bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Title Page

Title: Placental gene expression-based cell type deconvolution: Cell proportions drive preeclampsia gene expression differences

Authors: Kyle A Campbell[1], Justin A Colacino[2,3], Muraly Puttabyatappa[4], John F Dou[1], Elana R Elkin[2], Saher S Hammoud[5,6], Steven E Domino[6], Dana C Dolinoy[2, 3], Jaclyn M Goodrich[2], Rita Loch-Caruso[2], Vasantha Padmanabhan[2,3,4,6], Kelly M Bakulski[1]*

Affiliations:
[1]Epidemiology, School of Public Health, University of Michigan
[2]Environmental Health Sciences, School of Public Health, University of Michigan
[3]Nutritional Sciences, School of Public Health, University of Michigan
[4]Pediatrics, Michigan Medicine, University of Michigan
[5]Human Genetics, Michigan Medicine, University of Michigan
[6]Obstetrics and Gynecology, Michigan Medicine, University of Michigan
[7]Department of Urology, Michigan Medicine, University of Michigan

Abstract

The placenta mediates adverse pregnancy outcomes, including preeclampsia, which is characterized by gestational hypertension and proteinuria. Placental cell type heterogeneity in preeclampsia is not well-understood and limits mechanistic interpretation of bulk gene expression measures. We generated single-cell RNA-sequencing samples for integration with existing data to create the largest deconvolution reference of 19 fetal and 8 maternal cell types from placental villous tissue at term (n=15,532 cells). We deconvoluted eight published microarray case-control studies of preeclampsia (n=330). Deconvolution revealed excess extravillous trophoblasts and fewer mesenchymal cells. Adjustment for cellular composition reduced preeclampsia-associated differentially expressed genes (FDR<0.05) from 1,224 to 0, whereas pathway alterations exhibiting a metabolic adaptation to hypoxia were robust to cell type adjustment. Cellular composition explained 35.1% of the association between preeclampsia and *FLT1* overexpression. Our findings indicate substantial placental cellular heterogeneity in preeclampsia that predicts previously observed bulk gene expression differences. Our deconvolution reference lays the groundwork for cellular heterogeneity-aware investigation into placental dysfunction and adverse birth outcomes.

Keywords: Placenta, preeclampsia, extravillous trophoblast, FLT1, single-cell RNA sequencing

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Declarations

Acknowledgements: This research was supported by the Michigan Lifestage Environmental Exposures and Disease center (P30 ES017885), Michigan State University's Environmental Influences on Child Health Outcomes program (UG3 OD023285, UH3 OD023285), the University of Michigan M-cubed pilot grant program, and the Puerto Rico Testsite for Exploring Contamination Threats (P42 ES017198). K.A.C. was supported by the National Institutes of Health (T32 HG00040). J.A.C. was supported by the Ravitz Family Foundation, the Forbes Institute for Cancer Discovery at the University of Michigan Rogel Cancer Center, and the National Institutes of Health (R01 ES028802, U01 ES026697, and P30 CA046592). M.P. was supported by the National Institutes of Health (T32 ES007062). E.R.E was supported by the National Institutes of Health (T32 DK071212). K.M.B. was supported by research grants from the National Institute of Environmental Health Sciences (R01 ES025531; R01 ES025574). We also acknowledge the submitting investigators of previously published placental single-cell RNA sequencing data (dbGaP accession number phs001886.v1.p1), whose work was sponsored in part by the Perinatology Research Branch, Division of Obstetrics and Maternal-Fetal Medicine, Division of Intramural Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, U.S. Department of Health and Human Services.

Author contributions: Kyle Campbell: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing – Original Draft, Visualization **Justin A. Colacino:** Conceptualization, Methodology, Software, Validation, Investigation, Resources, Writing – Review & Editing, Visualization, Supervision, Project administration, Funding acquisition **Muraly Puttabyatappa:** Investigation, Writing – Review & Editing, Project administration **John Dou:** Software, Validation, Data Curation **Elana R. Elkin:** Writing – Review & Editing **Saher Sue Hammoud:** Methodology, Writing – Review & Editing, Funding acquisition **Steven E. Domino:** Methodology, Resources, Writing – Review & Editing, Project administration, Funding acquisition **Dana C. Dolinoy:** Writing – Review & Editing, Funding acquisition **Jaclyn M. Goodrich:** Writing – Review & Editing, Funding acquisition **Rita Loch-Caruso:** Conceptualization, Methodology, Resources, Writing – Review & Editing, Supervision, Project administration, Funding acquisition **Vasantha Padmanabhan:** Conceptualization, Resources, Writing – Review & Editing, Project administration, Funding acquisition **Kelly M. Bakulski:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Visualization, Supervision, Project administration, Funding acquisition

Competing interests: The authors declare no conflicts of interest/competing interests in the production of this work.

Materials and correspondence: Kelly M. Bakulski (bakulski@umich.edu)

Human Subject Data: This study did not collect data on human subjects. Study protocols for tissue collection were approved by the University of Michigan Institutional Review Board (HUM00017941, HUM00102038).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction

The public health burden of adverse pregnancy outcomes is substantial. An important example is preeclampsia, which affected 6.5% of all pregnant people in the United States in 2017 and is characterized by high maternal blood pressure and damage to other organ systems. Adverse pregnancy outcomes may lead to myriad health complications including elevated risk of chronic diseases throughout the life course [1]. The placenta, a temporary organ that develops early in pregnancy, promotes maternal uterine artery remodeling; mediates transport of oxygen, nutrients, and waste [2]; secretes hormones to regulate pregnancy; metabolizes various macromolecules and xenobiotics; and can serve as a selective barrier to some, but not all, pathogens and xenobiotics [3]. The executive summary of the *Placental Origins of Adverse Pregnancy Outcomes: Potential Molecular Target*s workshop recently concluded that most adverse pregnancy outcomes are rooted in placental dysfunction [4]. Despite this, the molecular underpinnings of placental dysfunction are poorly understood.

Placenta-specific cell types, including cytotrophoblasts, syncytiotrophoblasts, extravillous trophoblasts, and placental resident macrophage Hofbauer cells are all essential for placental development, structure, and function [5]. Dysfunction of these specific cell types likely plays a role in placental pathogenesis. For example, extravillous trophoblasts are responsible for invading into the maternal decidua early in pregnancy to remodel uterine arteries and increase blood flow to the placenta [2]. Inadequate or inappropriate invasion of extravillous trophoblasts has previously been implicated in preeclampsia etiology [6–8]. Despite some knowledge of the roles of specific placental cell types in the development of preeclampsia, relatively little is known about how individual cell types contribute to placental dysfunction.

Existing research models used to investigate the function and dysfunction of individual cell types are limited. Protocols to isolate primary placental cells for experimental research are restricted to one or few cell types [9–14]. Cell type-specific assays are costly and require special techniques or training resulting in small sample sizes and have not yet been scalable to large epidemiological studies [15–17]. Furthermore, placental cell lines such as BeWo, derived from choriocarcinoma [18], and HTR-8/SVneo, immortalized by SV40 [19], are typically derived by processes that alter the DNA of the cells, limiting their in vivo translatability.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Consequently, the characteristics of even healthy placental cell type function and especially their connections to adverse outcomes such as preeclampsia are incompletely understood.

Measures of gene expression in bulk placental tissue are used to better understand the biological mechanisms underlying adverse pregnancy outcomes [20–22] and are common in epidemiological studies [23]. Gene expression profiles differ systematically by cell type [24, 25]. Thus, bulk placental tissue-level gene expression measurements represent a convolution of gene expression signals from individual cells and cell types [26, 27]. Deconvolution refers to the bioinformatic process of estimating the distribution of cell types that constitute the tissue [28, 29]. Deconvoluting tissue-level gene expression profiles is essential to eliminate effects introduced by unmodeled cell type proportions [30] by disentangling shifts in cell type proportions from direct changes to cellular gene expression [31]. Reference-based deconvolution boasts biologically interpretable cell type proportion estimates with few modeling assumptions but relies on independently collected cell type-specific gene expression profiles as inputs [31]. Prior placental cell type-specific gene expression measures from term villous tissue [16] had a limited number of biological replicates and included neither technical replicates nor benchmarking against physically isolated placental cell types. A robust, accessible, and publicly available gene expression deconvolution reference is currently unavailable for healthy placental villous tissue.

To advance the field of perinatal molecular epidemiology, our goal was to develop an accessible and robust gene expression deconvolution reference for healthy placental villous tissue at term. We generated single-cell RNA-sequencing data with technical replicates for integration with existing cell type-specific placental gene expression data [16]. Additionally, we benchmarked these single-cell cell type-specific gene expression profiles against placental cell types isolated with more conventional fluorescence-activated cell sorting followed by RNA-sequencing. Finally, to apply our deconvolution approach and assess links between preeclampsia and placental cell types and their proportions, we applied our placenta cell type gene expression reference to deconvolute bulk placental tissues in a secondary data analysis of a case-control study [32] of preeclampsia.

Results

Single-cell gene expression map of healthy placental villous tissue

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

From healthy term placental villous tissue, 9,942 cells across a total of two biological replicates and two technical replicates were sequenced and analyzed. These data were combined with single-cell RNA-sequencing data of 6,313 cells from three healthy term villous tissue samples in a previously published study [16] (**Supplementary Table 1**). Cells were excluded if they were doublets or outliers in total RNA content, number of genes detected, or mitochondrial gene expression (**Supplementary Figures 1-2**). Fetal or maternal origin was determined by genetic variation in sequencing data. Fetal sex was determined by *XIST* expression (**Supplementary Figure 3**). The final analytic sample included 15,532 cells and 36,601 genes across five biological replicates, two of which had a technical replicate.

Uniform manifold approximation and projection [33] was used to visualize sequencing results in two dimensions (**Figure 1A**). Cells clustered into 19 fetal and 8 maternal cell types with 66.5% of all cells of fetal origin (**Table 1**). Observed placenta-specific trophoblast cell types included cytotrophoblasts, proliferative cytotrophoblasts, extravillous trophoblasts, and syncytiotrophoblasts. Proliferative cytotrophoblasts were distinguished by overexpression of genes related to the mitotic cell cycle ($p_{adj}=7.1\times10^{-64}$) (**Supplementary Figure 4**). Other fetal-specific cell types included mesenchymal stem cells, fibroblasts, endothelial cells, and Hofbauer cells (**Figure 1B**). Fibroblasts were distinguished from mesenchymal stem cells by their overexpression of type 1 collagen genes, e.g., *COL1A1* ($\log_2$-Fold-Change ($\log_2$FC) = 3.0, $p_{adj}=2.48\times10^{-16}$). Lymphocytes, B cells, and monocytes were generally of mixed maternal and fetal origin (**Figure 1B-C**).

Cell type-defining transcripts were identified by comparing the expression of a transcript in one cell type against that gene's average expression across all other cell types (**Supplementary Table 2**). *FLT1* expression was upregulated in extravillous trophoblasts ($\log_2$FC=3.28), cytotrophoblasts ($\log_2$FC=0.73), and proliferative cytotrophoblasts ($\log_2$FC=0.27). Trophoblast cell types had the largest and most diverse transcriptomes, characterized by the largest number of RNA molecules and detected genes per cell (**Supplementary Figure 5**). Cell-type defining biological processes were highlighted with functional enrichment analysis (**Supplementary Table 3**). For example, syncytiotrophoblasts were enriched for transcripts involved in hormone biosynthesis ($p_{adj}<0.001$) and growth hormone receptor signaling ($p_{adj}=0.003$). Technical replication in samples 1 and 2 appeared high in Uniform Manifold Approximation and Projection (UMAP) space (**Supplementary Figure 6A-B**). Indeed, the average intra-cluster gene expression between technical

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

replicates had an average Pearson correlation (mean ± standard deviation) of 0.94 ± 0.14 for sample 1 and 0.88 ± 0.20 for sample 2 (p-values<0.001).

