Gene Expression and DNA Methylation in Human Placental Cell Types: Applications to Deconvolution, Prenatal Metals Exposure, and Preeclampsia Molecular Epidemiology

by

Kyle A. Campbell

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Epidemiological Science)
in the University of Michigan
2023

Doctoral Committee:

Assistant Professor Kelly Bakulski, Chair
Professor Emerita Rita Loch-Caruso
Associate Professor Justin Colacino
Professor Emerita Vasantha Padmanabhan
Associate Professor Sung Kyun Park

Kyle A. Campbell

kyleac@umich.edu

ORCID iD: 0000-0002-1836-0341

Table of Contents

List of Tables

List of Figures

Abstract

The human placenta mediates adverse perinatal outcomes and current research examines molecular changes in bulk placental tissue. However, bulk molecular measures represent a convolution of signals across cell types, obscuring biological mechanisms and biasing study results.

Placental cell composition in preeclampsia is not well-understood and limits interpretation of bulk gene expression measures. In **Chapter 2**, I integrated a single-cell RNA-sequencing atlas of 19 fetal and 8 maternal cell types from placental villous tissue (n=9 biological replicates) at term (n=40,494 cells) and deconvoluted eight published microarray case-control studies of preeclampsia (n=173 controls, 157 cases). Preeclampsia was associated with excess extravillous trophoblasts (POR = 1.94, 95% CI [1.61, 2.34]) and fewer mesenchymal (POR = 0.79, 95% CI [0.73, 0.85]) and Hofbauer cells (POR = 0.67, 95% CI [0.59, 0.77]). Cellular composition mediated the association between preeclampsia and *FLT1* (37.8%, 95% CI [27.5%, 48.8%]) overexpression. My findings demonstrated placental cellular composition heterogeneity in preeclampsia drives previously observed bulk gene expression differences. This novel deconvolution reference allows for cellular composition-aware investigation into adverse perinatal outcomes.

To enable robust estimation of placental cell composition from bulk DNA methylation, in **Chapter 3**, I integrated a DNA methylation atlas (n=81) of 5 placental and 7 non-placenta cell types (p=192 fractions). Methylation was quantified via the

Illumina DNA methylation microarray (450k or EPIC), and common probes were quality filtered (n=407,628 DNA methylation sites). To identify cell type-discriminating DNA methylation sites, I ranked the top 50 hyper- and hypomethylated sites per cell type by F-test. To deconvolute bulk placental tissues, I applied robust partial correlation. Consistent with placental biology, bulk placental tissue (n=35) cell type proportion estimates (mean ± standard deviation) were predominately syncytiotrophoblast (57.8% ± 8.3%), stromal (20.6 ± 5.9%), cytotrophoblast (11.0% ± 4.1%), endothelial (7.5% ± 2.2%), and Hofbauer cells (1.5% ± 1.2%). This atlas can robustly estimate cell composition from placental DNA methylation data to detect unexpected non-placental cell types and improve casual inference.

Malapropos perinatal exposure to essential and toxic metals is widespread and linked to adverse outcomes. Healthy placental morphology is essential to pregnancy, but the relationship between metals exposure and placental cell composition is poorly understood. In Error! Reference source not found., I analyzed data from two prospective p regnancy cohorts: MARBLES (n=83) and EARLI (n=94). Urinary metals (p=18) concentrations were measured during early or late gestation. Placental DNA methylation was measured with EPIC or 450k microarrays. I estimated cellular composition via the **Chapter 3** reference. Demographics-adjusted beta regression models tested associations between metals concentrations (p=18) during early or late gestation and each placental cell type proportion (p=5). Results were meta-analyzed. A doubling in late gestation barium concentration was associated with a 0.2% increment in mean Hofbauer cell proportion to 2.5% (POR = 1.08, 95% CI [1.02, 0.14], q=0.25). Cadmium exposure was associated with a 3.0% decrement in mean syncytiotrophoblast

proportion to 59.5% (POR = 0.88, 95% CI [0.78, 0.98], q=0.31). Divalent metals exposure may disrupt placental structure, particularly among Hofbauer cells and syncytiotrophoblasts.

In conclusion, I applied state-of-the-art laboratory and bioinformatic approaches to develop cost-effective and scalable reference-based deconvolution methods for researchers to estimate cell composition in bulk placental tissue and I demonstrated the power of these novel approaches through applications to human populations. To provide valid, more causally interpretable results, future perinatal investigations should carefully consider placental cell composition using approaches developed in this dissertation.

Chapter 1 Introduction to Biological and Epidemiological Frameworks for Cell Types in Studies of Placental Gene Expression and Epigenetic Measures

This chapter has been adapted in part from a manuscript published in *Current Environmental Health Reports* (2020) [1].

Authors: Kyle A. Campbell[1], Justin A. Colacino[2], Sung Kyun Park[1,2], and Kelly M. Bakulski[1]
Affiliations:
[1]Department of Epidemiology, University of Michigan School of Public Health, University of Michigan, Ann Arbor, Michigan,
[2]Department of Environmental Health Sciences, University of Michigan School of Public Health, University of Michigan, Ann Arbor, Michigan

1.1 Early development and public health

The public health burden of adverse pregnancy outcomes is substantial. For example, the incidence of preterm birth ranges from 12-13% in the United States and is increasing [2]. The executive summary of the *Placental Origins of Adverse Pregnancy Outcomes: Potential Molecular Target*s workshop concluded that most adverse pregnancy outcomes are rooted in the placenta and its development [3]. Adverse pregnancy outcomes may even lead to chronic diseases throughout the life course of their offspring [4].The deployment of the Human Placenta Project by the National Institutes of Health prioritizes research into better understanding the important role the placenta may play in both perinatal and lifelong health [5]. A key gap in our knowledge is the molecular underpinnings of placental dysfunction. Advancements in understanding the molecular landscape of the placenta will lead to a better understanding of disease etiology and novel preventive and therapeutic interventions.

It is well understood that healthy placentation and placental function are essential to fetal development [6]. However, the developmental origins of health and disease hypothesis (DOHaD) posits that early-life exposures, even during gestation, affects biologic programming and lifelong health [7]. The placenta has emerged as a priority target organ in the DOHaD framework [4]. Advancements in understanding the role of perturbations to placental function or development during gestation, including environmental exposures, may explain the placenta's role in adverse perinatal or later life outcomes.

1.2 Placental structure and function

The temporary placenta develops early in pregnancy to promote maternal uterine artery remodeling; mediates transport of oxygen, nutrients, and wastes [8]; secretes hormones to regulate pregnancy; metabolizes various macromolecules and xenobiotics; and serves as a selective barrier to some pathogens and xenobiotics [6]. The fetal-maternal interface of the placenta is composed of tissue of fetal and maternal origin. The fetal chorionic plate includes the parenchymal villous tissue. The maternal basal plate (or decidua) is the site of placentation (**Figure 1.1**). The chorionic plate and basal plate circumscribe an intervillous space perfused with maternal blood. The branching, repeating tree-like villous tissue, called villi, protrude into the intervillous space, immersed in maternal blood (**Figure 1.1a**) [6, 8, 9]. Villi separate maternal and fetal blood by only three to four cell layers. Proper structural development and adequate functioning of placental villous tissue is critical to overall placental function.

Several placenta-specific cell types are responsible for the unique structure and function of the organ. The syncytiotrophoblast is a multi-nucleated, semi-continuous

syncytium covering the surface of villi and some parts of the basal and fetal plates (**Figure 1.1b**). The syncytiotrophoblast fulfills most placenta-specific functions [9]. Beneath the syncytiotrophoblast lies a layer of proliferative cytotrophoblasts that syncytialize to form and restore the syncytiotrophoblast [9, 10]. Cytotrophoblasts also differentiate to extravillous cytotrophoblasts that are responsible for invading the decidua and remodeling maternal spiral arteries to support the placenta and fetus [8].The Hofbauer cell is the resident tissue macrophage of the placenta. In addition to serving immune cell functions, Hofbauer cells may be essential throughout placental development to promote angiogenesis [11, 12]. To gain insight into placental function, placenta-specific cell types should be prioritized for interrogation.

1.3 Gene expression and cell function

DNA contains the information for the coding of proteins and RNA molecules responsible for cellular function, viability, and reproduction [13]. Gene expression represents the phenotype the cell is actively expressing. RNA sequencing provides direct information about gene expression by sequencing total RNA in the cell [14] and hence can be applied to study the placenta. For example, changes in placental villous tissue gene expression in response to *in utero* environmental exposures have been studied using this approach. Decreased placental expression of *BDNF* and *SYN1*—genes implicated in neurodevelopment—has been associated with exposure to fine particle air pollution [15]. However, the upstream regulation of gene expression is another important molecular outcome that provides insight into tissue health.

1.4 Epigenetics and DNA methylation

Epigenetics refers to the programming of cellular state, memory, or fate not attributable to changes in DNA sequence [16, 17]. Epigenetic modifications include DNA methylation, histone modifications, and non-coding RNAs [18]. Epigenetics are essential for several aspects of human development. First, DNA methylation undergoes dynamic changes during embryonic development. In the preimplantation embryo, the paternal genome experiences rapid, widespread DNA demethylation. Meanwhile, the maternal genome is passively demethylated to a lesser extent through replication without DNA methylation maintenance [19]. Second, during the reprogramming process, DNA methylation is maintained at specific locations in both the paternal and maternal genomes, termed genomic imprinting [20]. Third, X-chromosome inactivation is a dosage compensation mechanism that randomly transcriptionally silences one of two X chromosomes in females. [21]. Fourth, as tissues differentiate during embryogenesis, they acquire more specialized epigenetic marks. Broad regions of the epigenome are often regulated in concert and we observe larger-scale tissue-specific differentially methylated regions [22]. Epigenetic marks help to lock tissues in their differentiation state and maintain tissue identity. Appropriate epigenetic regulation is essential for healthy development and these processes can be dysregulated by environmental exposures.

DNA methylation, one of the most studied epigenetic mechanisms, describes the methylation of the fifth carbon of the nucleotide cytosine. The availability of DNA methylation microarrays have become a popular and relatively inexpensive approach to measure DNA methylation in large studies . DNA methylation undergoes drastic

reprogramming during mammalian development [23, 24] and is a key regulator of cellular differentiation [25]. Epigenetic dysregulation is a hallmark of multiple diseases, including cancer [26], neurodegeneration [27], and cardiovascular disease [28]. Critically, epigenetic modifications are labile to environmental influences, especially during gestation [29]. Understanding how environmental exposures impact epigenetic regulation during development likely will impact our understanding of disease etiology and identify novel methods for prevention and treatment [30].

1.5 Tissue specificity in environmental epigenetics

Different tissues have biologically determined epigenetic differences. When designing an epigenetic study, a crucial step is to determine the tissue of interest. A given tissue may be the “target” tissue for a disease process or exposure effects, while another tissue may be a measurable “surrogate” tissue for monitoring biomarkers associated with disease or exposure. Compelling arguments can be made for selecting target tissues (such as brain or lung) based on their direct links to disease. However, the postmortem timing of acquisition for many target tissues, scarcity of exposure and confounder data in most tissue banks, and often modest available sample sizes present challenges for target tissue research [31]. Epigenetic analyses on easily collected surrogate tissues (such as blood or saliva) may be less connected to the disease. Despite this, surrogate tissue research can provide valuable contributions related to etiologic timing through longitudinal sampling, identification of associations with environmental exposures, defining early biomarkers, developing translational utility, and even providing mechanistic insights in some cases [31]. The field of environmental epigenetics is strengthened by evidence from both surrogate and target tissue studies.

Traditionally, studies have investigated a single tissue of interest (either surrogate or target) at a time. Critical investigation of multiple tissues simultaneously is essential to identify similarities in epigenetic profiles across normal, diseased, and exposed tissues. Evidence on DNA methylation correlations across normal and diseased tissues vary. For example, correlation in DNA methylation signatures between surrogate and brain tissue was limited to few informative sites that varied by brain region [32]. Schizophrenia-associated DNA methylation signatures identified in blood and brain only overlapped at 7.9% of positions [33]. To specifically interrogate the utility of surrogate tissues in environmental epigenetics, the Toxicant Exposures and Responses by Genomic and Epigenomic Regulators of Transcription (TARGET) II consortium is studying epigenetic effects of exposure across tissues in perinatal mouse models [34]. Human population studies have also identified environment-associated epigenetic alterations that are detectable across tissues. For example, DNA methylation signatures associated with smoking exposure were first identified in cord blood [35]. A portion of the smoking signatures were then identified in adult blood [36], and a smaller portion were identified in adult lung tissue [37]. Testing epigenetic signatures across tissue types is a powerful approach to disentangle exposure- or disease-related systemic and tissue-specific alterations. A high degree of overlap in epigenetic signatures across target and surrogate tissues due to an exposure would provide support for the hypothesis that surrogate tissue types can provide relevant information about epigenetic alterations in target tissues. Epigenetic signatures specific to an exposure in a target tissue may represent unique effects that are informative of tissue-

specific regulation. Understanding the tissue specificity of environmental and disease epigenetics is an important and ongoing field.

This dissertation focuses on the term placenta as the target tissue of interest given its centrality to perinatal outcomes. Surrogate tissues may also be considered when attempting to study the target tissue in which the disease process is hypothesized to occur. Surrogate and target tissues have complementary advantages and disadvantages. Disease target tissues, such as the brain, may not be limited or not available for a given study design, though surrogate tissues may still provide insight into disease etiology or onset [31]. For example, circulating RNA and extracellular vesicle in maternal blood have been investigated as a source of information about the health status of the placenta or have even been implicated as a potential causal factor in preeclampsia [38–40]. The placenta is the organ of pregnancy and should be considered a target tissue to understand adverse perinatal outcomes. When selecting a tissue type for study, investigators must be thoughtful about the possible utility and scope of inferences in that tissue. When possible, a multi-tissue approach in sampling, biobanking, and measurements will allow for the most robust biological interrogations. Single tissue studies should make comparisons to publicly available data in multiple tissues to extend the reach of insights.

1.6 Cell type specificity complicates the study of gene expression and DNA methylation in complex tissues

The post-delivery placenta is an ideal target tissue for study because it has a direct role in fetal-maternal health and affords readily available, risk-free sampling. RNA sequencing and DNA methylation provide insight into the functional molecular

consequences of exposure. However, it is difficult to draw biological inferences from previously published study results because these measures represent averages over the tissue and across all cells and cell types present. Placental cell type differentiation is driven by a combination of transcription factor expression, epigenetic modifications, and physiologic conditions. This is a directed process required for adequate placental function that is better understood in mice than humans [8, 25, 41–43]. Pluripotent stem cells differentiate to myriad cell types by selective regulation of differentiation pathways [24, 25]. Therefore, RNA, DNA methylation, transcription factor, and other epigenetic profiles differ systematically by cell type [44, 45]. Complex cell type mixtures make up tissues. For example, whole blood contains many cell types including T-cells, B-cells, granulocytes, monocytes, and natural killer cells. Thus, tissue-level measures of DNA methylation, such as whole blood DNA methylation, reflect averages across all cells present (**Figure 1.2**). Tissue level measurements are therefore a convolution of many cell and cell type-specific molecular signals [46, 47].

Accounting for convolution due to cell type heterogeneity is critical to identify underlying biological mechanisms and facilitate proper interpretation. For brevity and simplicity, as well as the fact that epigenetic changes carry concomitant gene expression changes, the remainder of this section will mostly focus on epigenetics and DNA methylation. However, the topics and concepts discussed are equally applicable to direct measures of gene expression in addition to DNA methylation measures.

Given that tissue measures reflect DNA methylation averages across a mixture of cells, differences in DNA methylation by exposure or disease could have multiple underlying biological mechanisms. There are at least three biological scenarios that

lead to the same tissue-level DNA methylation signal (**Figure 1.3**). First, the exposure could have a uniform effect on DNA methylation across all cell types, leading to a substantial change in average DNA methylation (**Figure 1.3a**). For example, aging has a uniform direct effect on DNA methylation across cell types, termed the epigenetic "clock." DNA methylation at these positions strongly correlate with age across tissues and cell types [48, 49].

Alternatively, a difference in average tissue DNA methylation may be observed when vulnerable cell types exhibit a large shift in DNA methylation (**Figure 1.3b**). As an example, a small study of smoking and non-smoking healthy volunteers tested for differences in DNA methylation among sorted blood immune cell subpopulations. In smokers, two DNA methylation sites within the Growth Factor Independent 1 Transcriptional Repressor (*GFI1*) or F2R Like Thrombin or Trypsin Receptor 3 (*F2RL3*) genes were hypomethylated in granulocytes but not in peripheral blood mononuclear cells. Further, two sites within the Coproporphyrinogen Oxidase (*CPOX*) or G Protein-Coupled Receptor 15 (*GPR15)* genes were hypomethylated in peripheral blood mononuclear cells, including some T cell subtypes, but not in granulocytes [50]. Another small study of smokers and nonsmokers observed cell type-specific associations between smoking and DNA methylation in CD14+ monocytes, CD15+ granulocytes, CD19+ B cells, and CD2+ pan T cells [*30]. These results show cell types may have variable and specific DNA methylation susceptibility to environmental exposures.

In a third plausible scenario, the exposure has no direct effect on DNA methylation in any cell type. The apparent shift in average DNA methylation is attributable to a difference in cell type proportions between exposed and unexposed

individuals (**Figure 1.3c**). For example, in whole blood, cigarette smoking is associated with DNA methylation at a locus within *GPR15*. When considering blood immune cells separately, no direct effect of cigarette smoking was observed on *GPR15* methylation in GPR15+CD3+ T cells. Instead, smoking led to an increase in the relative proportion of GPR15+CD3+ T cells in whole blood [52]. Exposures can influence DNA methylation measures by causing a shift in cell type proportions, which can have important consequences in the tissue.

Tissue-level differences in DNA methylation can be biologically attributed to direct DNA methylation effects across all cells, direct DNA methylation effects in vulnerable cell types, or shifts in cell type heterogeneity. Each scenario represents a unique consequence of an exposure and warrants further investigation. Because bulk tissue-level measures of DNA methylation fail to resolve such biologically distinct mechanisms, observational studies are potentially fraught with incorrect conclusions and misinterpretation [53]. Applying methods to account for cell type heterogeneity is critical to identify underlying biological mechanisms and facilitate proper interpretation.

1.7 Methods for estimating or accounting for cell type heterogeneity

Accounting for cell type heterogeneity in DNA methylation data allows investigators to distinguish shifts in cellular heterogeneity from direct effects of an exposure on DNA methylation, both of which offer potential insights into disease etiology [46]. There are five main approaches to account for cellular heterogeneity: cellular separation, unbiased single-cell profiling, cell counting, and cellular deconvolution *in silico* by reference-based or reference-free methods (**Table 1.1**). Studies may elect to use one or more of these methods, based on their study design,

timing, tissue, and sample or measure availability. The advantages and trade-offs of each method are described below.

Direct physical cellular separation is a method to account for cell type heterogeneity in a mixed tissue that requires purifying cells or cell type subpopulations before measuring DNA methylation. Cell sorting technologies such as fluorescence-activated cell sorting or magnetic-activated cell sorting allow the user to isolate cellular subpopulations based on various stains, morphological characteristics, or expression of known cell type markers [54]. *A priori* knowledge of the distinguishing characteristics of cellular subpopulations present in the tissue is required, however, and represents a key limitation of this approach. Cells must also be processed and separated fresh at the time of sample collection or stored in a way to allow cell membranes to survive freeze-thaw. This can be achieved by using a cryopreservation blood tube or dissociation of solid tissues to a viable single-cell suspension, which is then cryopreserved prior to cell population separation and DNA methylation measurement. Investigators should be cautious as cell types may differentially survive processing. Following cell type separation, DNA methylation is measured in sorted cell types.

Single-cell epigenetics is an emerging technology that accommodates cell type heterogeneity. Single-cell approaches bypass the need for *a priori* cell type marker identification and generate single-cell epigenetic measures in an unbiased manner. These data can be aggregated at the cell type level using unbiased clustering to quantify epigenetic heterogeneity within and across cell types. Single-cell DNA methylation approaches are being rapidly developed and there is not a current consensus method. Current disadvantages include limited coverage and robustness,

labor requirements, and cost [55]. Single-cell technologies may even allow for mechanistic investigation of exposures and DNA methylation within individual cells or cell types and subtypes of tissues, organs, and organisms [56]. Like direct cellular separation, initial sample processing steps apply.

Direct cell counting methods, such as complete blood counting or histopathological cell counting, are used to quantify the relative abundance of cell types in a sample. DNA methylation measures are then made at the tissue level and investigators can adjust for the cell type counts in downstream analyses. This approach requires fresh samples or samples prepared for counting, such as fixed tissues. Direct cell counting allows investigators to test for exposure differences in cell type proportions. Unlike the previous two methods, however, cell counting offers no information about the direct effects of exposure on DNA methylation. Only cell type proportion estimates of a sample are available and can be used for adjustment or interpretation of a tissue DNA methylation measure.

Indirect cellular deconvolution is a class of methods to account for cell type heterogeneity via *in silico* estimation of cell type proportions. Deconvolution refers to the bioinformatic process of accounting for differences in intrasample cell type heterogeneity in tissues [57, 58]. Given that the previous three methods require specific laboratory preparation and processing at the time of sample collection, bioinformatic deconvolution is more commonly implemented in observational studies. To leverage DNA methylation data generated from heterogeneous tissues, two classes of deconvolution methods have been developed—reference-based and reference-free.

Reference-based methods are supervised and rely on independently collected cell type-specific DNA methylation profiles to estimate cell type proportions in a tissue sample. Advantages of the reference-based methods include: quantification of cell type proportions, biologically interpretable model components, and few model assumptions [46, 58, 59]. Reference panels are currently available for cord blood [60–64], umbilical cord tissue [64], adult blood [45, 65], frontal cortex (neuron vs. non-neuron) [66], and broadly epithelial versus fibroblast cell types [67] (**Table 1.2**). Disadvantages include a lack of demographically diverse reference samples, a limited number of reference panels, an assumption about constituent cell types, and challenges in identifying methylation sites and regions that discriminate cell types [46, 58, 68]. Similar to direct cell counts, cell type proportions estimated from reference-based deconvolution can be used in regression models when analyzing tissue DNA methylation measures. Reference-based methods would be an appropriate and powerful approach to implement when cell type references are available for a tissue of interest.

Reference-free methods are unsupervised methods to account for variation in DNA methylation data, including cell type heterogeneity. This category encompasses many algorithms that account for sources of variation that are unmeasured and unmodeled due biological sources of variation like cell type heterogeneity or nonbiological sources of variation, such as random noise or batch effects in an association study. Reference-free methods, like “surrogate variable analysis”, were originally developed for RNA expression deconvolution [69], and are now being applied in epigenome-wide association studies [58]. Advantages of unsupervised methods include no required *a priori* knowledge of tissue cell types, flexible modelling strategies,

and no required cell type references, allowing them to be used in any tissue. Disadvantages include the general inability to estimate intrasample cell type proportions and the large number of delicate model assumptions, including the assumption that the largest driver of variation is due to cell type proportion differences [58, 70, 71]. Following reference-free processing, depending on the specific method, investigators either implement exposure testing on the resulting adjusted DNA methylation matrix or they account for the reference-free “cell types” in regression models when analyzing tissue DNA methylation measures. Reference-free methods are only recommended for tissues lacking adequate references.

Cell type heterogeneity must be considered and accounted for in any epigenetic study of bulk tissue. Direct measures of DNA methylation should be prioritized over indirect methods to faithfully capture the DNA methylation state of each sample without reliance on imperfect indirect methods that “smooth over” inter-sample differences such as *in silico* deconvolution. Of direct approaches, single-cell molecular assays show the greatest promise because they unbiasedly account for cell type heterogeneity with the greatest resolution and can later be aggregated at the cell type or sub-cell type level if desired. These methods, however, are not yet widely available for observational human studies and are still subject to dissociation bias [72]. At this time, among indirect deconvolution approaches, referenced-based methods should be prioritized over reference-free methods. Reference-based deconvolution requires fewer assumptions and affords greater transparency and biological interpretability. A recent comparison of indirect deconvolution methods in the placenta found that reference-based methods consistently outperformed reference-free methods [73]. However, reference-free

deconvolution is invaluable when reference data are unavailable or biosampling logistics prevent cell sorting or single-cell approaches. Further studies of deconvolution algorithm performance and collection of high-quality and diverse DNA methylation reference profiles are required to advance indirect deconvolution approaches. Researchers should be transparent in reporting the assumptions and selection criteria for any cell type heterogeneity approach used.

1.8 Overview of recommendations

The public health burden of adverse pregnancy outcomes is substantial and most adverse pregnancy outcomes are rooted in the placenta and its development. Epigenetic modifications and concomitant gene expression regulation are essential for human development and are labile to environmental influences, especially during gestation. Epigenetic dysregulation is a hallmark of multiple diseases. Investigators routinely measure gene expression or DNA methylation in readily available tissues. However, tissues and cell types exhibit specific epigenetic patterning and heterogeneity between samples complicates studies of both biomarkers. Failure to account for cell type heterogeneity limits identification of biological mechanisms and biases study results. Tissue-level molecular measures represent a convolution of signals from individual cell types. Tissue-specific molecular investigation is an evolving field, and the use of disease-affected target, surrogate, or multiple tissues has inherent trade-offs and affects inference. Likewise, experimental and bioinformatic approaches to accommodate cell composition heterogeneity have varying assumptions and inherent trade-offs that affect inference.

The relationships between exposure, disease, tissue-level gene expression or DNA methylation, cell type-specific gene expression or DNA methylation, and cell composition heterogeneity must be carefully considered in study design and analysis. Causal diagrams can inform study design and analytic strategies. Properly addressing cell composition heterogeneity limits sources of potential bias, avoids misinterpretation of study results, and allows investigators to distinguish shifts in cell type proportions from direct changes to cellular epigenetic and gene expression programming, both of which provide insights into disease etiology and aid development of novel methods for prevention and treatment.

1.9 Current research needs

The placenta should be prioritized as a target organ in the DOHaD framework and biomedical and epidemiology research more generally to understand and prevent early life and lifelong morbidity and mortality. Cellular heterogeneity poses a critical challenge in applying molecular measures such as gene expression and DNA methylation in bulk tissue studies. This is particularly relevant to placenta-specific cell types that are not fully understood and directly related to placental function. Currently, deconvolution references for human term, placental villous tissue are limited or unavailable for gene expression and DNA methylation. Accessible, scalable, and robust placental deconvolution references and their application to perinatal studies are needed to advance molecular perinatal epidemiology.

1.10 Brief description of the dissertation chapters

Chapter 1 introduces the public health importance of molecular perinatal studies of the human placenta, the basic anatomy of the placenta, gene expression, epigenetics and DNA methylation, the limitations and challenges of studying molecular measures in complex tissues, cell composition heterogeneity as a major contributor to these limitations and challenges, techniques to address cellular heterogeneity in complex tissues, and current research needs in deploying reference-based deconvolution approaches to address cellular heterogeneity in the placenta.

Chapter 2 discusses my primary research efforts to address placental cellular heterogeneity in transcriptomic studies of the term human placenta by developing a single-cell RNA-sequencing based deconvolution reference, integrated from primary and previously published data. I apply this reference to assess how cell composition may mediate preeclampsia-associated gene expression differences in a secondary analysis of preeclampsia case-control studies.

Chapter 3 describes my primary research efforts to address placental cellular heterogeneity in epigenetic studies of the term human placenta. I developed a sorted cell type DNA methylation-based deconvolution reference panel that integrates primary and previously published cell type-specific DNA methylation profiles.

In Chapter 4, I apply the deconvolution reference developed in Chapter 3 to estimate term placental cell composition in two prospective birth cohorts. I then test the association between prenatal exposure to 18 metals and term placental cell composition during early or late gestation to understand how prenatal metals exposure may affect fetal-maternal health.

In Chapter 5, I discuss the utility of cellular heterogeneity in epidemiological and biological frameworks, recommendations, and considerations for conducting molecular epidemiology studies, and the future directions and public health significance of the research presented in this dissertation.

A unifying conceptual diagram of the primary research aims is presented in **Figure 1.4**. To address these challenging research gaps, this dissertation leverages advanced laboratory-based methods in primary tissue collection, dissociation, and cell type sorting, single-cell gene expression characterization, bulk gene expression and DNA methylation characterization. In addition, I integrate these laboratory-based findings with state-of-the-art computational approaches in single-cell and bulk gene expression analysis, DNA methylation analysis, creation and validation of molecular deconvolution references, application of deconvolution to epidemiologically model cell composition, general bioinformatics, bioinformatic data integration, meta-analysis, molecular epidemiology, and causal inference. Ultimately, this dissertation aims to provide scalable and cost-effective approaches and a template to demonstrate their capability to empower investigators to faithfully and carefully consider cell composition in perinatal epidemiology to improve casual interpretability of study results and reveal biological mechanisms of health and disease.

1.11 Figures and Tables

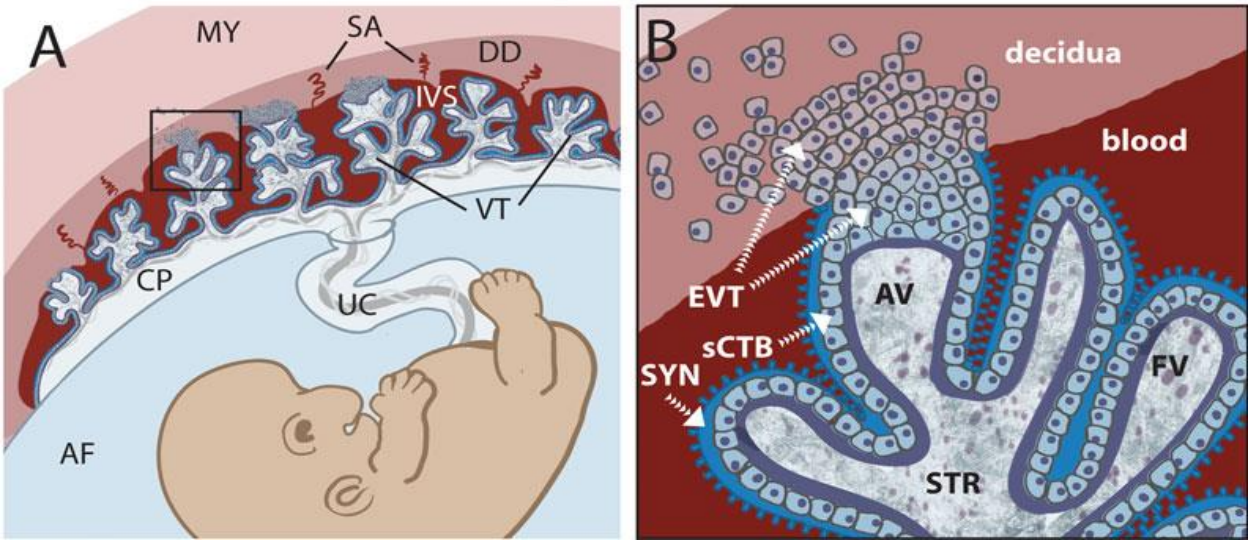


Figure 1.1 Placental structure around six weeks gestation. MY: myometrium, SA: spiral arteries, DD: decidua; IVS:intervillous space VT: villous tree, CP: chorionic plate, UC: umbilical cord, AF: amniotic fluid, AV: anchoring villi, STR: stroma, FV: floating villi, EVT: extravillous trophoblast, sCTB: cytotrophoblast, SYN: syncytiotrophoblast [74].

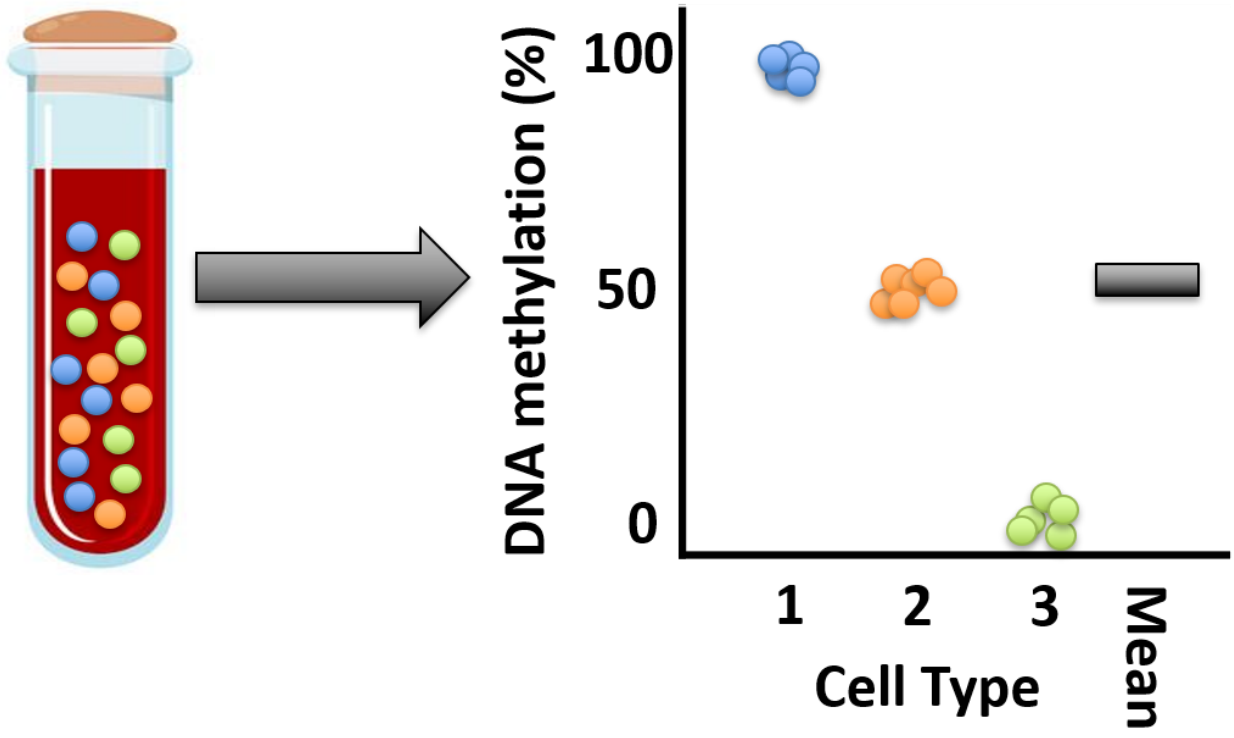


Figure 1.2 Conceptual model for understanding tissue measures of DNA methylation as a mixture of signals from cell types. A complex tissue such as whole blood is composed of many individual cell types. Individual circles represent cells, colored by cell type identity. In this example, three cell types compose the tissue. The investigator performs a tissue-level assessment of DNA methylation that averages across cell types. The black bar represents the aggregate observed tissue-level mean DNA methylation signal.

The cell type DNA methylation profiles are not observed. The investigator may incorrectly conclude that each genomic locus in each cell type in the sample is uniformly methylated at 50% if they do not consider the cell type heterogeneity of the sample.

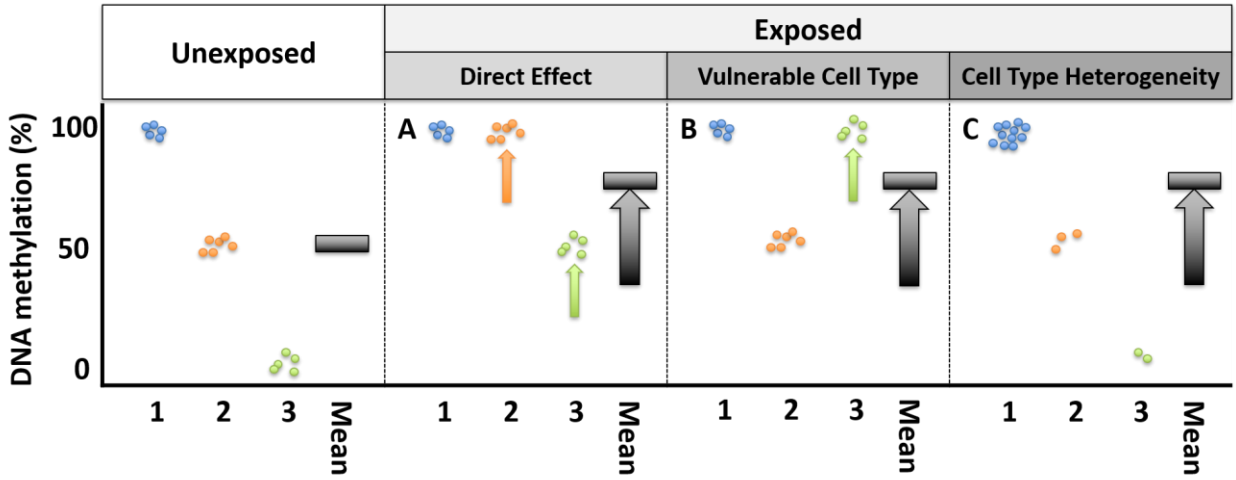


Figure 1.3 Conceptual model for differences in tissue DNA methylation by exposure status, considering tissues as mixtures of cell types. Individual circles represent cells, colored by three cell type identities. The black bar represents the observed tissue-level mean DNA methylation signal. ***A-C*** *represent distinct biological scenarios that could lead to the same exposure-related DNA methylation signal.* ***A.*** *The exposure uniformly increases DNA methylation in each cell type population, which increases the observed DNA methylation signal.* ***B.*** *The exposure directly increases DNA methylation in one vulnerable cell type.* ***C.*** *The exposure does not have a direct effect on DNA methylation and the observed increase in DNA methylation signal is completely mediated by differences in cell type proportion between the exposed and unexposed samples.*

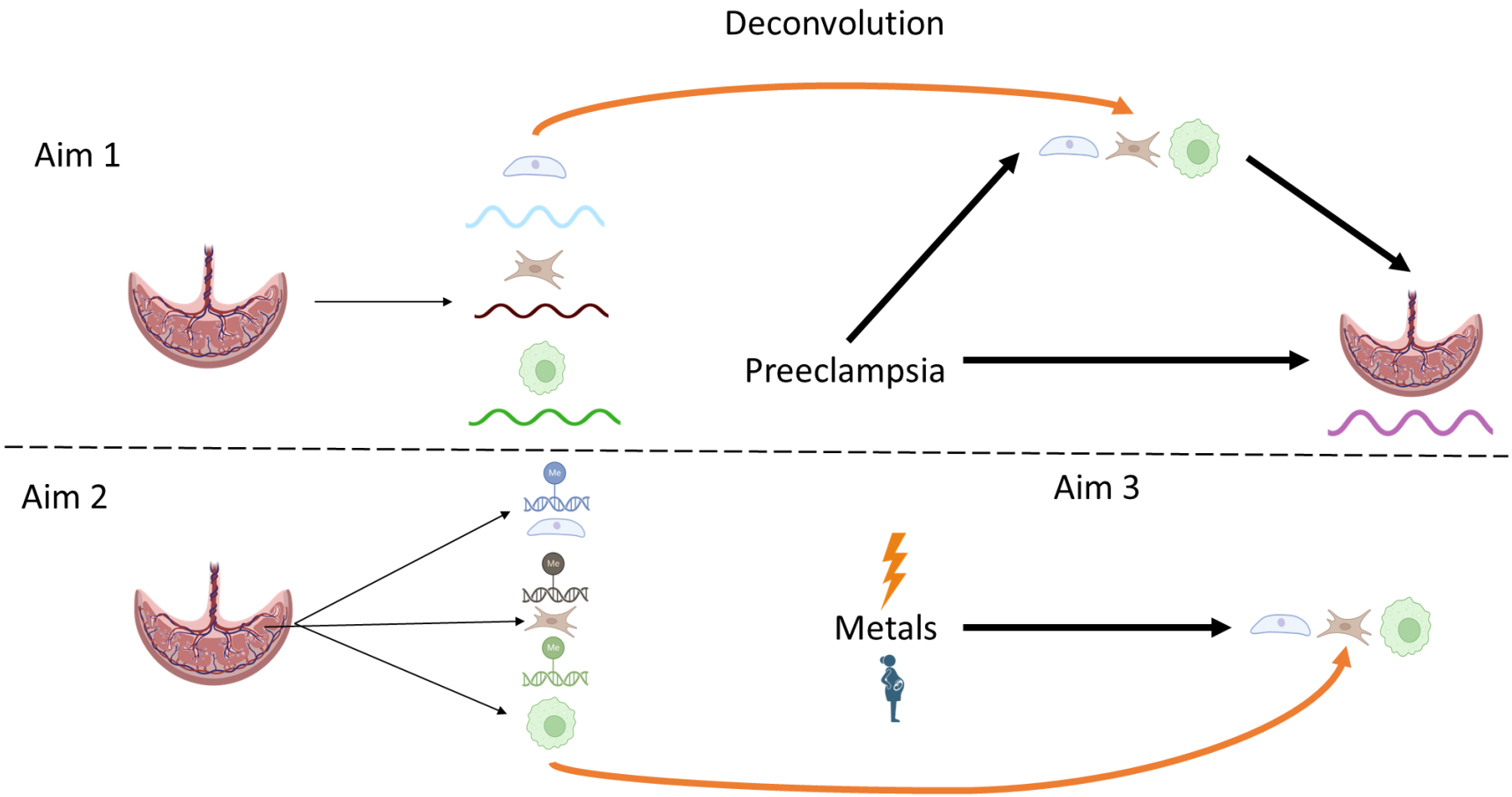


Figure 1.4 Conceptual overview of research aims. In aim 1, I characterized placental cell type-specific gene expression to deconvolute a secondary data analysis of preeclampsia. I then test the relationship between preeclampsia and gene expression, preeclampsia and term placental cell composition, and finally perform a unified mediation and interaction analysis to test whether placental cell composition mediates the effects of preeclampsia on gene expression. In aim 2, I characterized placental cell type-specific DNA methylation to create a deconvolution reference. In aim 3, I apply the deconvolution reference generated in aim 2 to test the association between early or late gestation prenatal metals exposure measured in urine and term placental cell composition in two prospective birth cohorts.

