including *Flt1*; and genes such as *PAPPA2* and *PRG2*, which are involved in regulating the insulin-like growth factor (IGF) pathway. This aligns with our previous understanding that precise coordination of trophoblast cell dynamics, including proliferation, differentiation, timing and extent of decidual invasion, is critical for ensuring sufficient arterial remodeling and placentation. Indeed, alterations in these mechanisms are known to be predisposing factors for preeclampsia and other placenta-associated pathologies[20,22].

Meanwhile, a number of maternal genes associated with preeclampsia are most highly expressed in the smooth muscle and endothelial cells encircling and lining the vessel walls in the maternal decidua (**Figure 3b**). This includes genes contributing to ECM formation, such as collagen IV subunit-encoding *COL4A1* and *COL4A2*, and *PXDN* (peroxidasin); genes regulating cell proliferation and growth, such as *INHBA* and *WWTR1* (*TAZ*); the inflammatory cytokine *IL-6*; endothelial nitric oxide synthase *NOS3*; and *Flt-1/sFlt-1*, which modulate angiogenesis. The selective enrichment of these genes in cells of the decidua vasculature reinforces our understanding that a number of vascular smooth muscle and endothelial cell functions are critical for ensuring proper maternal vascular adaptation in early pregnancy[199]. The dysregulation of processes such as tissue remodeling and vascular homeostasis in

these maternal cells represents another set of mechanisms potentially contributing to the pathogenesis of preeclampsia.

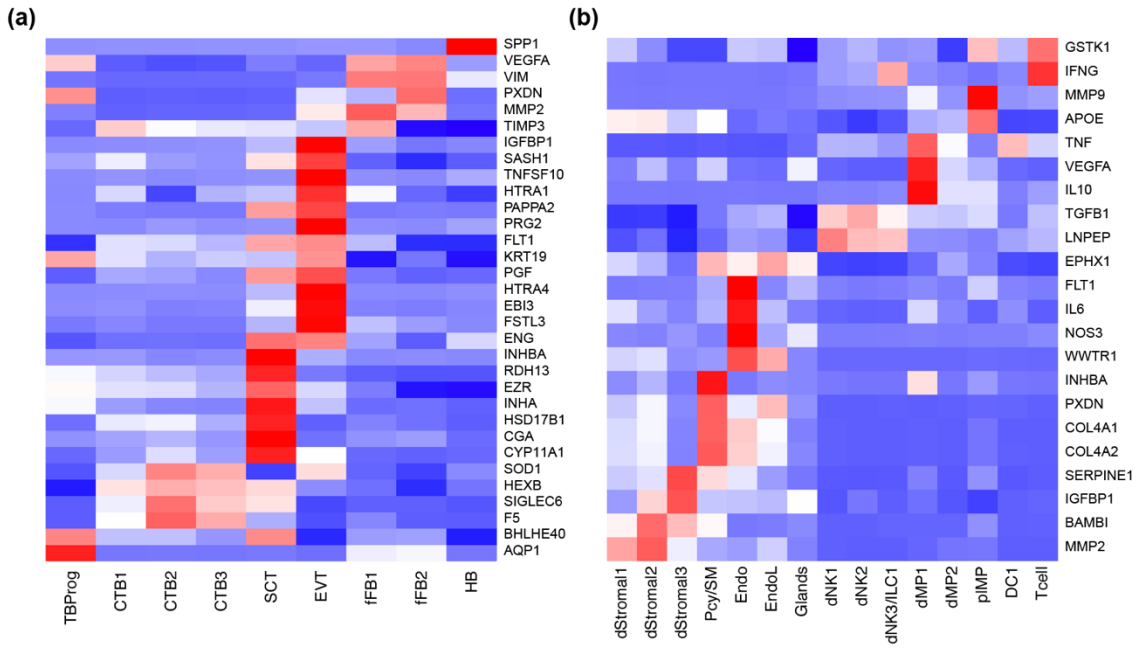

Figure 3. Preeclampsia-associated genes are preferentially expressed by fetal trophoblasts and by maternal endothelial and vascular smooth muscle cells. (a) Heatmap plotting the preeclampsia-linked genes that are known to be fetally expressed and that are significantly upregulated in specific fetal cell populations. Trophoblasts, particularly SCT and EVT, are shown to upregulate a number of these genes. **(b)** Heatmap plotting the preeclampsia-linked genes that are known to be maternally expressed and that are significantly upregulated in specific maternal cell populations. Vascular smooth muscle cells and endothelial cells exhibit enriched expression of a number of these genes.