Fluorescence-activated cell sorting of major placental cell types

We isolated bulk placental villous tissue, enriched syncytiotrophoblasts, and sorted five cell types (Hofbauer cells, endothelial cells, fibroblasts, leukocytes, extravillous trophoblasts, and cytotrophoblasts) via fluorescence-activated cell sorting (FACS) from four healthy term, uncomplicated Cesarean sections for bulk RNA sequencing, labelled samples 1 (same sample source as single-cell RNA sequencing sample 1), 6, 7, and 8 (**Supplementary Table 4**). Principal components analysis of sorted-cell bulk RNA sequencing revealed three loosely defined clusters along principal component (PC) 1: (i) composite tissue and syncytiotrophoblast, (ii) Hofbauer and fibroblast, and (iii) cytotrophoblast, extravillous trophoblast, and endothelial. Leukocyte samples were scattered along PC1 (**Supplementary Figure 8**). Comparison of the expression of a transcript in one cell type against that gene's average expression in all other cell types identified cell type-defining transcripts (**Figure 2**). 37,929 genes were tested. 746 genes were dropped from the syncytiotrophoblast contrast due to excessively low counts, low variability, or extreme outlier status. Large-scale gene expression differences were observed for each cell type (**Supplementary Table 5**). Functional analysis of cell type-defining transcripts revealed cell type-defining biological processes (**Supplementary Table 6**). For example, syncytiotrophoblasts were enriched for transcripts relevant to endothelium development, mesenchyme development, and vasculogenesis (p_{adj}<0.001). Normalized counts between sorted and single-cell RNA-sequencing cell types had an average Spearman correlation of 0.73 ± 0.09 (p<0.001) (**Supplementary Figure 9**).

Cell proportion deconvolution of bulk placental tissue dataset

Based on the single cell data, we created a placental signature gene matrix that incorporated an average of 10,044 differentially expressed genes for each of the 27 fetal and maternal cell types (**Supplementary Figure 10**). We applied this signature matrix to estimate cell proportions from bulk placental tissue in 157 preeclampsia cases and 173 controls in a dataset [32] compiled from eight previously published studies [32, 34–40]. Gestational age was 2.2 weeks younger in cases than controls (p-value<0.001, **Table 2**). All deconvoluted samples exhibited high goodness-of-fit between original bulk mixtures and the estimated cell

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

type proportion mixtures (p-values<0.001). Among the signature genes, original bulk and estimated mixtures had a Pearson correlation of 0.71 $\pm$ 0.03 and root mean square error of 0.74 $\pm$ 0.03 (**Supplementary Table 7**). Fetal naïve CD4+ T cells and fetal B cells were estimated to be at 0% abundance in all samples and were dropped from downstream analyses. Cytotrophoblasts were the most abundant estimated fetal cell type (29% $\pm$ 4%) followed by syncytiotrophoblasts (21% $\pm$ 4%) and naïve CD8+ T cells (9% $\pm$ 2%). The most common maternal cell types were naïve CD8+ T cells (3% $\pm$ 2%), natural killer cells (3% $\pm$ 2%), and plasma cells (3% $\pm$ 2%).

Differentially abundant cell type proportions in preeclampsia cases versus controls

To test for differences in cell proportions between preeclampsia cases and controls (**Supplementary Figure 11**), we fit beta regression models adjusted for each cell type proportion and study source, fetal sex, and gestational age. Fetal memory CD4+ T cells (p=0.03) and extravillous trophoblasts (p<0.001) were more abundant (**Figure 3**) in preeclampsia cases relative to controls. The unadjusted median extravillous trophoblast abundance was 5.3% among cases compared to 2.6% among controls. Mesenchymal stem cells (median percent composition in cases vs. controls, 4.9% vs. 6.2%), Hofbauer cells (6.4% vs. 7.9%), and fibroblasts (6.1% vs. 6.7%) were all less abundant among preeclampsia cases compared to controls (p<0.001). Among maternal cell types, natural killer cells (1.6% vs. 1.4%) were more abundant among preeclampsia cases compared to controls (p=0.002).

Differential expression between preeclampsia cases and controls attenuated by cell type proportion adjustment

To test whether microarray gene expression differences between preeclampsia cases and controls are partly driven by differences in cell type abundances, we fit differential gene expression models adjusted for covariates study source, fetal sex, and gestational age with and without adjustment for imputed cell type proportions. To reduce the number of model covariates and account for dependence between deconvoluted cell type proportions, we applied principal components analysis to the imputed cell type proportions. The first five PCs accounted for 77.8% of the variance and were added as additional covariates to form the cell type-adjusted model. Variation in PCs 1 and 2 was largely driven by syncytiotrophoblasts (53.3%) and cytotrophoblast (21.5%) proportions and did not readily separate cases from controls (**Supplementary**

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figures 12A, C). Variation in PC3 was largely driven by extravillous trophoblast proportions (69.1%) and to a lesser extent Hofbauer cells (13.5%) and mesenchymal stems cells (6.1%). PC3 provided some separation between cases and controls (**Supplementary Figures 12B, D**).

In the cell type-naïve base model, 1,224 genes were differentially expressed in preeclampsia cases versus controls (**Figure 4A**). Gene set enrichment analysis identified 81 overrepresented pathways in the base model (**Figure 5A**). Biological process pathways such as cellular respiration (q-value<0.001), translational termination (q<0.001), and related pathways were downregulated whereas cell type differentiation pathways such as cornification (q<0.001) and endothelial cell development (q=0.04) were upregulated. Remarkably, when the base model was additionally adjusted for the first five PCs of imputed cell type proportions, there were zero differentially expressed genes between preeclampsia cases and controls (**Figure 4B**). Of the cell type-adjusted results, 40 pathways were overrepresented (**Figure 5B**). Downregulation of translation termination (q<0.001), cellular respiration (q<0.001), and related pathways and upregulation of glycolytic process through fructose 6-phosphate (q=0.01) were robust to cell type proportion adjustment.

Differential expression of preeclampsia-associated gene *FLT1* mediated by placental cell type proportions

Overexpression of *FLT1* in placental tissue [41–44], detection of a soluble isoform of *FLT1* in maternal circulation [45, 46], and fetal genetic variants near *FLT1* [47] have implicated *FLT1* in preeclampsia etiology. Because we observed cell type-specific expression patterns of *FLT1* in trophoblasts, particularly in extravillous trophoblasts, we hypothesized that the observed attenuation of *FLT1* differential expression may be due in part to the differences in cell type proportions observed between preeclampsia cases and controls. To test this hypothesis, we applied a unified mediation and interaction analysis to quantify the proportion of *FLT1* expression differences mediated by deconvoluted cell type proportions. We did not observe an interaction between preeclampsia status and placental cell composition (average mediated effect percent difference = -3.8%, 95% CI [-23.4%, 15.7%]). 35.1% (95% CI [25.5%, 46.3%]) of the association between preeclampsia and *FLT1* expression was attributable to differences in placental cell composition between preeclampsia cases and controls (**Figure 6**).

Discussion

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

To create the largest, publicly available deconvolution reference of 19 fetal and 8 maternal cell type-specific gene expression profiles, we newly sequenced placental villous cells, integrated those results with data from a previously published study, and built a signature gene matrix for deconvolution. To assess reproducibility, we assayed single-cell gene expression profiles for two term placentas in technical replicates following cryopreservation of dissociated villous tissue. To validate single-cell placental cell type expression profiles, we created a novel fluorescence-activated cell sorting scheme to enrich and sequence RNA from five important placental cell types as well as syncytiotrophoblasts. We applied our results to deconvolute cell type proportions in a previously published epidemiologic microarray study of the pregnancy complication preeclampsia, revealing placental cell type proportion differences between preeclampsia cases and controls at term. We then showed that large gene expression differences between preeclampsia cases and controls (n=1,224 genes, $p_{adj}<0.05$) were markedly attenuated after adjustment for cell type proportions (n=0 genes, $p_{adj}<0.05$). Preeclampsia associated pathways that were robust to cell type proportion adjustment included downregulation of replication-related and cellular respiration pathways and upregulation of glycolysis through fructose 6-phosphate. Finally, to quantify the attenuation of differential expression of the preeclampsia biomarker *FLT1*, we applied mediation analysis to show that approximately 35% of the association between increased *FLT1* expression and preeclampsia was attributable to placental cell composition. Cell type proportions may be an important factor in gene expression differences in placental tissue studies.

By integrating our new single-cell RNA-sequencing results with those from a previously published study, our integrated dataset, to our knowledge, is the largest for healthy, term placental villous tissue to date. We document term cell type-specific gene expression patterns for well-characterized placental cell types, including syncytiotrophoblasts [9], cytotrophoblasts [12], and extravillous trophoblasts [13]. In addition, we provide gene expression markers for relatively understudied placental cell types such as endothelial cells, mesenchymal stem cells, and Hofbauer cells as well as maternal peripheral mononuclear cells recovered from the maternal-fetal interface. Compared to the previous analysis of the published samples [16] which relied on predominately sex-specific gene expression markers to differentiate proliferative from non-proliferative cytotrophoblasts, we show that functional enrichment analysis revealed broad upregulation of proliferation pathways in proliferative cytotrophoblasts. We replicated the findings of a previous study [9] that size exclusion

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

filtration can enrich syncytiotrophoblasts from gently digested villous tissue based on comparison to single-cell results. We developed a novel fluorescence-activated cell sorting scheme to simultaneously isolate viable cells from five major placental cell types. This protocol can be used to simultaneously isolate placental cell types for functional assays or other experiments. The low representation of some cell types such as trophoblasts in our single-cell RNA-sequencing results suggest that these cell types may be sensitive to a conventional cryopreservation protocol, an extension of dissociation bias that has been previously documented in other tissues [48]. Future studies may propose alternative approaches to perform unbiased single-cell RNA in placental tissues.

Our preeclampsia findings are consistent with prior pathophysiological understanding of the disorder, and we provide linked cell type and gene expression data in bulk tissue for the first time. Among preeclampsia cases, we observed an elevated proportion of extravillous trophoblasts and underrepresentation of stromal cell types, which may reflect an arrest in placental cell type differentiation and maturation following insufficient uterine spiral artery remodeling implicated in preeclampsia [49–51]. In the cell type-naïve differential expression model, consistent with previous findings, placentas from pregnancies with preeclampsia overexpressed *FLT1*, *LEP*, and *ENG* [41–44]. In our cell type-adjusted model, *FLT1* and *LEP* remained only nominally significant whereas *ENG* was not differentially expressed. Mediation analysis confirmed that a significant proportion of *FLT1* overexpression was attributable to changes in the cellular composition of the placenta. These results suggest that placental cell type proportion differences may be an overlooked factor in explaining the well-documented association between preeclampsia and *FLT1* expression [41–44]. Downregulation of replication-related and cellular respiration pathways and upregulation of glycolysis through fructose 6-phosphate was robust to cell type adjustment, suggesting intracellular changes to these pathways. Together, these enrichment results suggest a metabolic adaption to hypoxia. Placental hypoxia is characteristic of preeclampsia [52, 53]. Because oxygen tension is a critical factor in trophoblast differentiation, inappropriate oxygenation may partially explain the elevated proportion of extravillous trophoblasts, though regulators of this process such as *HIF1A* and *TGFB3* were not differentially expressed at the tissue level [54]. A recent single-cell RNA-sequencing case-control study of preeclampsia, however, identified upregulation of *TGFB1* in extravillous trophoblasts, potentially indicative of altered trophoblast differentiation or invasion [55,

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

56]. Consistent with our other findings, this study also observed a similar trend in cell type proportion differences and upregulation of *FLT1* in extravillous trophoblasts and *ENG* in syncytiotrophoblasts among between preeclampsia cases and controls [55]. Future work should consider and account for the cell type-specific expression patterns of genes that regulate placental development or are associated with preeclampsia to better understand preeclampsia etiology.

This study has several strengths. This is one of the first studies to profile the parenchymal healthy term villous tissue in the placenta and we integrate our dataset with samples from a previously published study to generate the largest cell type-specific placental villous tissue gene expression reference to date. Single-cell RNA-sequencing allows us to agnostically capture diverse placental cell types without *a priori* knowledge of cell types and their characteristics. To our knowledge, this is also the first study to demonstrate technical replication of single-cell RNA-sequencing in placental villous tissue. We verify our results with conventional RNA-sequencing of FACS-sorted placental cell types. We confirmed enrichment of syncytiotrophoblasts through size exclusion filtration is consistent with single-cell results. We also applied our findings to a large target deconvolution dataset of preeclampsia that contained placental measures from hundreds of participants across eight different studies. Most importantly, this is the first study of preeclampsia to account for cell type heterogeneity, a critical factor in bulk tissue assays, in an epidemiologic sample.