Table 1.1 Approaches to address cell type heterogeneity in DNA methylation studies.

Approach	Cellular Measure	DNA Methylation Measure	Advantages	Disadvantages	Examples
Cell type sorting	Direct	Cell type	Measures of DNA methylation within cell types Allows development of cell type-specific reference profiles	Requires careful sample processing Laborious sorting Cell type specificity limited to prior knowledge	Adherence Density Antibody-binding: • Flow Cytometry • Magnetic Bead Reviewed in [75]
Single-cell technologies	Direct	Single-cell	Measures of DNA methylation at single-cell resolution No *a priori* knowledge of cell types required	Requires careful sample processing Expensive Limited throughput Requires single-cell bioinformatics expertise	Approaches: many, reviewed in [55] Analysis software: BEAT [76], methylpy [77], LIGER [78], Melissa [79]
Cell counting	Direct	Tissue	Well-studied and classic experimental techniques	Requires careful sample processing Laborious sorting Subject to error Limited information	Microscope counting in fresh or histological samples Complete blood counts Other cell count procedures
Reference-based deconvolution	Indirect	Tissue	Cell type proportions estimates Few and transparent model assumptions with biological interpretability	Requires cell type-specific reference profiles Demographically diverse profiles are limited *A priori* knowledge of cell types required	Houseman's constrained projection [80], CIBERSORT [81], IDOL [82], EpiDISH [83]
Reference-free deconvolution	Indirect	Tissue	No *a priori* knowledge of cell types required No cell type-specific reference profiles required	Cell type proportion estimates generally unavailable Requires delicate model assumptions that can lack biological interpretability	SVA [69], ISVA [84], RUV [85], FaST-LMM-EWASher [86], RefFreeEWAS 1.0/2.0 [42, 86], ReFACTor [88], BayesCCE [89]

Table 1.2 Summary of cell type-specific reference panels available for reference-based indirect deconvolution.

Tissue	Sample Population	Number of Cell Types	Cell Types Measured	Sorting Technology	DNA Methylation Platform	Reference
Adult Blood	6 healthy male donors, aged 25-60	7	Neutrophils, Eosinophils Monocytes, B cells, NK cells, $CD8^+$ T cells, $CD4^+$ T cells	Magnetic-activated cell sorting	Illumina 450k microarray	Reinius et al., 2012 [45]
	37 healthy donors, aged 19-59	6	Neutrophils, Monocytes, B cells, NK cells, $CD8^+$ T cells, $CD4^+$ T cells	Magnetic-activated cell sorting	Illumina EPIC 850k microarray	Salas et al., 2018 [65]
Umbilical Cord Blood	11 healthy full-term singleton deliveries	6	Granulocytes, Monocytes, B cells, NK cells, $CD8^+$ T cells, $CD4^+$ T cells	Fluorescence-activated cell sorting	Illumina 450k microarray	Gervin et al., 2016 [61]
	17 healthy full-term singleton vaginal deliveries	7	Granulocytes, Monocytes, B cells, NK cells, $CD8^+$ T cells, $CD4^+$ T cells, Nucleated Red Blood cells	Magnetic-activated cell sorting	Illumina 450k microarray	Bakulski et al., 2016 [62]
	7 elective non-laboring Caesarean-section	7	Granulocytes, Monocytes, B cells, NK cells, $CD8^+$ T cells, $CD4^+$ T cells, Nucleated Red Blood cells	Fluorescence-activated cell sorting	Illumina 450k microarray	de Goede et al., 2016 [63]
	14 healthy singleton full-term deliveries from women aged 28-38	6	Granulocytes, Monocytes, B cells, NK cells, $CD8^+$ T cells, $CD4^+$ T cells	Magnetic-activated cell sorting	Illumina EPIC 850k microarray	Lin et al., 2018 [64]
	Mixed from above, 34 cell type fractions removed or reclassified from Bakulski, Gervin	7	Granulocytes, Monocytes, B cells, NK cells, $CD8^+$ T cells, $CD4^+$ T cells, Nucleated Red Blood cells	Mixed	Mixed	Gervin et al., 2019 [**39] combines Gervin , Bakulski, de Goede, and Lin
Umbilical Cord Tissue	14 healthy singleton full-term deliveries from women aged 28-38	3	Stromal, Endothelial, and Epithelial	Magnetic-activated cell sorting	Illumina EPIC 850k microarray	Lin et al., 2018 [64]

Frontal Cortex	29 post-mortem major depressive disorder cases and 29 matched controls	2	Neuron Non-neuron	Fluorescence-activated cell sorting	Illumina 450k microarray	Guintivano et al., 2013 [66]
Broadly Epithelial	ENCODE cell lines and Reinius et al., 2012 [45] dataset	3	11 Epithelial and 7 Fibroblast cell lines, adult blood immune cell types	Secondary Analysis	Illumina 450k microarray	Zheng et al., 2018 [67]

Chapter 2 Placental Cell Type Deconvolution Reveals that Cell Proportions Drive Preeclampsia Gene Expression Differences

This chapter has been adapted from a manuscript published in *Communications Biology* (2023) [90].

Authors: Kyle A Campbell[1], Justin A Colacino[2,3], Muraly Puttabyatappa[4], John F Dou[1], Elana R Elkin[2], Saher S Hammoud[5,6,7], 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, Ann Arbor, MI, 48105, USA.

[2]Environmental Health Sciences, School of Public Health, University of Michigan, Ann Arbor, MI, 48105, USA.

[3]Nutritional Sciences, School of Public Health, University of Michigan, Ann Arbor, MI, 48105, USA.

[4]Pediatrics, Michigan Medicine, University of Michigan, Ann Arbor, MI, 48105, USA.

[5]Human Genetics, Michigan Medicine, University of Michigan, Ann Arbor, MI, 48105, USA

[6]Obstetrics and Gynecology, Michigan Medicine, University of Michigan, Ann Arbor, MI, 48105, USA.

[7]Department of Urology, Michigan Medicine, University of Michigan, Ann Arbor, MI, 48105, USA.

2.1 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 (n=9 biological replicates) at term (n=40,494 cells). We deconvoluted eight

published microarray case-control studies of preeclampsia (n=173 controls, 157 cases). Preeclampsia was associated with excess extravillous trophoblasts and fewer mesenchymal and Hofbauer cells. Adjustment for cellular composition reduced preeclampsia-associated differentially expressed genes (log_2 fold-change cutoff=0.1, FDR<0.05) from 1,154 to 0, whereas downregulation of mitochondrial biogenesis, aerobic respiration, and ribosome biogenesis were robust to cell type adjustment, suggesting direct changes to these pathways. Cellular composition mediated a substantial proportion of the association between preeclampsia and *FLT1* (37.8%, 95% CI [27.5%, 48.8%]), *LEP* (34.5%, 95% CI [26.0%, 44.9%]), and *ENG* (34.5%, 95% CI [25.0%, 45.3%]) overexpression. Our findings indicate substantial placental cellular heterogeneity in preeclampsia contributes to previously observed bulk gene expression differences. This novel deconvolution reference lays the groundwork for cellular heterogeneity-aware investigation into placental dysfunction and adverse birth outcomes.

2.2 Introduction

The public health burden of adverse pregnancy outcomes is substantial. An important example is preeclampsia, which affected 6.5% of all live births in the United States in 2017 and is characterized by high maternal blood pressure, proteinuria, and damage to other organ systems [91]. Adverse pregnancy outcomes may lead to myriad health complications including elevated risk of chronic diseases throughout the life course [4]. The placenta, a temporary organ that develops early in pregnancy, promotes maternal uterine artery remodeling; mediates transport of oxygen, nutrients, and waste [8]; 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 [6]. The executive summary of the Placental Origins of Adverse Pregnancy Outcomes: Potential Molecular Targets workshop recently concluded that most adverse pregnancy outcomes are rooted in placental dysfunction [3]. 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 [9]. 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 [8]. Inadequate or inappropriate invasion of extravillous trophoblasts has previously been implicated in preeclampsia etiology [92–94]. 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 [95–100]. 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 [101–103]. Furthermore, placental cell lines such as BeWo, derived from choriocarcinoma [104], and HTR-8/SVneo, immortalized by SV40 [105], are typically derived by processes that alter the DNA of the cells, limiting their in vivo translatability. 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 [106–108] and are common in epidemiological studies [109]. Gene expression profiles differ systematically by cell type [44, 45]. Thus, bulk placental tissue-level gene expression measurements represent a convolution of gene expression signals from individual cells and cell types [46, 47]. Deconvolution refers to the bioinformatic process of estimating the distribution of cell types that constitute the tissue [57, 58]. Deconvoluting tissue-level gene expression profiles is essential to account for effects introduced by unmodeled cell type proportions [110] by disentangling shifts in cell type proportions from direct changes to cellular gene expression [1]. 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 [1]. Prior placental cell type-specific gene expression measures from term villous tissue [101, 102] 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 [101, 102]. 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 bulk RNA-sequencing. Finally, to 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 [111] of preeclampsia, including a mediation analysis of the preeclampsia-associated genes *FLT1*, *LEP*, and *ENG* that quantifies the role cellular composition plays in explaining bulk gene expression measures.

2.3 Methods

2.3.1 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. Pregnant women provided written informed consent for research use of discarded tissues. Study protocols for discarded tissue collection and research use were approved by the University of Michigan Institutional Review Board (HUM00017941, HUM00102038). 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 [98]. We subjected approximately 1g minced dissected villous tissue to the Miltenyi Tumor Dissociation Kit on the GentleMACS Octo Dissociator with Heaters (Miltenyi Biotec) to 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 to remove undigested tissue and subsequently 40μm [97, 98]. To collect a syncytiotrophoblast-enriched fraction, the fraction between 40μm and 100μm was washed from the 40μm strainers, adapting a previous protocol that collected syncytiotrophoblasts throughout this size range [95]. 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).

Previously published single-cell RNA-sequencing raw data of healthy, term placental villous tissue samples came from the Database of Genotypes and Phenotypes (Pique-Regi et. al, accession number phs001886.v1.p1 [112]) SRR10166478 (Sample 3), SRR10166481 (Sample 4), and SRR10166484 (Sample 5) [102]. The collection and use of human materials for the study were approved by the Institutional Review Boards of the Wayne State University School of Medicine. All participating women provided written informed consent prior to sample collection [102]. Additional previously published samples came the European Genome-Phenome Archive (Tsang et. al, accession number EGAS00001002449 [38]) (Samples 6-9) [101]. The study was approved by the Joint Chinese University of Hong Kong-New Territories East Cluster Clinical Research Ethics Committee, and informed consent was obtained after the nature and possible consequences of the studies were explained. Pregnant women were recruited from the Department of Obstetrics and Gynecology, Prince of Wales

Hospital, Hong Kong with informed consent; the subjects studied had consented to sequencing data archiving [101].

2.3.2 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). Paired end 110 base pair reads were sequenced on NovaSeq 6000 (Illumina).

2.3.3 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 [113]. Previously published single-cell RNA-sequencing raw data of healthy, term placental villous tissue samples from the Database of Genotypes and Phenotypes (Pique-Regi et. al, accession number phs001886.v1.p1) SRR10166478 (Sample 3), SRR10166481 (Sample 4), and SRR10166484 (Sample 5) [102] and from the European Genome-Phenome Archive (Tsang et. al, accession number EGAS00001002449) (Samples 6-9) [101] were processed identically. The freemuxlet program in the latest version (accessed 2021/12/05) of the ‘popscle’ package was used to assign fetal or

maternal origin and identify 736 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) [114]. Per cell quality control criteria were calculated using the quickQCPerCell() function (scater R package, version 1.18.6) with default settings [115] (**Figure 2.2**, **Figure 2.3**) and included total unique RNA transcripts (also called unique molecular identifiers), unique genes, and percentage of reads mapping to mitochondrial genes [116]. According to current recommended best practice, each batch was quality-controlled separately [117]. We excluded 6,497 low-quality outlier cells defined as cells with less than 500 unique RNA molecules, less than 200 unique genes, or that were outliers in mitochondrial gene mapping rate. Mitochondrial mapping outliers exceeded four median absolute deviations in samples 1 and 2 (mitochondrial reads > 9.2%) or three median absolute deviations in samples 3, 4, and 5 (mitochondrial reads > 8.9%) and samples 6, 7, 8C, 8P, 9C, and 9P (mitochondrial reads > 9.1%). To generate normalized gene expression data for visualizations and analyses that required normalization, single-cell gene counts were library-size normalized by dividing the number of counts by the total number of counts expressed in that cell, multiplied by a scale factor of 10,000, and log-transformed with the NormalizeData() function (Seurat R package, version 4.1.1).

2.3.4 Single-cell RNA-sequencing clustering and cluster annotation

Maternal and fetal cells were split into separate datasets for clustering. To integrate data from cells across study sources and visualize clustering results with uniform manifold projection [118], we used the mutual nearest neighbor batch correction approach via FastMNN from ‘SeuratWrappers’ with default settings (R package, version

0.3.0) [119, 120]. Supervised 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’ [121, 122]. A priori canonical cell type marker gene expression patterns and cluster marker genes were used to assign cell types to cell clusters (see results). Cells that fell outside cell type clusters or 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) [123]. 723 maternal-maternal or fetal-fetal putative doublets were excluded after integration and clustering. Using the manually annotated Michigan (this study) and Pique-Regi (phs001886.v1.p1) cell cluster labels as the reference data, Tsang sample (EGAS00001002449) cells were algorithmically annotated with ‘SingleR’ (R package, version 1.6.1) [124] with default settings, followed by manual review. Cells with low prediction certainty (assignment score lower than three median absolute deviations of all cells assigned) were excluded as putative maternal-maternal or fetal-fetal doublets. Fetal sex in Michigan (this study) samples was determined with average normalized *XIST* expression; fetal sex in Pique-Regi and Tsang samples was determined by annotation and confirmed with average normalized *XIST* expression (**Figure 2.4**). The final analytic sample included 40,494 cells and 36,601 genes across nine biological replicates, two of which had a technical replicate (Samples 1,2) and another two included peripheral subsampling (Samples 8,9).

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

Technical correlation was assessed by Spearman 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 single-cell gene expression counts [116, 122]. 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 and a $\log_2$ fold change cutoff of 0.25. Pairwise cluster markers were identified in ‘Seurat’ with the FindMarkers function with an identical testing regime. Overexpressed genes were ranked by decreasing $\log_2$ 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, and with the g:SCS multiple testing correction method applying a significance threshold 0.05 [125].

2.3.6 In silico testing of deconvolution performance

To test the performance and robustness of our placental single-cell RNA sequencing deconvolution reference, we randomly split our analytic single-cell RNA sequencing dataset into 50% training and 50% testing subsets with balanced cell type proportions [126]. We applied the test subset with the CIBERSORTx Docker container (accessed 2021-12-07) to create a signature gene expression matrix to test deconvolution performance with default settings [127]. To evaluate the reference’s robustness to fetal sex and ability to discriminate immune cell types of fetal versus maternal origin, we generated in silico pseudo-bulk test mixtures with known

distributions of fetal and maternal cells, as well as male and female placental cells. Test mixtures included all of the 50% testing data, only fetal cells from the test data, only maternal cells from the test data, only female fetal cells from the test data, or only male cells from the test data. For the female and male fetal cell test mixtures, the baseline distribution of maternal cells was maintained by randomly down-sampling the maternal cells and randomly down-sampling the male fetal cells to the number of female fetal cells. We used the signature matrix generated from the training data to estimate constituent cell type proportions in these test mixtures using CIBERSORTx with cross-platform S-mode batch correction and 50 permutations to evaluate imputation goodness-of-fit. Pearson correlations and root mean square error between the test set predicted and actual cell type proportions in the test mixtures were used to assess deconvolution performance.

2.3.7 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 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 (**Figure 2.12**) [96, 97, 99, 128–135]. Syncytiotrophoblast fragments were enriched from villous tissue digests. We isolated cell type fractions and whole villous tissue from four healthy term, uncomplicated Cesarean sections, labelled Sorted 1 (same sample source as single-cell RNA-sequencing sample 1), Sorted 2, Sorted 3, and Sorted 4. We subjected 24 cell type fractions with sufficient RNA content to RNA-sequencing, including two cytotrophoblast, one endothelial, three extravillous trophoblast, three fibroblast, four Hofbauer cell, four leukocyte, and two syncytiotrophoblast fractions, and five whole tissue samples (**Table 2.3**).

Detailed antibody information: FITC, marker CD9: Mouse IgG1-kappa, clone HI9a (2.5 µg/mL), Biolegend #312103, lot B188319, Biolegend #312104, lot B232916; isotype control: clone MOPC-21 Biolegend #400107, Lot B199152 (2.5 µg/mL). APC, marker CD45: Mouse IgG1-kappa, clone 2D1, Biolegend #368511, Lot B215062 (0.125 µg/mL); isotype control: clone MOPC-21, Biolegend #400121, lot B216780 (0.125 µg/mL). PE/CY-7, marker HLA-ABC: Mouse IgG2a-kappa, clone W6/32, Biolegend #311429, lot B188649, Biolegend #3111430, lot B238602 (0.44 µg/mL); isotype control: clone MOPC-173, Biolegend #400231, lot B209000 (0.44 µg/mL);. BV421, marker CD31: Mouse IgG1-kappa, clone WM59, Biolegend #303123, lot B204347, Biolegend #303124, lot B232010 (0.625 µg/mL); isotype control: clone MOPC-21, Biolegend

#400157, lot B225357 (0.625 µg/mL). PE, marker HLA-G: Mouse IgG2a-kappa, clone 87G, Biolegend #335905, lot B222326, Biolegend #335906, lot B199294 (5 µg/mL); isotype control clone MOPC-173, Biolegend #400211, lot B227641 (5 µg/mL). Mouse IgG1-kappa, clone MEM-G/9, Abcam #24384 Lot GR3176304-1 (2.5 µg/mL); isotype control: monoclonal, Abcam #ab81200, lot GR267131-1 (2.5 µg/mL). 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, followed by doublet discrimination (FSC-HxFSC-W and SSC-HxSSC-W). Ax750 was used to sort on viability. Extravillous trophoblasts were isolated based on Human Leukocyte Antigen-G (HLA-G) expression. Cytotrophoblasts are HLA-ABC negative. HLA-ABC positive cells were then subjected to a CD45/CD9 gate to isolate Hofbauer cells and a heterogeneous population of leukocytes. Finally, CD45-/CD9- population is sorted into the endothelial or fibroblast bins based on CD31 expression (**Figure 2.12**).

2.3.8 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 [136] and 'MultiQC' v1.7 [137]. Reads were aligned to the GRCh38.p12/ GENCODEv28 reference transcriptome using 'STAR' v2.6.0c with default settings [113]. featureCounts from 'subread' v1.6.1 was used to quantify and summarize gene expression with default settings [138].

2.3.9 Sorted placental cell type differential expression analysis and comparison to single-cell results

For visualizations or analyses that required normalized gene counts, sorted cell type gene counts were library-size normalized with the median ratio method using the counts() function (DESeq2 R package, version 1.32.0). As recommended [139], we excluded genes that were not present in at least three samples and did not have an expression of 10 library size-normalized counts. To visualize broad cell type-specific gene expression patterns, we used 'DESeq2's (R package, version 1.32.0) plotPCA() function with the regularized logarithm transformation, blinded to experimental design. Upregulated genes in each cell type were identified using the negative binomial linear model two-tailed Wald test in 'DESeq2' (R package, version 1.32.0) 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 and a $\log_2$ fold change cutoff of 1.2

[139]. 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 the universe, excluding electronically generated annotations, and with the default g:SCS multiple testing correction method applying significance threshold adjusted p-value of 0.05 [125]. To compare sorted and single-cell results, we tabulated unique overlapping differentially expressed genes and overrepresented pathways by cell type (**Table 2.4**) Peripheral fetal and maternal immune cell types from the single-cell RNA-sequencing data were collapsed to one leukocyte category, cytotrophoblast subtypes to one cytotrophoblast category, and mesenchymal stem cells and fibroblasts to one fibroblast category for this comparison.

We used the CIBERSORTx Docker container (accessed 2021-12-07) to create a signature gene expression matrix for deconvolution from the counts of the single-cell RNA-sequencing data with the following default parameters: differential expression q-value<0.01, no minimum gene expression cutoff, and a 300 gene feature selection floor and a 500 gene feature selection ceiling [127]. We used the signature matrix to estimate constituent cell type proportions in the 4 whole tissue (with 1 additional technical replicate) and 19 sorted or enriched cell type fractions using CIBERSORTx with cross-platform S-mode batch correction and 50 permutations to evaluate imputation goodness-of-fit. We collapsed the high-resolution single-cell cell type cluster labels to the seven cell type fractions we targeted for comparison with sorted cell type results.

2.3.10 Application: 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 [111]. We used the CIBERSORTx Docker container (accessed 2021-12-07) to create a signature gene expression matrix for deconvolution from the counts of the single-cell RNA-sequencing data with the following default parameters: differential expression q-value<0.01, no minimum gene expression cutoff, and a 300 gene feature selection floor and a 500 gene feature selection ceiling [127]. We used the signature matrix to estimate constituent cell type proportions in GSE75010 using CIBERSORTx with cross-platform S-mode batch correction and 50 permutations to evaluate imputation goodness-of-fit.

2.3.11 Application: 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 regression models adjusted for gestational age, sex, and study source [140] (**Supplementary Data 8**). 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. For modelling purposes, zero percent abundance estimates were transformed to $\frac{1}{2}/n$ where n is the number of observations (n = 330) .

Differential expression analysis was conducted in limma [141] with default linear models adjusted for gestational age, fetal sex, and study source with empirical Bayes standard error moderated t-test statistics. A cell type-adjusted model was built on the

base model adjusted for gestational age, fetal sex, and study source and additionally adjusted for the first five principal components of deconvoluted cell type proportions (**Supplementary Data 9**). Statistical significance was assessed at false discovery rate-adjusted q-value<0.05 and a $\log_2$ fold change cutoff of 0.1. Differentially expressed genes were descending-ranked by value of the moderated test statistic for gene set enrichment analysis in desktop version GSEA 4.1.0 with the GSEAPreranked tool with default settings against the c5.go.bp.v7.5.1.symbols.gmt gene set database [142, 143] (**Supplementary Data 10**). Principal components analysis was performed with prcomp() from 'stats-package' (R, version 4.0.5) without scaling and with default settings.

A unified mediation and interaction analysis [144] was conducted in 'CMAverse' (R package, version 0.1.0) [145] via the g-formula approach [146] to estimate causal randomized-intervention analogues of natural direct and indirect effects [147] through direct counterfactual imputation. The model was operationalized with preeclampsia status as the binary exposure, $\log_2$ transformed gene expression intensity 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.

2.3.12 Statistics and Reproducibility

Technical replication measured by average intra-cluster gene expression between technical replicates was tested via the two-tailed Spearman correlation test

within Samples 1 and 2 assessed across all 32,738 common genes. The number of cells contributing expression data for each cell type are available in **Table 2.2**. Single-cell cluster marker genes were identified in 'Seurat' with the FindAllMarkers function with default settings on single-cell gene expression counts [116, 122]. 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 false discovery rate-adjusted p-value less than 0.05 and a $\log_2$ fold change cutoff of 0.25 (n=40,494 cells). The final analytic sample included 40,494 cells and 36,601 genes across nine biological replicates, two of which had a technical replicate (Samples 1B and 2B) and another two included peripheral subsampling (Samples 8P and 9P). Pairwise cluster markers were identified in 'Seurat' with the FindMarkers function with an identical testing regime (n=6,132 cells for proliferative vs. non-proliferative cytotrophoblasts). 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 the universe, excluding electronically generated annotations, and with the default g:SCS multiple testing correction method applying significance threshold adjusted p-value of 0.05 [125]. Overexpressed genes per cell type cluster are available in **Supplementary Data 2** and ontology results in **Supplementary Data 3**. Overexpressed genes and related enrichment results comparing proliferative to non-proliferative cytotrophoblasts are available in **Supplementary Data 1**.

Upregulated genes in each cell type were identified using the negative binomial linear model two-tailed Wald test in 'DESeq2' (R package, version 1.32.0) 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 and a $\log_2$ fold change cutoff of 1.2 [139] (n=19 cell type fraction samples with breakdown by cell type available in **Table 2.3**). 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 the universe, excluding electronically generated annotations, and with the default g:SCS multiple testing correction method applying significance threshold adjusted p-value of 0.05 [125]. Differentially expressed genes per cell type available in **Supplementary Data 4** and number of differentially expressed genes is summarized in **Table 2.4**. Ontology results are available in **Supplementary Data 5**.

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) [111]. We used the CIBERSORTx Docker container (accessed 2021-12-07) to create a signature gene expression matrix for deconvolution from the counts of the single-cell RNA-sequencing data with the following default parameters: differential expression q-value<0.01, no minimum gene expression cutoff, and a 300 gene feature selection floor and a 500 gene feature selection ceiling [127]. We used the signature matrix to estimate

constituent cell type proportions in GSE75010 using CIBERSORTx with cross-platform S-mode batch correction and 50 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 [140]. Cell types imputed at zero percent abundance across all samples were excluded (n=2 excluded: fetal naïve CD4+ T cells and fetal GZMB+ Natural Killer cells). 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 [141] with default settings using linear models (n=14,651 genes) 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 and a $\log_2$ fold change cutoff of 0.1. Differentially expressed genes were descending-ranked by the value of the moderated test statistic for gene set enrichment analysis in desktop version GSEA 4.1.0 with the GSEAPreranked tool with default settings against the c5.go.bp.v7.5.1.symbols.gmt gene set database [142, 143].

A unified mediation and interaction analysis [144] was conducted in 'CMAverse' (R package, version 0.1.0) [145] via the g-formula approach [146] to estimate causal

randomized-intervention analogues of natural direct and indirect effects [147] through direct counterfactual imputation. The model (n=330) was operationalized with preeclampsia status as the binary exposure, normalized $\log_2$ gene expression signal intensity 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.

2.4 Results

2.4.1 Single-cell gene expression map of healthy placental villous tissue

A conceptual layout of the laboratory methods and analyses contained within this manuscript is provided in **Figure 2.1**. From healthy term placental villous tissue, 9,244 cells across a total of two biological replicates and two technical replicates were sequenced and analyzed (Michigan sample). These data were combined with single-cell RNA-sequencing data of 5,911 cells from three healthy term villous tissue samples in a previously published study (Pique-Regi sample) [102] and 25,339 cells from four healthy term villous tissue samples in another previously published study, two of which were subsampled with an additional peripheral placental villous tissue sample (Tsang sample) [101] (**Table 2.1**). Cells were excluded if they had low RNA content (<500 unique RNA molecules), few genes detected (<200), or were doublets or outliers in mitochondrial gene expression (**Figure 2.2**, **Figure 2.3**). Fetal or maternal origin of cells was determined by genetic variation in sequencing data. Fetal sex was determined by

XIST expression (**Figure 2.4**). The final analytic sample included 40,494 cells and 36,601 genes across nine biological replicates, two of which had a technical replicate and another two included peripheral subsampling.

Uniform manifold approximation and projection (UMAP) [118] was used to visualize sequencing results in two dimensions with mutual nearest neighbor batch correction [119] (**Figure 2.5**). Cells clustered into 19 fetal and 8 maternal cell types with 84.4% of all cells being of fetal origin (**Figure 2.5a**). Cell type clustering decisions balanced cluster stability, resolution, and biologic plausibility with prior knowledge. If desired, downstream analyses could collapse cell subtypes into a single, more general cell type cluster. We observed placenta-specific trophoblast cell types including cytotrophoblasts (*KRT7*), proliferative cytotrophoblasts (*KRT7*, *STMN1* and other proliferation-related genes) [148], extravillous trophoblasts (*HLA-G*) [131], and syncytiotrophoblasts (*PSG4* and other pregnancy-specific hormone genes) (**Figure 2.6a**) [149]. Proliferative cytotrophoblasts were distinguished from other cytotrophoblasts by overexpression of genes related to cytoplasmic translation ($p_{adj}=8.1x10^{-15}$) and mitotic sister chromatin segregation ($p_{adj}=1.5x10^{-12}$), indicative of their proliferative phenotype (**Figure 2.7**). Other fetal-specific cell types included mesenchymal stem cells (*COL1A1*^{lo}, *TAGLN*^{lo}, *LUM*^{hi}), fibroblasts (*COL1A1*^{hi}, *TAGLN*^{hi}, *LUM*^{lo}) [150], endothelial cells (*PECAM1*) [151], and Hofbauer cells (*CD163*) [96] (**Figure 2.5b**).

Fetal and maternal lymphocytes, B cells, and monocytes were also captured (**Figure 2.5b-c**). We observed fetal and maternal B cells (*CD79A*) [152], maternal plasma cells (*XBP1*, *IGHA* and other immunoglobulins) [153]. We also observed fetal

and maternal CD14+ monocytes (*CD14*$^{+}$/*FCGR3A*$^{-}$), maternal CD16+ monocytes (*CD14*$^{+}$/*FCGR3A*$^{+}$) [154], and a small population of fetal plasmacytoid dendritic-like cells (*FLT3*$^{+}$/*ITM2C*$^{+}$) [155, 156]. We further observed fetal and maternal natural killer cells (*NKG7*) and fetal GZMB+ or GZMK+ natural killer cell subtypes, and fetal natural killer T cells (*NKG7*$^{+}$/*CD3E*$^{+}$/*CD8A*$^{-}$) [157, 158]. Finally, we observed a variety of T cell subtypes: naïve CD4+ (*CCR7*, *CD3E*, *CD4*), naïve CD8+ (*CCR7*, *CD3E*, *CD8A*), memory CD4+ (*S100A4*, *CD3E*, *CD4*, *IL2*, *CCR7*^{lo}), and activated CD8+ T cells (*NKG7*, *CD3E*, *CD8A*) (**Figure 2.6b**) [159].

To identify upregulated genes in each cell type, we compared the expression of a gene in one cell type against that gene's average expression in all other cell types (**Supplementary Data 2**). Consequently, the same genes could be upregulated across several cell types of similar lineage. *FLT1* expression was highly upregulated in extravillous trophoblasts ($\log_2$ fold change (FC)=3.89, p_{adj}<0.001). Trophoblast cell types had the largest and most diverse transcriptomes, characterized by the largest number of unique RNA transcripts and detected genes per cell (**Figure 2.8**). Functional analysis of upregulated genes revealed cell type specific biological processes (**Supplementary Data 3**). For example, fetal extravillous trophoblasts were enriched for genes relevant to placental structure and function such as cell migration (p_{adj}<0.001) and response to oxygen levels (p_{adj}<0.001) and syncytiotrophoblasts were enriched for genes involved in steroid hormone biosynthetic process (p_{adj}<0.001). Technical replication in Michigan samples 1 and 2 appeared high in UMAP space (**Figure 2.9**). Indeed, the average intra-cluster gene expression between technical replicates had an

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

2.4.2 Single-cell RNA sequencing deconvolution reference exhibits excellent in silico performance

Based on the single cell data, we created a placental signature gene matrix that incorporated expression information across an algorithmically selected 5,229 signature genes to estimate the cellular composition of 27 fetal and maternal cell types from whole tissue gene expression data (**Figure 2.10**). To test the performance and robustness of this placental single-cell RNA sequencing deconvolution reference, we randomly split our analytic single-cell RNA sequencing dataset into 50% training and 50% testing subsets with balanced cell type proportions [126]. The same training dataset was used for each comparison; test mixtures were generated from the testing half of the dataset. Using a signature gene expression matrix generated from the training data, we estimated cell type composition in in silico pseudo-bulk testing data mixtures of known cell type composition with varying contributions of fetal vs. maternal origin cells and male vs. female fetal cells (**Figure 2.11**). In all mixtures, the 27 predicted and actual cell type proportions were correlated (p-value<0.001 for each test). In the primary deconvolution analysis of all cell types at their natural rates (n=20,242), estimated and actual cell type proportions had a Pearson correlation coefficient of 0.956 (95% CI [0.904, 0.980]). The worst performance was under the unrealistic scenario that the mixture was composed entirely of maternal cell types (n=3,162) with a Pearson correlation of 0.734 (95% CI [0.491, 0.871]) between actual estimated cell type proportions. Our deconvolution reference was also robust to fetal sex when only male

fetal cells (Pearson correlation = 0.893, 95% CI [0.776, 0.950]) were included (n=8,394), or only female fetal cells (Pearson correlation = 0.983, 95% CI [0.964, 0.993]) (n=8,394). Together, these results show that our reference panel can successfully deconvolute placental tissues, though some maternal cell types common to both mother and fetus may be erroneously labelled fetal in the absence of fetal cells of those cell types.

2.4.3 Fluorescence-activated cell sorting of major placental cell types yielded mixed cell type isolation results

We isolated whole 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 Sorted 1 (same sample source as single-cell RNA sequencing sample 1), Sorted 2, Sorted 3, and Sorted 4 (**Figure 2.12**, **Table 2.3**). For analysis, as recommended [139], we excluded 19,048 genes that were not present in at least 3 samples and an additional 865 genes that did not have a cumulative library size-normalized count of at least 10. Principal components analysis of whole-transcriptome sorted-cell bulk RNA sequencing normalized counts is provided in **Figure 2.13**.

To identify upregulated genes in each cell type, we compared the expression of a gene in one cell type against that gene's average expression in all other cell types (**Figure 2.14**). Consequently, the same genes could be upregulated across several cell types of similar lineage. All 38,468 uniquely mapping genes were tested. 746 genes were algorithmically 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 Data 4**). Functional analysis of upregulated genes revealed cell type specific biological processes (**Supplementary Data 5**). For example, syncytiotrophoblasts were enriched for genes relevant to placental structure and function such as angiogenesis, cell-substrate adhesion, and regulation of epithelial cell proliferation (p_{adj}<0.001). To compare sorted and single-cell differential expression and enrichment results, we tabulated the number of unique genes and pathways overlapping between the two analyses after collapsing the single-cell cell type cluster labels to the seven cell type fractions that we had targeted to isolate for downstream analyses (**Table 2.4**). On average, 15.0% of single-cell upregulated genes and 5.9% of enriched pathways were also identified among the sorted cell data. On average, 17.5% of sorted cell type upregulated genes and 39.2% of pathways were also identified among the single-cell data. Sorted endothelial cells results were limited due to the limited number of biological replicates.

We applied the single-cell deconvolution reference to estimate cell proportions in the 4 whole tissue (with 1 additional technical replicate) and 19 sorted or enriched cell type fractions. We collapsed the single-cell cell type cluster labels to the seven cell type fractions we targeted for isolation for downstream analyses (**Supplementary Data 6 – Sheet 1).** All deconvoluted samples exhibited high goodness-of-fit between original bulk mixtures and the estimated cell type proportion mixtures (p-values<0.001). Among the signature genes, original bulk and estimated cell type fractions had a Pearson correlation (mean $\pm$ standard deviation) of 0.73 $\pm$ 0.11 and root mean square error of 0.88 $\pm$ 0.04 (**Supplementary Data 6 – Sheet 2)**. Deconvolution results (mean $\pm$

standard deviation) suggest we successfully isolated fibroblast- (n=3, 74.7% ± 0.6%) and leukocyte-enriched (n=4, 82.3% ± 24.8%) cell type fractions. Other cell type targets were less successful (range 0%-26% estimated purity). The Hofbauer cell fraction was predicted to be mostly leukocytes (n=4, 91.5% ± 0.5%).