Meanwhile, examining the genes implicated in aberrant fetal growth and birth weight, we observed that interestingly, decidual stromal cell populations exhibit upregulation of a particularly large proportion of the maternally expressed genes which are implicated in fetal growth (**Figure 4a**). Most of these genes are members of the IGF signaling pathway, with dStromal1 expressing enriched levels of the insulin-like growth factors *IGF1* and *IGF2*, and dStromal2/3 expressing high levels of the genes encoding IGF binding proteins *IGFBP1-6*. Fetally expressed genes in the IGF pathway have also been linked to fetal growth and birth weight, and we observed that elevated expression of the fetal IGFs and their binding proteins is distributed across several placental cell populations (**Figure 4b**), with EVT and trophoblast progenitors expressing high levels of *IGFBP1, 2, 5, 7*; fetal fibroblasts exhibiting upregulated *IGF2* and *IGFBP3*/*7*; and Hofbauer cells (fetal macrophages) expressing high levels of *IGF1* and *IGFBP4*.

IGF1 and IGF2 binding to the IGF1 receptor activates the PI3K and MAPK signalling pathways to promote increased protein synthesis, cell growth and proliferation, and reduced apoptosis[200]. The IGF

binding proteins function to increase the bioavailability of IGFs and their accessibility to receptors, while also potentially synergizing with IGF to activate downstream pathways through binding to other cell surface receptors[201]. The different decidual stromal subsets and placental cell populations appear to preferentially express varying members of maternal and fetal genes in the IGF/IGFBP families, suggesting they occupy specific but complementary functional roles in modulating IGF signaling, and in turn, fetal growth.