This study also has several limitations. Although our cellular sample size comprised of 15,532 cells is relatively large compared to previous single-cell RNA-sequencing studies of term placental villous tissue, this dataset still represents a limited biologic replicate sample size. Our newly sequenced samples came from a convenience sample without available demographic information beyond healthy Cesarean-section status. Similarly, the sample size of FACS-sorted tissues was limited, and some cell type fractions were excluded due to low RNA quality. This study did not include placental tissues for single cell analysis from preeclamptic patients to confirm intracellular gene expression changes. Though deconvolution fit statistics were highly significant, we had no gold standard to verify accurate deconvolution of cell counts. Future studies may verify whether cell type proportions estimated in diseased or vaginally delivered tissues are robust to a deconvolution reference generated from healthy villous tissue delivered via Cesarean-section. Residual confounding may remain in our statistical models due to the limited number of common covariates across all eight preeclampsia

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

case-control studies. Due to the nature of villous tissue sampling, our study design is cross-sectional, limiting our ability to establish temporality between exposure and outcome to rule out reverse causation. As with any study conditioned on live birth, selection bias may affect our results. However, the effects of harmful exposures that lead to selection tend to be underestimated in these scenarios [57, 58]. Therefore, our results likely represent a conservative underestimate of the effects of preeclampsia on inappropriate cell composition and preeclampsia status on *FLT1* expression.

In summary, we provide a cell type-specific deconvolution reference via single-cell RNA-sequencing in the parenchymal placental term villous tissue. We verified this reference by developing a novel FACS scheme to characterize six major placental cell types with RNA-sequencing. We applied this deconvolution reference to an epidemiologic preeclampsia dataset to reveal biologically relevant shifts in placental cell type proportions between preeclampsia cases and controls. Once cell type proportion differences were accounted for, differential gene expression differences were markedly attenuated between preeclampsia cases and controls. Enrichment analysis revealed downregulation of replication-related and cellular respiration pathways and upregulation of glycolysis through fructose 6-phosphate in preeclampsia. A substantial proportion of the overexpression of the *FLT1* in preeclampsia was mediated by placental cell composition. This is the first study to evaluate cell type proportion differences in an epidemiological study of placental parenchymal tissue and preeclampsia, or genome-wide gene expression differences adjusting for cell type heterogeneity. These results add to the growing body of literature that emphasizes the centrality of cell type heterogeneity in molecular measures of bulk tissues. We provide a publicly available placental cell type-specific gene expression reference for term placental villous tissue to overcome this critical limitation.

Methods

Placental tissue collection and dissociation

Placentas were collected shortly after delivery from healthy, full term, singleton uncomplicated Cesarean sections at the University of Michigan Von Voigtlander Women's Hospital. Villous placental tissue biopsies were collected and minced for dissociation after cutting away the basal and chorionic plates and scraping villous tissue from blood vessels [12]. We subjected approximately 1g minced dissected villous tissue to the Miltenyi Tumor Dissociation Kit on the GentleMACS Octo Dissociator with Heaters (Miltenyi Biotec) to

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

yield single-cell suspensions of viable placental cells in 5μM StemMACS™ Y27632 (Miltenyi Biotec) in RPMI 1640 (Gibco) according to manufacturer's instructions for "soft" tumor type. Red blood cells were depleted using RBC lysis buffer (Biolegend) according to manufacturer's protocol A. Single-cell suspensions were size-filtered at 100μm and subsequently 40μm. To collect a syncytiotrophoblast-enriched fraction, the fraction between 40μm and 100μm was washed from the 40μm strainers [9]. Single-cell suspensions <40μm were cryogenically stored in 5μM StemMACS™ Y27632 90% heat-inactivated fetal bovine serum (Gibco)/10% dimethyl sulfoxide (Invitrogen). For each placenta, additional whole villous tissue samples were stored in RNALater (Qiagen).

Placental single-cell RNA sequencing

Villous tissue single-cell suspensions were thawed and sorted via fluorescence-activated cell sorting with LIVE/DEAD Near-IR stain (Invitrogen) for viability and forward-scatter and side-scatter profiles to eliminate cellular debris and cell doublets. Viability- and size-sorted single-cell suspensions were submitted to the University of Michigan Advanced Genomics Core for single-cell RNA sequencing. Single cells were barcoded, and cDNA libraries constructed on the Chromium platform (10X Genomics, Single Cell 3' v2 chemistry). Asymmetric paired-end 110 base pair reads were sequenced on NovaSeq 6000 (Illumina).

Single-cell RNA-sequencing preprocessing

Raw reads were processed, deconvoluted, droplet filtered, and aligned at the gene level with the Cell Ranger pipeline using default settings (v4.0.0, 10X Genomics) based on the GRCh38 GENCODEv32/Ensembl 98 reference transcriptome with STAR v2.5.1b [59]. Previously published single-cell RNA-sequencing raw data of healthy, term placental villous tissue samples (dbGaP ID phs001886.v1.p1) SRR10166478, SRR10166481, and SRR10166484 downloaded through the NCBI Sequence Read Archive were processed identically [16]. The *freemuxlet* function in the latest version (accessed 2021/12/05) of the 'popscle' package was used to assign fetal or maternal origin and identify 120 mosaic doublets for removal based on single nucleotide polymorphisms with minor allele frequency greater than 10% from the 1000 Genomes Phase 3 reference panel (released 2013/05/02) [60]. Per cell quality control criteria were total RNA molecules, unique genes, and percentage of reads mapping to mitochondrial genes [61]. We excluded 1,398 low-quality outlier cells defined as cells that exceeded four median absolute deviations in samples 1 and 2 or three median absolute deviations

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

in samples 3, 4, and 5 using the *quickQCPerCell* function in 'scater' (R package, version 1.18.6) with default settings [62] (**Supplementary Figures 1-2**).

Single-cell RNA-sequencing clustering and cluster annotation

Maternal and fetal cells were split into separate datasets for clustering. Mutual nearest neighbor batch correction by biological replicate with default settings was used to identify cell type clusters and visualize clustering results via uniform manifold projection [33] using *FastMNN* from 'SeuratWrapper' (R package, version 0.3.0) [63]. Iterative clustering and sub-clustering with 'Seurat' (R package, version 4.0.1) function *FindClusters* at different resolution parameters were evaluated using cluster stability via clustering trees in 'clustree' [64, 65]. *A priori* canonical cell type marker gene expression patterns and cluster marker genes were used to assign cell types to cell clusters [16, 66–71]. Cells that fell outside cell type clusters and outlying in doublet density calculated with *computeDoubletDensity* were removed as putative doublets and doublet clusters were identified with *findDoubletClusters* for removal in 'scDblFinder' (R package, version 1.4.0) [72]. 723 maternal-maternal or fetal-fetal putative doublets were excluded after integration and clustering. Fetal sex in phs001886.v1.p1 samples was determined by annotation and confirmed with *XIST* expression. The final analytic sample had 15,532 cells and 36,601 genes across five biological and two technical replicates.

Single-cell RNA-sequencing differential expression and biological pathway enrichment statistical analysis

Technical correlation was assessed by Pearson correlation after averaging the normalized expression for each gene by cluster and by technical replicate. Cluster marker genes were identified in 'Seurat' with the *FindAllMarkers* function with default settings on uncorrected single-cell gene expression counts [61, 65]. Specifically, including both maternal and fetal cell types, the expression level in each cell type cluster was compared against the average expression of that gene across all other cell types using the two-tailed Wilcoxon Rank Sum test with significance defined at a false discovery rate-adjusted p-value less than 0.05. Pairwise cluster markers were identified in 'Seurat' with the *FindMarkers* function with an identical testing regime. Overexpressed genes were ranked by decreasing log-fold change for functional enrichment analysis with 'gprofiler2' (R package, version 0.2.0, database version e102_eg49_p15_7a9b4d6) using annotated genes as

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

the universe, excluding electronically generated annotations, with the g:SCS multiple testing correction method applying significance threshold 0.05 [73].

Fluorescence-activated cell sorting of major placental cell types from villous tissue

Villous tissue single-cell suspensions were quickly thawed and stained with 5 fluorescently labeled antibodies (CD9-FITC, CD45-APC, HLA-A,B,C-PE/Cy7, CD31-BV421, and HLA-G-PE) as well as the LIVE/DEAD Near-IR stain (Invitrogen) to isolate 6 viable populations of placental cells by fluorescence activated cell sorting at the University of Michigan Flow Cytometry Core Facility. Initial flow cytometry experiments included fluorescence minus one, single color compensation, and isotype controls. Isotype controls were found to be the most conservative and were consequently included in all sorting experiments, as well as single-color compensation controls due to the large number of colors used in sorting. The six populations of cells were Hofbauer cells, endothelial cells, fibroblasts, leukocytes, extravillous trophoblasts, and cytotrophoblasts We developed a novel five-marker cell surface fluorescence activated cell sorting (FACS) scheme to sort cytotrophoblasts (HLA A,B,C-), endothelial cells (CD31+), extravillous trophoblasts (HLA-G+), fibroblasts (CD9+), Hofbauer cells (CD9-), and leukocytes (CD45+/CD9+) from villous tissue (**Supplementary Figure 6**) [10, 11, 13, 74–81]. Syncytiotrophoblast fragments were enriched from villous tissue digests. We isolated cell type fractions and composite villous tissue from four healthy term, uncomplicated Cesarean sections, labelled samples 1 (same sample source as single-cell RNA-sequencing sample 1), 6, 7, and 8. We subjected 24 cell type fractions with sufficient RNA content to RNA-sequencing, including five composite, two cytotrophoblast, one endothelial, three extravillous trophoblast, three fibroblast, four Hofbauer cell, four leukocyte, and two syncytiotrophoblast fractions (**Supplementary Table 4**).

Detailed antibody information: FITC, marker CD9: Mouse IgG1-kappa, clone HI9a, Biolegend #312103, lot B188319, Biolegend #312104, lot B232916; isotype control: clone MOPC-21 Biolegend #400107, Lot B199152. APC, marker CD45: Mouse IgG1-kappa, clone 2D1, Biolegend #368511, Lot B215062; isotype control: clone MOPC-21, Biolegend #400121, lot B216780. PE/CY-7, marker HLA-ABC: Mouse IgG2a-kappa, clone W6/32, Biolegend #311429, lot B188649, Biolegend #3111430, lot B238602; isotype control: clone MOPC-173, Biolegend #400231, lot B209000. BV421, marker CD31: Mouse IgG1-kappa, clone WM59, Biolegend #303123, lot B204347, Biolegend #303124, lot B232010; isotype control: clone MOPC-21,

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Biolegend #400157, lot B225357. PE, marker HLA-G: Mouse IgG2a-kappa, clone 87G, Biolegend #335905, lot B222326, Biolegend #335906, lot B199294; isotype control clone MOPC-173, Biolegend #400211, lot B227641. Mouse IgG1-kappa, clone MEM-G/9, Abcam #24384 Lot GR3176304-1; isotype control: monoclonal, Abcam #ab81200, lot GR267131-1. Validation information available on manufacturer's website under the catalog ID for each antibody.

A cut-off of 0.1% events was used to set a series of gates. Cells were first gated on size and granularity (FSC-HxSSC-H) to eliminate debris (not shown), followed by doublet discrimination (FSC-HxFSC-W and SSC-HxSSC-W) (not shown). Ax750 was used to sort on viability (not shown). Extravillous trophoblasts were isolated based on Human Leukocyte Antigen-G (HLA-G) expression (**Supplementary Figure 7A**). Cytotrophoblasts are HLA-ABC negative (**Supplementary Figure 7B**). HLA-ABC positive cells were then subjected to a CD45/CD9 gate to isolate Hofbauer cells and a heterogeneous population of leukocytes (**Supplementary Figure 7C**). Finally, CD45-/CD9- population is sorted into the endothelial or fibroblast bins based on CD31 expression (**Supplementary Figure 7D**).