2.4.4 Cell proportion deconvolution of bulk placental tissue preeclampsia dataset

We applied the single-cell deconvolution reference to estimate cell proportions from bulk placental tissue in 157 preeclampsia cases and 173 controls [111] compiled from eight previously published studies [111, 160–166]. Mean gestational age was 2.2 weeks younger in cases than controls (p-value<0.001, **Table 2.5**). All deconvoluted samples exhibited high goodness-of-fit between original bulk mixtures and the estimated cell type proportion mixtures (p-values<0.001). Among the signature genes, original bulk and estimated mixtures had a Pearson correlation (mean ± standard deviation) of 0.70 ± 0.04 and root mean square error of 0.73 ± 0.03 (**Supplementary Data 7**). Fetal naïve CD4+ T cells and fetal GZMB+ Natural Killer cells were estimated to be at 0% abundance in all samples and were dropped from downstream analyses. Cytotrophoblasts were the most abundant (mean ± standard deviation) estimated fetal cell type (27.9% ± 4.3%) followed by syncytiotrophoblasts (23.4% ± 5.0%) and mesenchymal stem cells (10.3% ± 3.3%). The most common maternal cell types were naïve CD8+ T cells (2.8% ± 1.5%), plasma cells (1.4% ± 0.7%), and B cells (1.3% ± 0.8%). A comparison of deconvoluted whole tissue cell type proportions among healthy individuals (**Figure 2.15**) between the microarray dataset GSE75010 (n=173 controls), our whole tissue bulk RNA-sequencing samples (Sorted samples 1-4), and the single-cell dataset compiled here (Single-cell samples 1-9) suggests syncytiotrophoblasts and

endothelial cells are underrepresented in the single-cell data. This is likely due to dissociation bias, which has been commonly observed in single-cell assays of other tissues [72]. Overall, the Pearson correlation of the average deconvoluted cell type proportion across the 27 cell types between healthy bulk RNA-sequencing and microarray controls was 0.80 (95% CI: [0.60, 0.91]).

2.4.5 Differentially abundant cell type proportions in preeclampsia cases versus controls

To test for differences in cell proportions between preeclampsia cases and controls (**Figure 2.16**), we fit beta regression models for each cell type proportion adjusted for study source, fetal sex, and gestational age to estimate the prevalence odds ratio for each cell type (**Supplementary Data 8**). Among fetal cell types, extravillous trophoblasts ($p<0.001$), memory CD4+ T cells ($p=0.007$), CD8+ activated T cells ($p=0.005$), and natural killer T cells ($p=0.006$) were more abundant (**Figure 2.17**) in preeclampsia cases relative to controls. The unadjusted median extravillous trophoblast abundance was 6.4% among cases compared to 2.1% among controls. Mesenchymal stem cells (median percent composition in cases vs. controls, 8.8% vs. 11.0%), Hofbauer cells (2.7% vs. 4.4%), and fetal naive CD8+ T Cells (4.2% vs. 4.5%) were all less abundant among preeclampsia cases compared to controls ($p<0.001$). Among maternal cell types, maternal plasma cells (1.6% vs. 1.2%) were more abundant among preeclampsia cases compared to controls ($p<0.001$).

2.4.6 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 linear differential gene expression models adjusted for covariates study source, fetal sex, and gestational age with and without adjustment for deconvoluted cell type proportions. To reduce the number of model covariates and account for dependence between deconvoluted cell type proportions, we applied principal components (PC) analysis to deconvoluted cell type proportions. The first five PCs accounted for 87.2% of the variance in deconvoluted cell type proportions and were added as additional covariates to form the cell type-adjusted model. Variation in PCs 1 and 2 was largely driven by syncytiotrophoblasts (33.8%), extravillous trophoblasts (33.5%), and cytotrophoblasts (15.3%) proportions and provided some separation between cases from controls (**Figure 2.18a, c**). Variation in PC3 was largely driven by cytotrophoblasts (50.1%) and to a lesser extent syncytiotrophoblasts (16.6%), mesenchymal stems cells (14.5%), and extravillous trophoblasts (13.7%) (**Figure 2.18b, d**).

In the cell type-naïve base models (n=14,651 genes, 173 controls, and 157 cases) adjusted for study source, gestational age, and fetal sex, 550 genes were differentially upregulated and 604 were downregulated in preeclampsia cases versus controls (**Figure 2.19a, Supplementary Data 9**). Gene set enrichment analysis of biological processes identified 41 overrepresented pathways in the base model (**Figure 2.20a, Supplementary Data 10**). Biological process pathways such as aerobic respiration (false discovery adjusted $q<0.001$), mitochondrial respiratory chain complex assembly ($q<0.001$), glutathione metabolism ($q=0.003$), and ribosome biogenesis ($q=0.001$) were overrepresented among downregulated genes. No pathways were

overrepresented among upregulated genes, though intermediate filament organization (q=0.26), keratinocyte differentiation (q=0.77), and endothelial cell development (q=0.42) had comparable enrichment scores. 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 2.19b, Supplementary Data 9**). Of the cell type-adjusted results, 19 pathways were overrepresented (**Figure 2.20b, Supplementary Data 10**). Downregulation of mitochondrial respiratory chain complex assembly (q<0.001), aerobic respiration (q=0.001), ribosome biogenesis (q=0.001) and glutathione metabolism (q=0.02) were overrepresented among downregulated genes. Detection of chemical stimulus involved in sensory perception of smell (q=0.04) and non-coding RNA processing (q=0.04) were also overrepresented pathways among downregulated genes. Neuroepithelial cell differentiation (q=0.04) was overrepresented among upregulated genes. Vascular endothelial growth factor receptor signaling pathway (q=0.15), of which *FLT1* is a member, had an enrichment score of 1.77 (up from 1.34, q=0.43 in the base model) but did not meet the q-value cutoff. Overall, downregulation of mitochondrial biogenesis, aerobic respiration, and ribosome biogenesis and related pathways were robust to cell type proportion adjustment.

2.4.7 Differential expression of preeclampsia-associated genes mediated by placental cell type proportions

Overexpression of *FLT1* in placental tissue [167–170], detection of a soluble isoform of *FLT1* in maternal circulation [171, 172], and fetal genetic variants near *FLT1* [173] 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 cellular composition (overall proportion attributable to interaction = -5.8%, 95% CI [-17.1%, 5.0%]). We therefore dropped interaction parameters from the model for the final analysis. In the model without interaction, 37.8% (95% CI [27.5%, 48.8%]) of the 1.05 (95% CI [0.89, 1.21]) $\log_2$ signal intensity increase in the association between preeclampsia and *FLT1* expression was attributable to differences in placental cell composition between preeclampsia cases and controls (**Figure 2.21**). Overexpression of *LEP* and *ENG* have also been associated with preeclampsia [167–170]. Mediation results were similar for *LEP* (total effect = 2.62 (95% CI [2.26, 2.97] $\log_2$ signal intensity increase; proportion mediated = 34.5% (95% CI [26.0%, 44.9%]) and *ENG* (total effect = 0.93 (95% CI [0.79, 1.07] $\log_2$ signal intensity increase; proportion mediated = 34.5% (95% CI [25.0%, 45.3%]).

2.5 Discussion

To create the largest, publicly available placental RNA 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 previously published studies, and built a signature gene matrix for deconvolution of bulk villous tissue gene expression data. In silico testing of our deconvolution reference demonstrated successful and robust deconvolution. To

compare single-cell placental cell type expression profiles to more conventional sorting methods, we created a fluorescence-activated cell sorting scheme to enrich and sequence RNA from five important placental cell types as well as syncytiotrophoblasts. Deconvolution of sorted cell type fractions with the single-cell deconvolution reference suggested most conventionally sorted cell types are far less pure than what can be accomplished with clustering and aggregation of single cell results, and at much lower cell type resolution. We applied the single-cell deconvolution reference to estimate 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 were markedly attenuated after adjustment for cell type proportions. Preeclampsia-associated pathways, including downregulation of mitochondrial biogenesis, aerobic respiration, and ribosome biogenesis were robust to cell type adjustment, suggesting direct changes to these pathways. Finally, to quantify the attenuation of differential expression of the preeclampsia biomarkers *FLT1*, *LEP*, and *ENG*, we applied mediation analysis to show cellular composition mediated a substantial proportion of the association between preeclampsia and *FLT1*, *LEP*, and *ENG* overexpression. Cell type proportions may be an important and often overlooked 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 and possibly only reference available 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 [95], cytotrophoblasts [98], and extravillous trophoblasts [99]. 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 [102] 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. The low representation of some cell types such as trophoblasts in our single-cell RNA-sequencing results from the Michigan study suggest that these cell types may be especially sensitive to dissociation and disintegrate before transcript capture, commonly referred to as dissociation bias [72]. Michigan samples 1 and 2 also included a cryopreservation step like those employed in large-scale epidemiological studies that may have exacerbated dissociation bias [174]; this applies to both single-cell and sorted cell type experiments. Future studies may propose alternative approaches to perform unbiased single-cell RNA sequencing in placental tissues; indeed, single nucleus RNA sequencing has been used to characterize syncytiotrophoblast and may be more appropriate to assay such cell types sensitive to dissociation procedures [175]. We verified that our deconvolution reference exhibited strong performance even with extremely imbalanced and unlikely real-world test mixture distributions by fetal sex and maternal cell type representation.

Our preeclampsia findings are consistent with prior pathophysiological understanding of the disorder, linking cell type proportion estimates and gene expression data in bulk tissue. Among preeclampsia cases, we observed an elevated proportion of extravillous trophoblasts and underrepresentation of stromal cell types, which may reflect an arrest in conventional placental cell type differentiation and maturation following insufficient uterine spiral artery remodeling implicated in preeclampsia [176–178]. A recent study of bulk placental gene expression across trimesters suggests that Hofbauer cells may more abundant in the 2nd trimester compared to the 3rd, possibly to support vasculogenesis, though this study involved a small deconvolution reference that contained a limited variety of cell types [179]. A better understanding of the evolution of temporal placental composition changes may yield greater insight into placental pathologies. In the cell type-naïve differential expression model, consistent

with previous findings, placentas from pregnancies with preeclampsia overexpressed *FLT1*, *LEP*, and *ENG* [167–170]. In our cell type-adjusted model, *FLT1* and *LEP* remained only nominally significant whereas *ENG* did not meet the nominal significance threshold. Mediation analysis confirmed that a significant proportion of *FLT1*, *LEP*, and *ENG* overexpression was attributable to differences 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*, *LEP*, and *ENG* expression [167–170]. Downregulation of mitochondrial biogenesis, aerobic respiration, and ribosome biogenesis were robust to cell type adjustment, indicating direct changes to these pathways beyond shifts in cell type abundance. Indeed, disruption of the mitochondrial fission-fusion cycle [180], malperfusion [181, 182], and inhibited protein synthesis secondary to endoplasmic reticulum stress [183, 184] have all previously been associated with preeclampsia. Interestingly, cell type adjustment increased the enrichment score results of vascular endothelial growth factor receptor signaling pathway, a mechanistic hypothesis in preeclampsia etiology [171, 181, 185, 186], from 1.36 to 1.77 (q=0.43 to q=0.15). This approach may reveal the biological mechanisms of other diseases beyond cellular composition differences. 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* [187] were not differentially expressed at the tissue level in our analysis. 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 [188, 189]. A similar study revealed decreased activity of gene network modules regulated by transcription factors *ATF3*, *CEBPB*, and *GTF2B* and decreased expression of *CEBPB* and *GTF2B* in preeclamptic extravillous trophoblasts compared to controls; follow-up *in vitro* experiments suggested *CEBPB* and *GTF2B* knockdown reduced extravillous trophoblast viability and invasion [190]. 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 between preeclampsia cases and controls [188]. Future work should consider and account for cell type proportions and 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. We profile the parenchymal healthy term villous tissue in the placenta and integrate our dataset with samples from previously published studies to generate the largest, to the best of our knowledge, cell type-specific placental villous tissue gene expression reference to date. Single-cell RNA-sequencing allowed us to agnostically capture diverse placental cell types without a priori knowledge of cell types and their characteristics and tabulate gene expression patterns at high resolution and specificity. Our in silico deconvolution tests demonstrated robust performance to even extreme distributions of maternal or sex of fetal cells. We demonstrate technical replication of single-cell RNA-sequencing in placental villous tissue. We were able to apply our findings to a large target deconvolution dataset of preeclampsia that contained placental measures from hundreds of participants across eight different studies. Most importantly, we evaluate cell type proportion differences in an epidemiological study of placental parenchymal tissue and preeclampsia, and genome-wide gene expression differences accounting for cell type heterogeneity, a critical limitation in bulk tissue assays.

This study also has several limitations. Although our cellular sample size comprised of 40,494 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 compared to epidemiologic scale studies. Our newly sequenced samples came from a convenience sample without available demographic information beyond uncomplicated and 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 or exhibited poor estimated

purity, likely complicated by degradation of cell surface markers from apoptosis characteristic of development and parturition [191, 192] and sample processing. This study did not include placental tissues for single cell analysis from preeclamptic patients to confirm intra-cell type gene expression changes. Despite excellent in silico performance, we had no external gold standard to verify deconvolution performance. This deconvolution reference may not be sensitive enough to discriminate between cell subtypes such as proliferative vs. non-proliferative cytotrophoblasts that are clearly delineated in the single-cell analysis; in such cases, investigators may collapse cell type proportions counts into a single major cell type group, such as cytotrophoblasts. 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 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 [193, 194]. Therefore, our results likely represent a conservative underestimate of the effects of preeclampsia on inappropriate cell composition and preeclampsia status on gene expression.

In summary, we provide a cell type-specific deconvolution reference via single-cell RNA-sequencing in the parenchymal placental term villous tissue. We demonstrated this reference was robust to different distributions of maternal and fetal sex through in silico validation testing. 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 bulk 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 mitochondrial biogenesis, aerobic respiration, and ribosome biogenesis were robust to cell type adjustment, suggesting direct changes to these pathways. A substantial proportion of the overexpression of the *FLT1*, *LEP*, and *ENG* in preeclampsia was mediated by placental cell composition. 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.

2.6 Figures and Tables

Table 2.1 Summary of single-cell RNA-sequencing sample characteristics and sequencing quality metrics. This study collected Samples 1-2. Samples 3-5 were downloaded from Pique-Regi et al., 2019 [102]. Samples 6-9P were download from Tsang et al., 2017 [101]. For Samples 1-2, A/B pairs are technical replicates. For Samples 8-9, C/P pairs represent centrally (C) or peripherally (P) sampled villous tissue from the same placenta.

		Pre-filtering Quality Control Metrics					Post-filtering Quality Control Metrics							
Sample	Fetal Sex	Droplets Sequenced	Total Unique RNA Molecules	Total Unique Genes Detected	Total Cells	Maternal - Fetal Doublets Removed	Unique RNA Molecules (Median)	Unique RNA Molecules (IQR)	Unique Genes (Median)	Unique Genes (IQR)	Percent Mitochondria Gene Expression (Median)	Percent Mitochondria Gene Expression (IQR)	Putative Doublets Removed	Cells in Final Analytic Sample
1A	F	737,280	15,329,288	32,738	2,573	28	4,021	2,717	1,247	442	4.1	1.98	116	2,214
1B	F	737,280	14,777,010	32,738	2,600	33	3,870	2,521	1,189	426	4.06	1.85	134	2,280
2A	M	737,280	14,306,604	32,738	2,544	25	3,988	2,448	1,171	430	2.86	1.48	92	2,292
2B	M	737,280	14,799,594	32,738	2,740	29	3,875	2,410	1,157	427	2.85	1.56	109	2,458
3	M	737,280	17,075,126	36,601	1,907	0	3,556	6,830	1,292	1,781	2.76	2.89	105	1,620
4	F	737,280	28,250,436	36,601	2,653	4	5,352	12,664	1,833	2,869	3.73	3.79	119	2,081
5	M	737,280	29,693,207	36,601	2,456	1	6,544	10,889	2,186	2,569	2.25	1.94	136	2,210
6	M	737,280	71,470,103	36,601	6,018	401	1,639	1,996	733	716	3.02	4.2	619	3,765
7	M	737,280	18,472,614	36,601	16,968	206	1,617	3,237	687	844	2.66	2.91	3,860	10,679
8C	F	737,280	32,463,561	36,601	4,918	5	1,880	1,986	793	631	3.09	3.72	533	3,726
8P	F	737,280	62,317,586	36,601	2,284	3	1,302	1,608	585	567	3.35	4.24	219	1,651
9C	F	737,280	20,847,761	36,601	3,137	0	2,173	2,242	875.5	690	1.92	2.31	247	2,508
9P	F	737,280	64,935,823	36,601	3,612	1	2,073.5	2,088	879	648	2.06	2.97	208	3,010

Table 2.2 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. Proportions represent the overall proportion of that cell type in the dataset or among cells of only fetal or maternal origin. The final analytic sample included 40,494 cells and 36,601 genes across nine biological replicates, two of which had a technical replicate (Samples 1B and 2B) and another two included peripheral subsampling (Samples 8P and 9P).

Distribution of cell types by sample and fetal/maternal origin

Cell Type	1A	1B	2A	2B	3	4	5	6	7	8C	8P	9C	9P	Count	Overall Proportions	Fetal Origin	Maternal Origin
Fetal B Cells	109	149	195	229	38	27	15	0	35	9	1	1	1	809	2.0%	2.4%	-
Fetal CD14+ Monocytes	232	187	174	197	45	23	42	1	11	3	0	1	0	916	2.3%	2.7%	-
Fetal CD8+ Activated T Cells	220	205	148	162	78	76	6	0	29	1	2	0	12	939	2.3%	2.7%	-
Fetal Cytotrophoblasts	3	7	0	0	350	859	841	261	374	82	203	163	359	3,502	8.6%	10.3%	-
Fetal Endothelial Cells	6	11	22	15	1	4	3	1,058	180	181	84	236	137	1,938	4.8%	5.7%	-
Fetal Extravillous Trophoblasts	0	0	0	0	6	222	3	234	3,336	251	97	5	2	4,156	10.3%	12.2%	-
Fetal Fibroblasts	1	0	0	3	25	17	50	335	246	179	46	75	51	1,028	2.5%	3.0%	-
Fetal GZMB+ Natural Killer	10	7	7	7	48	16	6	1	9	3	1	0	4	119	0.3%	0.3%	-
Fetal GZMK+ Natural Killer	7	13	7	11	106	90	10	0	4	2	0	0	0	250	0.6%	0.7%	-
Fetal Hofbauer Cells	23	27	115	136	107	114	266	240	607	932	326	49	174	3,116	7.7%	9.1%	-
Fetal Memory CD4+ T Cells	26	41	42	45	68	54	28	11	8	3	1	5	0	332	0.8%	1.0%	-
Fetal Mesenchymal Stem Cells	53	46	54	60	377	11	501	1,207	4,127	1,806	606	1,633	1,755	12,236	30.2%	35.8%	-
Fetal Naive CD4+ T Cells	222	215	225	213	30	55	25	10	62	5	6	7	2	1,077	2.7%	3.2%	-
Fetal Naive CD8+ T Cells	47	30	100	117	17	37	9	1	12	0	4	0	0	374	0.9%	1.1%	-
Fetal Natural Killer T Cells	30	34	22	32	51	42	8	0	1	1	0	0	0	221	0.5%	0.6%	-
Fetal Nucleated Red Blood Cells	0	0	1	1	10	3	0	2	14	1	1	1	0	34	0.1%	0.1%	-
Fetal Plasmacytoid Dendritic Cells	3	1	2	7	21	12	7	11	16	4	3	10	7	104	0.3%	0.3%	-
Fetal Proliferative Cytotrophoblasts	1	0	1	0	111	185	282	334	680	141	211	231	453	2,630	6.5%	7.7%	-
Fetal Syncytiotrophoblast	7	18	6	3	6	73	0	34	133	3	9	54	38	384	0.9%	1.1%	-
Maternal B Cells	118	144	347	339	2	22	0	0	6	2	2	0	0	982	2.4%	-	15.5%

Maternal CD14+ Monocytes	271	282	181	212	21	6	8	0	122	3	1	0	1	1,108	2.7%	-	17.5%
Maternal CD8+ Activated T Cells	449	435	266	277	33	47	4	0	41	29	6	20	2	1,609	4.0%	-	25.4%
Maternal FCGR3A+ Monocytes	97	114	40	48	31	9	78	25	521	67	40	11	10	1,091	2.7%	-	17.2%
Maternal Naive CD4+ T Cells	52	42	56	66	11	20	8	0	31	7	1	3	1	298	0.7%	-	4.7%
Maternal Naive CD8+ T Cells	74	95	181	172	8	41	5	0	11	0	0	0	0	587	1.4%	-	9.3%
Maternal Natural Killer Cells	114	139	68	60	19	8	3	0	56	10	0	3	1	481	1.2%	-	7.6%
Maternal Plasma Cells	39	38	32	46	0	8	2	0	7	1	0	0	0	173	0.4%	-	2.7%
Count	2,214	2,280	2,292	2,458	1,620	2,081	2,210	3,765	10,679	3,726	1,651	2,508	3,010	40,494	100.0%	34,165	6,329

Table 2.3 Fluorescence-activated cell sorting and RNA-sequencing quality control results. Each sample was sorted into six cell type populations with a matched whole tissue sample. Cell count describes the total number of FACS-sorted cells. Table columns describe total RNA given in nanograms with RNA integrity index score (RIN), fastQC pass/fail, and whether the sample was sequenced in the paired-end (PE) or single-end (SE) and included in the experiment based on RIN score and total RNA (all sequenced samples included).
**Matched with sample 1 from the single-cell RNA-sequencing assay.*

Sample ID	Cell Type	Cell Count	Total RNA (ng)	RIN	fastQC	Sequenced (PE/SE)
Sorted 1*	Syncytiotrophoblast	N/A	0.2	1	N/A	Dropped
	Hofbauer	2.00E+05	4.2	5	Pass	SE
	Leukocyte	1.57E+05	6.9	7.3	Pass	SE
	Extravillous Trophoblast	8.35E+03	1.1	6.4	Pass	SE
	Cytotrophoblast	2.83E+05	1.8	8.1	Pass	SE
	Fibroblast	1.55E+04	34	7.3	Pass	SE
	Endothelial Cells	7.60E+03	3	2.8	N/A	Dropped
	Whole Tissue	N/A	108	4.9	Pass	SE
	Whole Tissue	N/A	209	7.8	Pass	PE
Sorted 2	Syncytiotrophoblast	N/A	0.2	1	N/A	Dropped
	Hofbauer	1.57E+05	1.6	4.9	Pass	SE
	Leukocyte	1.37E+05	63	8.3	Pass	SE
	Extravillous Trophoblast	3.60E+03	1.1	6	Pass	SE
	Cytotrophoblast	8.00E+04	<0.1	<0.1	N/A	Dropped
	Fibroblast	1.06E+04	35	8.1	Pass	SE
	Endothelial Cells	2.30E+03	2	8.1	Pass	SE
	Whole Tissue	N/A	332	8.7	Pass	SE
Sorted 3	Syncytiotrophoblast	N/A	255	4	Pass	SE

	Hofbauer	1.02E+05	45	9.1	Pass	SE
	Leukocyte	8.30E+04	50	8.5	Pass	SE
	Extravillous Trophoblast	2.50E+04	7.2	2.8	Pass	SE
	Cytotrophoblast	3.72E+05	3.7	2.4	N/A	Dropped
	Fibroblast	3.61E+03	2.9	1.3	N/A	Dropped
	Endothelial Cells	2.97E+03	3	2.5	N/A	Dropped
	Whole Tissue	N/A	9572	5.3	Pass	SE
Sorted 4	Syncytiotrophoblast	N/A	107	6.5	Pass	PE
	Hofbauer	8.40E+04	87	6.7	Pass	PE
	Leukocyte	2.20E+04	0.82	7.6	Pass	PE
	Extravillous Trophoblast	4.50E+04	2.4	8	Pass	PE
	Cytotrophoblast	2.56E+05	1.5	6.9	Pass	PE
	Fibroblast	1.70E+04	1.7	7.2	Pass	PE
	Endothelial Cells	1.20E+04	0.14	1	Pass	Dropped
	Whole Tissue	N/A	298	8.2	Pass	PE

Table 2.4 Number of overlapping differentially upregulated genes and overrepresented biological process pathways between the single- and sorted cell type differential expression and enrichment testing analyses.

Cell Type	Single-cell Differentially Expressed Genes	Sorted Differentially Expressed Genes	Overlapping Genes	Percentage overlap (single-cell denominator)	Percentage overlap (sorted denominator)
Cytotrophoblast*	2011	366	13	0.6%	3.6%
Endothelial Cells	769	251	1	0.1%	0.4%
Extravillous Trophoblasts	693	1073	18	2.6%	1.7%
Fibroblasts**	649	3883	310	47.8%	8.0%
Hofbauer Cells	297	2210	77	25.9%	3.5%
Leukocytes***	3378	2262	486	14.4%	21.5%
Syncytiotrophoblast	206	1154	28	13.6%	2.4%
Mean:	1,143	1,600	133	15.0%	5.9%

Cell Type	Single-cell Overrepresented Pathways	Sorted Overrepresented Pathways	Overlapping Pathways	Percentage overlap (single-cell denominator)	Percentage overlap (sorted denominator)
Cytotrophoblast*	98	32	12	12.2%	37.5%
Endothelial Cells	243	6	0	0.0%	0.0%
Extravillous Trophoblasts	59	18	8	13.6%	44.4%
Fibroblasts**	142	51	25	17.6%	49.0%
Hofbauer Cells	242	343	160	66.1%	46.6%
Leukocytes***	1031	53	51	4.9%	96.2%
Syncytiotrophoblast	13	233	1	7.7%	0.4%
Mean:	261	105	37	17.5%	39.2%

*Single-cell subtypes cytotrophoblasts and proliferative cytotrophoblasts collapsed to single category
**Single-cell subtypes fibroblasts and mesenchymal stem cells collapsed to single category
***Single-cell peripheral immune cell subtypes collapsed to single category

Table 2.5 Demographic characteristics of eight previously published bulk microarray placental gene expression case-control studies (accessed through GSE75010) for deconvolution application testing.

Descriptives statistics of microarray preeclampsia case-control studies

	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%)	

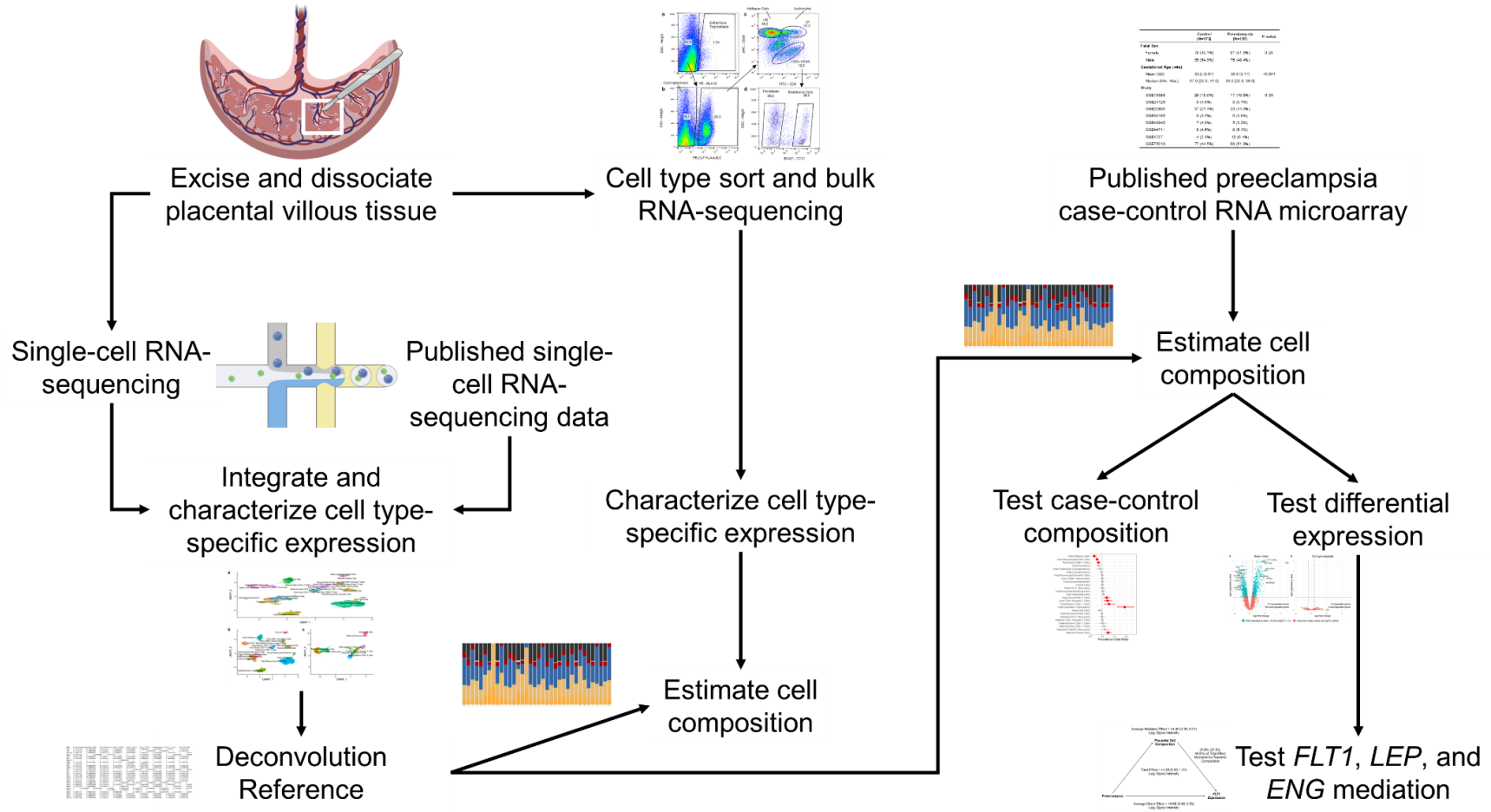


Figure 2.1 Conceptual layout of the laboratory methods and analyses contained within this manuscript. Created with BioRender.com.

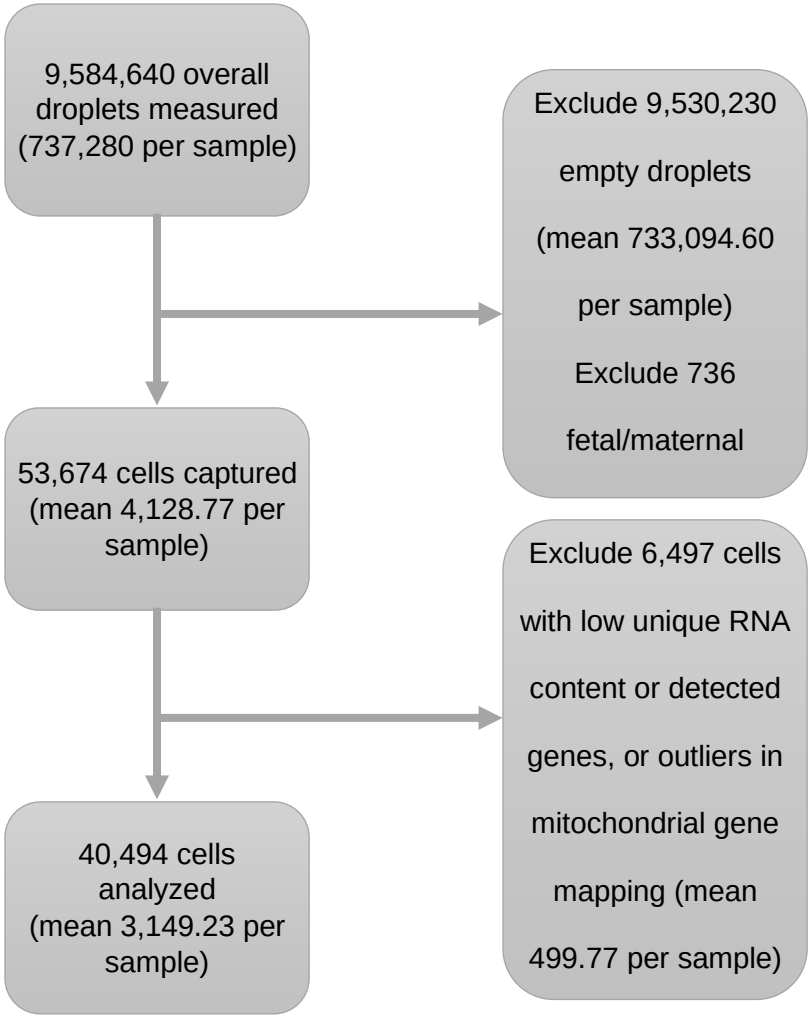


Figure 2.2 Placental single-cell RNA sequencing quality control pipeline

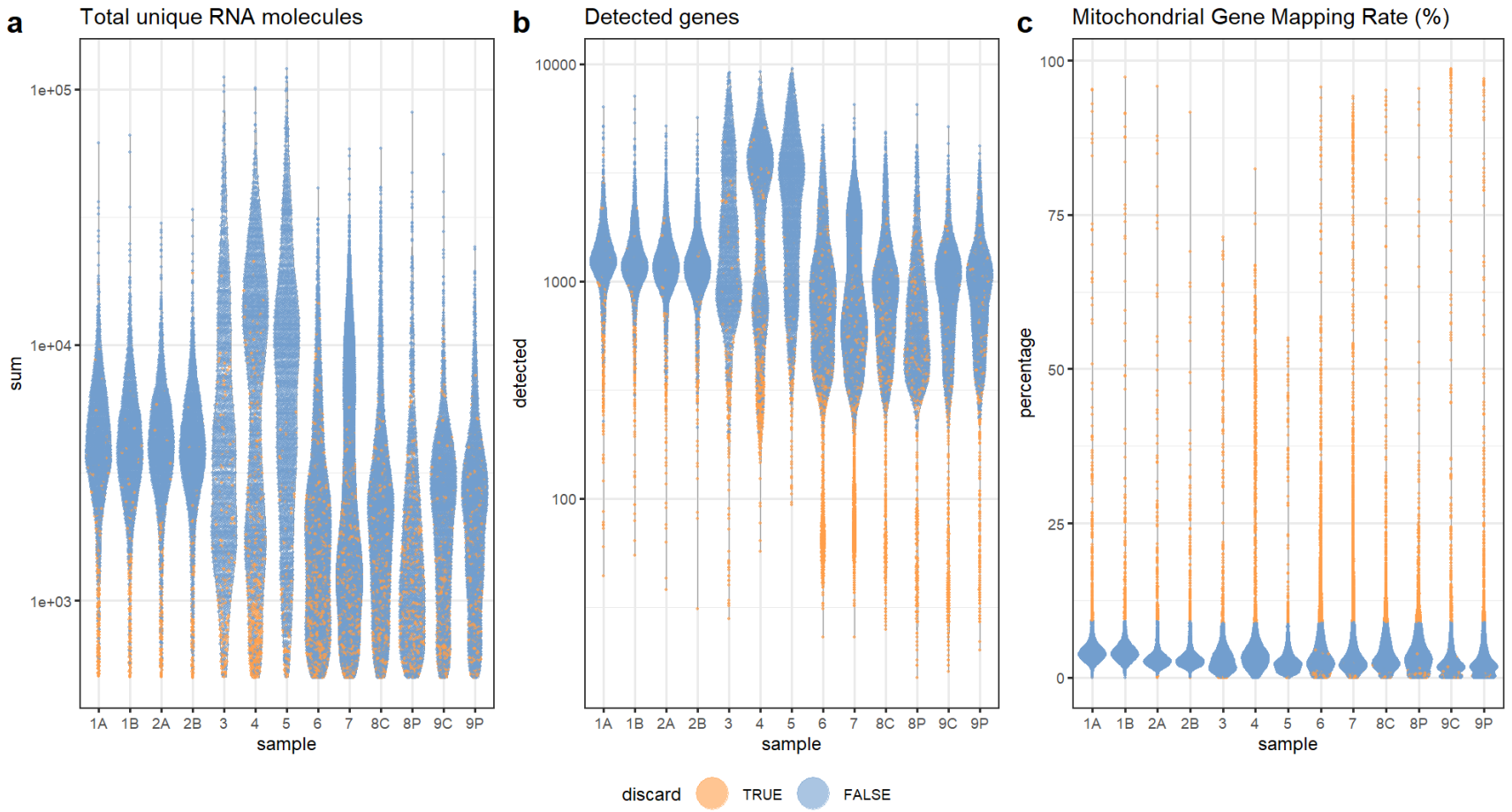


Figure 2.3 Placental single cell RNA sequencing of quality metrics by sample, visualized using violin plots. Orange cells were discarded based on outlier status on any of the following metrics: (a) total unique RNA transcripts (also called unique molecular identifiers) < 500, (b) number of genes expressed < 200, or (c) outliers in percent mitochondrial genes expressed.