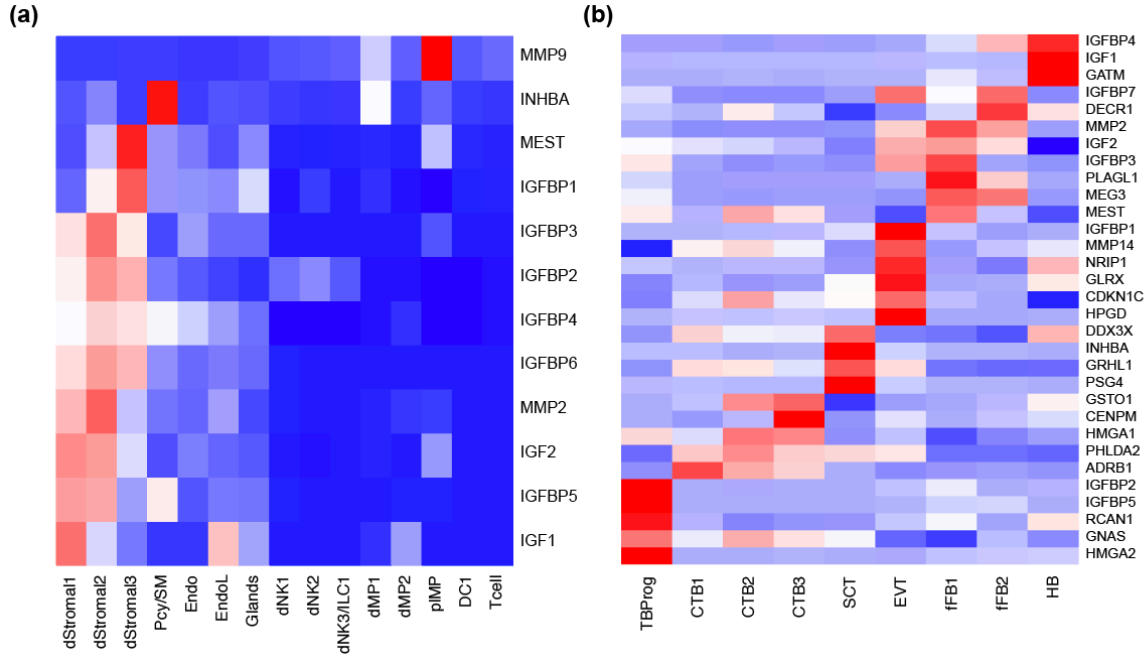

Figure 4. Genes associated with aberrant fetal growth or birth weight are preferentially expressed by maternal stromal cells and by fetal trophoblasts, fibroblasts, and Hofbauer cells. (a) Heatmap plotting fetal growth-associated genes that are known to be maternally expressed and that are significantly upregulated in specific maternal cell populations. The decidual stromal cell subsets (dStromal1-3) upregulate a number of insulin-like growth factors (IGFs) and IGF-binding proteins (IGFBPs) which comprise the IGF signaling axis. **(b)** Heatmap plotting genes linked to fetal growth that are known to be fetally expressed and that are significantly upregulated in specific fetal cell populations. Trophoblast, fibroblast, and Hofbauer cells collectively upregulate genes associated with the IGF axis.

5.4 Discussion

In this chapter I presented our analyses intersecting existing data on genes associated with pregnancy complications or fertility-related conditions with decidual and placental scRNA-seq data. We sought to explore how the cell type-specific resolution of gene expression afforded by scRNA-seq enables better understanding of which cell types are potentially driving the pathology of diseases or the development of these conditions.

Among the insights we derived from these analyses are that trophoblast subsets exhibit enriched expression of many of the fetally expressed genes implicated in preeclampsia (**Figure 3a**). This is consistent with existing knowledge that trophoblast growth and differentiation are tightly linked to the success of placentation and to the likelihood of developing preeclampsia. Collectively, the trophoblast-upregulated genes include *HSD17B1*, which regulates estradiol signaling, and *CGA*, a component of glycoprotein hormones, underscoring the notion that trophoblast cell functions are hormonally regulated and that imbalances in hormones themselves or their signaling could contribute to deficient placentation. Other enriched genes include *SASH1*, *TNFSF10* (*TRAIL*), *HTRA1*, *INHBA/INHA*, and *SOD1*, which collectively modulate the balance between cell growth, proliferation, and apoptosis, suggesting the key role of proper trophoblast development and differentiation in placentation. Trophoblast upregulation of genes such as *HTRA4*[202] and *EZR* also point to the importance of properly timed and regulated trophoblast cell adhesion, migration, and invasion at the maternal-fetal interface. Finally, the overexpression of vascular endothelial and insulin-like growth factor-involved genes such as *FLT1*, *PAPPA2*, and *PRG2* suggests that trophoblast roles in tissue and vascular remodeling are also of consequence. Overall, the intersection of fetal preeclampsia-associated genes with the placental cell types at the maternal-fetal interface enabled us to pinpoint the trophoblasts as the cell populations in which a number of these genes are most highly expressed, and in which their likely functional consequences are most apparent. At the same time, this analysis also reinforces the myriad factors underlying trophoblast cell dynamics and the importance of their coordinated regulation for proper trophoblast function, and in turn, the various means by which trophoblast cells may contribute to the pathogenesis of preeclampsia.

Meanwhile, vascular smooth muscle and endothelial cells exhibit upregulation of many of the maternally expressed genes associated with preeclampsia (**Figure 3b**), underscoring the roles of proper maternal vascular function and adaptation in ensuring successful placentation and influencing the risk of pregnancy complications linked to these processes. Smooth muscle cells express high levels of collagen IV genes *COL4A1* and *COL4A2* as well as *PXDN*, a peroxidase involved in crosslinking collagen IV chains and organizing the ECM[203,204], suggesting their consequential roles in mediating the tissue remodeling functions which are critical for regulation of EVT invasion and maternal arterial transformation. Other highly expressed genes include *WWTR1* (*TAZ*), a transcriptional coactivator in the Hippo pathway which modulates endothelial cell proliferation, metabolism, and vascular