Bulk placental tissue and sorted placental cell type RNA extraction and sequencing

Approximately 2mg of bulk RNALater-stabilized (Qiagen) bulk villous tissue was added to 350µL 1% β-mercaptoethanol (Sigma-Aldrich) RLT Buffer Plus (Qiagen) to Lysing Matrix D vials (MP Biomedicals). Samples were disrupted and homogenized on the MP-24 FastPrep homogenizer (MP Biomedicals) at 6m/s, setting MP24x2 for 35s. For the homogenized bulk villous tissue, syncytiotrophoblast-enriched fraction, and sorted cell types, RNA extraction was completed according to manufacturer's instructions using the AllPrep DNA/RNA Mini Kit (Qiagen) and stored at -80°C. RNA samples were submitted to the University of Michigan Advanced Genomics Core for RNA sequencing. Ribosomal RNAs were depleted with RiboGone (Takara) and libraries were prepared with the SMARTer Stranded RNA-Seq v2 kit (Takara). Paired- or single-end 50 base pair reads were sequenced on the HiSeq platform (Illumina). Raw RNA reads were assessed for sequencing quality using 'FastQC' v0.11.5 [82] and 'MultiQC' v1.7 [83]. Reads were aligned to the GRCh38.p12/ GENCODEv28 reference transcriptome using 'STAR' v2.6.0c with default settings [59]. featureCounts from 'subread' v1.6.1 was used to quantify and summarize gene expression with default settings [84].

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Validation testing: Sorted placental cell type differential expression analysis and comparison to single-cell results

Cell type-defining transcripts were identified using the negative binomial linear model two-tailed Wald test in 'DESeq2' (R package, version 1.30.1) adjusted for biological replicate using default settings with contrasts comparing the expression of a gene in one cell type against the average expression across all other cell types at a false discovery rate-adjusted p-value less than 0.05 [85]. Differentially expressed genes for each contrast were descending-ranked by absolute value of the test statistic for gene set enrichment analysis in desktop version GSEA 4.1.0 with the GSEAPreranked tool with default settings [86, 87]. Sorted- and single-cell read counts were appropriately library-normalized and log-transformed. Diverse fetal and maternal immune cell types from the single-cell RNA-sequencing data were summed and collapsed to one leukocyte category, cytotrophoblast subtypes to one cytotrophoblast category, and mesenchymal stem cells and fibroblasts to one fibroblast category. Normalized sorted cell type gene expression was compared to normalized single-cell RNA-sequencing results using Pearson and Spearman correlation coefficients to validate single-cell expression results.

Application testing: Bulk placenta gene expression dataset and CIBERSORTx deconvolution

Bulk placental tissue microarray gene expression (previously batch-corrected and normalized) from eight preeclampsia case-control studies was downloaded from the NCBI Gene Expression Omnibus (accession number GSE75010) for deconvolution [32]. We used the online version of CIBERSORTx (cibersortx.standford.edu, accessed 2021-01-25) to create a signature gene expression matrix for deconvolution with the following parameters: differential expression q-value<0.01, no minimum gene expression cutoff, and a 2000 gene feature selection ceiling [88]. We used the signature matrix to estimate constituent cell type proportions in GSE75010 using CIBERSORTx with cross-platform S-mode batch correction and 100 permutations to evaluate imputation goodness-of-fit.

Application testing: Preeclampsia case-control differential cell type abundance, differential gene expression statistical analysis, and mediation analysis

To test for differences in estimated cell type proportions between preeclampsia cases and controls, estimated cell type proportions for GSE75010 were regressed on preeclampsia case-control status using beta

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

regression models adjusted for gestational age, sex, and study source [89]. Statistical significance was assessed using the two-tailed Wald test applying a nominal significance threshold of 0.05. Cell types imputed at zero percent abundance across all samples were excluded.

Differential expression analysis was conducted in *limma* [90] with default linear models adjusted for gestational age, fetal sex, and study source. A cell type-adjusted model was built on the base model additionally adjusted for the first five principal components of deconvoluted cell type proportions. Principal components analysis was performed with *prcomp* from 'stats-package' (R, version 4.0.5) without scaling and default settings. Statistical significance was assessed at false discovery rate-adjusted q-value<0.05. Differentially expressed genes were descending-ranked by absolute value of the test statistic for gene set enrichment analysis in desktop version GSEA 4.1.0 with the GSEAPreranked tool with default settings [86, 87].

A unified mediation and interaction analysis [91] was conducted in 'CMAverse' (R package, version 0.1.0) [92] via the g-formula approach [93] to estimate causal randomized-intervention analogues of natural direct and indirect effects [94] through direct counterfactual imputation. The model was operationalized with preeclampsia status as the binary exposure, normalized *FLT1* expression as the continuous outcome, and the first five principal components of deconvoluted cell type proportions as continuous mediators. Baseline covariates included fetal sex and study source. Continuous gestational age was included as a confounder of the mediator-outcome relationship affected by the exposure. Confidence intervals were bootstrapped with 1000 boots with otherwise default settings. Statistical tests were two-tailed and interpreted at a p-value significance threshold of 0.05.

Statistical Information

Technical replication measured by average intra-cluster gene expression between technical replicates was tested via the two-tailed Pearson correlation test within Samples 1 and 2 assessed across all 36,601 genes. The number of cells contributing expression data for each cell type are available in **Table 1**. Single-cell cluster marker genes were identified in 'Seurat' with the *FindAllMarkers* function with default settings on uncorrected single-cell gene expression counts [61, 65]. Specifically, including cells from both maternal and fetal cell types, the expression level in each cell type cluster was compared against the average expression of that gene across all other cell types using the two-tailed Wilcoxon Rank Sum test with significance defined at a

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

false discovery rate-adjusted p-value less than 0.05 (n=15,532 cells). Pairwise cluster markers were identified in 'Seurat' with the *FindMarkers* function with an identical testing regime (n=2,835 cells for proliferative vs. non-proliferative cytotrophoblasts). Overexpressed genes (498.37 $\pm$ 400.82 genes across 27 cell type clusters or 746 differentially expressed genes for proliferative vs. non-proliferative cytotrophoblasts) were ranked by decreasing log-fold change for functional enrichment analysis with 'gprofiler2' (R package, version 0.2.0, database version e102_eg49_p15_7a9b4d6) using annotated genes as the universe, excluding electronically generated annotations, with the g:SCS multiple testing correction method applying significance threshold 0.05 [73]. Overexpressed genes per cell type cluster are available in **Supplementary Table 2**.

Cell type-defining transcripts were identified using the negative binomial linear model two-tailed Wald test in 'DESeq2' (R package, version 1.30.1) adjusted for biological replicate using default settings with contrasts comparing the expression of a gene in one cell type against the average expression across all other cell types at a false discovery rate-adjusted p-value less than 0.05 [85] (n=19 cell type fraction samples with breakdown by cell type available in **Supplementary Table 4**). Overexpressed genes for each contrast (1,261 $\pm$ 866.83 genes across 7 cell types) were descending-ranked by absolute value of the test statistic for gene set enrichment analysis in desktop version GSEA 4.1.0 with the GSEAPreranked tool with default settings [86, 87]. Differentially expressed genes per cell type available in **Supplementary Table 5** and number of differentially expressed genes is summarized in **Figure 2**. Sorted- and single-cell read counts were appropriately library-normalized and log-transformed. Diverse fetal and maternal immune cell types from the single-cell RNA-sequencing data were summed and collapsed to one leukocyte category, cytotrophoblast subtypes to one cytotrophoblast category, and mesenchymal stem cells and fibroblasts to one fibroblast category. Normalized sorted cell type gene expression (n=19 cell type fraction samples) was compared to normalized single-cell RNA-sequencing (n=15,484 cells) results across 7 cell types using two-tailed Pearson and Spearman correlation coefficient tests to validate single-cell expression results across 27,490 common genes (**Supplementary Figure 9**). Statistical tests were interpreted at a p-value significance threshold of 0.05.

Bulk placental tissue microarray gene expression (previously batch-corrected and normalized) from eight preeclampsia case-control studies was downloaded from the NCBI Gene Expression Omnibus (GSE75010) for deconvolution (n=330) [32]. We used the online version of CIBERSORTx

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

(cibersortx.standford.edu, accessed 2021-01-25) to create a signature gene expression matrix for deconvolution with the following parameters: differential expression q-value<0.01, no minimum gene expression cutoff, and a 2000 gene feature selection ceiling [88]. We used the signature matrix to estimate constituent cell type proportions in GSE75010 using CIBERSORTx with cross-platform S-mode batch correction and 100 permutations to evaluate imputation goodness-of-fit.

To test for differences in estimated cell type proportions between preeclampsia cases and controls (n=330), estimated cell type proportions for GSE75010 were regressed on preeclampsia case-control status using beta regression models (n=25 cell type proportion outcomes) adjusted for gestational age, sex, and study source [89]. Cell types imputed at zero percent abundance across all samples were excluded (n=2 excluded). Statistical significance was assessed using the two-tailed Wald test applying a nominal significance threshold of 0.05.

Differential expression analysis was conducted in *limma* [90] with default linear models adjusted for gestational age, fetal sex, and study source (n=330). A cell type-adjusted model was built on the base model additionally adjusted for the first five principal components of deconvoluted cell type proportions. Principal components analysis was performed with *prcomp* from 'stats-package' (R, version 4.0.5) without scaling and default settings. Statistical significance was assessed at false discovery rate-adjusted q-value<0.05. Differentially expressed genes were descending-ranked by absolute value of the test statistic for gene set enrichment analysis in desktop version GSEA 4.1.0 with the GSEAPreranked tool with default settings [86, 87].

A unified mediation and interaction analysis [91] was conducted in 'CMAverse' (R package, version 0.1.0) [92] via the g-formula approach [93] to estimate causal randomized-intervention analogues of natural direct and indirect effects [94] through direct counterfactual imputation. The model (n=330) was operationalized with preeclampsia status as the binary exposure, normalized *FLT1* expression as the outcome, and the first five principal components of deconvoluted cell type proportions as continuous mediators. Baseline covariates included fetal sex and categorical study source. Continuous gestational age was included as a confounder of the mediator-outcome relationship affected by the exposure. Confidence intervals were bootstrapped with 1000 boots with otherwise default settings. Statistical tests were two-tailed and interpreted at a p-value significance threshold of 0.05.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Data Availability

The cell type signature matrix and related files to deconvolute bulk gene expression measures are available through Github (https://github.com/bakulskilab). Raw placental single-cell RNA-sequencing and raw placental bulk RNA-sequencing generated by this study are available in the Gene Expression Omnibus repository (accession number GSE182381). The placental single-cell RNA-sequencing data that support the findings of this study are available in the Database of Genotypes and Phenotypes (accession number phs001886.v1.p1) [16]. The preeclampsia case-control microarray data that support the findings of this study are available in Gene Expression Omnibus repository (accession number GSE75010) [32].

Code Availability

All scripts to perform preprocessing, analyses, and deconvolution is available (https://github.com/bakulskilab).