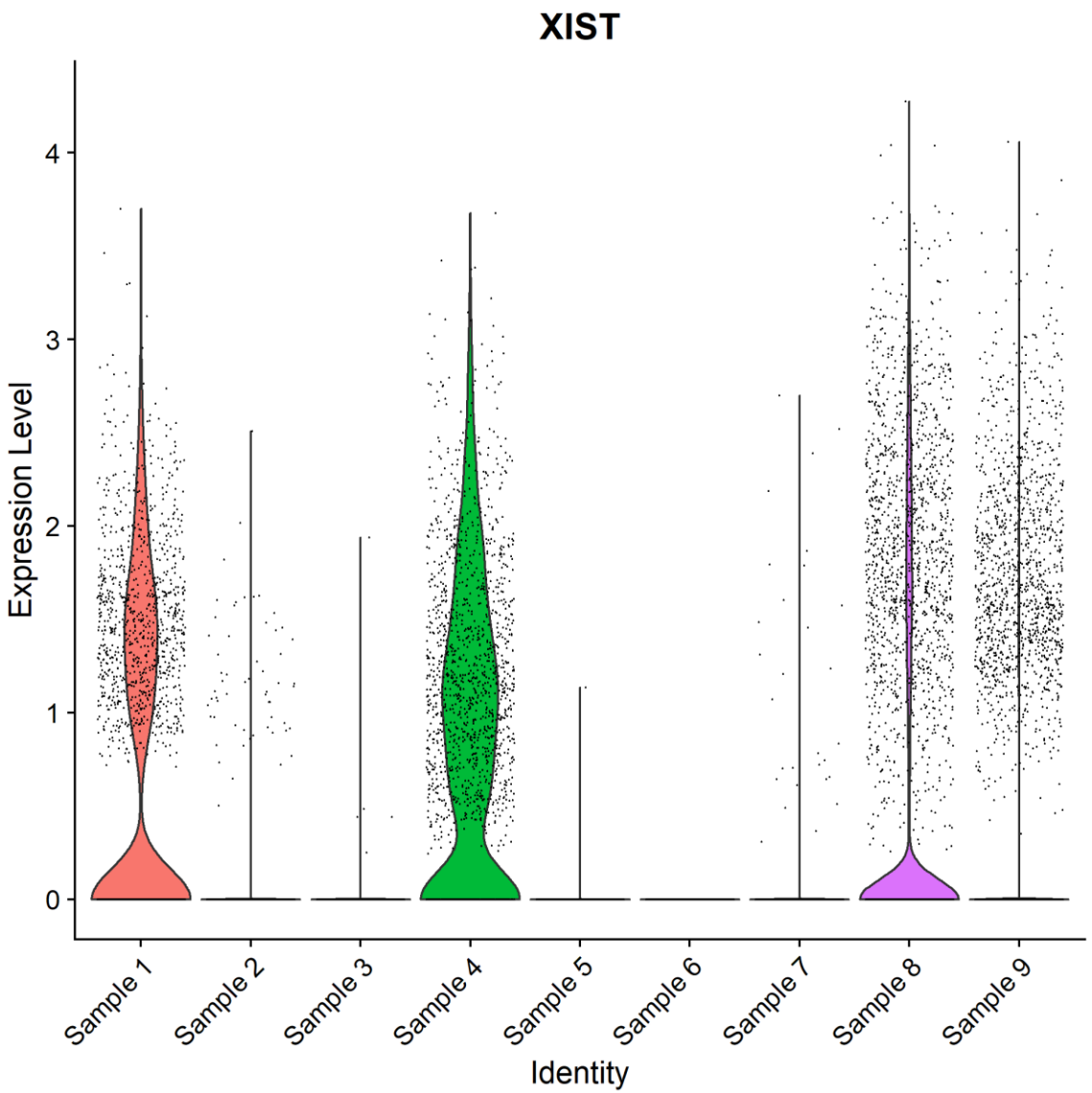


Figure 2.4 Library size-normalized and log-transformed XIST expression in fetal origin cells by biological replicates identifies Sample 1 as female due to high XIST expression, Sample 2 as male, and confirms fetal sex annotation for Samples 3-9.

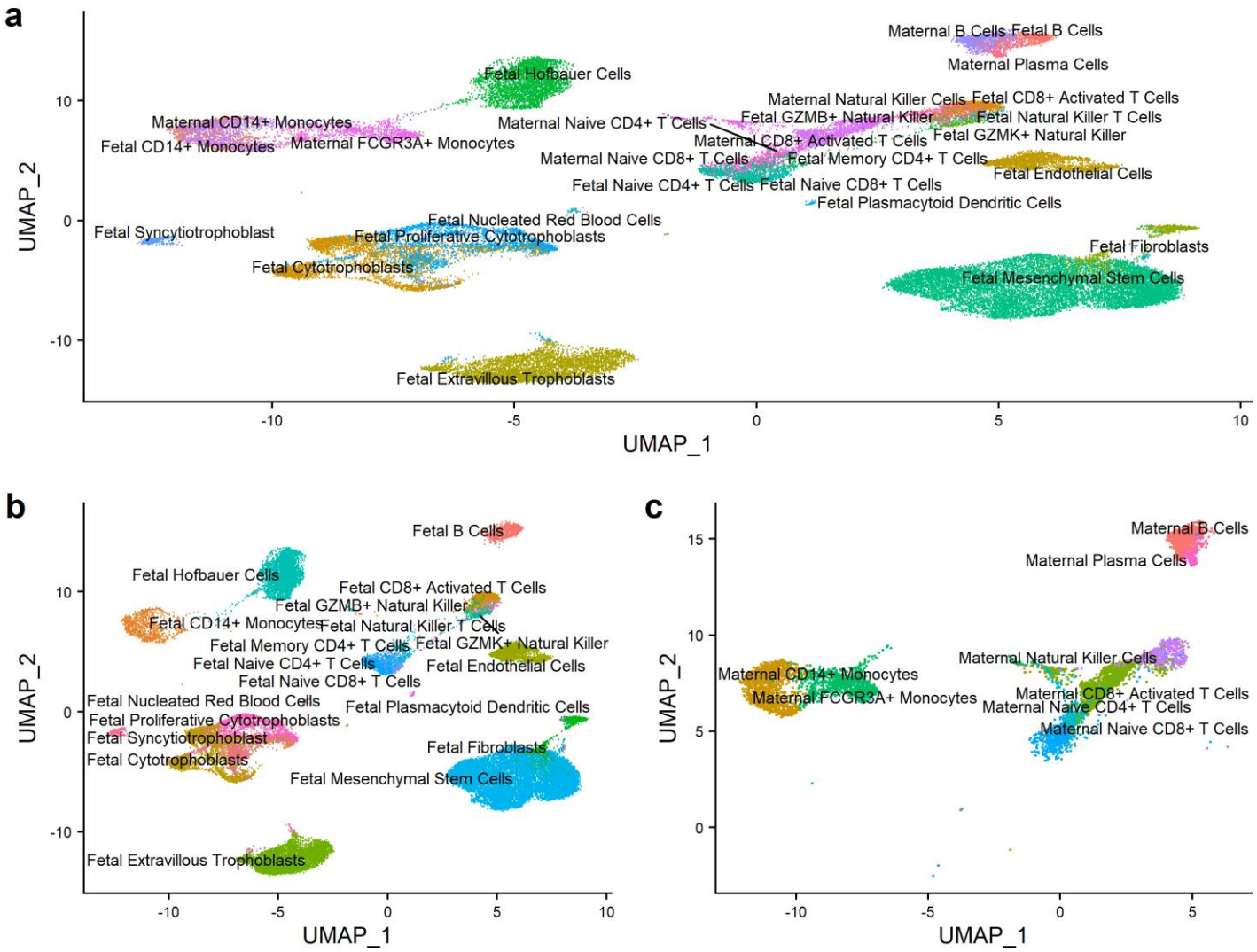


Figure 2.5 Integrated single-cell gene expression map of healthy placental villous tissue. (a) Uniform Manifold Approximation and Projection (UMAP) plot of all cells (n=40,494), with each cell colored by cell type cluster. (b) UMAP plot of fetal cells only (n=34,165), with each cell colored by cell type cluster. (c) UMAP plot of maternal cells only (n=6,329), with each cell colored by cell type cluster.

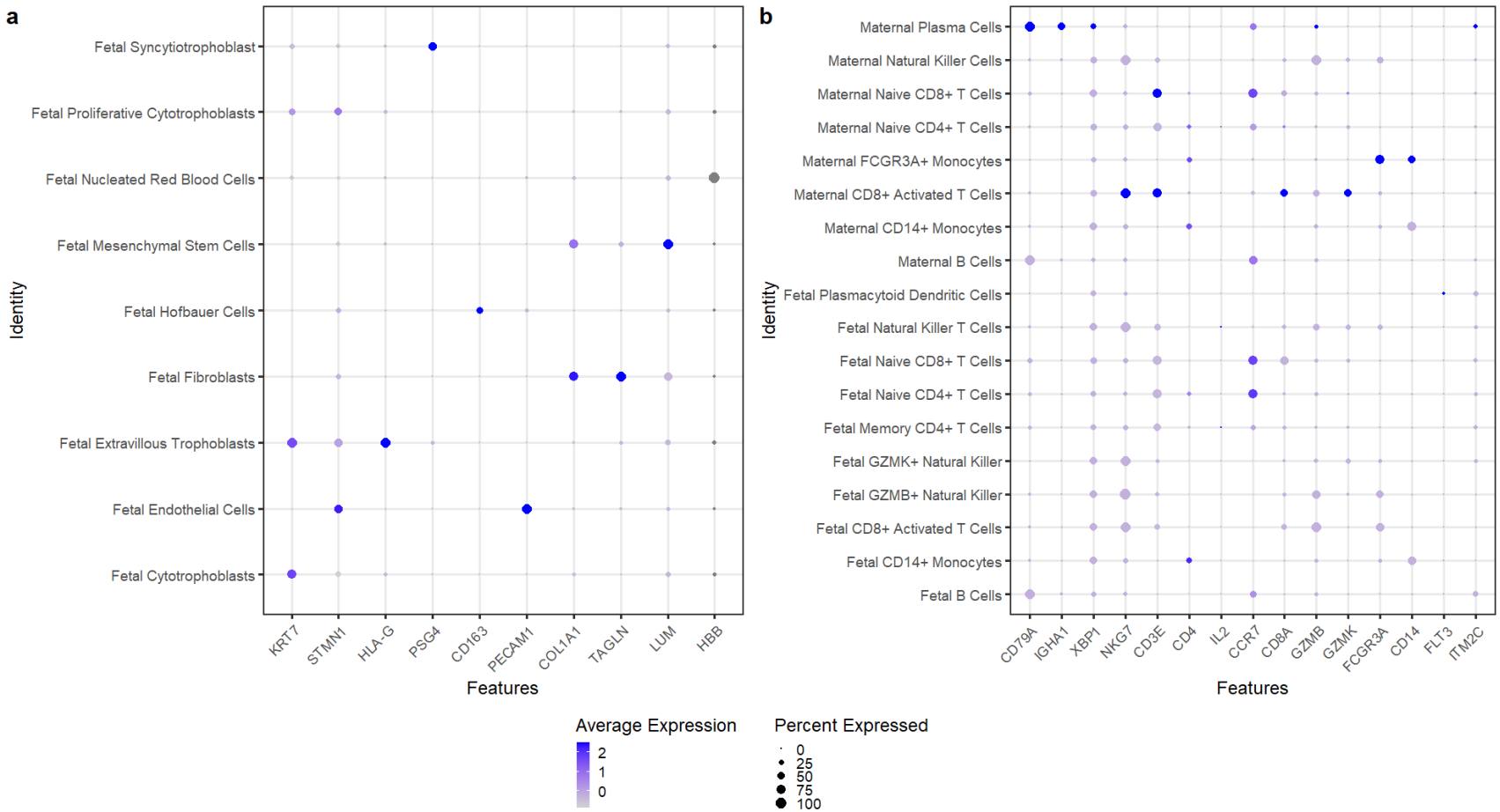
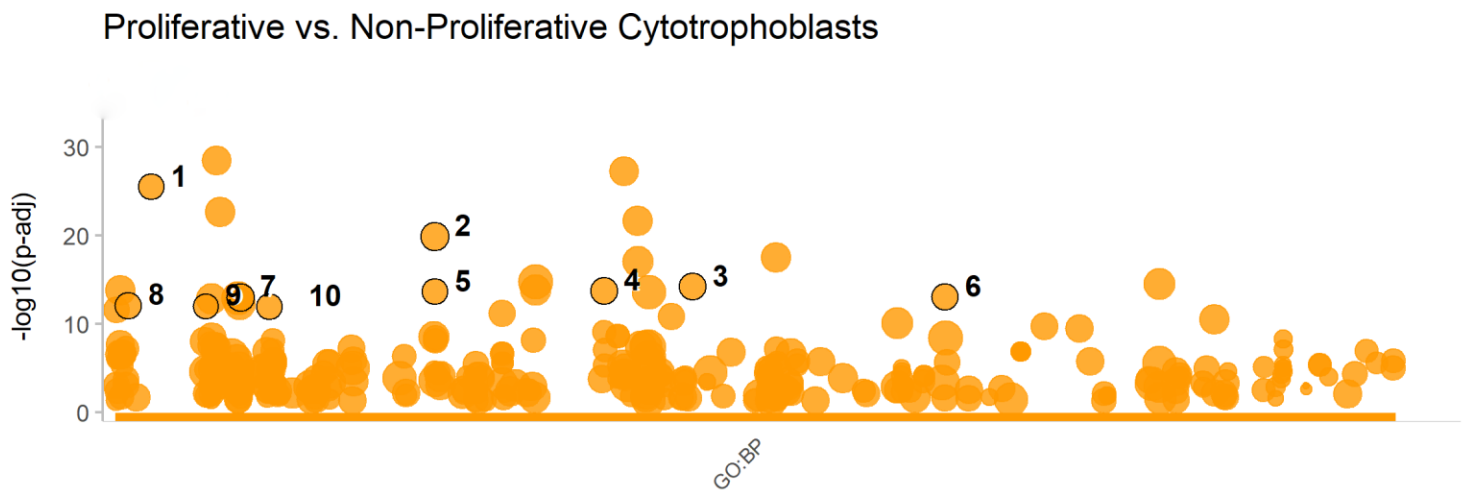


Figure 2.6 Dot plots of cell type marker genes used to annotate cell clusters to cell describing average normalized gene expression types (color darkness) and percentage of cells in a cell type cluster expressing that gene (point size). (a) Placental tissue cell types (b) Peripheral blood cell types.



id	source	term_id	term_name	term_size		p_value
1	GO:BP	GO:0002181	cytoplasmic translation	144		3.4e-26
2	GO:BP	GO:0022613	ribonucleoprotein complex biogenesis	353		1.3e-20
3	GO:BP	GO:0046034	ATP metabolic process	226		6.5e-15
4	GO:BP	GO:0042254	ribosome biogenesis	241		1.9e-14
5	GO:BP	GO:0022618	ribonucleoprotein complex assembly	163		2.4e-14
6	GO:BP	GO:0071826	ribonucleoprotein complex subunit organization	170		9.3e-14
7	GO:BP	GO:0007059	chromosome segregation	301		1.0e-13
8	GO:BP	GO:0000819	sister chromatid segregation	190		8.4e-13
9	GO:BP	GO:0006119	oxidative phosphorylation	120		9.9e-13
10	GO:BP	GO:0009060	aerobic respiration	154		1.1e-12

g:Profiler (biit.cs.ut.ee/gprofiler)

Figure 2.7 Top biological process gene ontology enrichment results with between 15 and 500 annotated genes for proliferative vs. non-proliferative cytotrophoblasts overexpressed differential expression results.

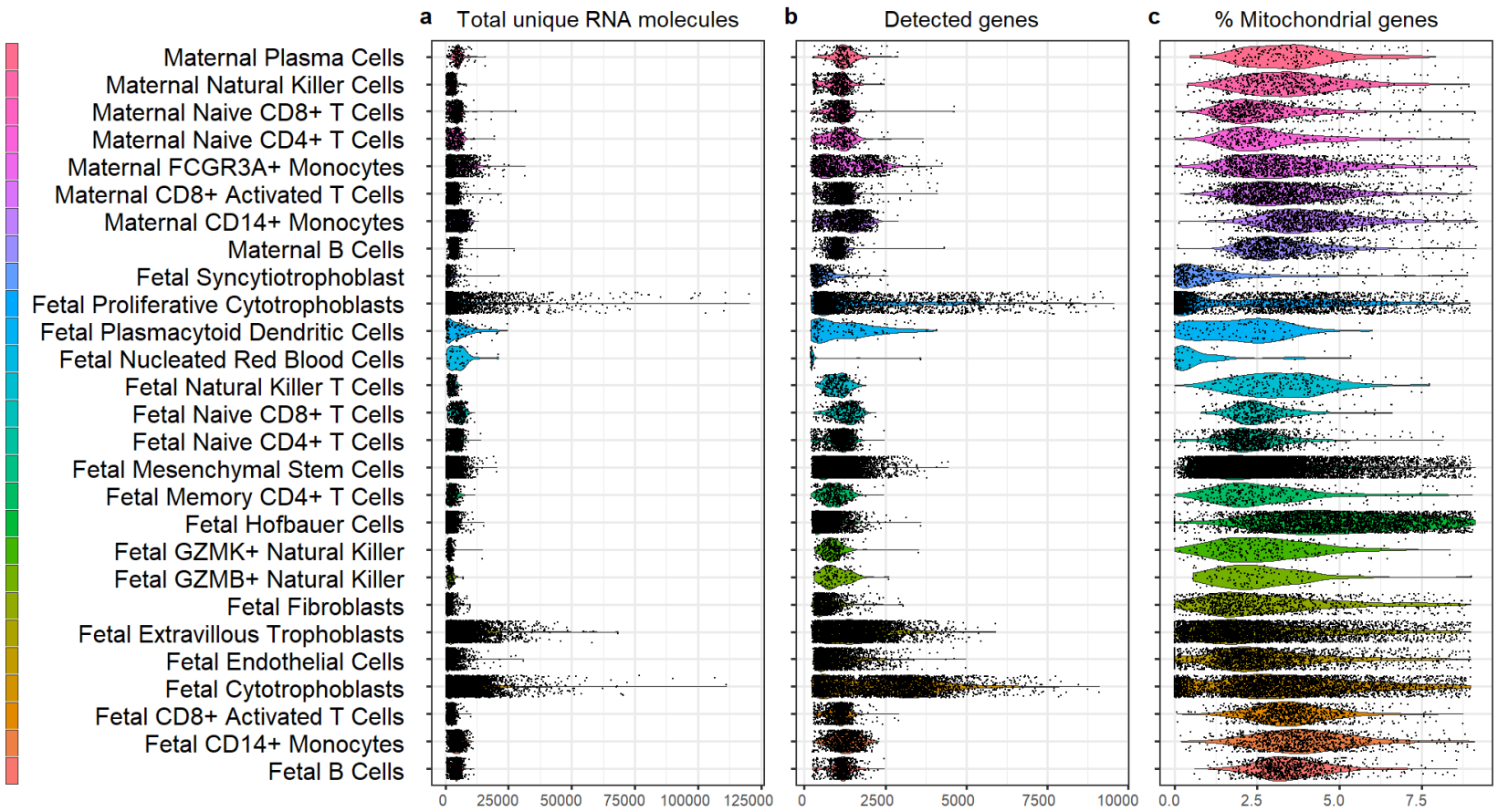


Figure 2.8 Placental single cell RNA sequencing of quality metrics by cluster, visualized using violin plots. (a) Number of genes expressed, (b) total unique RNA molecules, and (c) percent mitochondrial genes expressed.

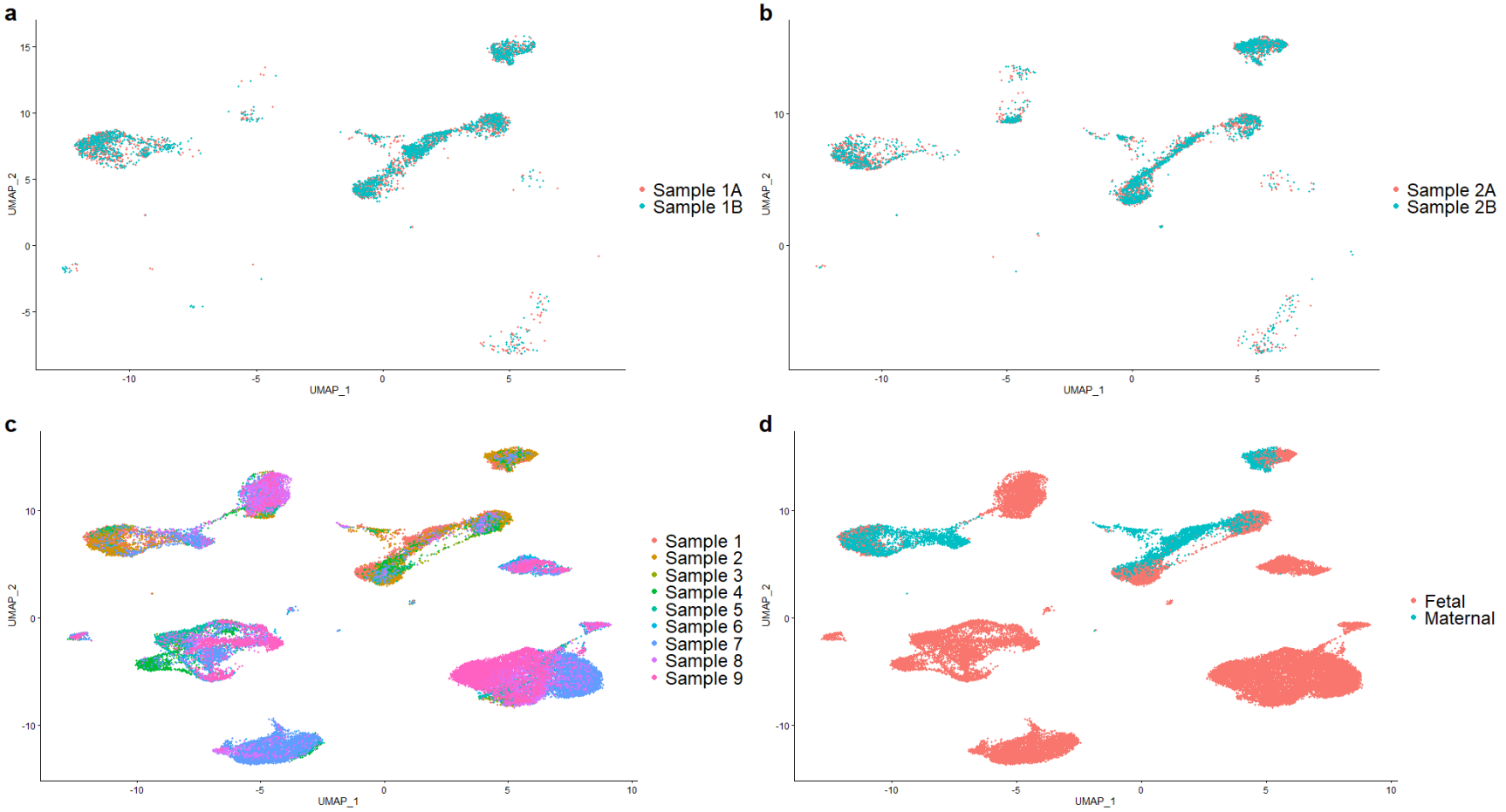


Figure 2.9 Uniform Manifold Approximation and Projection (UMAP) plots colored by key variables. (a) Technical replication in Sample 1 with points colored by technical replicate. (b) Technical replication in Sample 2 with points colored by technical replicate. (c) Biological replicates identified by point color with collapsed technical replicates. (d) Fetal/Maternal origin assignment by point color.

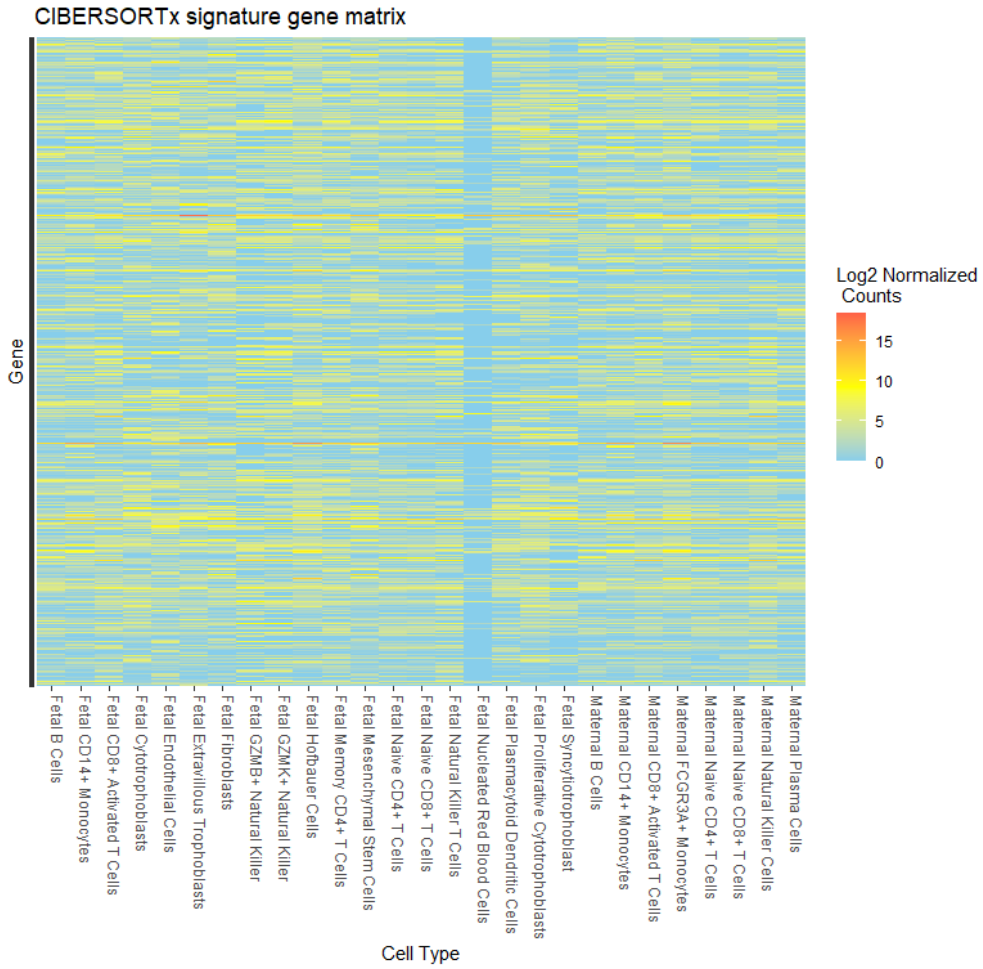


Figure 2.10 Heatmap of signature gene expression matrix in log2-transformed library size-normalized counts (counts per million) generated and used to deconvolute bulk placental tissue dataset. Cell types are encoded on the y-axis and genes are located along the x-axis. Blue indicates low expression of a gene and red represents high expression.

.

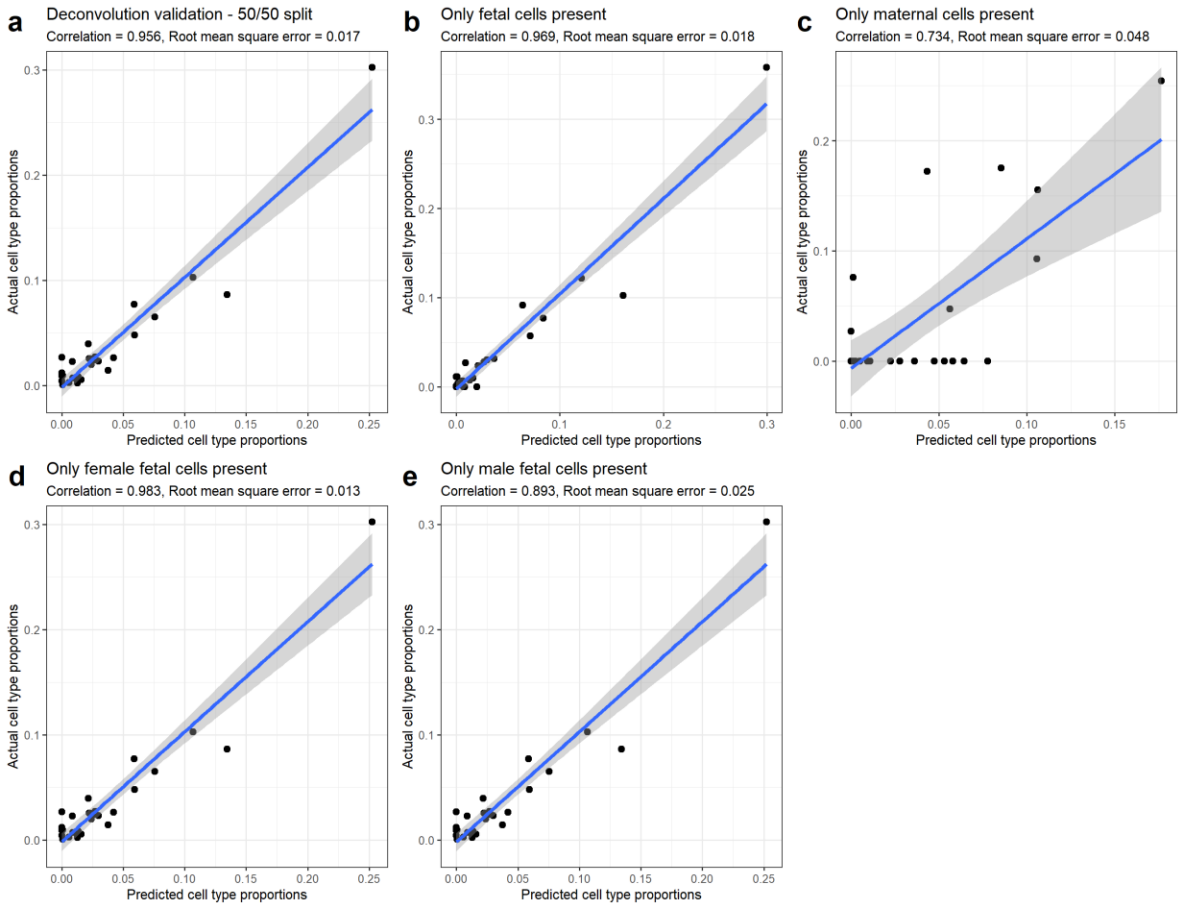


Figure 2.11 In silico placental deconvolution testing. Scatter plots summarizing the performance of our single-cell deconvolution reference using CIBERSORTx with in silico mixtures of single-cell libraries from a 50/50 training/test split of the integrated single-cell RNA-seq dataset (n=40,494). The same training dataset was used for each comparison; test mixtures were generated from the testing half of the dataset. Predicted deconvoluted cell type proportions for each of the 27 cell types are encoded on the x-axis. Actual cell type proportions from the test dataset are encoded on the y-axis. Correlation coefficients and root mean square error measures are presented for each comparison. A linear line of best fit overlays the results. The grey shaded area represents the 95% confidence intervals around the simple linear regression estimates. (a) The test mixture is the test half of the single-cell dataset (n=20,242). (b) The test mixture sampled only fetal cells (n=17,080). (c) The test mixture sampled only maternal cells (n=3,162). (d) The test mixture sampled only female fetal cells (n=8,394). (e) The test mixture sampled only male fetal cells (n=8,394).

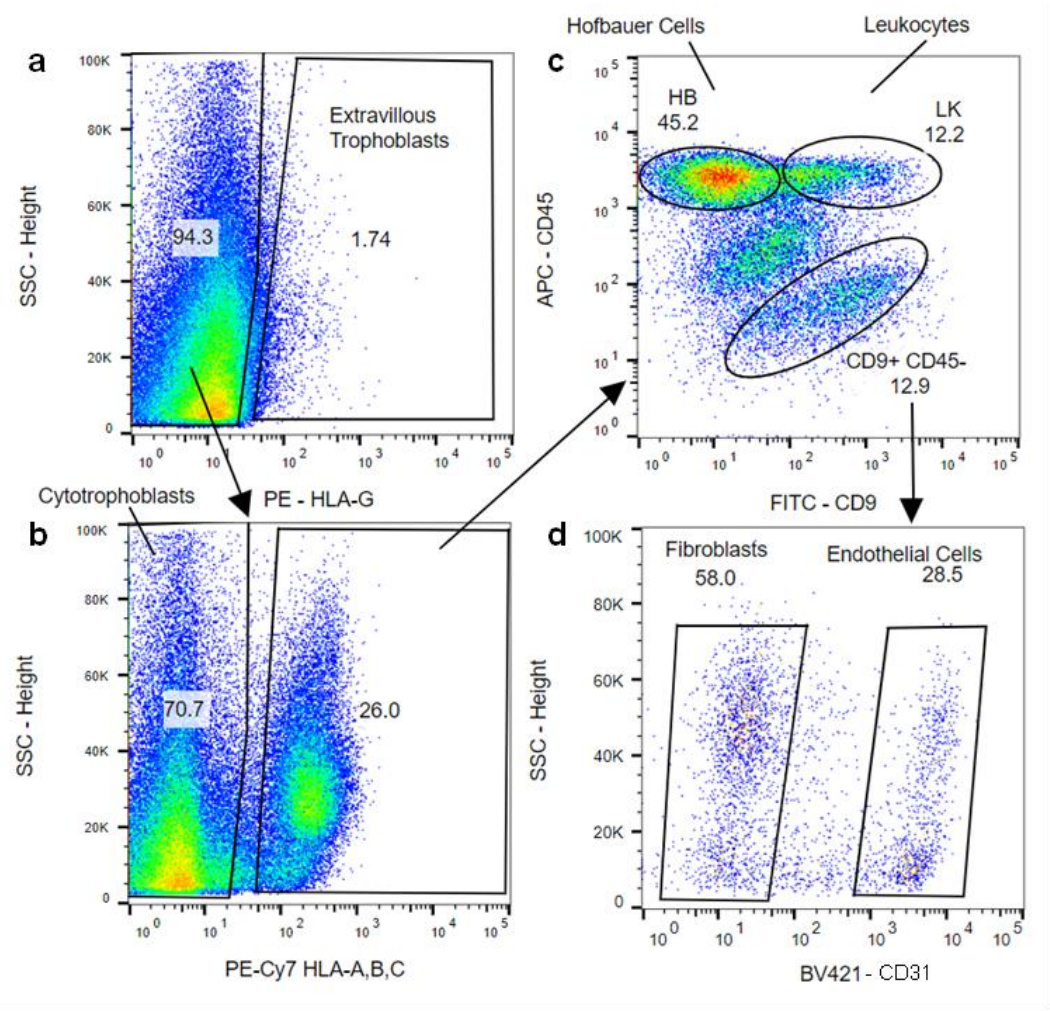


Figure 2.12 A representative FACS sort. Gating strategy: (a) HLA-G/PE to positively select extravillous trophoblasts; (b) HLA-ABC/PE-CY7 to negatively enrich for cytotrophoblasts; (c) CD9/FITC by CD45/APC to positively select for Hofbauer cells and leukocytes; (d) CD31/BV421 to distinguish endothelial cells from CD31- fibroblasts.

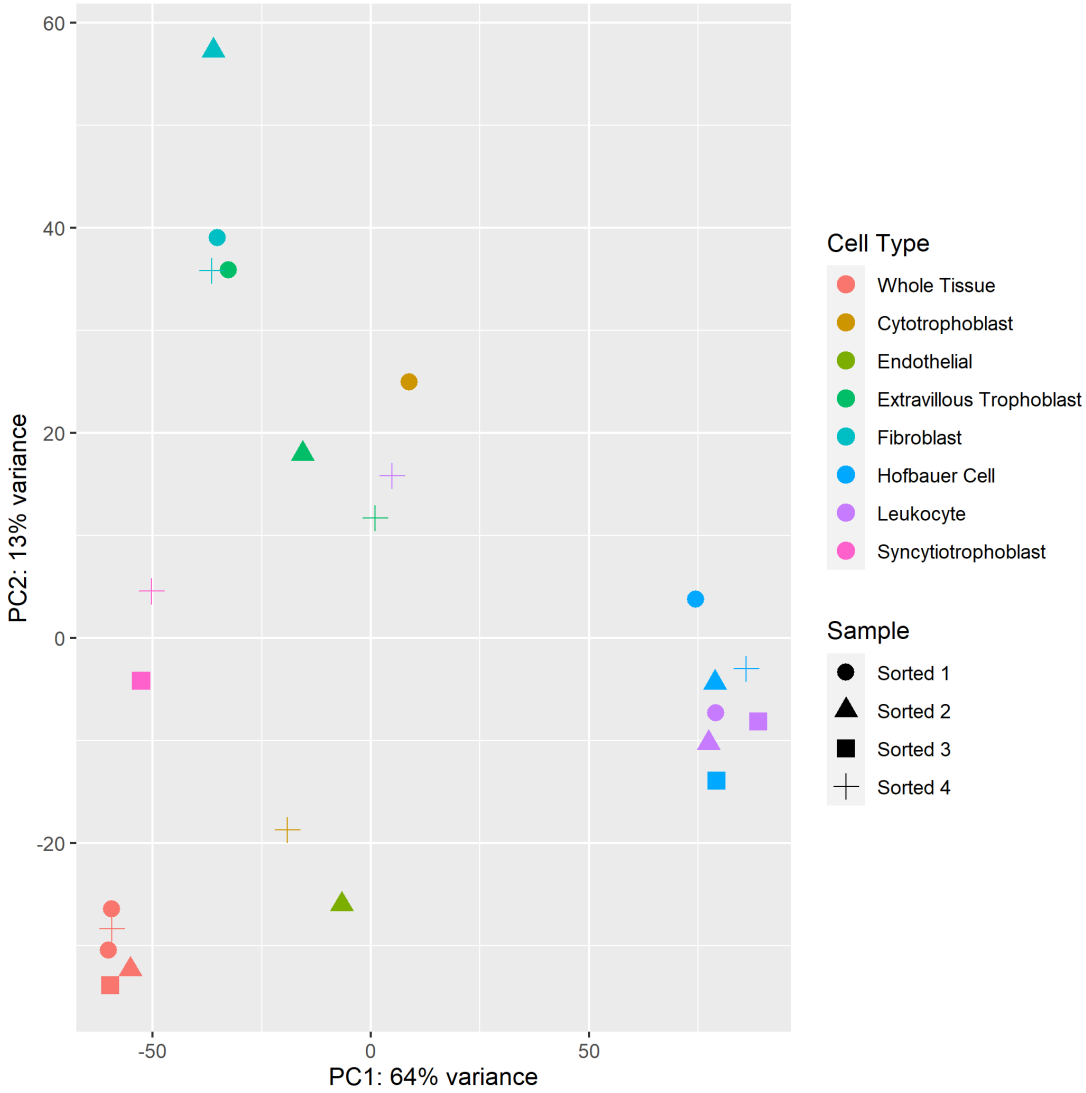


Figure 2.13 Principal components plot of fluorescence-activated cell sorting bulk RNA-sequencing results on sorted placenta samples. Point colors encode cell type. Shape denotes sample source.