remodeling[205]; *NOS3* (endothelial nitric oxide synthase), which synthesizes nitric oxide, a vasodilator and regulator of smooth muscle relaxation and leukocyte adhesion[206]; *IL-6*, a cytokine linked to vascular inflammation and endothelial dysfunction[207]; and *Flt-1/sFlt-1*, which bind VEGF/PIGF to either promote or inhibit angiogenesis, respectively. Collectively, these genes suggest that that fine-tuned regulation of local vascular tone, vascular growth and maintenance, and levels of pro-angiogenic and anti-angiogenic factors, along with the limitation of damaging inflammatory responses, all play important roles in controlling vascular smooth muscle and endothelial cell dynamics, and consequently, the various potential routes by which vascular dysfunction can occur and predispose the onset of preeclampsia. Taking the maternal and fetal preeclampsia-linked genes and the cell populations expressing them in aggregate highlights the complementary functional roles assumed by maternal and fetal cells and the intricate regulation of their functions that is likely necessary for facilitating proper placentation. Involved in this coordination are likely an array of interactions and mechanisms that upon dysregulation, can contribute to the complex pathogenesis of preeclampsia.

Additionally, in our examination of maternally expressed genes associated with abnormal fetal growth and birth weight, our analysis highlights the roles of decidual stromal cells as the main producers of components of the IGF axis (**Figure 4a**). We also show that trophoblasts, fibroblasts, and Hofbauer cell populations collectively express fetal IGFs and IGFBPs (**Figure 4b**). In the pregnancy context, IGFs, their binding proteins, and their receptors have been established to be pivotal to fetal development and growth. The IGF axis has been suggested to mediate nutrient availability and allocation to the fetoplacental unit, through regulation of maternal metabolism as well as various aspects of placentation[208]. IGF levels and signaling have been linked to stimulus of trophoblast growth and invasion[209] and to the regulation of placental substrate transport and hormone secretion[210,211]. Previously, only IGFBP1 was a well-documented stromal cell secretory product of relevance to IGF signaling[10,212], but our results suggest that stromal cells may in fact play a major role in mediating this important mechanism of fetal growth at the maternal-fetal interface. The elevated expression of IGFs and IGFBPs by both decidual and placental cells reinforces that IGFs potentially exert their effects through both autocrine and paracrine signaling or via both maternal and fetal circulation, highlighting the potential complexity of IGF regulation and how its modulation by multiple cell types indicates its ability to influence fetal growth through multiple modes of action.

Collectively, our results demonstrate the utility of leveraging scRNA-seq data to map disease or phenotype-associated genes identified through studies such as GWAS or gene expression profiling to specific cell types. They highlight the potential cellular functions and mechanisms of these genes, while clarifying in turn the consequential functions and processes of their associated cell types. One caveat of our approach, however, is that it is limited by the scope of gene curation and by the robustness of included studies. In the case of evaluating GWAS results, many implicated loci or risk variants fall within non-coding regions that cannot be conclusively mapped to candidate genes; thus, it may be difficult to further prioritize or otherwise study such variants based on the scRNA-seq data. Finally, in this current form of the analysis, we have not directly taken into account the fact that many gene and cellular functions are mediated through cell-cell interactions. This is particularly relevant, as our analyses strongly suggest that different cell types at the maternal-fetal interface collectively play overlapping and complementary roles in driving disease pathology or the development of specific phenotypes. A method for inferring cell-cell interactions from scRNA-seq data via the examination of ligand-receptor pairs expressed by cell populations has been recently developed (www.CellPhoneDB.org) and has been applied to further profile specific cell populations and predict communication among cell types in the decidua and placenta (Vento-Tormo R, Efremova E et al., unpublished). Intersection of the genes linked to pregnancy-related diseases or phenotypes with the ligand-receptor interactions predicted by CellPhoneDB would introduce a new layer of understanding for how cellular crosstalk at the maternal-fetal interface, rather than isolated cell types, ultimately contributes to disease processes.

Overall, this analysis facilitated an increased understanding of the development of pregnancy- and fertility-related conditions at the highly informative resolution of specific cell populations or cell states rather than merely at the tissue level. It also enabled the preliminary correlation of gene functions to the cellular functions differentiating individual cell populations, and in turn, the mapping of cell types to specific conditions. Ultimately, such analyses could aid in the development of more targeted, mechanistically-informed therapies and improved strategies for managing infertility or complications of pregnancy. These paired analyses of disease-associated genes and scRNA-seq data could be similarly applied to studying other tissues and their associated pathologies.