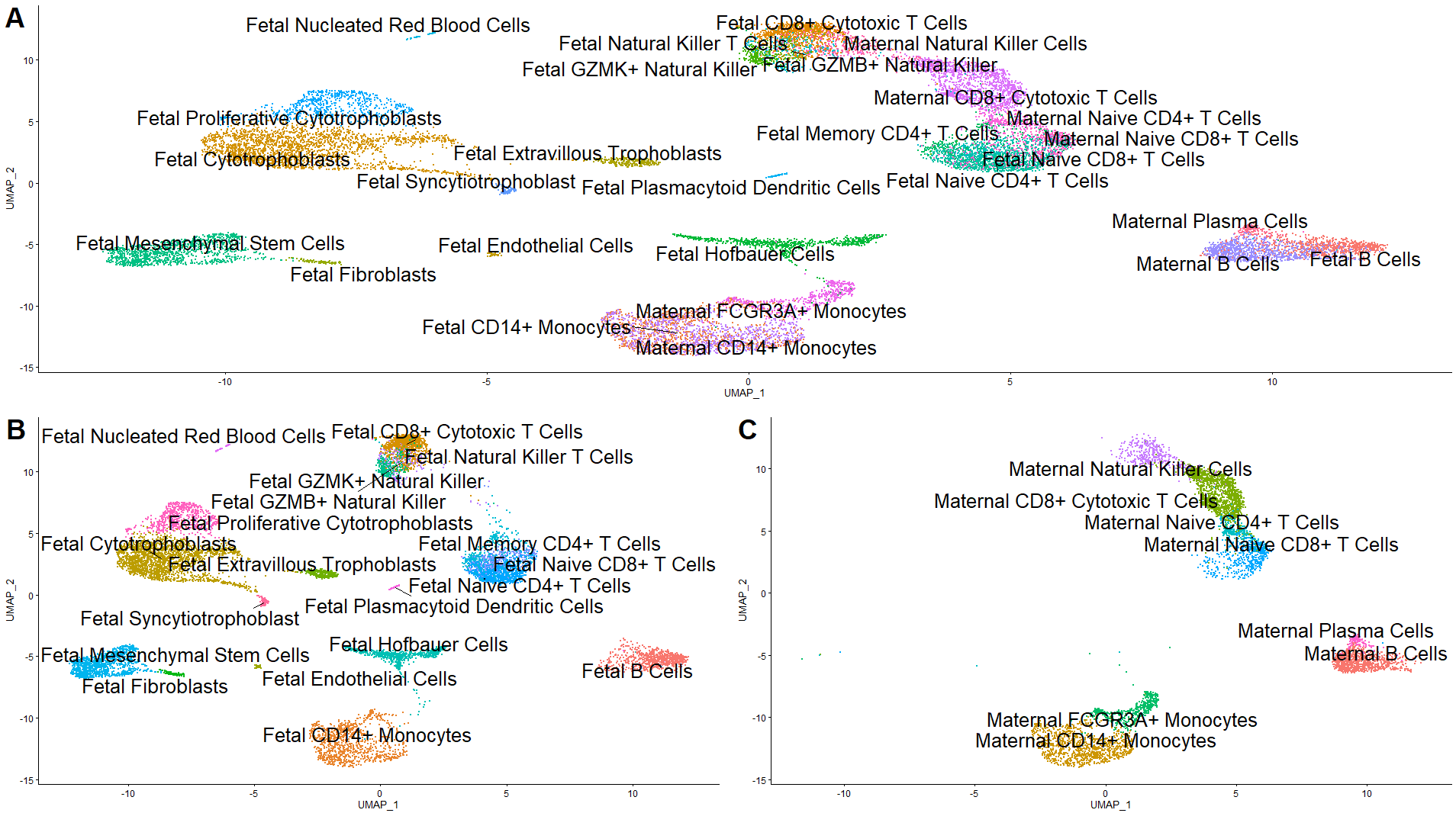

Figure 1. (A) Uniform Manifold Approximation and Projection (UMAP) plot of all cells, with each cell colored by cell type cluster. (B) UMAP plot of fetal cells only, with each cell colored by cell type cluster. (C) UMAP plot of maternal cells only, with each cell colored by cell type cluster.

Table 1. Number of cells captured by single-cell RNA sequencing in the final analytic dataset for each cell type by sample source. Overall cell composition by cell count provided for each cell type.

Cell Type	1A	1B	2A	2B	3	4	5	Fetal (%)	Maternal (%)
Fetal B Cells	111	152	199	233	39	28	16	7.5%	-
Fetal CD14+ Monocytes	240	196	186	204	52	28	43	9.2%	-
Fetal CD8+ Cytotoxic T Cells	220	210	148	163	79	76	6	8.7%	-
Fetal Cytotrophoblasts	4	9	1	1	378	921	891	21.4%	-
Fetal Endothelial Cells	5	11	23	15	1	5	6	0.6%	-
Fetal Extravillous Trophoblasts	0	0	0	0	11	229	6	2.4%	-
Fetal Fibroblasts	1	1	0	3	26	17	50	0.9%	-
Fetal GZMB+ Natural Killer	10	7	7	7	48	16	7	1.0%	-
Fetal GZMK+ Natural Killer	7	13	8	11	107	90	10	2.4%	-
Fetal Hofbauer Cells	26	26	112	134	117	121	277	7.9%	-
Fetal Memory CD4+ T Cells	26	42	43	47	70	54	29	3.0%	-
Fetal Mesenchymal Stem Cells	54	47	56	65	387	13	512	11.0%	-
Fetal Naive CD4+ T Cells	225	216	227	217	30	56	25	9.6%	-
Fetal Naive CD8+ T Cells	52	38	108	127	17	38	9	3.8%	-
Fetal Natural Killer T Cells	32	37	24	35	56	43	8	2.3%	-
Fetal Nucleated Red Blood Cells	1	0	1	0	25	6	15	0.5%	-
Fetal Plasmacytoid Dendritic Cells	5	4	2	7	24	18	8	0.7%	-
Fetal Proliferative Cytotrophoblasts	0	1	1	0	124	200	304	6.1%	-
Fetal Syncytiotrophoblasts	4	15	3	2	6	80	0	1.1%	-
Maternal B Cells	117	143	348	339	1	21	0	-	18.6%
Maternal CD14+ Monocytes	284	298	187	212	22	6	8	-	19.5%
Maternal CD8+ Cytotoxic T Cells	433	419	225	248	22	38	3	-	26.7%
Maternal FCGR3A+ Monocytes	120	137	50	69	36	11	84	-	9.7%
Maternal Naive CD4+ T Cells	51	36	47	58	9	19	3	-	4.3%
Maternal Naive CD8+ T Cells	69	94	178	168	6	32	3	-	10.6%
Maternal Natural Killer Cells	110	138	64	59	12	8	3	-	7.6%
Maternal Plasma Cells	36	37	30	46	0	7	2	-	3.0%

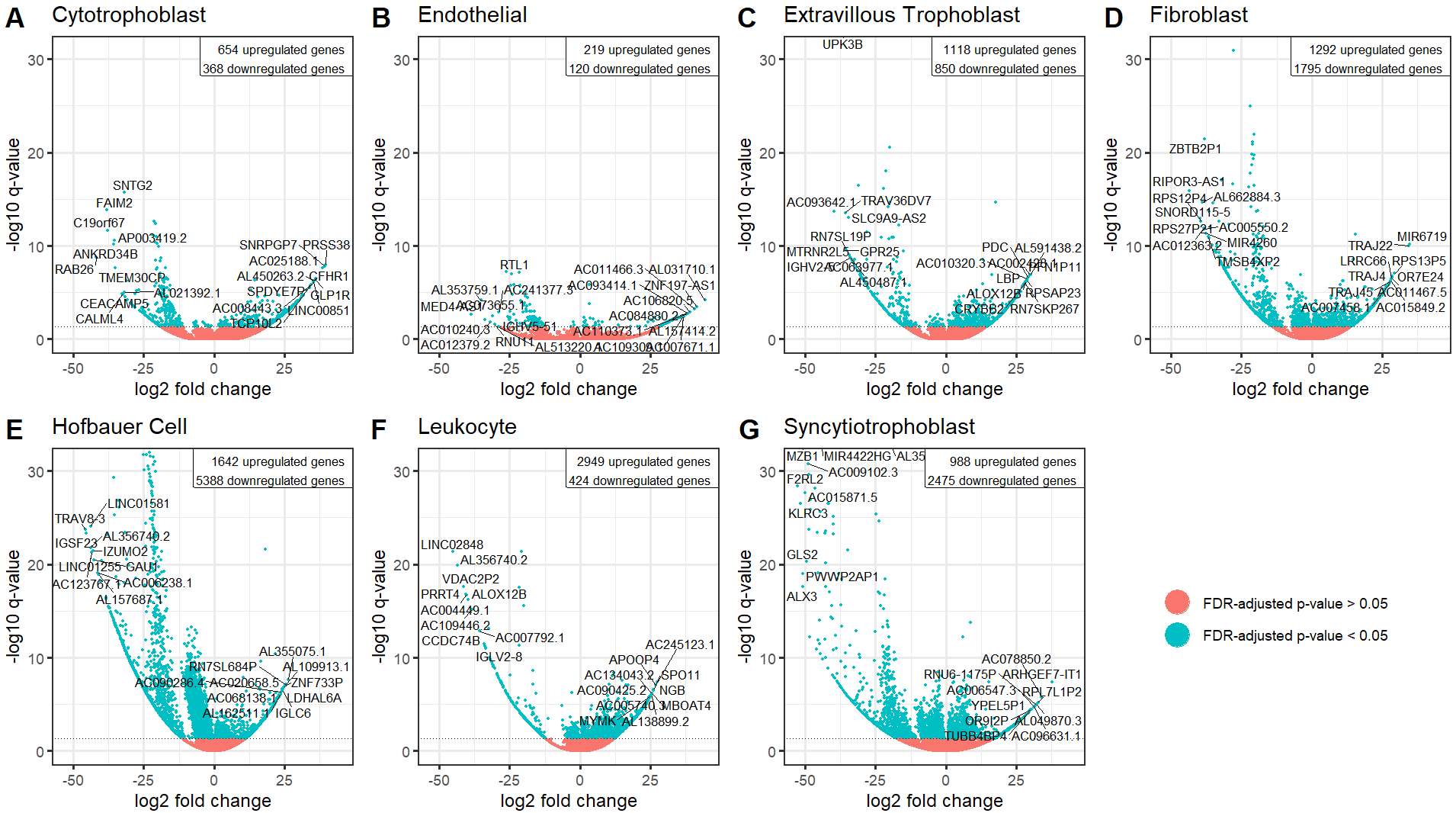

Figure 2. Volcano plots for fluorescence-activated-cell-sorted bulk RNA-seq differential expression in one cell type against average gene expression across other cell types. The y-axis encodes -log10 transformation of the false discovery-controlled q-value, with the cut-off for statistical significance at 0.05. The x-axis encodes log2 fold change of gene expression for the contrast of interest. The upper-right inset describes the number of differentially regulated genes per contrast. 37,929 genes were tested. 746 genes were dropped from the syncytiotrophoblast contrast by DESeq2's default automatic filtering algorithm due to excessively low counts, low variability, or extreme outlier status. (A) Cytotrophoblast. (B) Endothelial cell. (C) Extravillous trophoblast. (D) Fibroblast. (E) Hofbauer cell. (F) Leukocyte. (G) Syncytiotrophoblast.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 2. Demographic characteristics of eight previously published bulk microarray placental gene expression case-control studies (accessed through GSE75010) for deconvolution application testing.