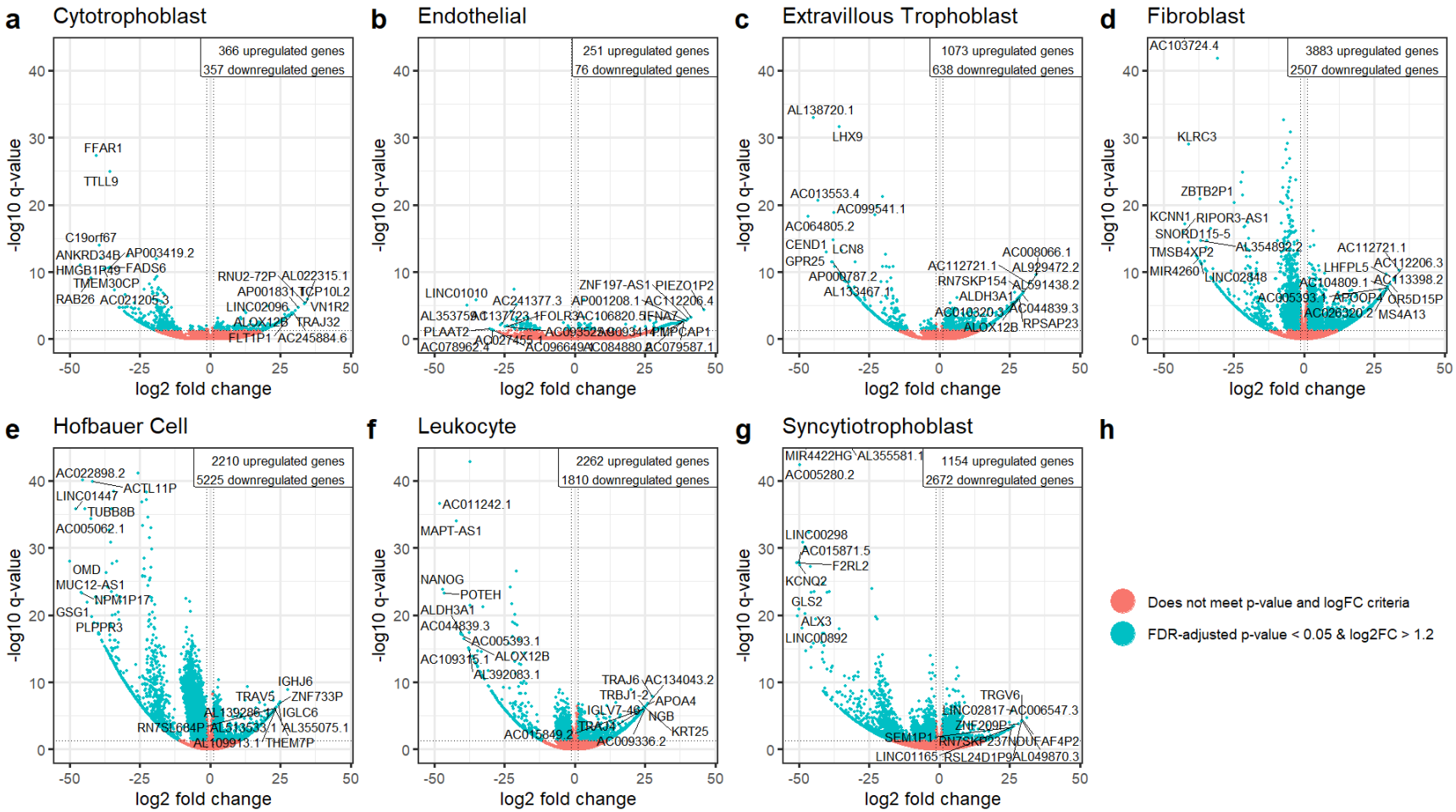


Figure 2.14 Volcano plots for fluorescence-activated-cell-sorted bulk RNA-seq differential expression in one cell type against average gene expression across other cell types. The y-axis encodes -log10 transformation of the false discovery-controlled q-value, with the cut-off for statistical significance 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

.

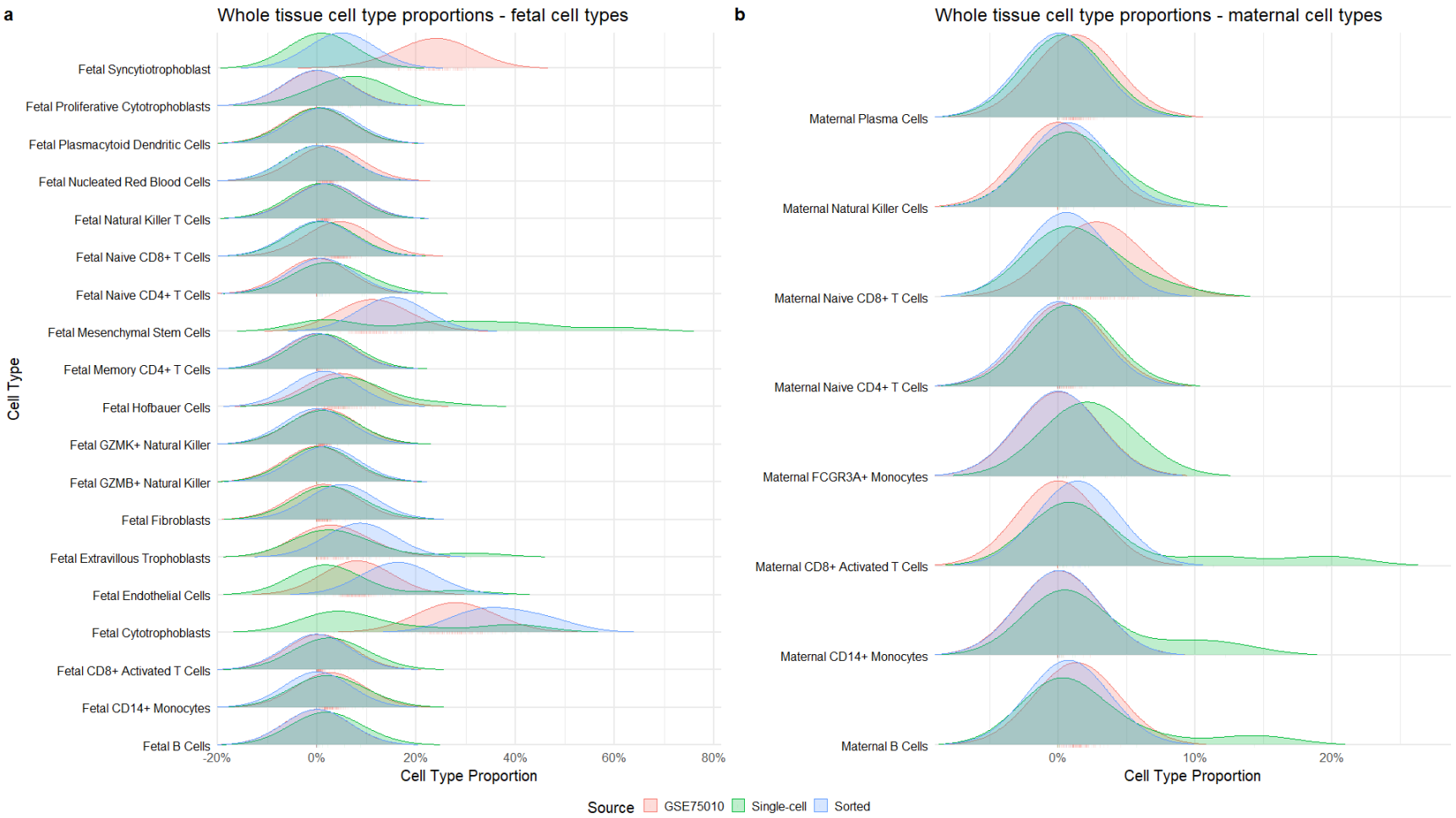


Figure 2.15 Distribution of estimated cell type proportions in whole placental villous tissue from Michigan samples and GSE75010 controls compared to the number of single cells captured in the single-cell RNA sequencing datasets. Density distribution is colored by study source. (a) Fetal cell types. (b) Maternal cell types.

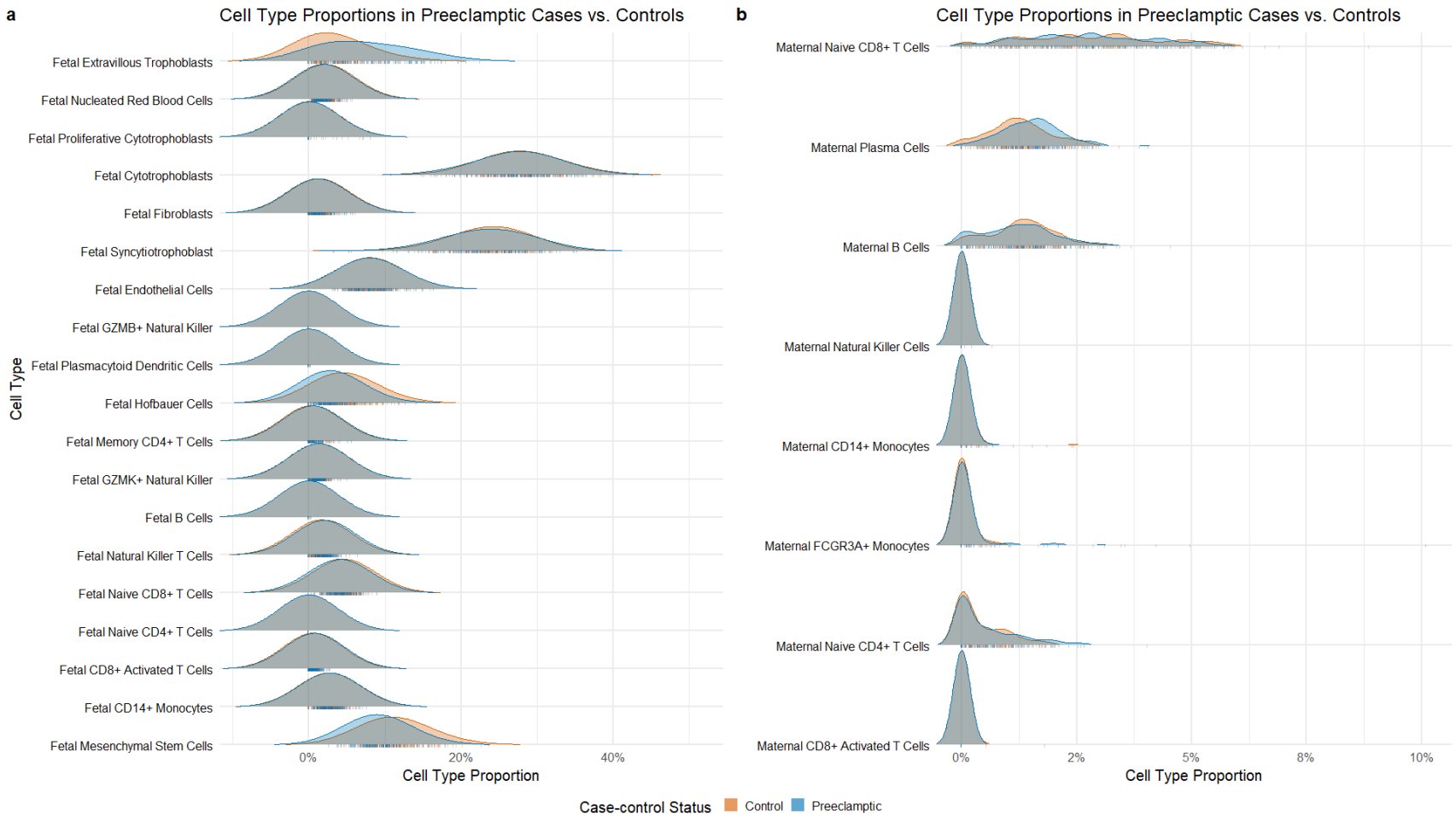


Figure 2.16 Distribution of estimated cell type proportions in preeclamptic cases versus controls. Density distribution is colored by case-control status. (a) Fetal cell types. (b) Maternal cell types.

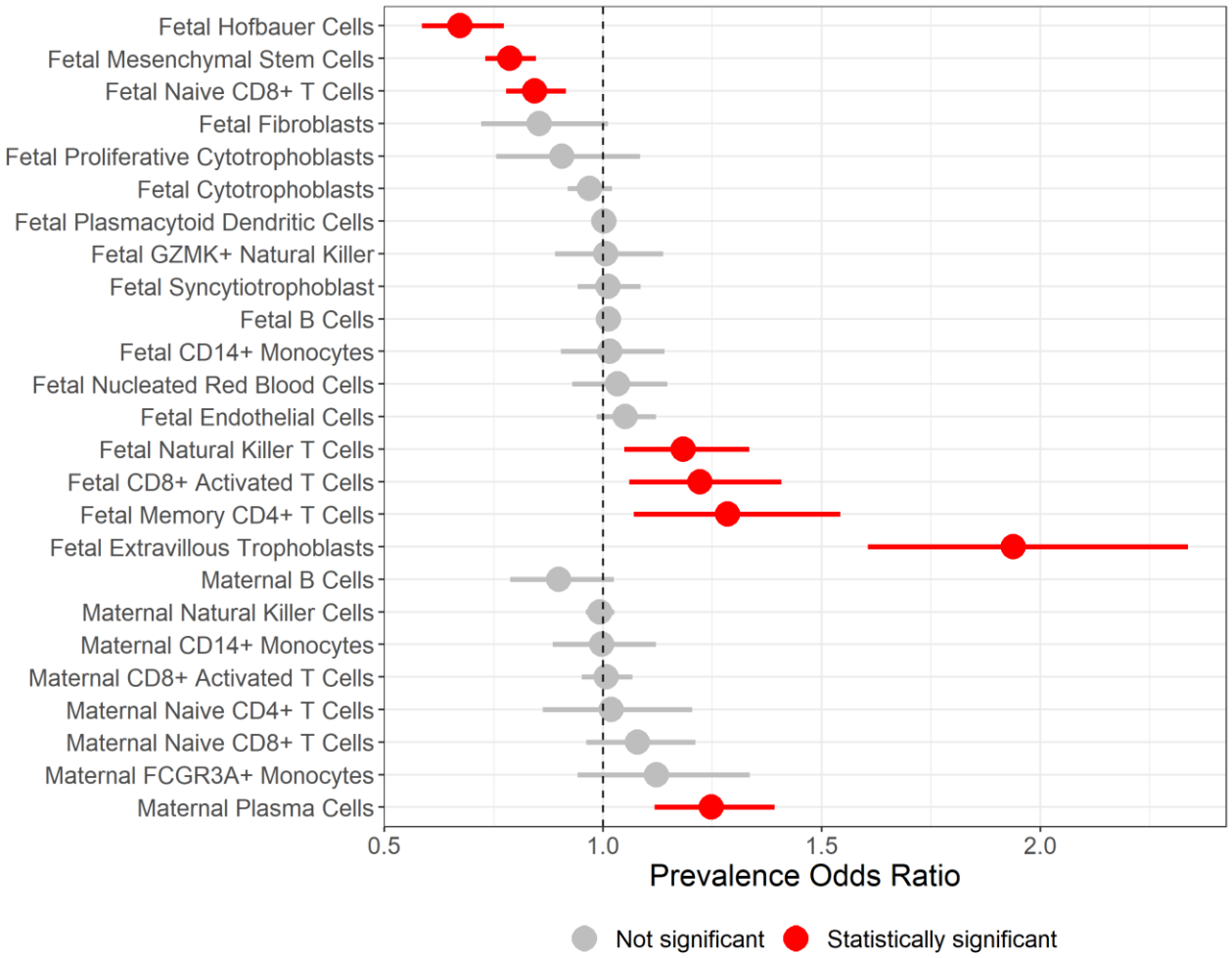


Figure 2.17 Preeclampsia case-control status and cell type proportion differential abundance analysis. Forest plot of multivariate beta regression models' prevalence odds ratio estimates adjusted for study source, gestational age, and fetal sex tested for a difference in each cell type's proportions in cases versus controls (n=157 cases, 173 controls). Horizontal lines indicate the range of the 95% confidence interval.

Figure 2.18 Principal component (PC) results of estimated cell type proportions. Contribution refers to the relative proportions (expressed as a percentage) of the variation in a principal component attributable to an individual cell type. (a) PC1 and PC2 dimension loadings are largely driven by fetal syncytiotrophoblasts and fetal extravillous trophoblasts. (b) PC1 and PC3 loadings are largely driven by fetal syncytiotrophoblasts, fetal extravillous trophoblasts, and fetal cytotrophoblasts. (c) Individual observations projected onto PC1xPC2, with observations colored and shape-coded by preeclampsia case-control status. (d) Individual observations projected onto PC1xPC3, with observations colored and shape-coded by preeclampsia case-control status.

.

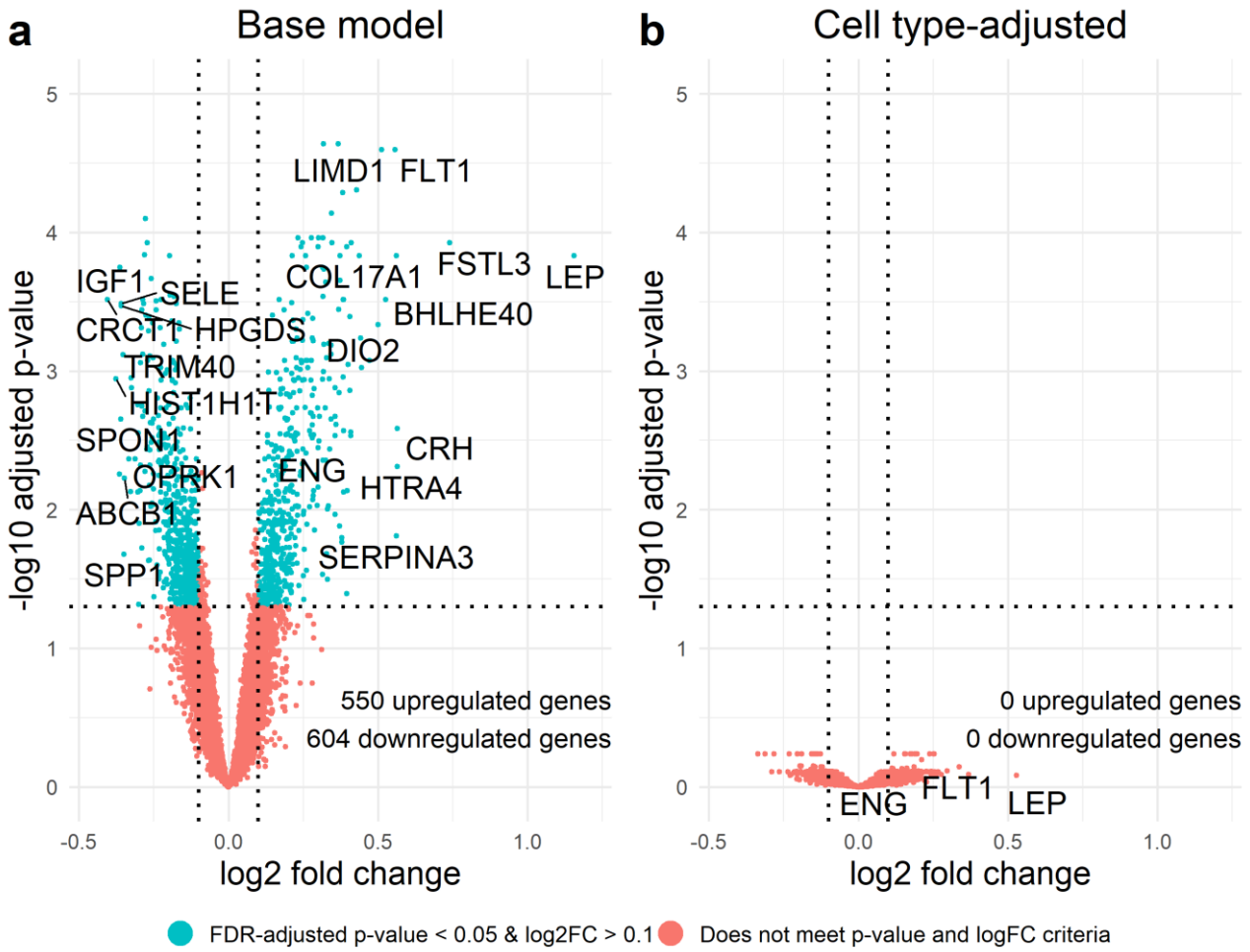


Figure 2.19 Preeclampsia case-control differential expression analysis. Volcano plots comparing differentially expressed genes in samples from 153 preeclampsia cases versus 173 healthy controls across two models: (a) the base model adjusted for covariates fetal sex, study source, and gestational age and (b) the model adjusted for fetal sex, study source, and gestational age and 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, LEP, and ENG are labelled as genes of interest in preeclampsia.

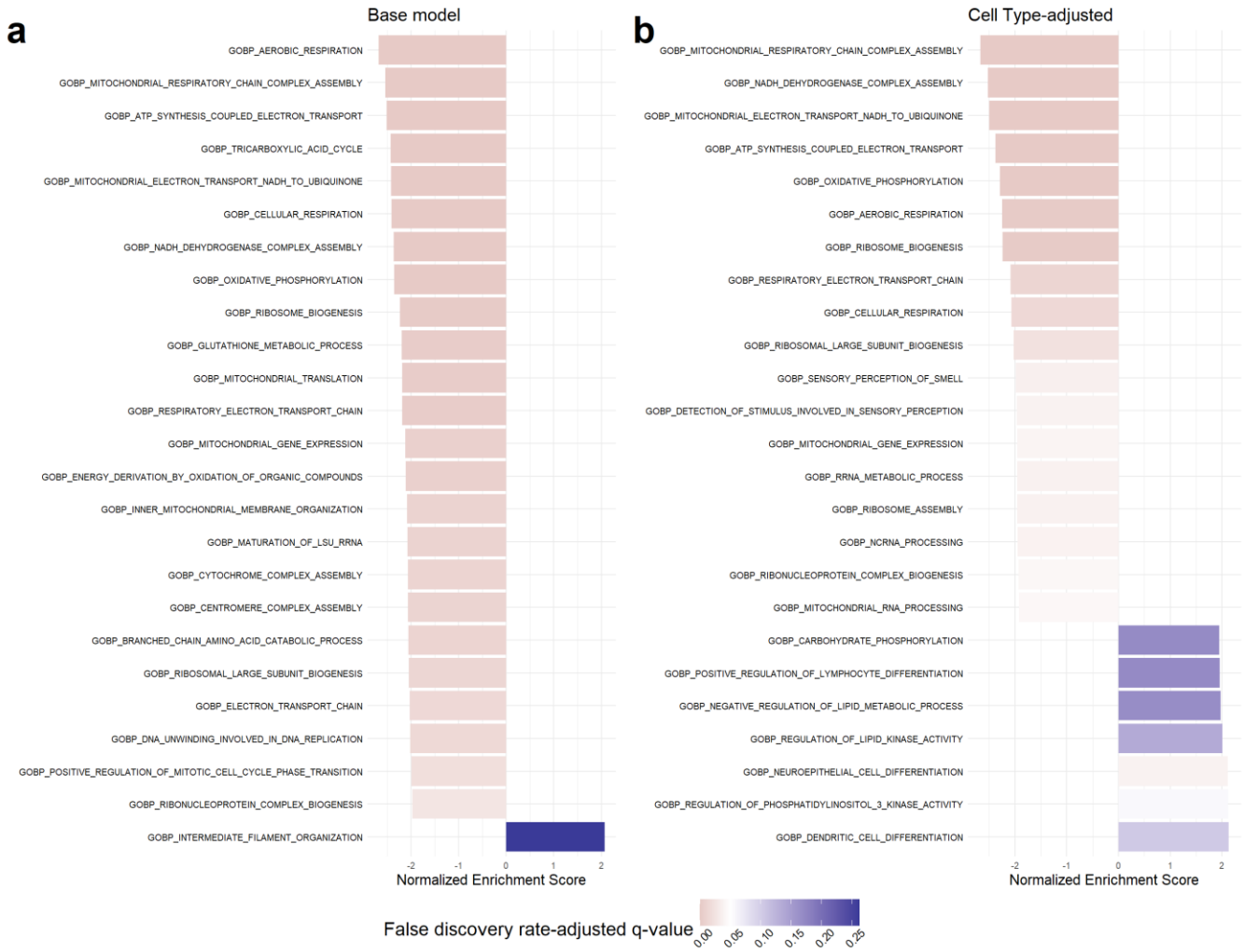


Figure 2.20 Preeclampsia case-control differential expression enrichment analysis. Top Gene Set Enrichment Analysis pathways from the Gene Ontology: Biological Processes database results for the differential expression analysis by preeclampsia case-control status. Results arranged by descending magnitude 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.

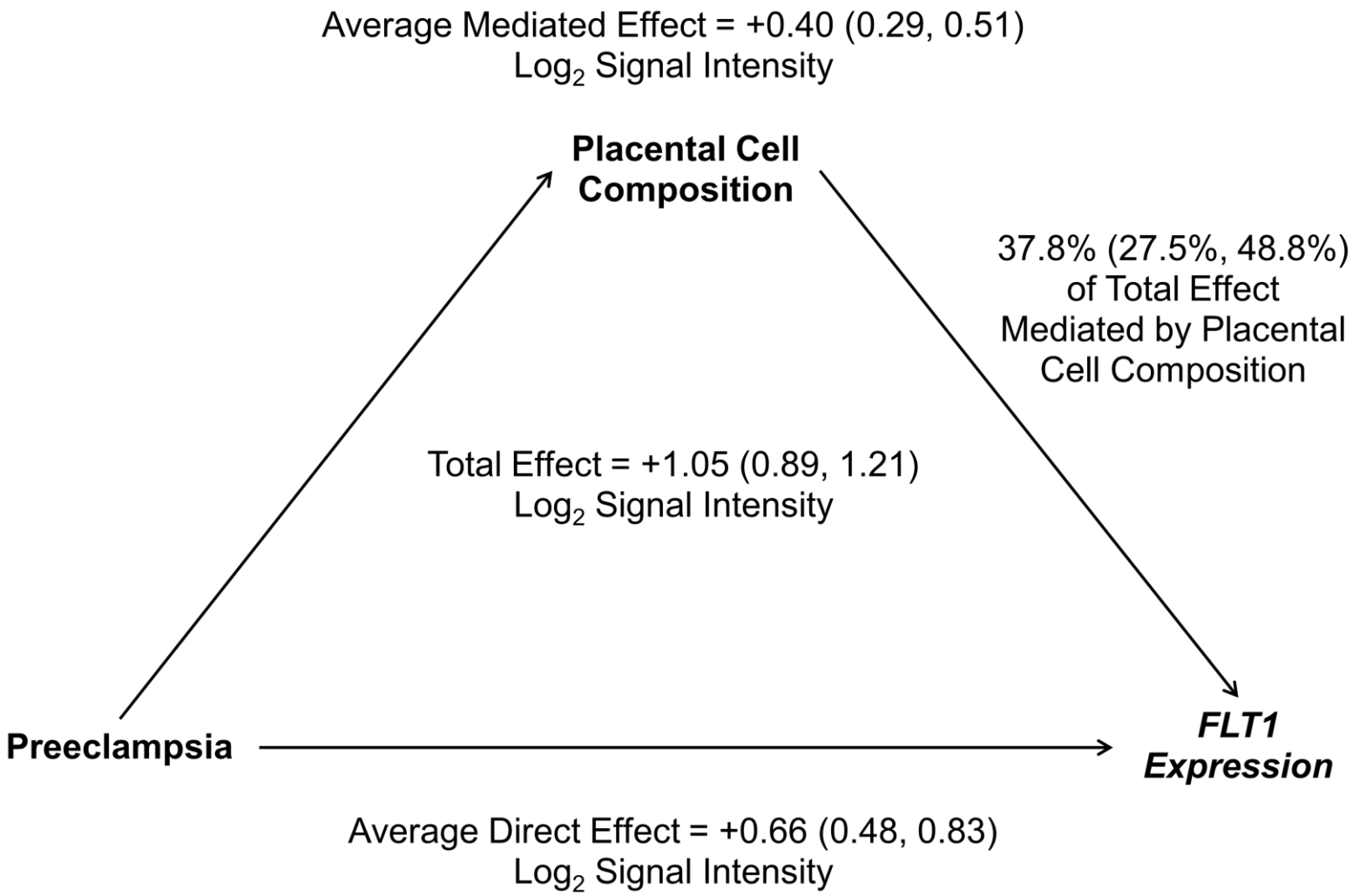


Figure 2.21 Placental cell composition as a mediator of FLT1 expression. Mediation of FLT1 gene expression by placental cell type composition (n=157 cases, 173 controls). Placental cell composition was operationalized as first five principal components of estimated cell type proportions. 95% confidence intervals are provided after effect estimates for each model parameter. The same framework was also applied with LEP or ENG expression as the outcome.

Chapter 3 Placental and Immune Cell DNA Methylation Reference Panel for Bulk Tissue Cell Composition Estimation in Epidemiological Studies

Authors: Campbell KA[1], Colacino JA[2,3], Dolinoy DC[2,3], Loch-Caruso R[2], Padmanabhan V[2,3,4,6], and Bakulski KM[1]

Affiliations:

[1]Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48105, USA.

[2]Environmental Health Sciences, School of Public Health, University of Michigan, Ann Arbor, MI 48105, USA.

[3]Nutritional Sciences, School of Public Health, University of Michigan, Ann Arbor, MI 48105, USA.

[4]Pediatrics, Michigan Medicine, University of Michigan, Ann Arbor, MI 48105, USA.

[5]Human Genetics, Michigan Medicine, University of Michigan, Ann Arbor, MI 48105, USA.

[6]Obstetrics and Gynecology, Michigan Medicine, University of Michigan, Ann Arbor, MI 48105, USA.

3.1 Abstract

DNA methylation perturbations in bulk placental tissue have been linked to adverse perinatal outcomes. To distinguish mechanistic DNA methylation changes from cell composition differences, robust cell type-specific placental DNA methylation profiles are needed. We analyzed 192 new and previously published DNA methylation profiles from 12 cell types, including cytotrophoblasts (n=32), endothelial cells (n=19), Hofbauer cells (n=26), stromal cells (n=29), syncytiotrophoblasts (n=4), six types of adult lymphocytes (n=36), and nucleated red blood cells (n=11), as well as 35 bulk placental tissue samples. Methylation was quantified via the Illumina DNA methylation microarray (450k or EPIC), and common probes were quality filtered (n=407,628 DNA methylation sites). To identify cell type-discriminating DNA methylation sites, we ranked the top 50 hyper- and hypomethylated sites per cell type by F-test, yielding 1,101 unique sites. To

estimate the cell composition of bulk placental tissue, we applied the robust partial correlation deconvolution algorithm using these sites. Consistent with placental biology, bulk placental tissue cell type proportion estimates (mean ± standard deviation) from methylation measures were predominately syncytiotrophoblast (57.8% ± 8.3%), stromal (20.6% ± 5.9%), cytotrophoblast (11.0% ± 4.1%), endothelial (7.5% ± 2.2%), Hofbauer cells (1.5% ± 1.2%), and CD4+ T cells (0.89% ± 0.84%). Other cell types had mean estimates less than 0.5%. This cell type DNA methylation reference panel can robustly estimate cell composition from placental DNA methylation data in epidemiological studies to detect unexpected non-placental cell types, reveal biological mechanisms, and improve casual inference. Future work will present a software package for community use so that future studies may estimate cell composition in placental tissue for testing with exposure or disease.

3.2 Introduction

Prenatal environmental or other exposures during gestation may induce perturbations to placental epigenetic programming, including DNA methylation, that ultimately result in disease [29]. Epigenetic dysregulation is characteristic of diseases such as cancer [26], neurodegeneration [27], and cardiovascular disease [28]. A recent review identified that between 2016 and 2021, 28 studies linked air pollution, chemical, metals, psychosocial, and smoking exposures to the placental epigenome [195]. A critical limitation of these studies, however, is a failure to account for cell type heterogeneity [47, 195]. In fact, two studies identified an association between prenatal mercury exposure and DNA methylation in fetal cord blood as well as a shift in cord blood cell type proportions, precluding biological interpretation as to the causal factor

[196, 197]. Bioinformatic reference-based deconvolution has emerged as a promising strategy to account for cell heterogeneity in bulk tissue assays by estimating cell type proportions from previously collected cell type-specific molecular profiles [46, 58, 59]. However, placental cell type-specific DNA methylation references are limited.

DNA methylation characterization of term placental cell types is limited [198]. Several studies have individually profiled DNA methylation of cytotrophoblasts and Hofbauer cells at term [199, 200]. One study conducted whole genome DNA methylation characterization of cytotrophoblasts from two term placentas [200]. DNA methylation from term Hofbauer cells was characterized on the Illumina Infinium Human Methylation27 BeadChip microarray [199]. This platform measured DNA methylation at ~27,000 sites across the genome, which is much lower coverage than the ~850,000 sites now available on the EPIC version of this platform. Only recently has one study profiled trophoblasts, stromal, endothelial cells, and Hofbauer cells from fractions of placental cells enriched using fluorescence-activated cell sorting, and a separate filter-enriched syncytiotrophoblast fraction on the Illumina EPIC DNA methylation microarray [201]. However, maternal peripheral immune cells and the highly epigenetically distinct nucleated red blood cells present only in fetal cord blood may also be intermingled with placental cell types in villous tissue samples [60]. Indeed, a DNA methylation epigenome-wide association study in fetal cord blood found that nucleated red blood cells explained most of the association between gestational age and DNA methylation [202]. A robust placental deconvolution reference requires additional placental cell type-specific DNA methylation profiles, and must account for other epigenetically distinct cell types present at the fetal-maternal interface.

This study aims to address the current limitations of existing approaches to account for cellular heterogeneity in DNA methylation studies of the human placenta. Here, we generate primary cell type-specific placental DNA methylation profiles for major placental cell types isolated via magnetic bead-activated cell sorting and integrate the results with existing placental, adult peripheral immune cell, and nucleated red blood DNA methylation profiles to create a unified placental DNA methylation deconvolution panel for reference-based deconvolution. Placental cell type-specific genome-wide DNA methylation panels can then be implemented in numerous existing or future studies that utilize placental villous tissue samples at term to improve precision, detect non-placental cells, reduce sources of potential bias due to cell type heterogeneity, and improve biological inference.

3.3 Methods

3.3.1 Placental tissue sample collection and dissociation

We collected placentas shortly after delivery from healthy, full term, singleton uncomplicated Cesarean sections at the University of Michigan Von Voigtlander Women's Hospital between February 2019 and March 2020. All study participants provided informed consent. This study protocol was approved by the University of Michigan Institutional Review Board (#HUM00017941). Villous placental tissue samples were collected, washed in phosphate buffered saline (PBS, Invitrogen #10010049), blotted dry, and minced from the maternal-facing side of placenta the after trimming away the basal and chorionic plates and scraping tissue from blood vessels until at least 10g of villous tissue were collected. Approximately 200mg of villous tissue was set aside and snap-frozen in liquid nitrogen.

We subjected isolated fetal villous tissue to sequential enzymatic digestions followed by discontinuous density gradients using slight modifications on established methods to isolate cytotrophoblasts, fibroblasts, and Hofbauer cells [96, 98]. Enzymatic digestions were carried out in a Trypsin-DNase I solution containing 0.25% trypsin (Corning, #25-050-CI), 0.20% DNase I (Roche, #10104159001), 25mm N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES, Millipore Sigma, H4034-25G), 2.0mM $CaCl_2$ (Millipore Sigma, #C7902-500G), and 0.8mM $MgSO_4$ (Millipore Sigma, M2643-500G) in Hanks' balanced salt solution (HBSS, Invitrogen #14185052) at 37°C with shaking at 150rpm in a large shaker bath [96]. After 15 minutes, the partially digested villous tissue was strained through a 104μm mesh with a pestle (Millipore Sigma, #CD1-1KT) and rinsed with 3mL heat-inactivated fetal bovine serum (FBS, Corning, # MT35016CV) and 2 rounds of 10mL DF medium, consisting of 10% FBS, 1% antibiotic-antimycotic (Invitrogen, #15240062), in Dulbecco's Modified Eagle Medium/Ham's F-12 (DMEM/F12 medium, Invitrogen, #12634010) to quench 15mL of Trypsin-DNase I digestion solution and facilitate straining.

The second and third digests were performed for 30 minutes each, but were otherwise identical to the first digestion. The supernatant from each digest was filtered sequentially through another 100μm filter and a 70μm filter. PBS washes were used as needed to facilitate straining. Adapting a previously published protocol [95], the 70μm filters were immediately inverted into a sterile petri dish and washed with DF medium to produce a syncytiotrophoblast-enriched fraction that was pooled from each of the three digests and cryopreserved in a cryopreservation solution containing 10% dimethyl sulfoxide (DMSO, Millipore Sigma #D8418) in FBS in liquid nitrogen. The supernatant

from the second and third digests were combined and loaded onto a 30%/35%/45%/50% Percoll discontinuous gradient consisting of 90% Percoll (GE Healthcare, #17089101) mixed with 25mM HEPES in HBSS in the appropriate concentrations. We performed density gradient centrifugation by centrifuging the column at 1000rcf for 20 minutes at room temperature with no brake [96]. To create a cytotrophoblast-enriched fraction, we collected the cell layer between the 35% and 45% fraction, which was cryopreserved in 10% DMSO in FBS in liquid nitrogen [96].

To isolate a Hofbauer cell and fibroblast-enriched fraction, we subjected the tissue to a final 1mg/mL Collagenase A, 0.2mg/mL DNase I 20mL enzymatic digestion in an RPMI medium containing 25mM HEPES, 10% FBS, and 1% antibiotic/antimycotic for 1 hour at 37°C. The digestion supernatant was strained through a 104µm mesh with a pestle (Millipore Sigma, #CD1-1KT) and rinsed with 3mL FBS and 2 rounds of 10mL RPMI medium to quench 20mL of Collagenase A-DNase I digestion solution and facilitate straining. We performed sequential density gradient centrifugations by centrifuging columns at 1200rcf for 20 minutes at room temperature with no brake. We loaded the single cell suspension onto a 20%/40% Percoll discontinuous gradient and collected cells from the 20%/40% interface. We loaded these cells onto a second 20%/25%/30%/35% discontinuous Percoll gradient. Cells were collected and pooled from the 20%/25%, 25%/30%, and 30%/35% interfaces. These cells were cryopreserved in 10% DMSO in FBS in liquid nitrogen.

3.3.2 Magnetic activated bead cell type sorting

Cryopreserved single cell suspensions were quickly thawed at 37°C, size-filtered at 30µm to eliminate debris, and resuspended in 4°C DF medium for the cytotrophoblast

fraction or RPMI medium for the Hofbauer cell/fibroblast fraction. Both fractions were subsequently resuspended in a 4°C BSA rinsing buffer, consisting of 0.5% Bovine Serum Albumin stock solution (Miltenyi Biotec, #130-091-376) in autoMACS washing solution (Miltenyi Biotec, #130-092-987). All media and cell suspensions were kept on ice for the remainder of the magnetic-activated sorting experiment. Following manufacturer's instructions, the Miltenyi Biotec octoMACS platform (Miltenyi Biotec, #130-042-108) was used to magnetically sort the target placental cell types cytotrophoblasts, Hofbauer cells, and fibroblasts. Briefly, the cytotrophoblast fraction was incubated for 10min at 4°C in the dark with 1:11 REAfinity™ allophycocyanin-conjugated anti-HLA-ABC antibody (APC-anti-HLA-ABC, Miltenyi Biotec, #130-101-467, lot #5191002371). The Hofbauer/fibroblast fraction was incubated for 10min at 4°C in the dark with 1:11 REAfinity™ allophycocyanin-conjugated anti-EGFR antibody (APC-anti-EGFR, Miltenyi Biotec, #130-110-529, lot #5190107002). Each fraction was washed with BSA rinsing buffer and incubated for 15min at 4°C in the dark with 1:5 anti-allophycocyanin MACS® MicroBeads (Miltenyi Biotec, #130-090-855, lot #5190809239). Each fraction was loaded onto an MS column (Miltenyi Biotec, #130-042-201) for magnetic separation. The EGFR- Hofbauer/fibroblast fraction was washed and incubated for 10min at 4°C in the dark with 1:11 REAfinity™ phycoerythrin-conjugated anti-CD10 antibody (anti-CD10-PE, Miltenyi Biotec, #130-114-5-2, lot #5190109168). The Hofbauer/fibroblast fraction was then washed and incubated for 15min at 4°C in the dark with 1:5 anti-phycoerythrin MACS® MicroBeads (Miltenyi Biotec, #130-105-639). The Hofbauer/fibroblast fraction was loaded onto an MS column

for magnetic separation. The CD10- fraction is enriched for fibroblasts and the CD10+ fraction is enriched for Hofbauer cells [96].

3.3.3 DNA Extraction

To collect whole tissue DNA samples, approximately 15mg of snap-frozen villous tissue was added to 180µL Buffer ATL (Qiagen, #939011) and 20µL proteinase K (Qiagen, #19131) to Lysing Matrix D vials (MP Biomedicals, #116913050-CF). Samples were homogenized on the MP-24 FastPrep homogenizer (MP Biomedicals, #116004500) at 6m/s, setting MP24x2 for 35s.The cryogenically stored cell type fractions were quickly thawed at 37°C for less than five minutes. DNA was extracted from each fraction using the DNeasy Blood and Tissue Kit (Qiagen, #69504, lot #157020715, 1630203096, 563011175) according to manufacturer's instruction for each magnetically separated sample, the syncytiotrophoblast-enriched sample, and the bulk whole villous tissue sample after homogenization.

3.3.4 DNA methylation measurement

DNA was submitted to the University of Michigan Epigenomics Core for quality assessment and bisulfite conversion. DNA quantity was measured with the Qubit High Sensitivity dsDNA assay (ThermoFisher Scientific, #Q32854). DNA quality was assessed with the TapeStation genomic DNA kit (Agilent, #5067-5365). Samples too dilute to be compatible with the kit were first concentrated using AMPure Beads (Beckman Coulter, #A63880) and re-quantified. An aliquot of 250ng DNA from each sample was bisulfite converted using the EZ DNA Methylation kit (Zymo Research, #D5001) according to manufacturer's instructions for Illumina DNA methylation arrays.

Samples were then sent to the University of Michigan Advanced Genomics Core for hybridization to the Infinium MethylationEPIC BeadChip v1.0 array (Illumina, #WG-317-1003), washing, and scanning according to the manufacturer's instructions. Samples were randomized across columns and rows (within plates) to reduce potential confounding due to batch effects.

3.3.5 DNA methylation data preprocessing and quality control

In addition to the data generated in this study, we also downloaded data from three previously published cell type-specific DNA methylation studies of the adult blood immune cells (GEO accession GSE110554, accessed via FlowSorted.Blood.EPIC R package v1.99.5 [203]) B cells, CD4+ T cells, CD8+ T cells, monocytes (Mono), natural killer cells (NK), and neutrophils (Neu) [65], nucleated red blood cells from umbilical cord blood (GEO accession, accessed via FlowSorted.CordBloodCombined.450k R package v1.8.0 [204]) [60], and placental cells (GEO accession GSE159526, accessed via ExperimentHub R package v2.0.0 [205]) [201]. Raw DNA methylation data were preprocessed and managed via the minfi R package v1.38.0 [206] and the ewastools R package v1.7 [207].

Datasets generated on the Illumina 450k microarray were preprocessed and quality controlled separately from those generated on the Illumina EPIC DNA methylation microarray. We preprocessed raw DNA methylation data with Noob background correction and dye-bias normalization. We excluded samples with low-intensity values below 10.5 relative fluorescence units or mismatched genotype or sex (n=0 samples excluded). For the EPIC array samples (n=189), we excluded 1 probe with an average bead count less than 5, 19,760 probes with >5% failure rate for

detection p-value >0.01, 23,448 cross-reactive probes [208], and 17,767 sex-specific probes. For the 450k array samples (n=132), we excluded 20,024 probes with >5% failure rate for detection p-value >0.01, 14,151 cross-reactive probes [208], and 9,646 sex-specific probes. Bead hybridization information was unavailable for the 450k umbilical cord samples and only the nucleated red blood cells (n=11) were retained for downstream analysis. Finally, to correct for type I vs. type II probe bias, we also performed Beta MIxture Quantile dilation normalization [209] via the ChAMP R package v2.22.0 [210]. All downstream DNA methylation analyses used beta values, or the methylation rate for a given site.

To make our results generalizable to the Illumina 450k array, we subset the EPIC array samples to only those DNA methylation sites shared across both arrays for all downstream analyses. To assess cell fraction purity, we assessed unsupervised DNA methylation profile clustering and predicted cell type lineage from an external reference [67]. To reduce dimensionality for unsupervised DNA methylation clustering, we selected the top 10% highest variance DNA methylation sites and performed principal components dimension reduction. To integrate data from cell fractions across study sources and visualize clustering results with uniform manifold projection [118], we used the mutual nearest neighbor batch correction approach [119] via FastMNN in the SeuratWrappers R package, version 0.3.0 [120]. Iterative clustering and sub-clustering with the Seurat R package v4.0.1 [122] function FindClusters at different resolution parameters were evaluated using cluster stability via clustering trees with clustree R package v0.4.4 [121]. We predicted the lineage composition (epithelial, immune cell, or stromal) of each sample with the robust partial correlation algorithm and the using the

centEpiFibIC.m dataset [67] implemented in EpiDISH R package v2.8.0 [83]. Samples that clustered inconsistently with other cell types and had less than 70% predicted epithelial lineage for cytotrophoblasts and syncytiotrophoblast or less than 50% predicted stromal or immune cell identity for other cell types were excluded from downstream analysis. 7 syncytiotrophoblast samples from this study and 1 syncytiotrophoblast sample from Yuan et al. [201] were excluded on this basis. We visualized sample and probe inclusion using a flow chart. For all included samples, we used a table to describe the number and frequency of samples by study source, fetal sex, cell type, Illumina array type, and cell separation method.

3.3.6 Characterization of cell type fractions with differential methylation analysis and biological process enrichment

To characterize cell type fractions, we assessed cell type-specific DNA methylation patterns by fitting linear models adjusted for sex with empirical Bayes standard error moderation in the *limma* R package v 3.48.3 [211]. To identify cell type-specific differentially methylated sites, we compared percent methylation values in one cell type against the average across other cell types. We used false discovery adjusted q-values to account for multiple comparisons. We visualized differences in DNA methylation between cell types using volcano plots of the average methylation difference and the $-\log_{10}$(q-value). To be considered biologically meaningful, we instituted an absolute percent methylation difference threshold cutoff of 10%. Statistical significance was assessed at a false discovery adjusted q-value < 0.001. We visualized overlap in significant sites across cell types we used an upset plot, separately for hyper- and hypomethylated sites in **Figure 3.4**.

To characterize the biological significance of differential methylation results, we performed gene ontology enrichment for the top 10,000 hyper- and hypomethylated sites per cell type with the missMethyl R package v1.26.1 [212]. This approach identifies overrepresented genes and their ontologies among differentially methylated sites while accounting for the number of sites tested per gene as well as probes associated with multiple genes. We visualized overlap in enriched gene ontologies using an upset plot in **Figure 3.5**.

3.3.7 Creation and application of a placental DNA methylation deconvolution reference

To identify cell type-discriminating DNA methylation sites, we ranked the top 50 each of hyper- and hypomethylated sites per cell type by F-test using the minfi R package v1.38.0 [206]. At these probes, we visualized DNA methylation levels across cell types using a heat map in **Figure 3.6**. These cell type-discriminating DNA methylation sites were used as the cell type DNA methylation references for deconvolution of whole tissue placental villous samples present in the analyzed datasets. We used the robust partial correlation algorithm implemented in EpiDISH R package v2.8.0 [83] to perform deconvolution. We visualized estimated cell proportions in whole tissue using a scatter plot in **Figure 3.7**. For each cell type, we calculated the estimated proportion median and quartiles.

3.4 Results

3.4.1 Placental cell type-specific DNA methylation profiles

After quality control procedures (Figure 3.1 DNA methylation quality control inclusion/exclusion flow chart.**Figure 3.1**), we analyzed DNA methylation data across 407,628 sites from 192 new and previously published DNA methylation profiles from 12 cell type-specific fractions, including cytotrophoblasts (n=32), endothelial cells (n=19), Hofbauer cells (n=26), stromal cells (n=29), syncytiotrophoblast-enriched fractions (n=4) [201], six types of adult lymphocytes (n=36) [65], and nucleated red blood cells (n=11) [60], as well as 35 bulk placental tissue samples (**Table 3.1**). We analyzed the placental datasets with DNA methylation profiles from adult blood peripheral immune cells and umbilical cord nucleated red blood cells, which may also be present in placental samples, especially those collected in large epidemiologic or biobanking studies. We further subset the DNA methylation sites to only those that are shared across EPIC and 450k platforms to make our deconvolution reference applicable to datasets generated from either platform. Background-corrected, probe type bias-corrected, and normalized DNA methylation proportions (sometimes called beta values) for each DNA methylation site were used for downstream analyses.

We performed several quantitative and qualitative tests to assess cell fraction purity. Consistent with successful cell type isolation, unsupervised clustering revealed that DNA methylation samples clustered by their target cell type cell type. We excluded 7 syncytiotrophoblast samples from this study and 1 from the Yuan et al. dataset [201] from the analysis. These samples clustered inconsistently with other syncytiotrophoblast samples and had a predicted epithelial lineage <70%. All other samples clustered consistently and had >70% predicted epithelial lineage for cytotrophoblasts and syncytiotrophoblast and greater than 50% predicted stromal, epithelial, or immune cell

identity for their appropriate lineages (**Figure 3.2**). Also consistent with successful cell type isolation, UMAP projection shows cell type lineages colocalize.

3.4.2 Characterization of cell type fractions with differential methylation analysis and biological process enrichment

To identify placental cell type-defining DNA methylation patterns and provide additional evidence of successful cell type isolation, we performed a differential methylation analysis. To identify cell type-specific differentially methylated sites, we compared percent methylation values in one cell type against the average across other cell types. The number of hyper- and hypomethylated sites per placental cell type ranged from 28,467 to 98,029 (**Figure 3.3**). Trophoblast subtypes had the largest number of differentially methylated sites. Trophoblast subtypes (cytotrophoblasts and syncytiotrophoblasts) also had the largest number of overlapping differentially methylated sites (**Figure 3.4**). Hofbauer cells had the largest number of unique differentially methylated sites.

3.4.3 Characterization of cell type fractions biological process enrichment of differentially methylated sites

The number of enriched pathways among placental cell types ranged from 411 for Hofbauer cells to 212 for syncytiotrophoblasts. Cell type-defining biological processes were highlighted by enriched biological processes unique to each cell type. The number of biological processes unique to placental cell types ranged from 30 for stromal cells to 11 for syncytiotrophoblast (Error! Reference source not found.). For e xample, enriched unique Hofbauer cell pathways included vascular process in the

circulatory system (p_{adj}=0.004), positive regulation of IL-10 production (p_{adj}=0.004), and lamellipodium organization (p_{adj}=0.004). Enriched pathways unique to endothelial cells included endothelium development (p_{adj}=0.007), regulation of cell junction assembly (p_{adj}=0.007), and basement membrane organization (p_{adj}=0.008).

3.4.4 Placental DNA methylation deconvolution reference

To create a deconvolution reference for placental villous tissue deconvolution, we ranked the top 50 hyper- and hypomethylated sites per cell type by F-test. This approach identifies cell type defining DNA methylation sites for application in deconvolution. We identified 1,101 unique cell type defining sites Figure 3.6. Using these sites, we used the robust partial correlation algorithm to estimate the cellular composition of the 35 villous whole tissue samples collected in this study and Yuan et al. [201] (Error! Reference source not found.). Consistent with placental biology, cell c omposition median (25^{th} percentile, 75^{th} percentile) estimates were 59.2% (52.7%, 63.1%) syncytiotrophoblast, 20.1% (16.4%, 24.0%) stromal, 10.2% (8.1%, 13.0%) cytotrophoblast, 7.2% (6.3%, 8.5%) endothelial, and 1.4% (0.6%, 2.1%) Hofbauer cells. Other cell types had median estimates of less than 1%. Only CD4+ T cells and natural killer cells had median estimates above 0%.

3.5 Discussion

To provide the most reliable, generalizable, and versatile placental DNA methylation deconvolution reference to date, we generated primary cell type-specific DNA methylation profiles for placental cell types and integrated them with previously published samples. We successfully generated primary cell type-specific DNA

methylation profiles for placental cytotrophoblasts, fibroblasts, and Hofbauer cells. We uniformly processed and integrated these samples with previously published DNA methylation profiles of placental endothelial cells, Hofbauer cells, trophoblasts, stromal cells, and syncytiotrophoblasts, as well as nucleated red blood cells from umbilical cord blood and B cells, CD4+ T cells, CD8+ T cells, monocytes, natural killer cells, and neutrophils from adult blood. Differential methylation and gene ontology enrichment analyses characterized cell type-defining DNA methylation patterns and their associated biological processes. Finally, we confirmed in matched villous tissue samples, that deconvoluted cell type proportions were consistent with expected placental biology. This cell type DNA methylation reference panel can robustly estimate cell composition from placental DNA methylation data in epidemiological studies to detect unexpected non-placental cell types, reveal biological mechanisms, and improve casual inference.

The application of deconvolution techniques to address cell heterogeneity in bulk tissue molecular assays and especially its application to the placenta and DNA methylation is a developing field with limited findings. Prior studies were generally limited to one placental cell type and assessed DNA methylation on superseded microarray platforms [198–200]. Yuan et al., 2021 represented the most ambitious study to date in developing placental cell type-specific DNA methylation profiles for a deconvolution application [201]. Our independently collected samples were isolated using complementary alternative sorting and antigen marker strategies. Nonetheless, unsupervised clustering resulted in cell type-specific clustering as expected and deconvoluted bulk tissue proportions were consistent with Yuan et al., 2021's estimates [201]. Despite the alternative cell sorting strategies, cytotrophoblasts clustered with

trophoblasts, fibroblasts clustered with stromal cells, and Hofbauer cells clustered with Hofbauer cells.

There are several strengths to this study. We generated placental cell type-specific DNA methylation profiles using established protocols to isolate placental cytotrophoblasts, fibroblasts, and Hofbauer cells [96, 98]. The additional independent samples we collected using magnetic activated cell sorting allowed for sample comparisons with other placental cell types to compare within and across study to identify sample outliers with unsupervised clustering and improve confidence in reference DNA methylation profiles, including using an externa resource to predict lineage identity [67]. By limiting our selection of DNA methylation microarray probes used for deconvolution to only those shared on the 450k and EPIC array platforms, our results will be generalizable to data generated on either platform. The inclusion of nonplacental immune cells in the deconvolution panel allows for the detection of intermingled nonplacental cell types, providing feedback for investigators on placental sampling techniques and identifying potentially cryptic or unexpected cell distribution patterns in study samples. A similar approach has proved beneficial in using DNA methylation data to verify sample-of-origin single nucleotide polymorphism and sex checks in DNA methylation studies to detect sample swaps or contamination [207].

There are also limitations to this study. Despite applying an existing syncytiotrophoblast isolation protocol [95], we were unable to characterize syncytiotrophoblast samples that met our strict inclusion criteria. Our ability to confirm deconvolution performance was hampered by a lack of a gold standard reference. This is an ongoing limitation in the field. *In situ* cell counting strategies provide little

information about the overall tissue and placental cell sorting techniques are laborious and ill-suited to isolate the syncytiotrophoblast. Further developments in single-cell epigenetics approaches may ameliorate this limitation. Single-nuclei methods may enable more specific characterization of the syncytiotrophoblast [175]. Because current reference-based deconvolution approaches preclude the opportunity to transparently account for batch or confounding effects, particularly when combining data from different studies, our results may be biased by such factors. However, our unsupervised clustering approach highlighted cell type identity as the major driver of DNA methylation variation.

In conclusion, this study developed a robust and flexible DNA methylation reference panel for term placental bulk villous tissue. We provide newly sequenced cell type-specific placental DNA methylation profiles and integrated our samples with existing data, including non-placental cell types that may be present in bulk villous tissue samples. Our independent samples isolated with a complementary approach allowed us to verify and biologically characterize placental cell type-defining DNA methylation profiles. Epigenetics is a promising and increasingly investigated approach to link environmental exposures to biological mechanisms of placental disease and dysfunction. The deconvolution approach developed here advances perinatal epidemiology by allowing investigators to model cell composition in complex study questions to address critical limitations of tissue-level DNA methylation measures.

3.6 Figures and Tables

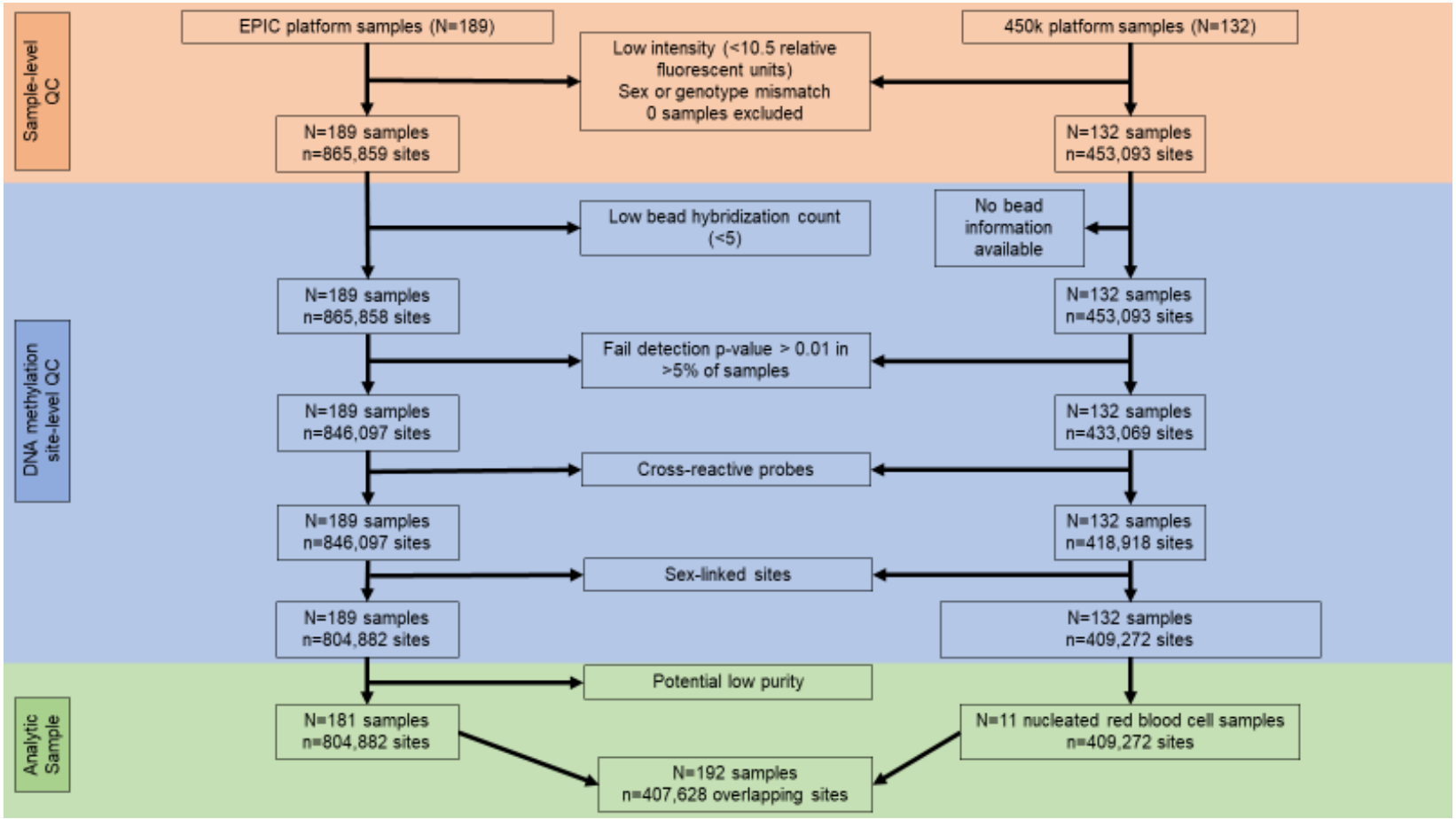


Figure 3.1 DNA methylation quality control inclusion/exclusion flow chart.

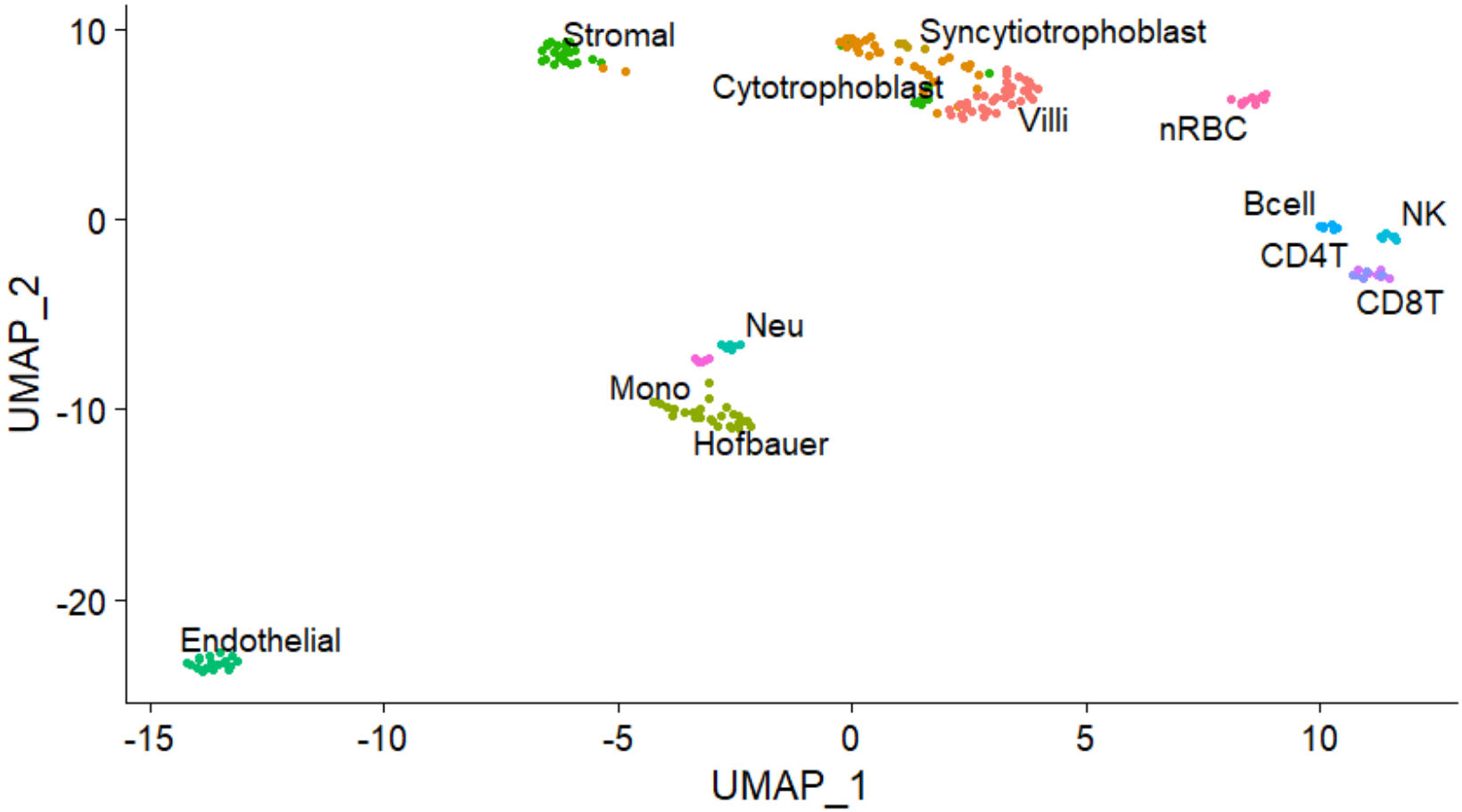


Figure 3.2 Low-dimensional representation of variable DNA methylation sites across all cell type and whole tissue samples shows samples cluster by cell type. Abbreviations: monocytes (Mono), natural killer cells (NK), neutrophils (Neu), and nucleated red blood cells (nRBC).

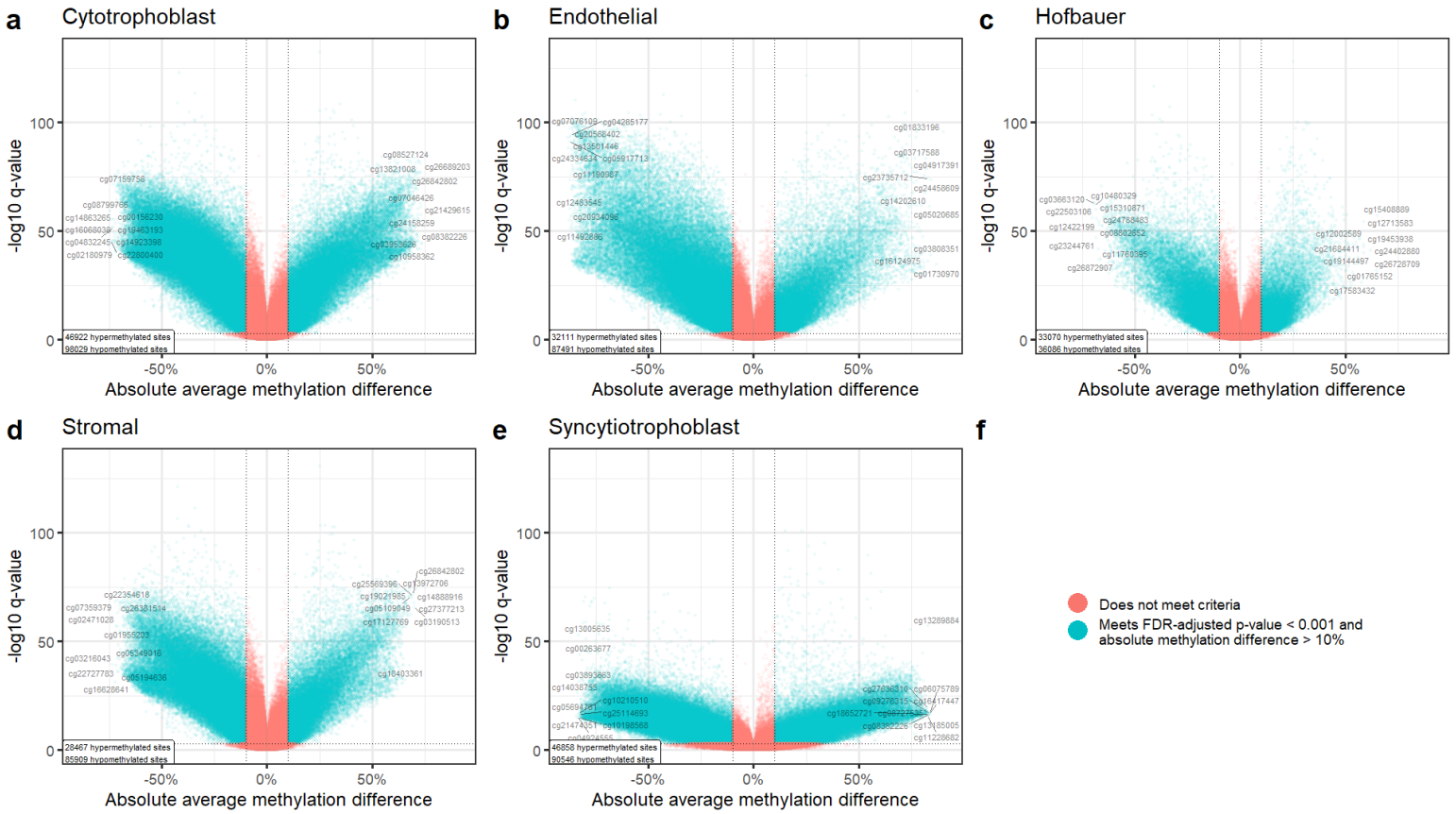


Figure 3.3 Placental cell type differential DNA methylation. Volcano plots comparing site-specific DNA methylation levels in one cell type against the average across all other cell types.

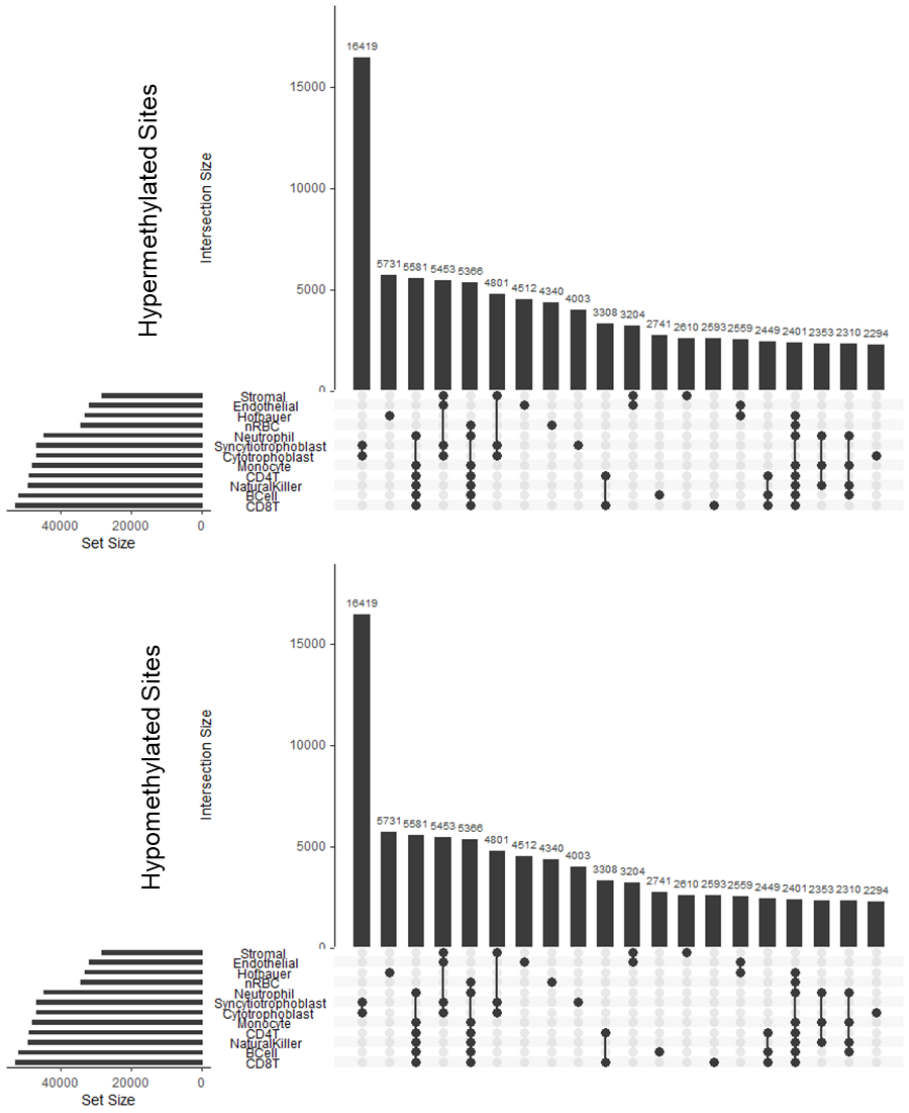


Figure 3.4 Upset plots of differential methylation results. Abbreviations: monocytes (Mono), natural killer cells (NK), neutrophils (Neu), and nucleated red blood cells (nRBC).

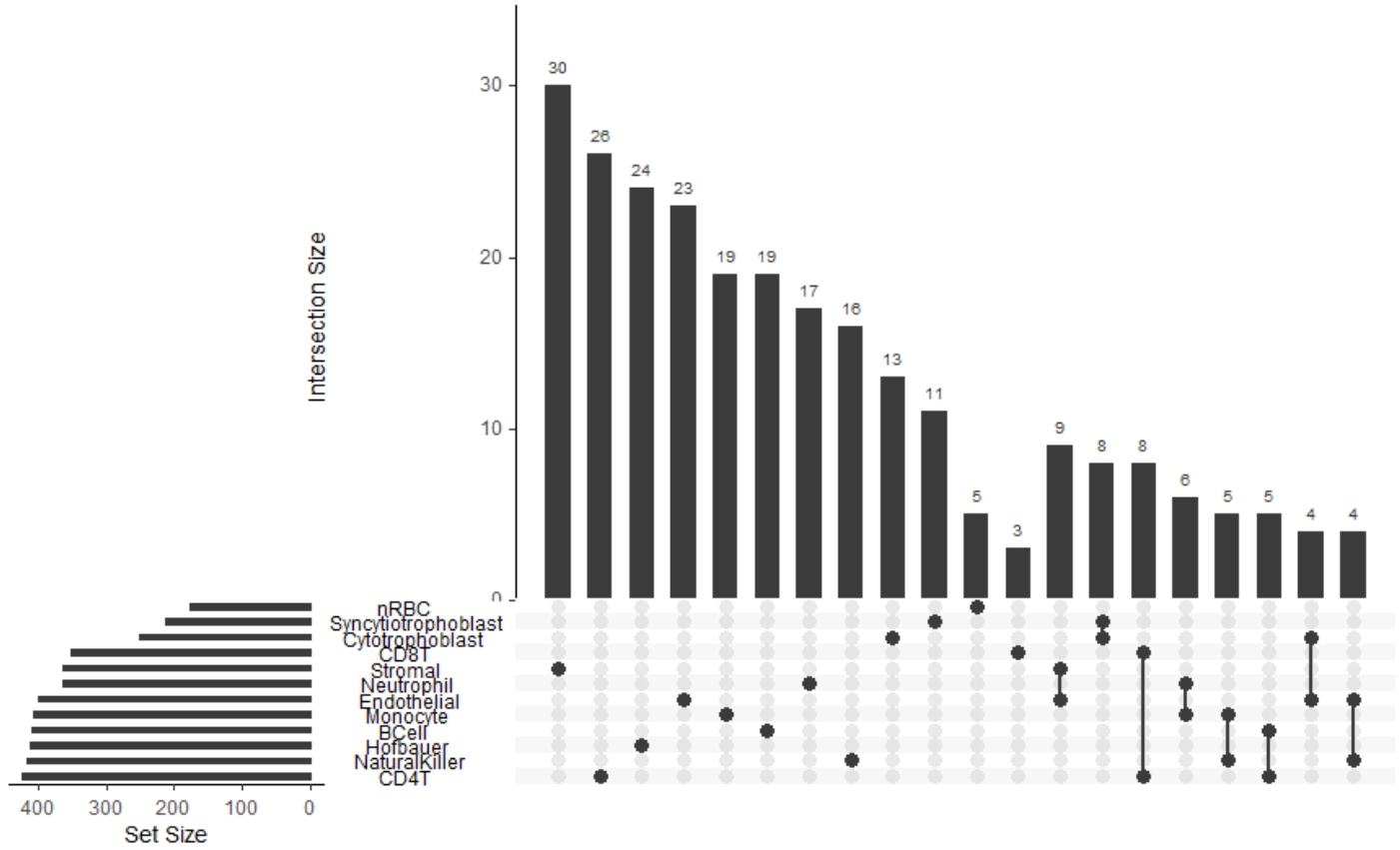


Figure 3.5 Upset plot of unique enriched pathways. Abbreviations: monocytes (Mono), natural killer cells (NK), neutrophils (Neu), and nucleated red blood cells (nRBC).

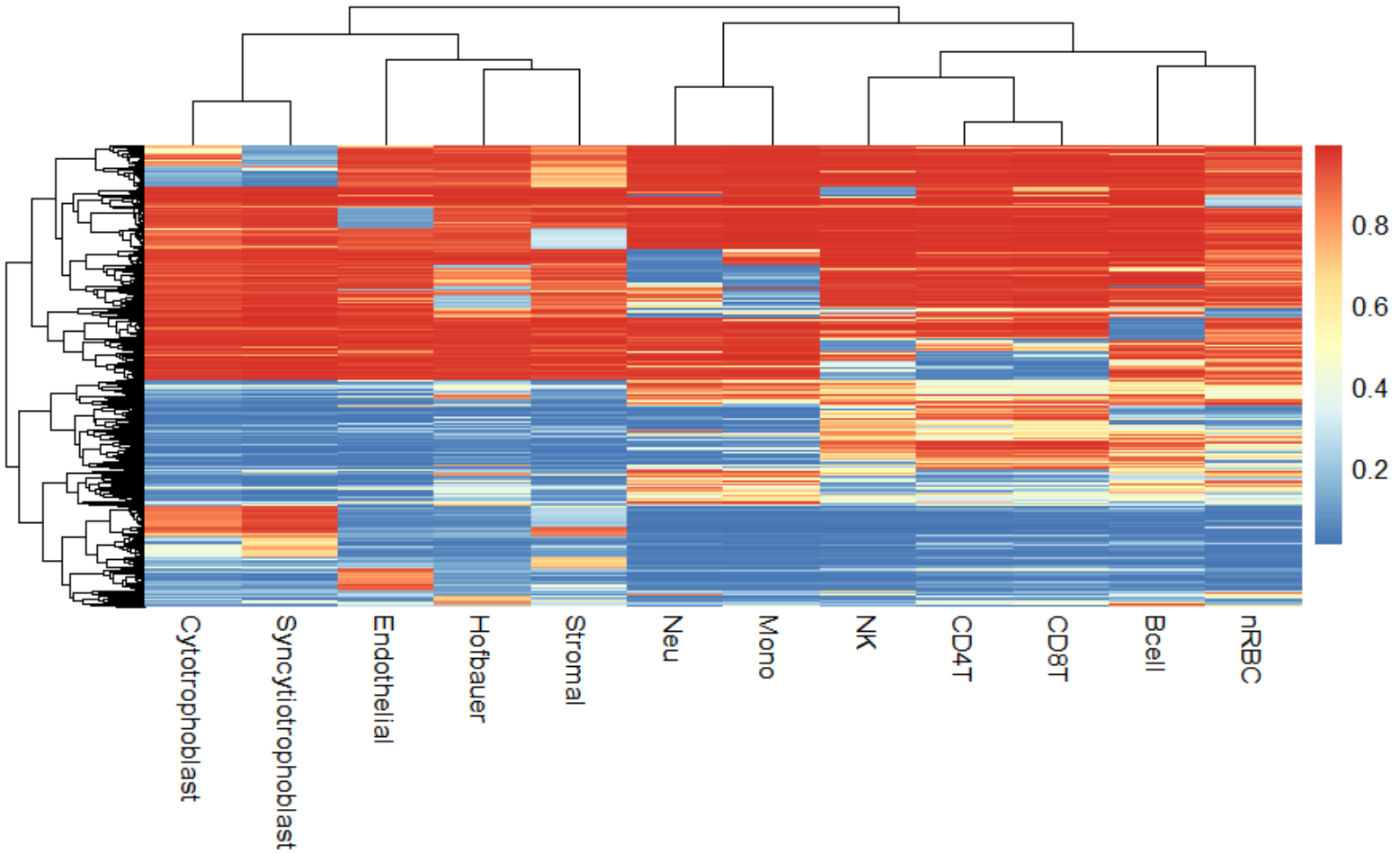


Figure 3.6 Heatmap of the cell type defining probes used for deconvolution. Color gradient represents percent methylation measure. Abbreviations: monocytes (Mono), natural killer cells (NK), neutrophils (Neu), and nucleated red blood cells (nRBC).