	Control (N=173)	**Preeclampsia (N=157)**	**P-value**
Fetal Sex			
Female	78 (45.1%)	81 (51.6%)	0.28
Male	95 (54.9%)	76 (48.4%)	
Gestational Age (wks)			
Mean (SD)	35.2 (3.97)	33.0 (3.17)	<0.001
Median [Min, Max]	37.0 [25.0, 41.0]	33.0 [25.0, 39.0]	
Study			
GSE10588	26 (15.0%)	17 (10.8%)	0.39
GSE24129	8 (4.6%)	8 (5.1%)	
GSE25906	37 (21.4%)	23 (14.6%)	
GSE30186	6 (3.5%)	6 (3.8%)	
GSE43942	7 (4.0%)	5 (3.2%)	
GSE44711	8 (4.6%)	8 (5.1%)	
GSE4707	4 (2.3%)	10 (6.4%)	
GSE75010	77 (44.5%)	80 (51.0%)	

Bivariate batch with Kruskal-Wallis ANOVA (regular ANOVA homogeneity of variances violated) for continuous variables and Chi-square test for categorical outcomes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

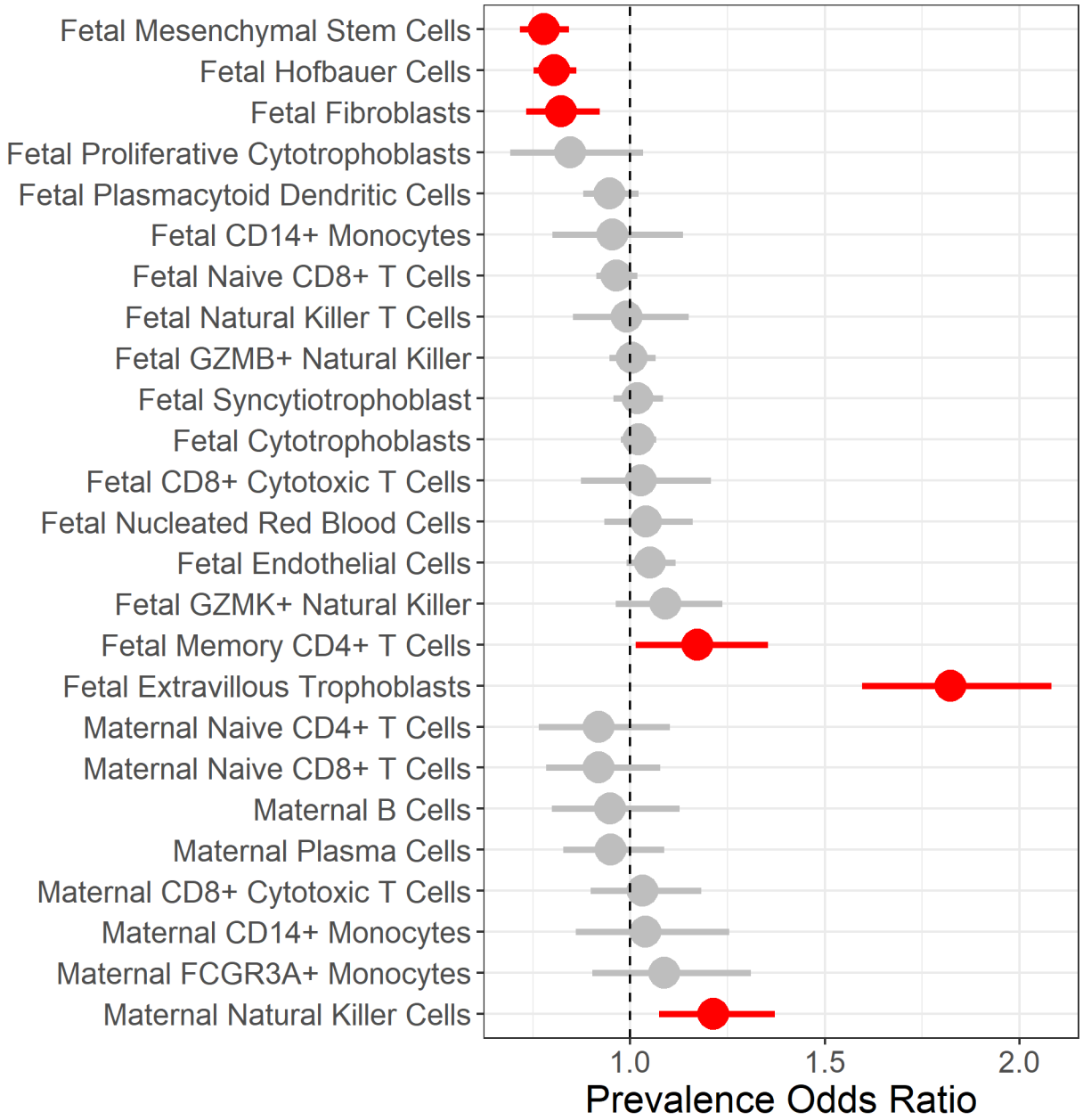

Figure 3. Forest plot of multivariate beta regression models' prevalence odds ratio adjusted for study source, gestational age, and fetal sex tested for a difference in each cell type's proportions in cases versus controls. Horizontal lines indicate the range of the 95% confidence interval.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

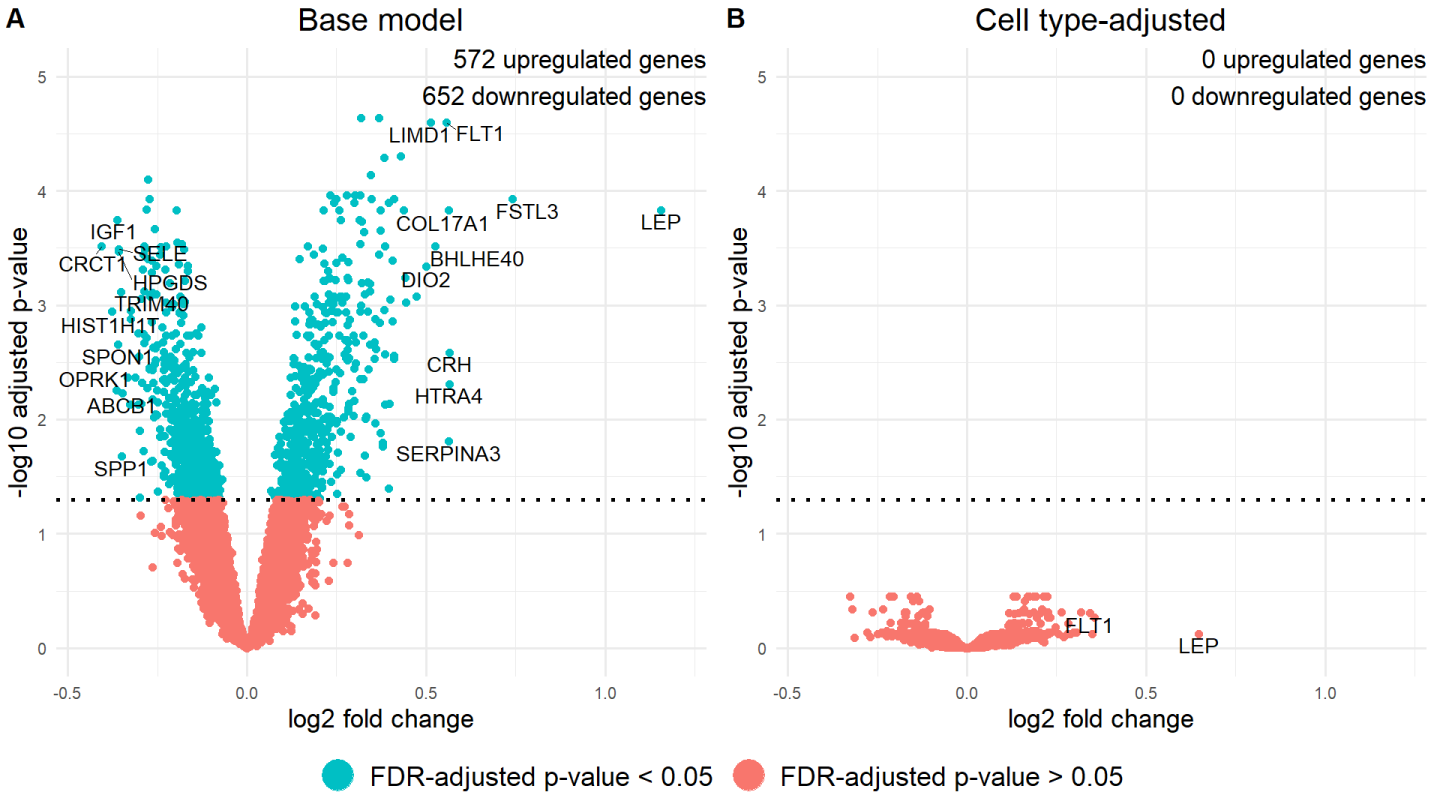

Figure 4. Volcano plots comparing differentially expressed genes in samples from preeclampsia cases versus healthy controls identified as significant by two models: (**A**) the base model adjusted for covariates fetal sex, study source, and gestational age and (**B**) the model additionally adjusted for the first five principal components of estimated cell type proportions. Dotted line represents a false discovery rate-adjusted q-value of 0.05. *FLT1* and *LEP* are labelled as a gene of interest in preeclampsia or as an outlier in log2 fold change, respectively.

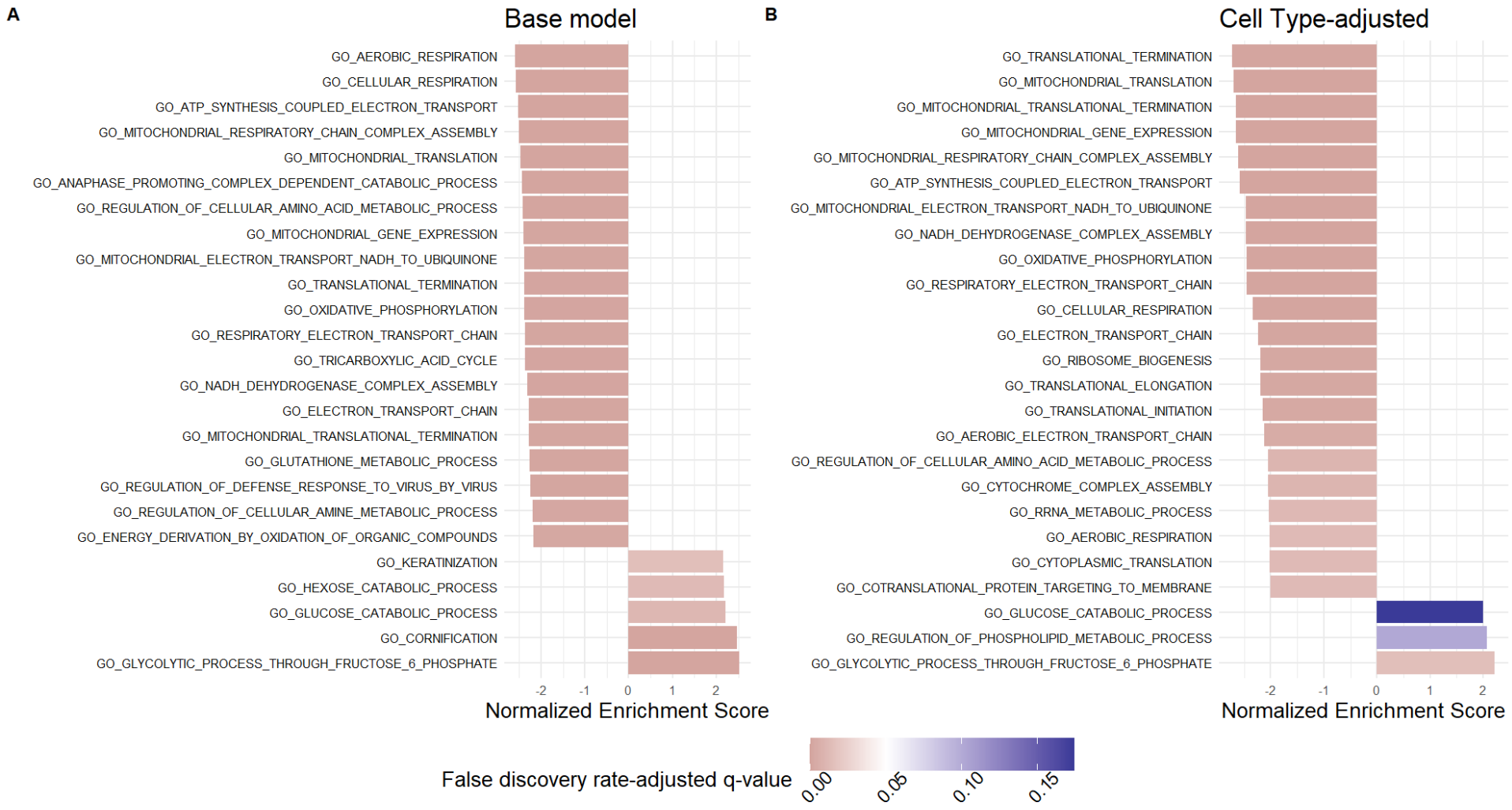

Figure 5. Top Gene Set Enrichment Analysis pathways from the Gene Ontology: Biological Processes database results for the differential expression analysis by preeclampsia case-control status. Results arranged by descending magnitude of the absolute value of the normalized enrichment score. Pathways colored red are significant at a false discovery rate-adjusted (FDR) q-value of 0.05 whereas pathways in blue are statistically insignificant. (**A**) Top pathways from the cell type-unadjusted analysis. (**B**) Top pathways from the cell type-adjusted analysis