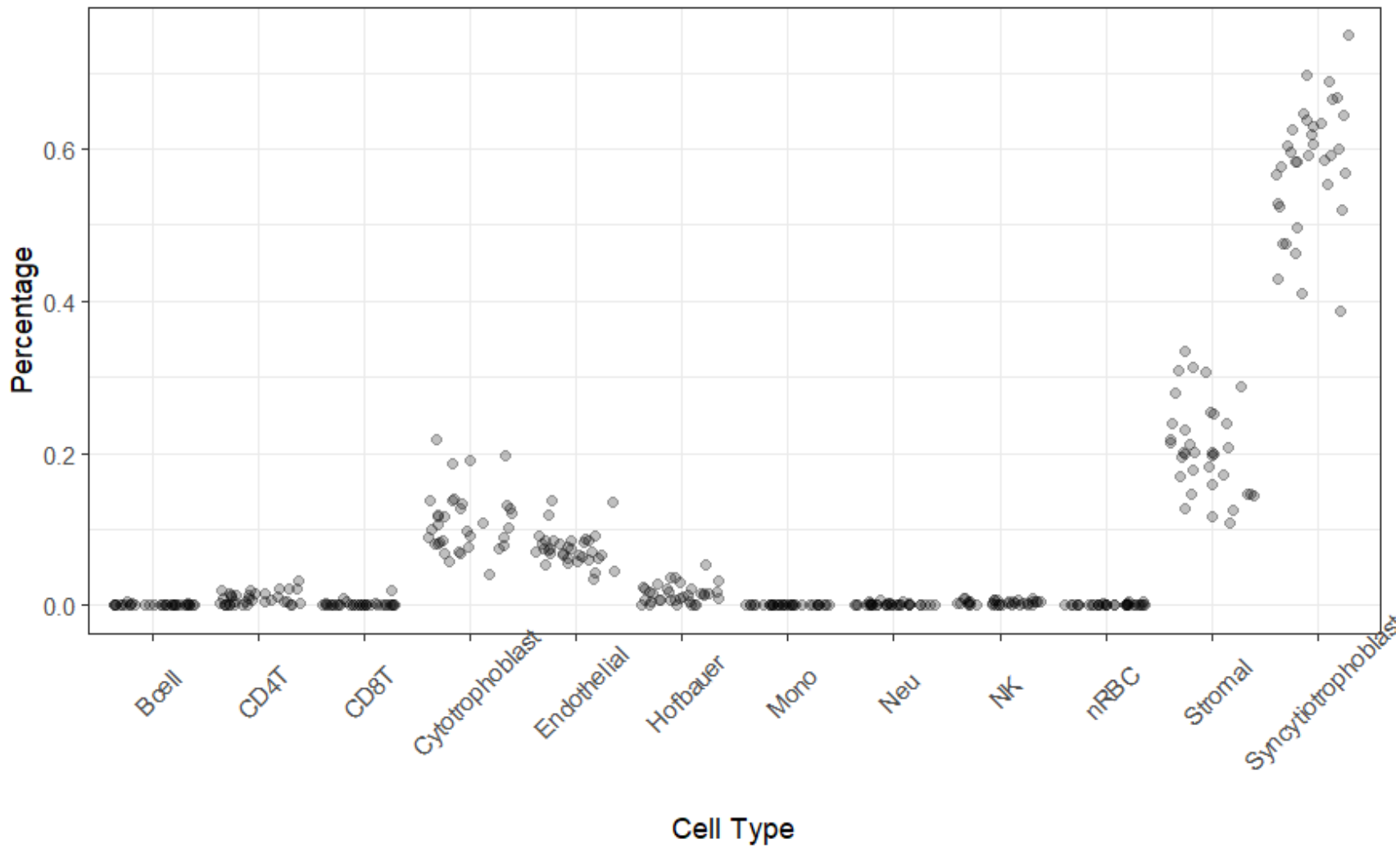


Figure 3.7 Reference-based cell composition estimates from whole placental villous tissue samples. Abbreviations: monocytes (Mono), natural killer cells (NK), neutrophils (Neu), and nucleated red blood cells (nRBC).

Table 3.1 DNA methylation samples used for development of deconvolution reference. Abbreviations: monocytes (Mono), natural killer cells (NK), neutrophils (Neu), and nucleated red blood cells (nRBC).

Characteristic	**Campbell**, N = 47[1]	**Yuan**, N = 98[1]	**Salas**, N = 36[1]	**FlowSorted.CordBlood.450k**, N = 11[1]
Fetal Sex				
F	35 (74%)	52 (53%)	6 (17%)	8 (73%)
M	12 (26%)	46 (47%)	30 (83%)	3 (27%)
Cell Type				
Villi	16 (34%)	19 (19%)	0 (0%)	0 (0%)
Cytotrophoblast	13 (28%)	19 (19%)	0 (0%)	0 (0%)
Syncytiotrophoblast	0 (0%)	4 (4.1%)	0 (0%)	0 (0%)
Hofbauer	8 (17%)	18 (18%)	0 (0%)	0 (0%)
Stromal	10 (21%)	19 (19%)	0 (0%)	0 (0%)
Endothelial	0 (0%)	19 (19%)	0 (0%)	0 (0%)
Neu	0 (0%)	0 (0%)	6 (17%)	0 (0%)
NK	0 (0%)	0 (0%)	6 (17%)	0 (0%)
Bcell	0 (0%)	0 (0%)	6 (17%)	0 (0%)
CD4T	0 (0%)	0 (0%)	6 (17%)	0 (0%)
CD8T	0 (0%)	0 (0%)	6 (17%)	0 (0%)
Mono	0 (0%)	0 (0%)	6 (17%)	0 (0%)
nRBC	0 (0%)	0 (0%)	0 (0%)	11 (100%)
Illumina Array Type				
IlluminaHumanMethylation450k	0 (0%)	0 (0%)	0 (0%)	11 (100%)
IlluminaHumanMethylationEPIC	47 (100%)	98 (100%)	36 (100%)	0 (0%)
Cell Separation Method				
FACS	0 (0%)	98 (100%)	36 (100%)	7 (64%)
MACS	47 (100%)	0 (0%)	0 (0%)	4 (36%)

[1] n (%)

Chapter 4 Metals Exposure During Pregnancy is Associated with Term Placental Cell Composition in Two Prospective Pregnancy Cohorts

Authors: Campbell KA[1], Bakulski KM[1], Dou JF[1], Feinberg JI[2], Croen LA[3], Newschaffer CJ[4], Hertz-Picciotto I[5], Schmidt RJ[6], Fallin MD[7]
Affiliations:
[1]Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48105, USA.
[2]Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205, USA.
[3]Division of Research, Kaiser Permanente, Oakland, CA 94612, USA.
[4]Department of Biobehavioral Health, Penn State University, State College, 325 Health and Human Development Building, University Park, PA, 16802, USA.
[5]Department of Public Health Sciences and The MIND Institute, School of Medicine, University of California-Davis, UC Davis School of Medicine - Medical Sciences 1C, Suite 123, Davis, CA, 95616, USA.
[6]Department of Public Health Sciences and The MIND Institute, School of Medicine, University of California-Davis, UC Davis School of Medicine - Medical Sciences 1C, Suite 123, Davis, CA, 95616, USA.
[7]Department of Mental Health, Johns Hopkins University, 624 N. Broadway, Hampton House 850, Baltimore, MD, 21205, USA.

4.1 Abstract

Malapropos exposure to essential and toxic metals during pregnancy and to the placenta is widespread and linked to adverse perinatal outcomes. Healthy placental morphology is essential to pregnancy, but little is known about the effects of *in utero* metals exposure on placental cell composition at term. We analyzed samples from two prospective pregnancy cohort studies, the Markers of Autism Risk in Babies and Learning Early Signs (MARBLES n=83) and Early Autism Risk Longitudinal Investigation (EARLI) (n=94). Urinary metals (n=18) concentrations were measured during early and late gestation. Placental DNA methylation at birth was measured with

EPIC (MARBLES) or 450k (EARLI) Illumina microarrays. We estimated placental cellular composition via deconvolution with the DNA methylation external reference developed in **Chapter 3**. Beta regression models tested associations between metals concentrations (p=18) at each time point and each placental cell type proportion (p=5), adjusted for sex, gestational age, maternal race/ethnicity, maternal age, and maternal education. Primary effect estimates were meta-analyzed across cohorts at each time point. Consistent with placental biology, median (Q1, Q3) cell composition in EARLI was 63.2% (59.0%, 68.0%) syncytiotrophoblast, 17.1% (15.1%, 19.4%) stromal, 7.8% (5.9%, 10.7%) cytotrophoblast, 5.6% (4.4%, 7.6%) endothelial, and 1.2% (0.3%, 2.3%) Hofbauer cells. In MARBLES, estimated cell composition was 60.1% (54.6%, 68.0%) syncytiotrophoblast, 18.6% (13.3%, 22.3%) stromal, 5.6% (3.0%, 9.2%) cytotrophoblast, 5.9% (4.1%, 7.6%) endothelial, and 2.9% (2.2%, 4.4%) Hofbauer cells. On average, 69.8% of metals concentrations were positively correlated within each cohort and timepoint. Early gestation zinc median concentration (Q1, Q3) was 15.9 µg/L (135.7 µg/L, 490.4 µg/L). At the mean cytotrophoblast cell type proportion across both cohorts of 7.8%, a doubling in early gestation zinc concentration was associated with a 0.8% mean decrement in cytotrophoblasts to 7.0% (POR = 0.89, 95% CI [0.82, 0.97], q=0.26). Late gestation barium median concentration was 2.4 µg/L (1.2 µg/L, 4.2 µg/L). At the mean Hofbauer cell type proportion across both cohorts of 2.3%, a doubling in late gestation barium concentration was associated with a 0.2% increment in Hofbauer cells to 2.5% (POR = 1.08, 95% CI [1.02, 0.14], q=0.25). Late gestation nickel median concentration was 4.6 µg/L (2.8 µg/L, 7.0 µg/L). Similarly, a doubling in late gestation nickel concentration was associated with a 0.4% increment in Hofbauer cell proportion

to 2.7% (POR = 1.16, 95% CI [1.05, 1.27], q=0.25). Late gestation cadmium median concentration was 0.1 µg/L (0.0 µg/L, 0.2 µg/L). At the mean syncytiotrophoblast cell type proportion across both cohorts of 62.5%, exposure to cadmium concentrations above the limit of detection was associated with a 3.0% mean decrement in syncytiotrophoblasts to 59.5% (POR = 0.88, 95% CI [0.78, 0.98], q=0.31). Late gestation copper median concentration was 9.2 µg/L (6.0 µg/L, 13.5 µg/L). At the mean endothelial cell type proportion across both cohorts of 6.1%, a doubling in late gestation copper concentration was associated with a 0.7% increment in endothelial cells to 6.8% (POR = 1.11, 95% CI [1.04, 1.18], q=0.24). We provide evidence that prenatal urinary metals concentrations are associated with term placental cell composition, which may have implications for increased disease risk related to disruption of placental structure as well as for molecular studies of the placenta that fail to account for potential cell composition differences.

4.2 Introduction

Prenatal exposure to known developmentally toxic heavy metals such as lead, cadmium, and mercury is associated with adverse health effects [213]. Similarly, overexposure or underexposure to essential metals such as selenium and manganese during pregnancy have been associated with adverse reproductive health outcomes [214–217]. Exposure to these metals is ubiquitous worldwide and has been demonstrated in pregnant women [218–221]. At varying concentrations, each of these metals has been discovered in maternal and fetal bio-samples, including the placenta [218–221]. Manganese and selenium cross the placental barrier, representing direct exposure to the developing fetus [222–224]. Lead and certain forms of mercury also

readily cross the placental barrier; conversely, cadmium and inorganic mercury do not appear to readily cross the placental barrier but do accumulate in the placenta, even serving as a metals exposure sink during pregnancy [225]. The placenta is directly exposed to each of these metals. How or if these or other essential or toxic metals impact placental structure, especially prenatally, is not well understood.

The placenta is the hallmark organ of gestation, fulfilling many core organ functions for the developing fetus as well as remodeling and maintaining the fetal-maternal environment to sustain gestation [6, 8]. Healthy placental morphology is essential to placental function [6, 8]. Direct toxicological data on heavy metals exposure in animal models and limited evidence from human studies have identified associations between malapropos exposure to toxic and essential metals and disrupted placental morphology and function [226–228]. Existing data in humans is largely limited to gross histopathological evaluation of the placenta and small sample sizes due to the challenges of assessing metals exposure and placental morphology at scale, requiring substantial expertise, time, and cost. Reference-based *in silico* deconvolution has emerged as a cost-effective way to quantify the cell type composition from molecular assays of bulk tissue samples [46, 58, 59]. The application of recently available cell type-specific placental DNA methylation reference profiles (**Chapter 3**) may elucidate relationships between placental cell composition and metals concentrations at scale.

Assessing the effects of perinatal concentrations of toxic and essential metals on term placental cell composition could provide insight into how metals may affect placental structure and subsequent function. Importantly, metals could be targeted for remediation or mitigating therapies to reduce the harmful effects of metals exposure on

a public health scale. Here, in two prospective birth cohorts, we test the hypothesis that prenatal concentrations of essential or toxic metals as measured by maternal urinary metals concentrations during early or late gestation affect the abundance of cell types in the term placenta as estimated via the reference-based DNA methylation deconvolution approach developed in **Chapter 3**.

4.3 Methods

4.3.1 Study sample description

We analyzed samples from the Markers of Autism Risk in Babies and Learning Early Signs (MARBLES) and the Early Autism Risk Longitudinal Investigation (EARLI) prospective birth cohorts. Each cohort recruits pregnant women who have already given birth to a proband already diagnosed with autism spectrum disorder and follows the participants for a postnatal 3 years.

The Early Autism Risk Longitudinal Investigation (EARLI): EARLI [229] is a family history-enriched [230] prospective cohort study. Pregnant women who have had a previous child (proband) with an autism spectrum disorder diagnosis are recruited and followed until the sibling of the proband reaches three years of age, referred to as the sibling (n=233 mother-child dyads) between November 2009 and March 2012. Inclusion criteria included: have had a biological child diagnosed with autism spectrum disorder, competent to communicate in English (or Spanish at two sites), 18 years of age or older, live within two hours of a study site, and no more than 28 weeks pregnant. Exposure assessment was performed during prenatal care visits with sibling bio-samples collected at delivery. EARLI study sites include Philadelphia (Drexel University), Baltimore (Johns Hopkins University), San Francisco/Oakland (Northern

California Kaiser), and Sacramento (University of California at Davis). Pregnant women were recruited from early intervention and special education systems and specialty clinics. In San Francisco/Oakland, participants were recruited from Kaiser Permanente enrollees. In Sacramento, participants were recruited from children receiving services through the California Department of Developmental Services.

Markers of Autism Risk in Babies and Learning Early Signs (MARBLES): MARBLES [231] is also a family history-enriched [230] prospective cohort study. Pregnant women or women planning another pregnancy who have had a previous child (proband) with an autism spectrum disorder diagnosis are recruited and followed until the sibling of the proband reaches three years of age, referred to as the sibling (n=389 women, 425 children) between December 2006 and July 2016. Pregnant women whose children would be related as a half-sibling or a closer relative to a proband were also recruited. Participants were recruited from children receiving services through the California Department of Developmental Services. Other inclusion criteria included competent to communicate in English, 18 years of age or older, and live within 2.5 hours of the Davis/Sacramento California study area. Exposure assessment was performed during prenatal care visits with sibling bio-samples collected at delivery.

4.3.2 Exposure assessment and operationalization

Metal exposure measures in both cohort studies were collected from first void urine samples collected during prenatal care visits. In EARLI, urinary bio-samples were collected at two time points during the 1st trimester or early 2nd trimester and during the 3rd trimester. In MARBLES, urinary bio-samples were collected at two time points during the 2nd and 3rd trimesters. We collapsed EARLI metals measures during 1st trimester or

early 2nd trimester into one timepoint called early gestation. In MARBLES, 2nd trimester samples were also classified as early gestation. In EARLI and MARBLES, we refer to 3rd trimester samples as late gestation. We visualized the distributions of metals concentrations by time point and cohort using violin plots in **Figure 4.3**.

Inductively coupled plasma mass spectrometry was used to characterize multiplexed metals levels across 22 metals: antimony, arsenic, barium, cadmium, cesium, chromium, cobalt, copper, lead, manganese, mercury, molybdenum, nickel, selenium, thallium, tin, tungsten, and zinc. Mass spectrometry was carried out by NSF International using the CDC Laboratory Procedure Manual's Urine Multi-Element ICP-DRC-MS #3018.3, modified to increase multiplex capacity on the Thermo Scientific iCAP RQ. Urine specific gravity was assayed using an ATAGO model PAL-10S refractometer, using water as the control sample. We accounted for urinary dilution by including specific gravity as a covariate in the statistical analysis . Metal concentrations below the limit of detection were imputed as the limit of detection divided by the square root of two [232]. Metal concentrations (μg/L) were $\log_2$ transformed and modelled continuously to account for skewed metal concentrations and provide interpretable effect estimates as a doubling in exposure concentration. Alternatively, metals with below limit of detection rates >30% were dichotomized as exposed versus unexposed for statistical analysis. We visualized the inclusion and operationalization of metals using a flow chart.

4.3.3 Placental bio-sample collection and DNA methylation measurements

In EARLI, full thickness tissue punches from a centrally located placental cotyledon and whole umbilical cord blood were collected at delivery. Forceps were used

to extract placental tissue from the fetal side of the placenta, which was archived alongside the umbilical cord blood samples at the Johns Hopkins Biological Repository at -80ºC. DNA was collected from placental samples via the DNeasy Tissue Kit (Qiagen) and quantified with a Nanodrop (ThermoFisher). Cord blood DNA was extracted via the DNA Midi kit (Qiagen) and quantified with a Nanodrop (ThermoFisher). Normalized DNA aliquots were sent to the Center for Inherited Disease Research (Johns Hopkins University) for bisulfite conversion and cleanup with the EZ DNA methylation gold kit (Zymo Research) according to manufacturer's instructions. DNA methylation was assessed via the Infinium HumanMethylation450 BeadChip microarray (Illumina) [233]. Samples were randomly assigned across plates and methylation control gradient and between-plate duplicated tissue controls were used.

In MARBLES, placental tissue was immediately collected from the fetal chorionic villus and frozen. Samples were stored at -80ºC in the UC Davis biorepository. Placental DNA was extracted with the Gentra Puregene kit (Qiagen). Placental DNA was bisulfite converted and cleaned via the EZ DNA methylation gold kit (Zymo Research) and sent to the John Hopkins SNP Center, a processing center shared with the Center for Inherited Disease Research (Johns Hopkins University). DNA methylation was measured via the Infinium HumanMethylationEPIC BeadChip microarray (Illumina). Samples were randomly assigned across plates and methylation control gradient and between-plate duplicated tissue controls were used.

4.3.4 Placental DNA methylation preprocessing and quality control

EARLI placenta (n=133) and cord blood samples (n=175) were run together in two batches on the 450k array and so were preprocessed together. We processed raw

Illumina image files with minfi, R package v1.30.0 in R, v3.6 [206]. We applied Noob background correction and dye-bias normalization [234]. Samples with discordant DNA methylation-predicted and actual fetal sex were excluded (placenta, n=0 excluded; cord blood, n=3 excluded) from downstream preprocessing steps. We excluded 661 microarray probes with >5% failure rate for detection p-value >0.01, 29,153 cross-reactive probes [235], and 48 Y chromosome probes. After exclusion of 5 non-singleton births, there were 128 EARLI placenta samples across 455,650 DNA methylation sites available for analysis. All downstream DNA methylation analyses used beta values, or the methylation rate for a given site.

MARBLES placental samples (n=92) measured on the EPIC microarray were preprocessed separately from the EARLI samples. We processed raw Illumina image files with minfi, R package v1.30.0 in R, v3.6 [206]. We applied Noob background correction and dye-bias normalization [234]. No samples had discordant DNA methylation-predicted and actual fetal sex. We excluded 1699 microarray probes with >5% failure rate for detection p-value >0.01, 43,068 cross-reactive probes [236], and 84 Y chromosome probes. We excluded 2 subsequent siblings from downstream analysis. 90 MARBLES placenta samples across 821,008 DNA methylation sites were available for analysis. All downstream DNA methylation analyses used beta values, or the methylation rate for a given site.

4.3.5 Outcome assessment and operationalization

Placental cell composition was the primary outcome of interest. Cell composition proportions were estimated via deconvolution using the deconvolution reference developed in **Chapter 3.4.4** via the robust partial correlation algorithm implemented in

EpiDISH R package, v2.8.0 [83] in R, v3.5.3. In EARLI, 1096 of 1101 deconvolution reference DNA methylation sites were available for analysis. In MARBLES, 1092 of 1101 deconvolution reference DNA methylation sites were available for analysis. EARLI and MARBLES datasets were deconvoluted separately on this basis. Cell proportions were estimated for cytotrophoblasts, endothelial cells, Hofbauer cells, stromal cells, syncytiotrophoblasts, nucleated red blood cells, B cells, $CD4^+$ T cells, $CD8^+$ T cells, monocytes, natural killer cells, and neutrophils. We visualized the distribution of estimated placental cell composition by cohort using a scatter plot in Figure 4.4.

4.3.6 Potential confounders and covariates

Both EARLI and MARBLES collected extensive longitudinal behavioral, medical history, health, and sociodemographic information through maternal self-report questionnaires. We considered fetal sex, gestational age, maternal race/ethnicity, maternal age, and maternal education to be potential confounders of the relationship between prenatal metals concentrations and term placental cell composition. Gestational age was operationalized continuously in weeks, maternal race/ethnicity was dichotomized as white vs. non-white, maternal age was operationalized continuously in years, and maternal education was dichotomized as less than a bachelor's degree vs. a bachelor's degree or higher. We described the sample and time point specific distributions of continuous covariates using mean and standard deviation and the distributions of categorical variables using count and frequency in **Table 4.1** via gtsummary R package v1.7.0 [237].

4.3.7 Analytic sample inclusion and exclusion criteria

From participants with available placental DNA methylation information, we excluded individuals from non-singleton births, subsequent siblings born to the same mother, participants missing metals concentrations, and participants missing any covariate information. We visualized sample inclusion using a flow chart in **Figure 4.1**.

4.3.8 Statistical analysis

All analyses were conducted in R v3.5.3 unless otherwise noted. Parametric and nonparametric statistical tests were used as appropriate to test differences between cohort sample characteristics. Chi-square tests were used to test differences between categorical variables. We used Spearman correlation matrices with the corrplot R package v0.92 [238] to test exposure-exposure and exposure-outcome bivariate associations within cohort samples at each time point. To test the association between individual cell type proportions and individual urinary metal concentrations at each time point, cell type-specific beta regression models (n=12 cell types) adjusted for maternal age (continuous), maternal education (binary), maternal race/ethnicity (binary), maternal urinary specific gravity (continuous), fetal gestational age (continuous), and fetal sex (binary) were used. The prevalence (a proportion, here) odds ratio (POR) was calculated by exponentiating beta regression model coefficients in the betareg R package v3.1-4 [239] as the primary effect measure of association. For continuous exposure measures (μg/L), concentrations were $\log_2$ transformed such that a 1-unit increase in the exposure corresponds to a doubling in metal concentration. For dichotomized metals measures, each observation was dichotomized as exposed if above limit of detection or unexposed if below the limit of detection. All statistical tests were interpreted, and all confidence intervals calculated, at an alpha level of 0.05 unless

otherwise noted. We visualized the results using forest plots separated by metal, cohort, and timepoint (**Figure 4.7-Figure 4.11**). Finally, due to the similar in cohort characteristic, we applied fixed effect inverse variance weighted meta-analysis across cohorts within gestation timepoint and assessed heterogeneity with Cochran's Q test and I^2 via the metagen R package v6.2-1 [240] in R v4.2.2 [241]. Meta-analyzed p-values were corrected for multiple comparisons with Benjamini-Hochberg false discovery rate (FDR) adjustment [242]; top hits were considered for FDR < 0.33. If top hits exhibited evidence of heterogeneity via a Cochran's Q test p-value < 0.05 and I^2 > 50%, we reran the metanalytic model as a random effects model and reported the result.

4.4 Results

4.4.1 Analytic sample description

In EARLI, of 133 participants with placental DNA methylation measures, we excluded 5 with non-singleton births, 28 that had missing metals exposure information, and 6 missing covariate information, leaving 94 observations for analysis. All 94 had information exposure information captured during the 1st trimester or early 2nd trimester (early gestation) and 79 had exposure information captured during the 3rd trimester (late gestation). In MARBLES, of 92 participants with placental DNA methylation measures, we excluded 2 observations that represented subsequent siblings from the same family, 7 participants with missing metals exposures, and no participants were missing covariate information, leaving 83 participants for analysis. Of these, 59 had exposure information captured in the 2nd trimester (early gestation), and 79 had exposure information captured during the 3rd trimester (late gestation) (Error! Reference source n

ot found.). EARLI and MARBLES covariate characteristics were comparable across fetal sex, gestational age, maternal race/ethnicity, maternal age, and maternal education. Specific gravity metrics were lower in the MARBLES sample compared to the EARLI sample (**Table 4.1**).

4.4.2 Exposure distribution and operationalization

Of the 22 multiplexed urinary metals measured, we excluded 4 (Be, Pt, U, V) that had less than 10% of samples measured above the limit of detection. Of the remaining 18 metals, we dichotomized 5 (Sb, Cd, Cr, Pb, W) that had more than 30% of samples measured below the limit of detection. Exposed status was defined as above limit of detection and unexposed was defined as below limit of detection. These criteria led to consistent cut points and exclusion results across the EARLI and MARBLES metals samples **(Figure 4.2)**. The final sample included 18 metals for analysis for both cohort samples **(Figure 4.3**). In EARLI, chromium and mercury were detected at higher concentrations in early compared to late gestation samples whereas the opposite was observed with cobalt and thallium (**Table 4.2**). In MARBLES, barium, cobalt, and copper were detected at higher concentrations in late compared to early gestation samples (**Table 4.3**).

4.4.3 Outcome assessment and operationalization

We successfully deconvoluted placental DNA methylation measures for all 94 EARLI and 83 MARBLES samples across 12 cell types. In EARLI, median (Q1, Q3) cell composition in EARLI was 63.2% (59.0%, 68.0%) syncytiotrophoblast, 17.1% (15.1%, 19.4%) stromal, 7.8% (5.9%, 10.7%) cytotrophoblast, 5.6% (4.4%, 7.6%) endothelial,

and 1.2% (0.3%, 2.3%) Hofbauer cells. In MARBLES, estimated cell composition was 60.1% (54.6%, 68.0%) syncytiotrophoblast, 18.6% (13.3%, 22.3%) stromal, 5.6% (3.0%, 9.2%) cytotrophoblast, 5.9% (4.1%, 7.6%) endothelial, and 2.9% (2.2%, 4.4%) Hofbauer cells. (**Figure 4.4**). Deconvolution estimates were consistent with expected placental biology: the majority of tissue cell composition was estimated to be trophoblast and stromal in origin. Median composition estimates for peripheral immune and nucleated red blood cells were less than 1%. B cells, cytotrophoblasts, CD4+ T cells, and monocytes were more abundant in EARLI than MARBLES. Hofbauer cells, Natural killer cells, and neutrophils were more abundant in MARBLES than EARLI (**Table 4.4**). We now focus on placental cell types for the remainder of the manuscript because the non-placental cell types examined were estimated to be of very small proportions, cohort sampling techniques actively avoided their collection, and they are not central to our hypothesis that *in utero* metals concentrations affects placental cell composition at term.

4.4.4 Bivariate associations between metals concentrations and term placental cell type proportions

We examined exposure-exposure and exposure-outcome bivariate associations within cohort samples during early gestation (**Figure 4.5**) and late gestation (**Figure 4.6**) using Spearman correlation matrices. Metals measures were generally positively correlated within cohorts during early and late gestation time points. Of 153 unique pairwise metals concentrations comparisons, 118 were positively and 2 negatively correlated in EARLI during early gestation while 109 were positively and 1 negatively correlated during late gestation. In MARBLES, 95 were positively and 1 negatively

correlated during early gestation while 105 were positively and 0 negatively correlated during late gestation.

Bivariate associations between metals concentrations during early and late gestation and cell composition were observed in both cohorts. In EARLI, early gestation levels of manganese, molybdenum, and selenium and late gestation levels of cadmium, cesium, and chromium were inversely associated with syncytiotrophoblast proportion abundance. Higher early gestation mercury levels were associated with lower Hofbauer cell proportion abundance while late gestation cesium, nickel, and thallium levels were associated with higher Hofbauer cell abundance. Higher early gestation molybdenum, late gestation cesium, cobalt, copper, nickel, and selenium levels were associated with higher endothelial proportion abundance. In MARBLES, higher early gestation lead and tin and higher late gestation barium concentrations were associated with lower syncytiotrophoblast proportions abundance. Higher early gestation barium, cesium, copper, lead, molybdenum, nickel, selenium, and tin concentrations were associated with higher Hofbauer cell proportion abundance. Higher early gestation tin was also associated with higher stromal and endothelial proportion abundances. Higher late gestation barium, cobalt, lead, nickel, and zinc concentrations were associated with higher Hofbauer cell proportion abundance. Higher late gestation molybdenum concentration was associated with higher cytotrophoblast proportion abundance. Higher late gestation lead concentration was associated with higher stromal proportion abundance. Higher late gestation tin concentration was associated with higher endothelial proportion abundance. The less abundant non-placental cell types examined

had mixed associations with metals concentrations in both cohorts during early and late gestation.

4.4.5 Association testing between early gestation metals concentrations and term placental cell type proportions

To identify adjusted associations between early or late gestation metals concentrations and term placental cell type proportions, we used beta regression to model cell type proportions individually adjusting for the potential confounders maternal age, maternal education, maternal race/ethnicity, fetal gestational age, fetal sex, and urinary specific gravity to account for urine concentration. The primary effect estimate was the prevalence odds ratio (POR) associated with a doubling in metal concentration exposure or exposed vs. unexposed if the metal concentration was dichotomized at the limit of detection.

In EARLI, of toxic metals, a doubling in early gestation barium concentration (POR = 1.13, 95% CI [1.03, 1.25]) and thallium level (POR = 1.18, 95% CI [1.01, 1.38] were associated with increased cytotrophoblast abundance (**Figure 4.7**). A doubling in barium concentration was associated with decreased Hofbauer cell abundance (POR = 0.90, 95% CI [0.81, 0.99]. A doubling in mercury concentration was associated with decreased abundance of stromal cells (POR = 0.93, 95% CI [0.88, 0.99]). Of essential metals, copper was inversely associated with endothelial cell abundance (POR = 0.83, 95% CI [0.72, 0.95). Manganese was inversely associated with syncytiotrophoblast abundance (POR = 0.93, 95% CI [0.88, 0.99]) and positively associated with stromal abundance (POR = 1.07, 95% CI [1.01, 1.14]). Selenium was positively associated with cytotrophoblast abundance (POR = 1.19, 95% CI [1.03, 1.38]).

In MARBLES, a doubling in tin concentration during early gestation was positively associated with endothelial cell abundance (POR = 1.12, 95% CI [1.03, 1.21]) (**Figure 4.8**). Also in MARBLES, a doubling in zinc concentration during early gestation was associated with a decrease in cytotrophoblast abundance (POR = 0.82, 95% CI [0.70, 0.97]) and positively associated with Hofbauer cell abundance (POR = 1.13, 95% CI [1.02, 1.25]).

Finally, we meta-analyzed early gestation cell composition results across both cohorts for each metal (**Figure 4.9**). Zinc was the only top hit in the early gestation metals meta-analysis. Early gestation zinc median concentration (Q1, Q3) across both cohorts was 15.9 µg/L (135.7 µg/L, 490.4 µg/L). At the mean cytotrophoblast cell type proportion across both cohorts of 7.8%, a doubling in early gestation zinc concentration was associated with a 0.8% decrement to 7.0% (POR = 0.89, 95% CI [0.82, 0.97], q=0.26). There was little evidence of heterogeneity in zinc effect estimates between the two cohorts (p = 0.20, I^2 = 38%).

4.4.6 Association testing between late gestation metals concentrations and term placental cell type proportions

In EARLI, of nonessential metals, a doubling in cesium (POR = 1.19, 95% CI [1.00, 1.40]) and nickel (POR = 1.17, 95% CI [1.00, 1.36]) concentration during late gestation was associated with an increased abundance of Hofbauer cells (**Figure 4.10**). Exposure to lead concentration above the limit of detection was associated with an increased abundance of endothelial cells (POR = 1.22, 95% CI [1.00, 1.49]). Higher cadmium concentration was associated with a decreased abundance of syncytiotrophoblast (POR = 0.87, 95% CI [0.76, 0.99]). Higher tungsten concentration

was associated with a decreased abundance of cytotrophoblasts (POR = 0.51, 95% CI [0.28, 0.91]). Of essential metals, a doubling in selenium concentration was associated with decreased stromal (POR = 0.91, 95% CI [0.83, 0.99]) and increased endothelial cell abundance (POR = 1.12, 95% CI [1.00, 1.25]).

In MARBLES, of nonessential metals, a doubling in barium concentration during late gestation was inversely associated with syncytiotrophoblast abundance (POR = 0.94, 95% CI [0.88, 0.98) and associated with increased Hofbauer cell (POR = 1.13, 95% CI [1.05, 1.22]) and stromal cell abundances (POR = 1.10, 95% CI [1.01, 1.21]) (**Figure 4.11**). Higher nickel concentration was associated with higher Hofbauer cell abundance (POR = 1.16, 95% CI [1.02, 1.31]). Tin concentration was associated with higher endothelial cell abundance (POR = 1.08, 95% CI [1.02, 1.14]). Exposure to cadmium concentration above the limit of detection was associated with increased Hofbauer cell abundance (POR = 1.28, 95% CI [1.01, 1.63]). Of essential metals, a doubling in cobalt (POR = 1.12, 95% CI [1.00, 1.26] or zinc (POR = 1.10, 95% CI [1.00, 1.20] concentration was associated with an increase in Hofbauer cell abundance. Copper was associated with increased endothelial cell abundance (POR = 1.11, 95% CI [1.02, 1.21]).

Finally, we meta-analyzed late gestation cell composition results across both cohorts for each metal (**Figure 4.12**). We observed 5 top hits in the late gestation metals meta-analysis. Late gestation barium median concentration (Q1, Q3) across both cohorts was 2.4 µg/L (1.2 µg/L, 4.2 µg/L). At the mean Hofbauer cell type proportion across both cohorts of 2.3%, a doubling in late gestation barium concentration was associated with a 0.2% increment to 2.5% (POR = 1.08, 95% CI

[1.02, 0.14], q=0.25). Similarly, a doubling in late gestation nickel concentration was associated with a 0.4% increment in expected Hofbauer cell proportion to 2.7% (POR = 1.16, 95% CI [1.05, 1.27], q=0.25). At the mean syncytiotrophoblast cell type proportion across both cohorts of 62.5%, exposure to cadmium concentrations above the limit of detection was associated with a 3.0% decrement to 59.5% (POR = 0.88, 95% CI [0.78, 0.98], q=0.31). At the mean endothelial cell type proportion across both cohorts of 6.1%, a doubling in late gestation copper concentration was associated with a 0.7% increment to 6.8% (POR = 1.11, 95% CI [1.04, 1.18], q=0.24). At the mean cytotrophoblast cell type proportion across both cohorts of 7.8%, exposure to tungsten concentrations above the limit of detection was associated with a 2.5% decrement to 5.3% (POR =0.67, 95% CI [0.40, 0.94], q=0.31). We did observe evidence of heterogeneity in tungsten/cytotrophoblast effect estimates between the two cohorts (p = 0.05, I^2 = 74%) with the majority of the weighting coming from the EARLI cohort (74.7%). In a random effects meta-analysis model, exposure to tungsten concentrations above the limit of detection was not associated with changes in cytotrophoblast cell type proportions (POR =0.78, 95% CI [0.18, 1.39]).

4.5 Discussion

In summary, this study investigated the relationships between *in utero* urinary metals concentrations during early or late gestation to 18 nonessential or essential metals and estimated term placental cell composition from DNA methylation microarrays in two prospective birth cohorts with meta-analysis. At the mean cytotrophoblast cell type proportion across both cohorts of 7.8%, a doubling in early gestation zinc concentration was associated with a 0.8% decrement to 7.0% (POR =

0.89, 95% CI [0.82, 0.97], q=0.26). At the mean Hofbauer cell type proportion across both cohorts of 2.3%, a doubling in late gestation barium concentration was associated with a 0.2% increment to 2.5% (POR = 1.08, 95% CI [1.02, 0.14], q=0.25). Similarly, a doubling in late gestation nickel concentration was associated with a 0.4% increment in expected Hofbauer cell proportion to 2.7% (POR = 1.16, 95% CI [1.05, 1.27], q=0.25). At the mean syncytiotrophoblast cell type proportion across both cohorts of 62.5%, exposure to cadmium concentrations above the limit of detection was associated with a 3.0% decrement to 59.5% (POR = 0.88, 95% CI [0.78, 0.98], q=0.31). At the mean endothelial cell type proportion across both cohorts of 6.1%, a doubling in late gestation copper concentration was associated with a 0.7% increment to 6.8% (POR = 1.11, 95% CI [1.04, 1.18], q=0.24). We provide evidence that prenatal metals levels may affect term placental cell composition which has implications for placental function and disease related to placental dysfunction as well as for molecular studies of the placenta that fail to account for potential cell composition differences.