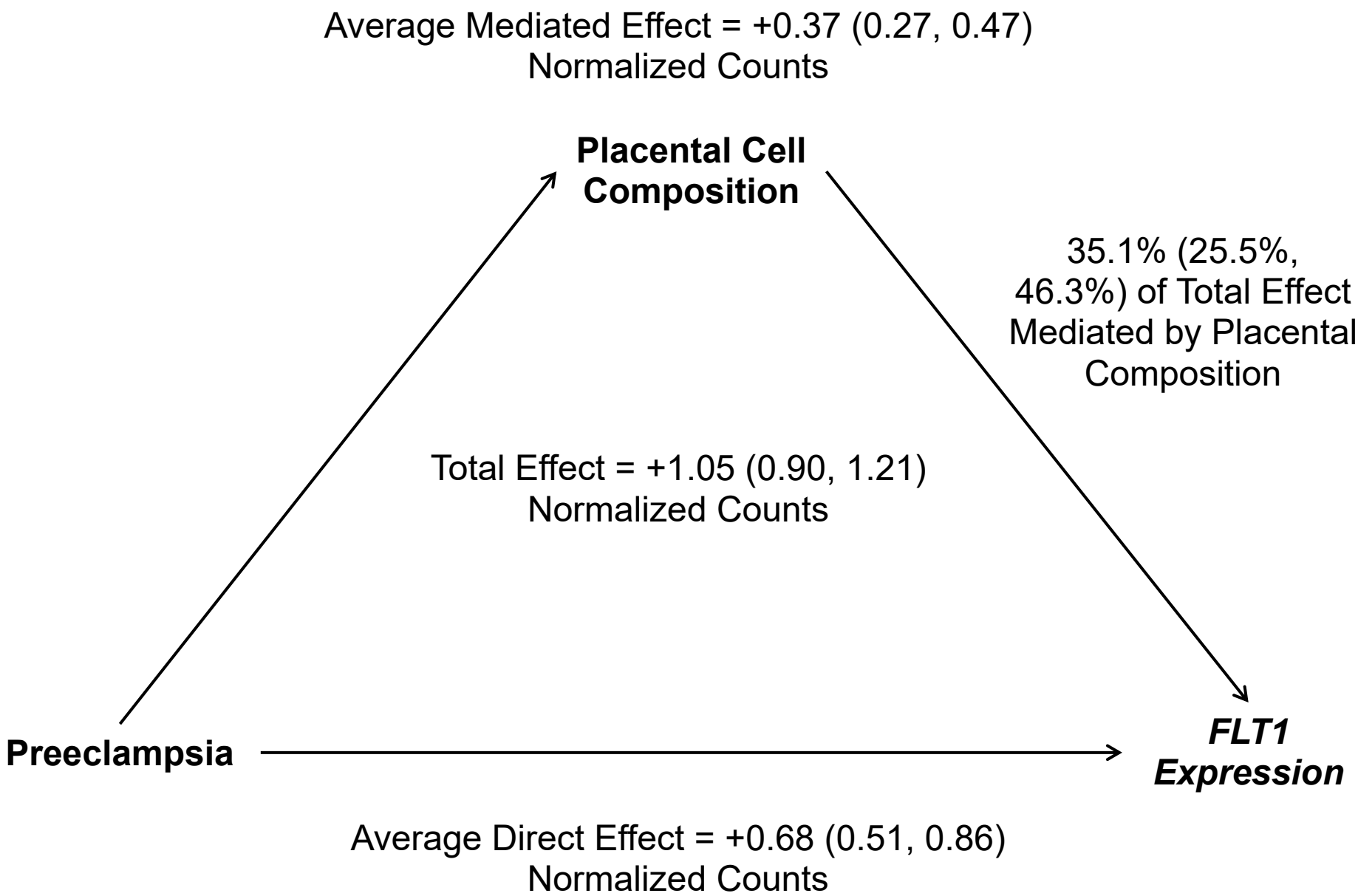

Figure 6. Mediation by placental cell type composition. Placental cell composition was operationalized as first five principal components of estimated cell type proportions. 95% confidence intervals are provided after effect estimates for each model parameter.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

1. Barker DJP, Thornburg KL (2013) Placental programming of chronic diseases, cancer and lifespan: A review. Placenta 34:841–845. https://doi.org/10.1016/j.placenta.2013.07.063

2. Maltepe E, Fisher SJ (2015) Placenta: The Forgotten Organ. Annual Review of Cell and Developmental Biology 31:523–552. https://doi.org/10.1146/annurev-cellbio-100814-125620

3. Gude NM, Roberts CT, Kalionis B, King RG (2004) Growth and function of the normal human placenta. Thrombosis Research 114:397–407. https://doi.org/10.1016/j.thromres.2004.06.038

4. Ilekis JV, Tsilou E, Fisher S, et al (2016) Placental origins of adverse pregnancy outcomes: potential molecular targets: an Executive Workshop Summary of the Eunice Kennedy Shriver National Institute of Child Health and Human Development. American Journal of Obstetrics and Gynecology 215:S1–S46. https://doi.org/10.1016/j.ajog.2016.03.001

5. Castellucci M, Kaufmann P (2006) Basic Structure of the Villous Trees. In: Benirschke K, Kaufmann P, Baergen R (eds) Pathology of the Human Placenta. Springer New York, New York, NY, pp 50–120

6. Lyall F, Robson SC, Bulmer JN (2013) Spiral Artery Remodeling and Trophoblast Invasion in Preeclampsia and Fetal Growth Restriction. Hypertension 62:1046–1054. https://doi.org/10.1161/HYPERTENSIONAHA.113.01892

7. Naicker T, Khedun SM, Moodley J, Pijnenborg R (2003) Quantitative analysis of trophoblast invasion in preeclampsia. Acta Obstetricia et Gynecologica Scandinavica 82:722–729. https://doi.org/10.1080/j.1600-0412.2003.00220.x

8. Kaufmann P, Black S, Huppertz B (2003) Endovascular Trophoblast Invasion: Implications for the Pathogenesis of Intrauterine Growth Retardation and Preeclampsia. Biology of Reproduction 69:1–7. https://doi.org/10.1095/biolreprod.102.014977

9. Yabe S, Alexenko AP, Amita M, et al (2016) Comparison of syncytiotrophoblast generated from human embryonic stem cells and from term placentas. PNAS 113:E2598–E2607. https://doi.org/10.1073/pnas.1601630113

10. Tang Z, Tadesse S, Norwitz E, et al (2011) Isolation of Hofbauer Cells from Human Term Placentas with High Yield and Purity. American Journal of Reproductive Immunology 66:336–348. https://doi.org/10.1111/j.1600-0897.2011.01006.x

11. Li L, Schust DJ (2015) Isolation, purification and in vitro differentiation of cytotrophoblast cells from human term placenta. Reprod Biol Endocrinol 13:. https://doi.org/10.1186/s12958-015-0070-8

12. Petroff MG, Phillips TA, Ka H, et al (2006) Isolation and culture of term human trophoblast cells. Methods Mol Med 121:203–217

13. Hirano T, Higuchi T, Ueda M, et al (1999) CD9 is expressed in extravillous trophoblasts in association with integrin α3 and integrin α5. Mol Hum Reprod 5:162–167. https://doi.org/10.1093/molehr/5.2.162

14. Robin C, Bollerot K, Mendes S, et al (2009) Human Placenta Is a Potent Hematopoietic Niche Containing Hematopoietic Stem and Progenitor Cells throughout Development. Cell Stem Cell 5:385–395. https://doi.org/10.1016/j.stem.2009.08.020

15. Tsang JCH, Vong JSL, Ji L, et al (2017) Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics. PNAS 114:E7786–E7795. https://doi.org/10.1073/pnas.1710470114

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

16. Pique-Regi R, Romero R, Tarca AL, et al (2019) Single cell transcriptional signatures of the human placenta in term and preterm parturition. eLife 8:. https://doi.org/10.7554/eLife.52004

17. Ma Y, Krikun G, Abrahams VM, et al (2007) Cell Type-specific Expression and Function of Toll-like Receptors 2 and 4 in Human Placenta: Implications in Fetal Infection. Placenta 28:1024–1031. https://doi.org/10.1016/j.placenta.2007.05.003

18. Pattillo RA, Gey GO (1968) The establishment of a cell line of human hormone-synthesizing trophoblastic cells in vitro. Cancer Res 28:1231–1236

19. Graham CH, Hawley TS, Hawley RG, et al (1993) Establishment and characterization of first trimester human trophoblast cells with extended lifespan. Exp Cell Res 206:204–211. https://doi.org/10.1006/excr.1993.1139

20. Brew O, Sullivan MHF, Woodman A (2016) Comparison of Normal and Pre-Eclamptic Placental Gene Expression: A Systematic Review with Meta-Analysis. PLOS ONE 11:e0161504. https://doi.org/10.1371/journal.pone.0161504

21. Lekva T, Lyle R, Roland MCP, et al (2016) Gene expression in term placentas is regulated more by spinal or epidural anesthesia than by late-onset preeclampsia or gestational diabetes mellitus. Scientific Reports 6:1–12. https://doi.org/10.1038/srep29715

22. Delahaye F, Do C, Kong Y, et al (2018) Genetic variants influence on the placenta regulatory landscape. PLOS Genetics 14:e1007785. https://doi.org/10.1371/journal.pgen.1007785

23. McHale CM, Zhang L, Thomas R, Smith MT (2013) Analysis of the transcriptome in molecular epidemiology studies. Environmental and Molecular Mutagenesis 54:500–517. https://doi.org/10.1002/em.21798

24. Meissner A, Mikkelsen TS, Gu H, et al (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454:766–770. https://doi.org/10.1038/nature07107

25. Reinius LE, Acevedo N, Joerink M, et al (2012) Differential DNA Methylation in Purified Human Blood Cells: Implications for Cell Lineage and Studies on Disease Susceptibility. PLOS ONE 7:e41361. https://doi.org/10.1371/journal.pone.0041361

26. Holbrook JD, Huang R-C, Barton SJ, et al (2017) Is cellular heterogeneity merely a confounder to be removed from epigenome-wide association studies? Epigenomics 9:1143–1150. https://doi.org/10.2217/epi-2017-0032

27. Jaffe AE, Irizarry RA (2014) Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biology 15:R31. https://doi.org/10.1186/gb-2014-15-2-r31

28. Shen-Orr SS, Gaujoux R (2013) Computational deconvolution: extracting cell type-specific information from heterogeneous samples. Current Opinion in Immunology 25:571–578. https://doi.org/10.1016/j.coi.2013.09.015

29. Teschendorff AE, Zheng SC (2017) Cell-type deconvolution in epigenome-wide association studies: a review and recommendations. Epigenomics 9:757–768. https://doi.org/10.2217/epi-2016-0153

30. Hunt GJ, Freytag S, Bahlo M, Gagnon-Bartsch JA (2019) dtangle: accurate and robust cell type deconvolution. Bioinformatics 35:2093–2099. https://doi.org/10.1093/bioinformatics/bty926

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

31. Campbell KA, Colacino JA, Park SK, Bakulski KM (2020) Cell Types in Environmental Epigenetic Studies: Biological and Epidemiological Frameworks. Curr Envir Health Rpt 7:185–197. https://doi.org/10.1007/s40572-020-00287-0

32. Leavey K, Benton SJ, Grynspan D, et al (2016) Unsupervised Placental Gene Expression Profiling Identifies Clinically Relevant Subclasses of Human Preeclampsia. Hypertension 68:137–147. https://doi.org/10.1161/HYPERTENSIONAHA.116.07293

33. McInnes L, Healy J, Melville J (2018) UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv:180203426 [cs, stat]

34. Sitras V, Paulssen RH, Grønaas H, et al (2009) Differential placental gene expression in severe preeclampsia. Placenta 30:424–433. https://doi.org/10.1016/j.placenta.2009.01.012

35. Nishizawa H, Ota S, Suzuki M, et al (2011) Comparative gene expression profiling of placentas from patients with severe pre-eclampsia and unexplained fetal growth restriction. Reprod Biol Endocrinol 9:107. https://doi.org/10.1186/1477-7827-9-107

36. Tsai S, Hardison NE, James AH, et al (2011) Transcriptional profiling of human placentas from pregnancies complicated by preeclampsia reveals disregulation of sialic acid acetylesterase and immune signalling pathways. Placenta 32:175–182. https://doi.org/10.1016/j.placenta.2010.11.014

37. Meng T, Chen H, Sun M, et al (2012) Identification of differential gene expression profiles in placentas from preeclamptic pregnancies versus normal pregnancies by DNA microarrays. OMICS 16:301–311. https://doi.org/10.1089/omi.2011.0066

38. Xiang Y, Cheng Y, Li X, et al (2013) Up-regulated expression and aberrant DNA methylation of LEP and SH3PXD2A in pre-eclampsia. PLoS One 8:e59753. https://doi.org/10.1371/journal.pone.0059753

39. Blair JD, Yuen RKC, Lim BK, et al (2013) Widespread DNA hypomethylation at gene enhancer regions in placentas associated with early-onset pre-eclampsia. Mol Hum Reprod 19:697–708. https://doi.org/10.1093/molehr/gat044

40. Nishizawa H, Pryor-Koishi K, Kato T, et al (2007) Microarray analysis of differentially expressed fetal genes in placental tissue derived from early and late onset severe pre-eclampsia. Placenta 28:487–497. https://doi.org/10.1016/j.placenta.2006.05.010

41. Enquobahrie DA, Meller M, Rice K, et al (2008) Differential placental gene expression in preeclampsia. Am J Obstet Gynecol 199:566.e1-566.11. https://doi.org/10.1016/j.ajog.2008.04.020

42. Várkonyi T, Nagy B, Füle T, et al (2011) Microarray Profiling Reveals That Placental Transcriptomes of Early-onset HELLP Syndrome and Preeclampsia Are Similar. Placenta 32:S21–S29. https://doi.org/10.1016/j.placenta.2010.04.014

43. Vennou KE, Kontou PI, Braliou GG, Bagos PG (2020) Meta-analysis of gene expression profiles in preeclampsia. Pregnancy Hypertension 19:52–60. https://doi.org/10.1016/j.preghy.2019.12.007

44. Sitras V, Fenton C, Acharya G (2015) Gene expression profile in cardiovascular disease and preeclampsia: A meta-analysis of the transcriptome based on raw data from human studies deposited in Gene Expression Omnibus. Placenta 36:170–178. https://doi.org/10.1016/j.placenta.2014.11.017

45. Luttun A, Carmeliet P (2003) Soluble VEGF receptor Flt1: the elusive preeclampsia factor discovered? J Clin Invest 111:600–602. https://doi.org/10.1172/JCI18015

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

46. Maynard SE, Min J-Y, Merchan J, et al (2003) Excess placental soluble fms-like tyrosine kinase 1 (sFlt1) may contribute to endothelial dysfunction, hypertension, and proteinuria in preeclampsia. J Clin Invest 111:649–658. https://doi.org/10.1172/JCI17189

47. McGinnis R, Steinthorsdottir V, Williams NO, et al (2017) Variants in the fetal genome near FLT1 are associated with risk of preeclampsia. Nature Genetics 49:1255–1260. https://doi.org/10.1038/ng.3895

48. Hedlund E, Deng Q (2018) Single-cell RNA sequencing: Technical advancements and biological applications. Molecular Aspects of Medicine 59:36–46. https://doi.org/10.1016/j.mam.2017.07.003

49. Raymond D, Peterson E (2011) A Critical Review of Early-Onset and Late-Onset Preeclampsia. Obstetrical & Gynecological Survey 66:497–506. https://doi.org/10.1097/OGX.0b013e3182331028

50. Brosens IA, Robertson WB, Dixon HG (1972) The role of the spiral arteries in the pathogenesis of preeclampsia. Obstet Gynecol Annu 1:177–191

51. Meekins JW, Pijnenborg R, Hanssens M, et al (1994) A study of placental bed spiral arteries and trophoblast invasion in normal and severe pre-eclamptic pregnancies. BJOG: An International Journal of Obstetrics & Gynaecology 101:669–674. https://doi.org/10.1111/j.1471-0528.1994.tb13182.x

52. Staff AC (2019) The two-stage placental model of preeclampsia: An update. Journal of Reproductive Immunology 134–135:1–10. https://doi.org/10.1016/j.jri.2019.07.004

53. Soleymanlou N, Jurisica I, Nevo O, et al (2005) Molecular Evidence of Placental Hypoxia in Preeclampsia. The Journal of Clinical Endocrinology & Metabolism 90:4299–4308. https://doi.org/10.1210/jc.2005-0078

54. Caniggia I, Mostachfi H, Winter J, et al (2000) Hypoxia-inducible factor-1 mediates the biological effects of oxygen on human trophoblast differentiation through TGFbeta(3). J Clin Invest 105:577–587. https://doi.org/10.1172/JCI8316

55. Zhang T, Bian Q, Chen Y, et al (2021) Dissecting human trophoblast cell transcriptional heterogeneity in preeclampsia using single-cell RNA sequencing. Mol Genet Genomic Med e1730. https://doi.org/10.1002/mgg3.1730

56. Cheng J-C, Chang H-M, Leung PCK (2013) Transforming Growth Factor-β1 Inhibits Trophoblast Cell Invasion by Inducing Snail-mediated Down-regulation of Vascular Endothelial-cadherin Protein. J Biol Chem 288:33181–33192. https://doi.org/10.1074/jbc.M113.488866

57. Bruckner TA, Catalano R (2018) Selection in utero and population health: Theory and typology of research. SSM Popul Health 5:101–113. https://doi.org/10.1016/j.ssmph.2018.05.010

58. Whitcomb BW, Schisterman EF, Perkins NJ, Platt RW (2009) Quantification of collider-stratification bias and the birthweight paradox. Paediatr Perinat Epidemiol 23:394–402. https://doi.org/10.1111/j.1365-3016.2009.01053.x

59. Dobin A, Davis CA, Schlesinger F, et al (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635

60. (2021) statgen/popscle. Center for Statistical Genetics

61. Luecken MD, Theis FJ (2019) Current best practices in single-cell RNA-seq analysis: a tutorial. Molecular Systems Biology 15:e8746. https://doi.org/10.15252/msb.20188746

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

62. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R | Bioinformatics | Oxford Academic. https://academic.oup.com/bioinformatics/article/33/8/1179/2907823. Accessed 28 Jan 2021

63. Haghverdi L, Lun ATL, Morgan MD, Marioni JC (2018) Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nature Biotechnology 36:421–427. https://doi.org/10.1038/nbt.4091

64. Zappia L, Oshlack A (2018) Clustering trees: a visualization for evaluating clusterings at multiple resolutions. Gigascience 7:. https://doi.org/10.1093/gigascience/giy083

65. Stuart T, Butler A, Hoffman P, et al (2019) Comprehensive Integration of Single-Cell Data. Cell 177:1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031

66. Vento-Tormo R, Efremova M, Botting RA, et al (2018) Single-cell reconstruction of the early maternal-fetal interface in humans. Nature 563:347–353. https://doi.org/10.1038/s41586-018-0698-6

67. Liu Y, Fan X, Wang R, et al (2018) Single-cell RNA-seq reveals the diversity of trophoblast subtypes and patterns of differentiation in the human placenta. Cell research 28:819–832. https://doi.org/10.1038/s41422-018-0066-y

68. Villani A-C, Satija R, Reynolds G, et al (2017) Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 356:. https://doi.org/10.1126/science.aah4573

69. Baboolal TG, Boxall SA, Churchman SM, et al (2014) Intrinsic multipotential mesenchymal stromal cell activity in gelatinous Heberden's nodes in osteoarthritis at clinical presentation. Arthritis Res Ther 16:R119. https://doi.org/10.1186/ar4574

70. Zhao Y, Li X, Zhao W, et al (2019) Single-cell transcriptomic landscape of nucleated cells in umbilical cord blood. Gigascience 8:giz047. https://doi.org/10.1093/gigascience/giz047

71. Bulmer JN, Morrison L, Johnson PM (1988) Expression of the proliferation markers Ki67 and transferrin receptor by human trophoblast populations. J Reprod Immunol 14:291–302. https://doi.org/10.1016/0165-0378(88)90028-9

72. Pierre-Luc (2021) plger/scDblFinder

73. Raudvere U, Kolberg L, Kuzmin I, et al (2019) g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Research 47:W191–W198. https://doi.org/10.1093/nar/gkz369

74. Schmon B, Hartmann M, Jones CJ, Desoye G (1991) Insulin and Glucose Do not Affect the Glycogen Content in Isolated and Cultured Trophoblast Cells of Human Term Placenta. J Clin Endocrinol Metab 73:888–893. https://doi.org/10.1210/jcem-73-4-888

75. Boyd AW (1987) Human Leukocyte Antigens: An Update on Structure, Function and Nomenclature. Pathology 19:329–337. https://doi.org/10.3109/00313028709103879

76. Blaschitz A, Weiss U, Dohr G, Desoye G (2000) Antibody Reaction Patterns in First Trimester Placenta: Implications for Trophoblast Isolation and Purity Screening. Placenta 21:733–741. https://doi.org/10.1053/plac.2000.0559

77. Gonen-Gross T, Goldman-Wohl D, Huppertz B, et al (2010) Inhibitory NK Receptor Recognition of HLA-G: Regulation by Contact Residues and by Cell Specific Expression at the Fetal-Maternal Interface. PLoS One 5:. https://doi.org/10.1371/journal.pone.0008941

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

78. Kaplan A, Chung K, Kocak H, et al (2008) Group B streptococcus induces trophoblast death. Microbial Pathogenesis 45:231–235. https://doi.org/10.1016/j.micpath.2008.05.003

79. Zozzaro-Smith PE, Bushway ME, Gerber SA, et al (2015) Whole mount immunofluorescence analysis of placentas from normotensive versus preeclamptic pregnancies. Placenta 36:1310–1317. https://doi.org/10.1016/j.placenta.2015.09.001

80. Coukos G, Makrigiannakis A, Amin K, et al (1998) Platelet-endothelial cell adhesion molecule-1 is expressed by a subpopulation of human trophoblasts: a possible mechanism for trophoblast-endothelial interaction during haemochorial placentation. Mol Hum Reprod 4:357–367. https://doi.org/10.1093/molehr/4.4.357

81. Cervar-Zivkovic M, Stern C (2011) Trophoblast Isolation and Culture. In: The Placenta. John Wiley & Sons, Ltd, pp 153–162

82. Andrews S (2010) FastQC: A quality control tool for high throughput sequence data.

83. Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32:3047–3048. https://doi.org/10.1093/bioinformatics/btw354

84. Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930. https://doi.org/10.1093/bioinformatics/btt656

85. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8

86. Subramanian A, Tamayo P, Mootha VK, et al (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. PNAS 102:15545–15550. https://doi.org/10.1073/pnas.0506580102

87. Mootha VK, Lindgren CM, Eriksson K-F, et al (2003) PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature Genetics 34:267–273. https://doi.org/10.1038/ng1180

88. Newman AM, Steen CB, Liu CL, et al (2019) Determining cell type abundance and expression from bulk tissues with digital cytometry. Nature Biotechnology 37:773–782. https://doi.org/10.1038/s41587-019-0114-2

89. Grün B, Kosmidis I, Zeileis A (2012) Extended Beta Regression in R: Shaken, Stirred, Mixed, and Partitioned. Journal of Statistical Software 48:1–25. https://doi.org/10.18637/jss.v048.i11

90. Smyth GK (2005) limma: Linear Models for Microarray Data. In: Gentleman R, Carey VJ, Huber W, et al (eds) Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Springer New York, New York, NY, pp 397–420

91. VanderWeele TJ (2014) A unification of mediation and interaction: a four-way decomposition. Epidemiology 25:749–761. https://doi.org/10.1097/EDE.0000000000000121

92. Shi B, Choirat C, Coull BA, et al (2021) CMAverse: A Suite of Functions for Reproducible Causal Mediation Analyses. Epidemiology 32:e20–e22. https://doi.org/10.1097/EDE.0000000000001378

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454041; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

93. Robins J (1986) A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. Mathematical Modelling 7:1393–1512. https://doi.org/10.1016/0270-0255(86)90088-6

94. VanderWeele TJ, Vansteelandt S, Robins JM (2014) Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. Epidemiology 25:300–306. https://doi.org/10.1097/EDE.0000000000000034