Our findings are consistent with previous direct toxicological evidence that prenatal exposure to cadmium affects placental morphology in animal models and indirect evidence through reduced neonatal development in humans indicative of placental dysfunction [226]. Consistent with our observation of decreased syncytiotrophoblast abundance with late gestation cadmium concentrations, a 2013 study of 14 pregnant women revealed increased syncytial knotting and fibrinoid deposits in women with placental cadmium level ≥ 15 mg/kg, and placental cadmium levels were positively correlated with urinary levels [227]. Zinc was associated with placental cell composition in early gestation. In pregnant mice, a 26% reduction in zinc levels altered

placental morphogenesis leading to an 8% reduction in fetal and placental weights, likely by reducing arterial blood pressure [228]. Further research, particularly in human populations, is required to fully understand the apparent links between prenatal metals concentrations and placental cell composition.

The heterogeneity in effect estimates at different gestational exposure timings may be due to differing windows of cell-specific biological susceptibility to *in utero* metals concentrations. Differences in cytotrophoblast abundance were only associated with early gestation zinc concentrations (POR = 0.89, 95% CI [0.82, 0.97], q=0.26). This may reflect the lasting effects of metal levels during early placental development, which is characterized primarily by trophoblast proliferation and differentiation [243]. Hofbauer, endothelial, and syncytiotrophoblast proportions were uniquely associated with late gestation metal levels. Hofbauer cells are thought to play a key role in placenta morphogenesis and homeostasis, particularly in vasculogenesis [12, 244]. A recent review of *in vitro* and animal models posited vasculogenesis as a target of teratogenic metal toxicity [245]. Our results particularly implicate the divalent cations barium, cadmium, nickel, copper, and zinc, which can be transported by divalent metal transporter-1 (DMT-1) [246]. Indeed, DMT-1 has been detected immunohistochemically in primary human placental syncytiotrophoblast, Hofbauer cells, and endothelial cells, representing potential direct exposure of these cell types to divalent metals in late gestation, all of which were associated with late gestation divalent metal concentrations in this analysis [247]. The same study demonstrated that DMT-1 expression and abundance was increased in cadmium-exposed pregnant women, highlighting metal response elements found in genes regulating essential and nonessential metals

concentrations such as the metallothioneins *ZnT-1* and *MTF-1* [247]. Maternal barium exposure has been recently linked to congenital heart defects, orofacial clefts, and neural tube defects [248–250]. Altered placental cell composition and morphology may be a consequence of or response to metals insults and may increase subsequent disease risk by altering placental structure and possibly function.

There are several strengths to this investigation. This study provides a pioneering example of applying a scalable reference-based deconvolution to large, well-characterized prospective birth cohort studies, leveraging existing molecular data to investigate novel questions about placental morphology. We ruled out excessive comingling of nonplacental immune cells in the EARLI and MARBLES DNA methylation samples by using an updated deconvolution reference panel developed in **Chapter 3**. Exposure assessment was a strength of the study, capturing specific gravity and multiplexed metals measures across 18 metals in both early and late gestation to allow for the detection of potential heterogeneous effect estimates at different gestational timings across many metals.

There are also limitations to this study. Urinary metals concentrations, in particular of essential metals, are reflective of metabolism in addition to exogeneous exposure and so must be interpreted with caution. Metal co-exposures may confound the association between an individual metal and cell composition outcomes. Alternative metals mixture modelling strategies could be employed to quantify joint mixture exposures with tradeoffs between model performance and interpretability of model results [251]. However, metals measures were positively correlated within time point and cohort, generally reflecting a common exposure mixture to all metals

simultaneously. Our statistical models did not account for potential nonlinear associations, which could be particularly relevant for the essential metals tested [252]. However, the distributions of metals observed in our study did not indicate deficiencies of essential metals. Our results may not be generalizable to the general population because both EARLI and MARBLES employ an autism spectrum disorder enriched-risk design.

This study investigated the relationship between prenatal metals concentrations during early or late gestation and placental cell composition at term, estimated using a novel reference panel to deconvolute bulk DNA methylation data. We observed associations between essential, nonessential, and toxic metals in early and late gestation and term placental cell type proportions that are essential to placental function. Our results demonstrate the utility of deploying scalable deconvolution approaches to epidemiologic samples to understand the effects of environmental exposures on placental morphology. Further research is required to elucidate the etiology of adverse perinatal outcomes downstream of disturbed placental morphology and concomitant dysfunction to develop new interventions and prioritize environmental exposures for public health prevention efforts.

4.6 Figures and Tables

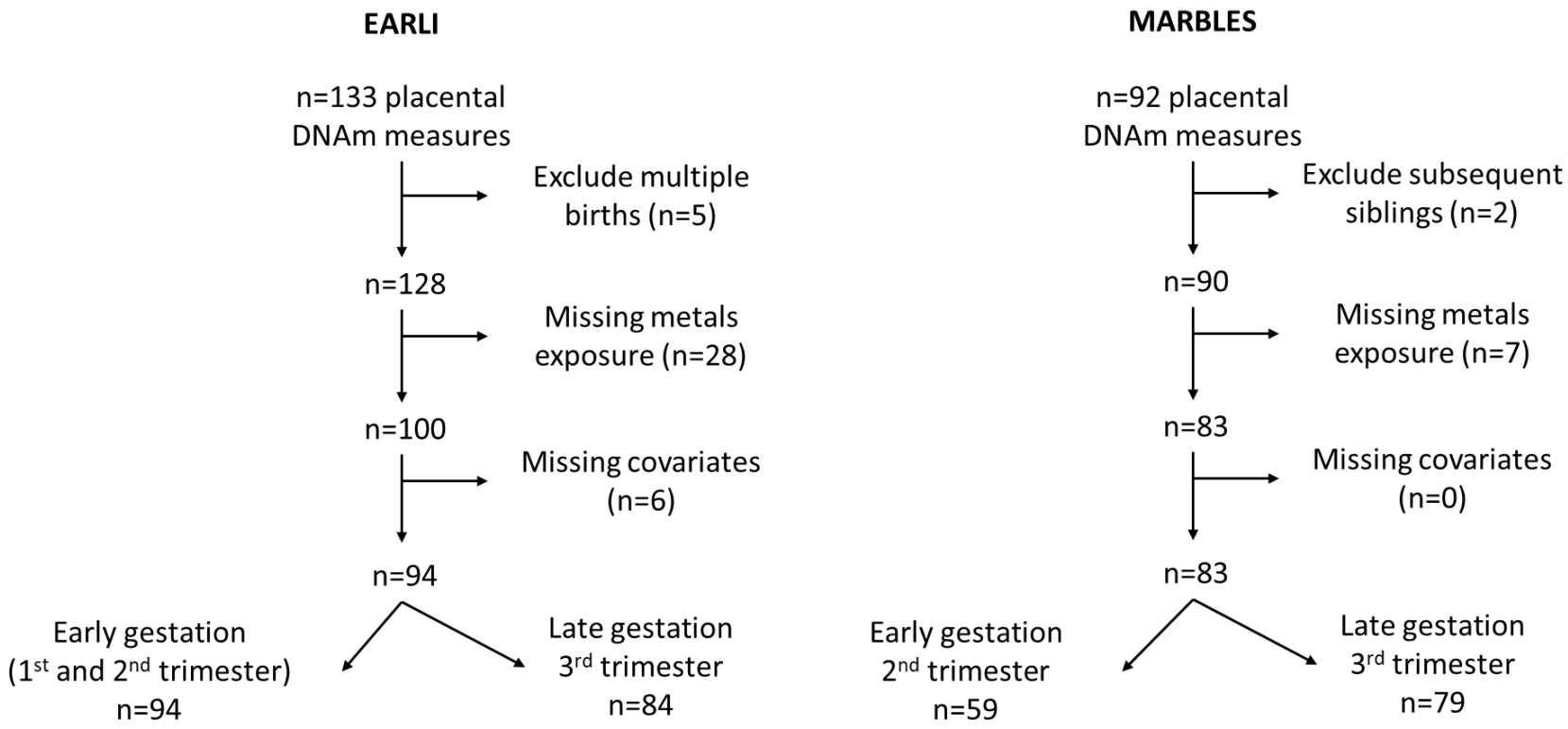


Figure 4.1 Participant inclusion/exclusion flow chart.

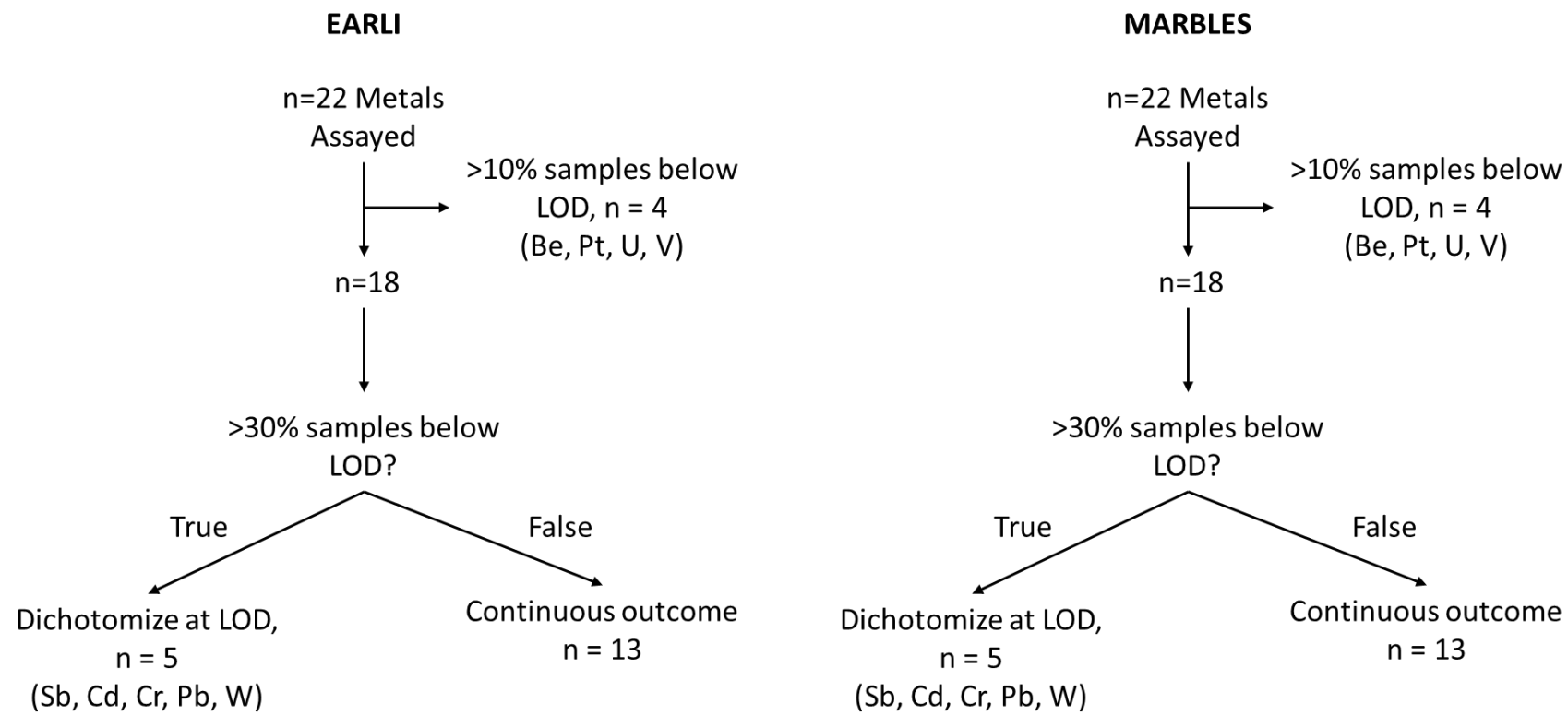


Figure 4.2 Inclusion/exclusion criteria and operationalization flow chart for prenatal urinary metals exposures.

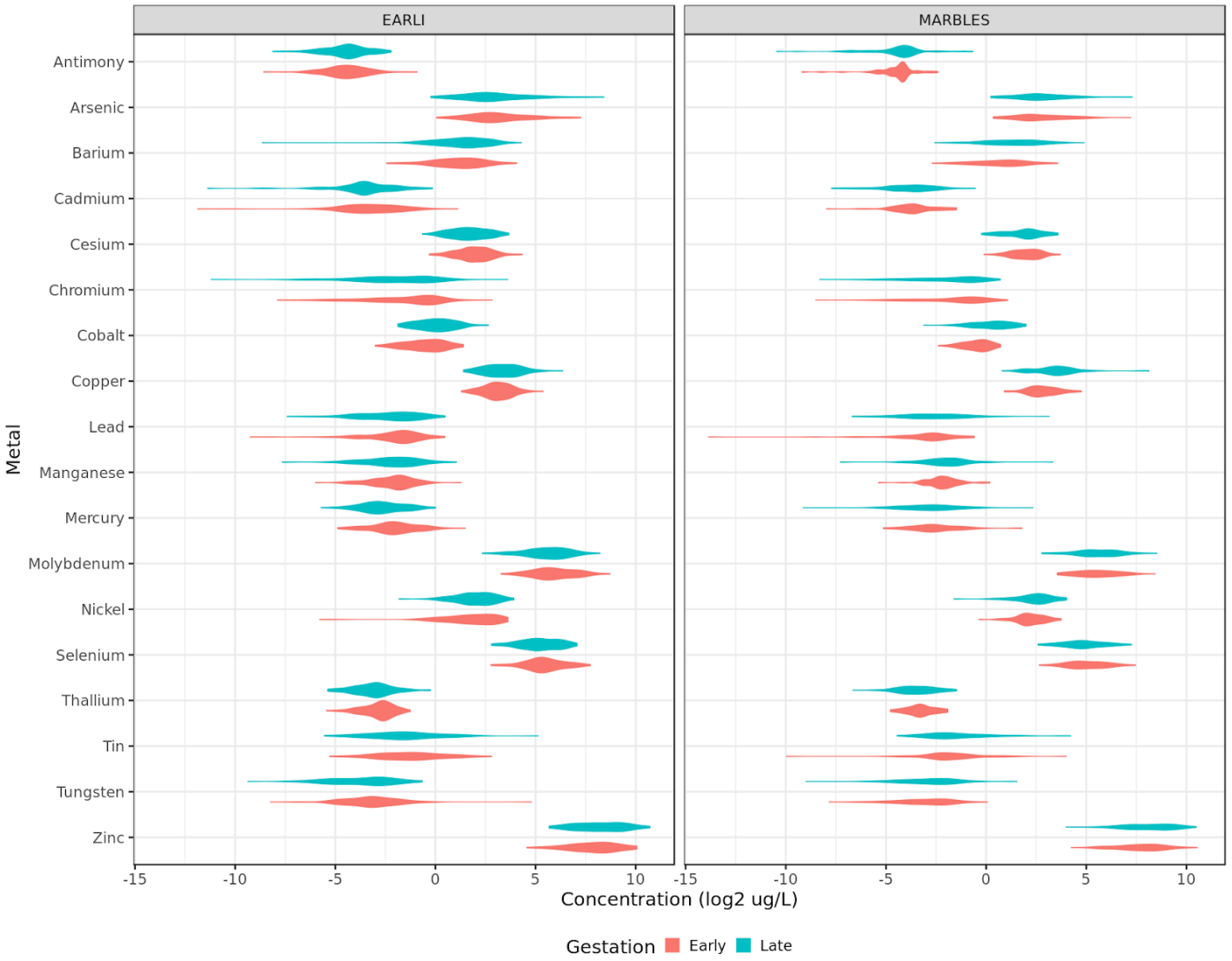


Figure 4.3 Exposure distribution between Early Autism Risk Longitudinal Investigation (EARLI) and Markers of Autism Risk in Babies and Learning Early Signs (MARBLES) cohort samples. Metals concentrations were measured in urine bio-samples via inductively coupled mass spectrometry.

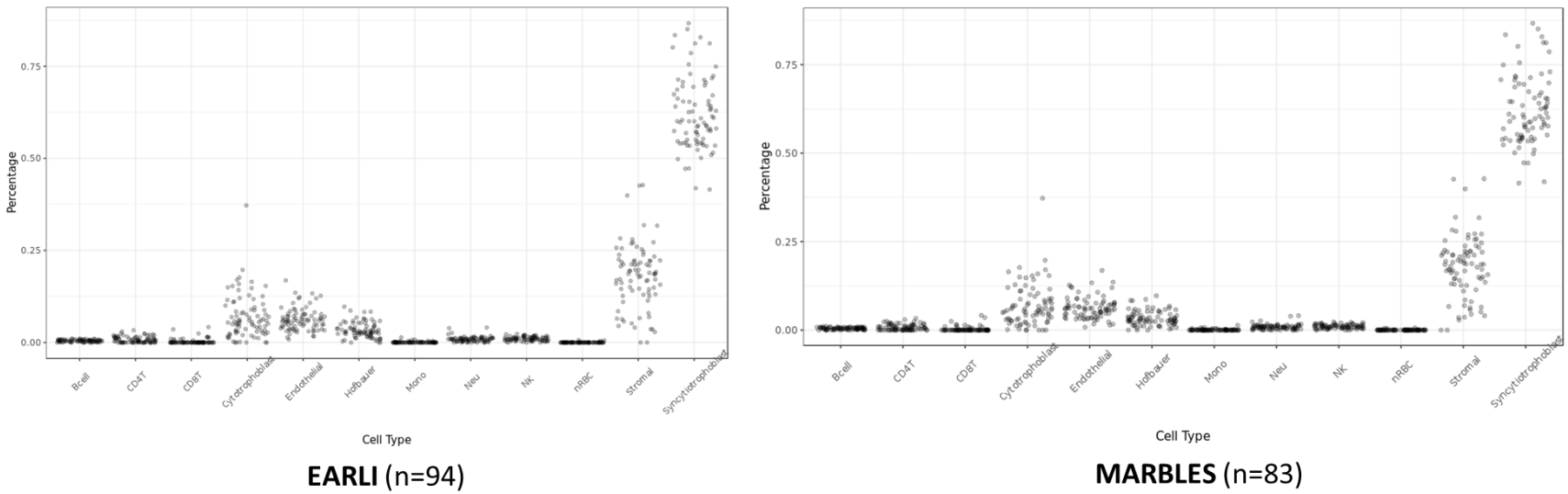


Figure 4.4 Deconvolution cell type proportions estimates for placental DNA methylation samples by cohort sample. Abbreviations: monocytes (Mono), natural killer cells (NK), neutrophils (Neu), and nucleated red blood cells (nRBC).

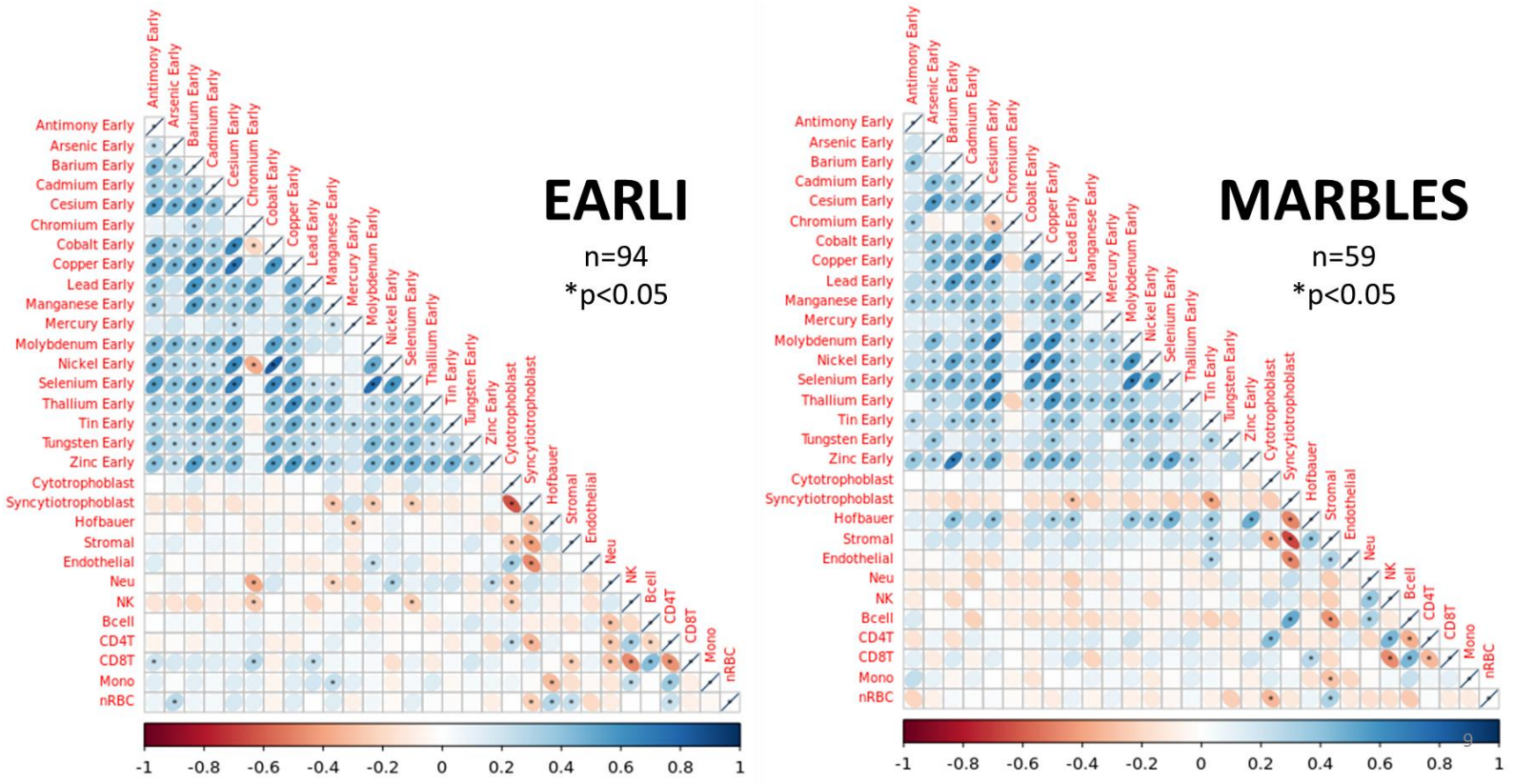


Figure 4.5 Bivariate associations with Spearman correlation plots for early pregnancy metals exposure across EARLI and MARBLES cohorts. Significant correlations with p-value < 0.05 denoted with asterisk. Color bar corresponds to Spearman correlation estimates. Abbreviations: monocytes (Mono), natural killer cells (NK), neutrophils (Neu), and nucleated red blood cells (nRBC).

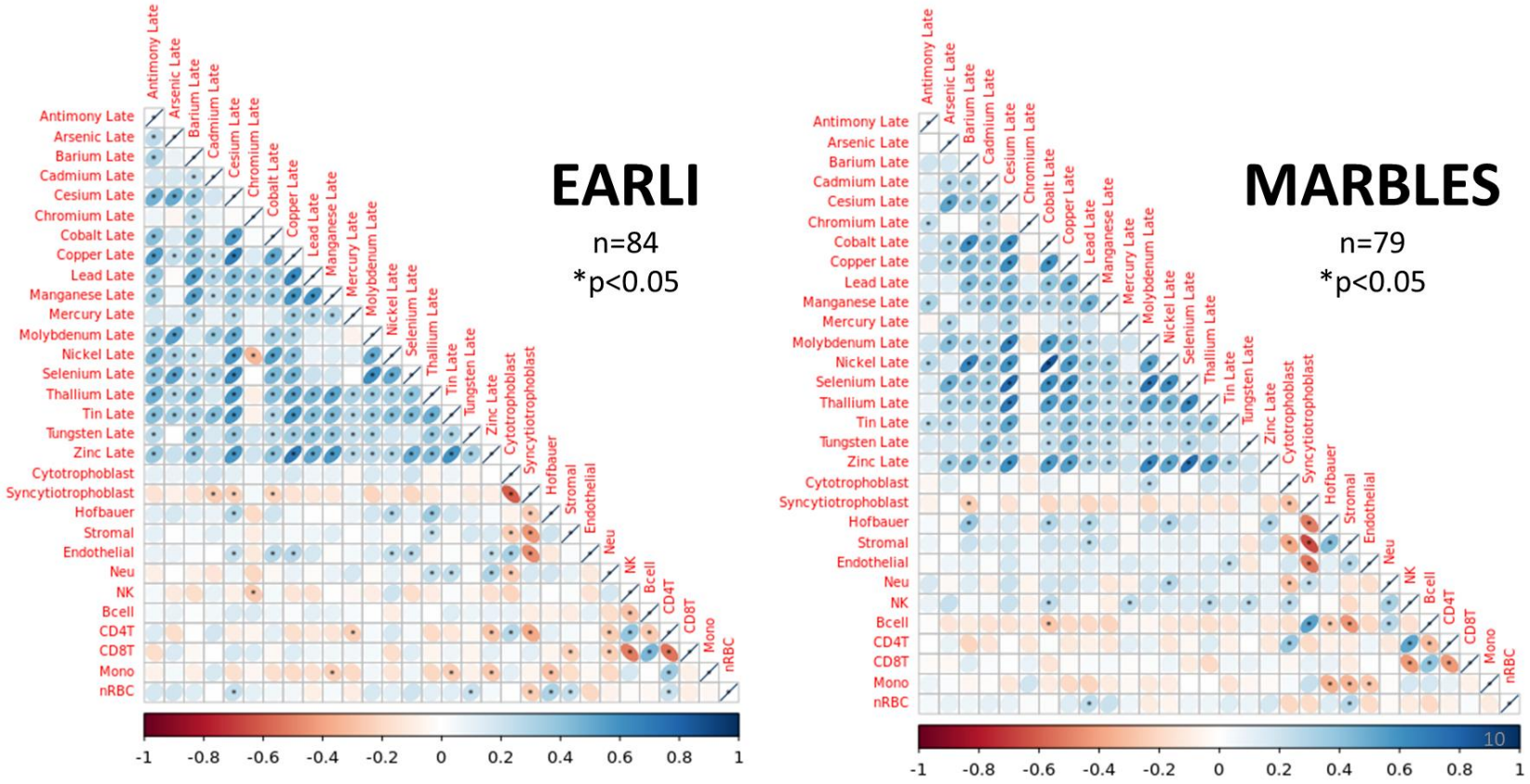


Figure 4.6 Bivariate associations with Spearman correlation plots for late pregnancy metals exposure across EARLI and MARBLES cohorts. Significant correlations with p-value < 0.05 denoted with asterisk. Color bar corresponds to Spearman correlation estimates. Abbreviations: monocytes (Mono), natural killer cells (NK), neutrophils (Neu), and nucleated red blood cells (nRBC).

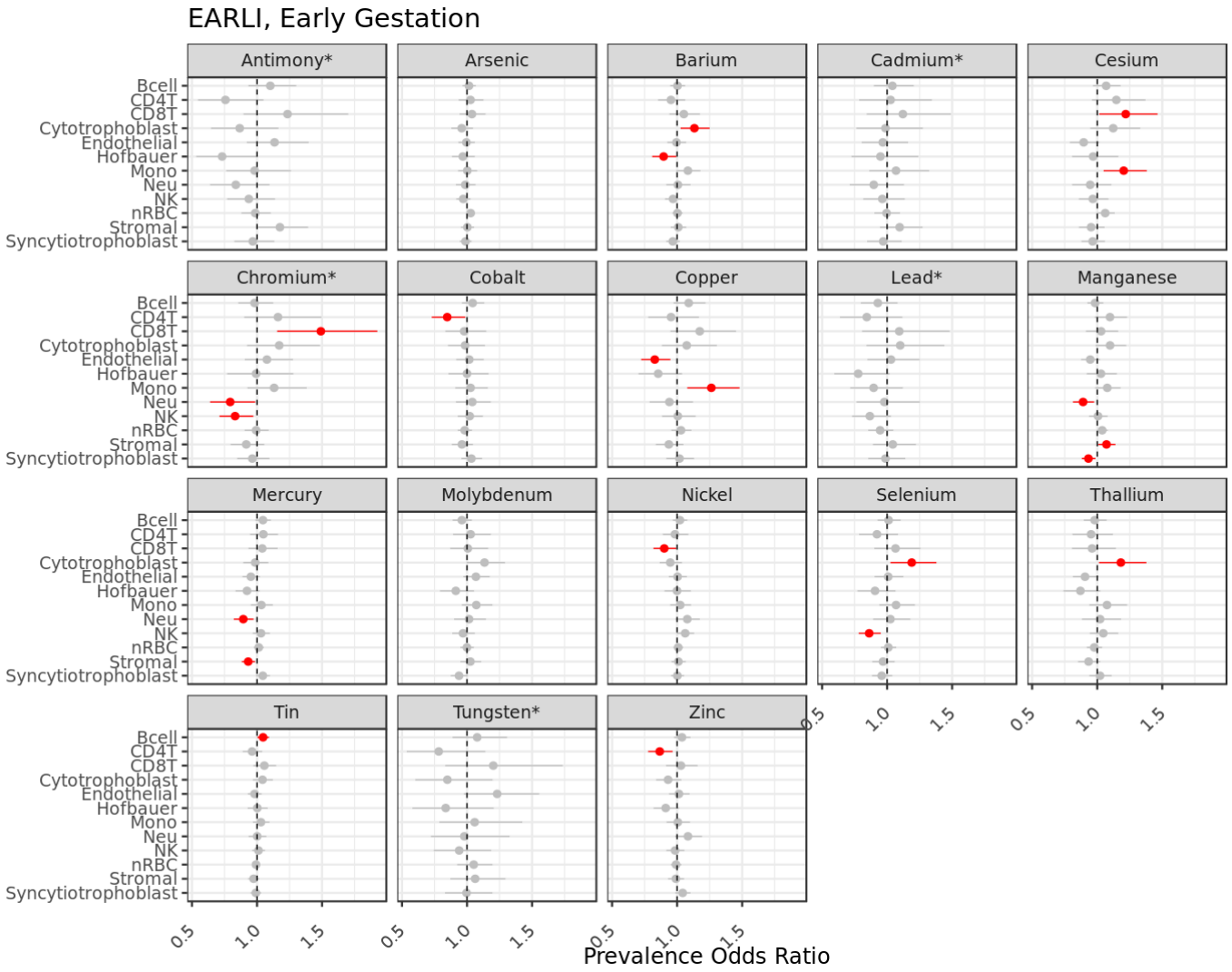


Figure 4.7 Forest plot of prevalence odds ratio estimates of cell composition changes in EARLI for a doubling in early gestation exposure concentration or exposed versus unexposed for the metals marked with an asterisk. Line segments represent 95% confidence intervals. Red color denotes statistical significance. Abbreviations: monocytes (Mono), natural killer cells (NK), neutrophils (Neu), and nucleated red blood cells (nRBC).

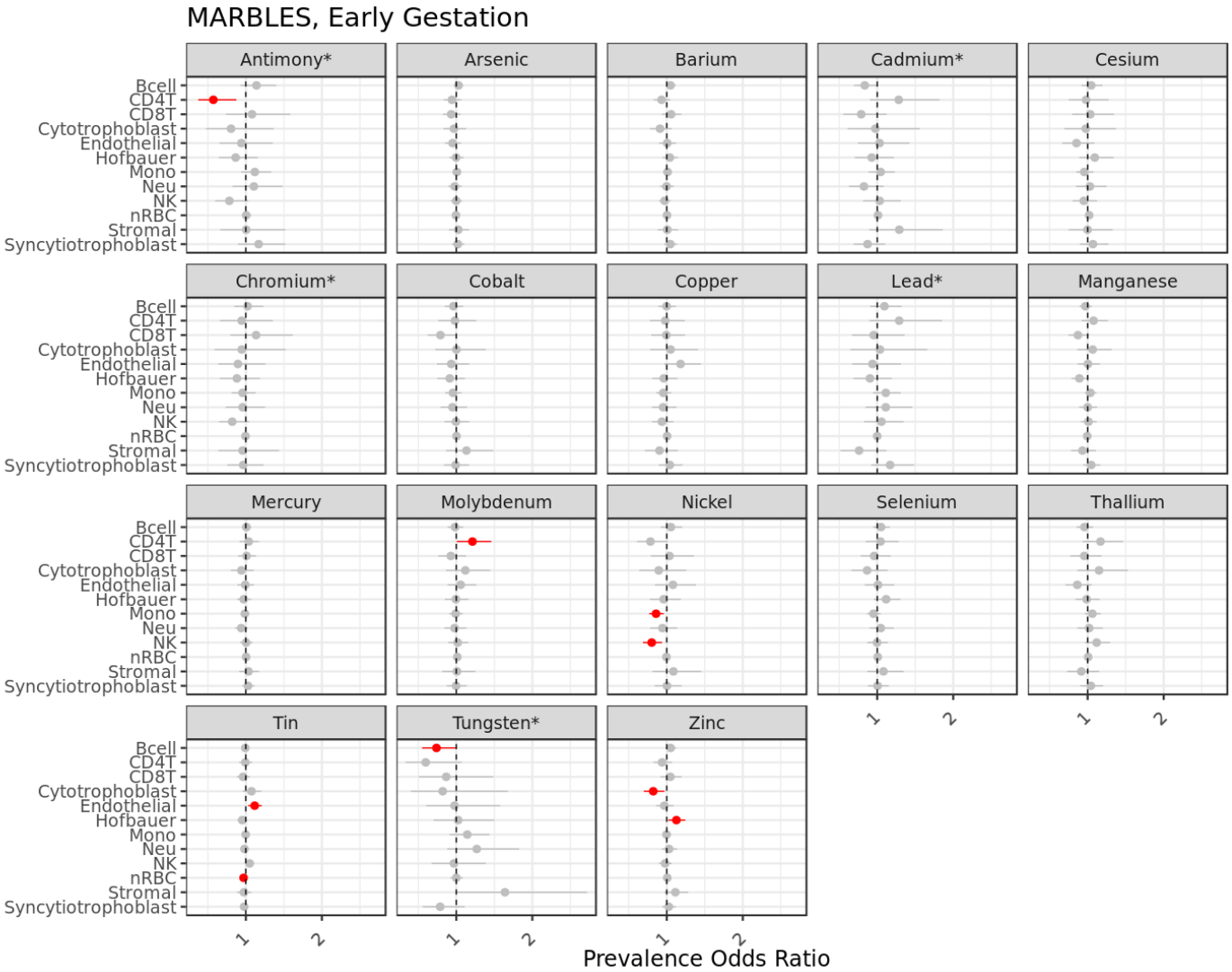


Figure 4.8 Forest plot of prevalence odds ratio estimates of cell composition changes in MARBLES for a doubling in early gestation exposure concentration or exposed versus unexposed for the metals marked with an asterisk. Line segments represent 95% confidence intervals. Red color denotes statistical significance. Abbreviations: monocytes (Mono), natural killer cells (NK), neutrophils (Neu), and nucleated red blood cells (nRBC).

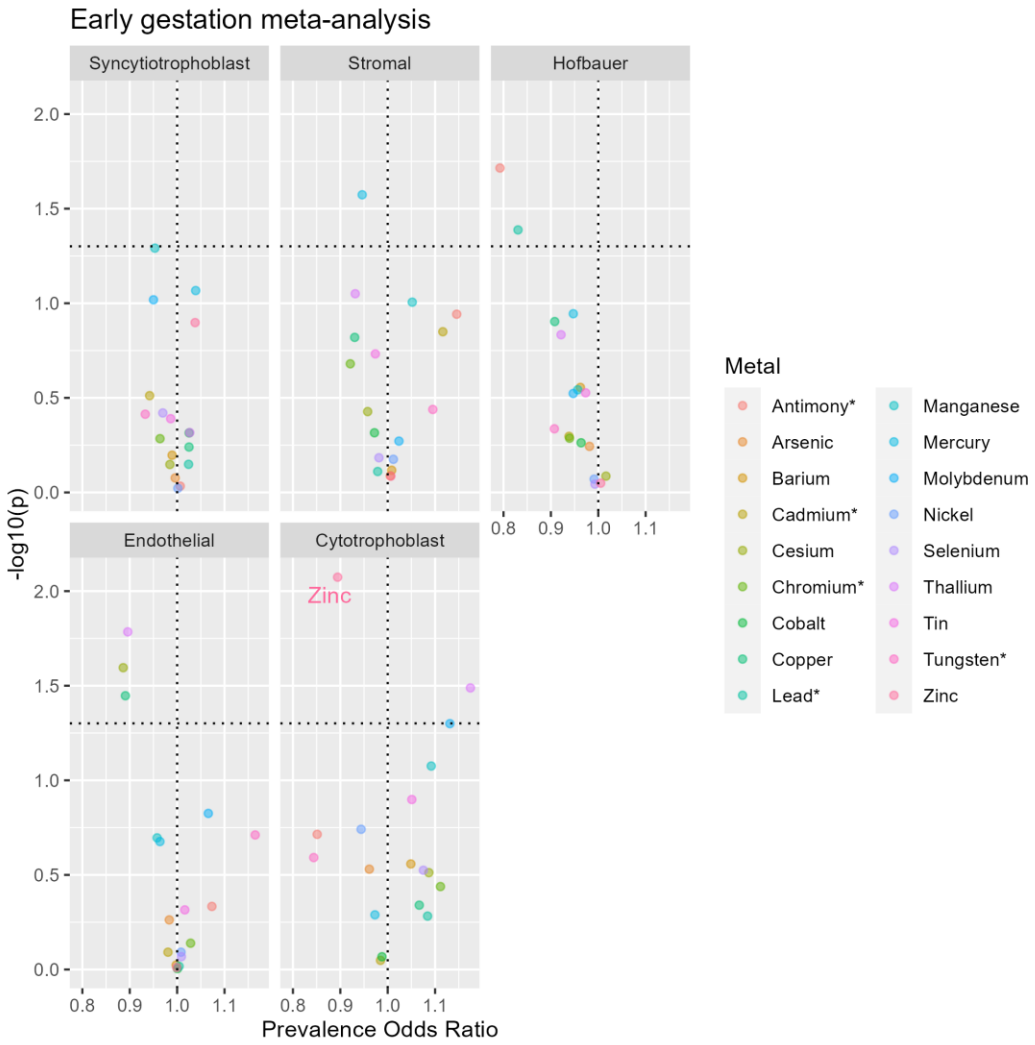


Figure 4.9 Volcano plot of inverse variance weighted meta-analyze prevalence odds ratio estimates of cell composition changes for a doubling in early gestation exposure concentration or exposed versus unexposed for the metals marked with an asterisk. Horizontal dashed line indicates nominal p-value of 0.05. Vertical dashed line indicates a null effect estimate of 1. Labelled points indicate a top hit with false discovery rate-adjusted q-value < 0.33.

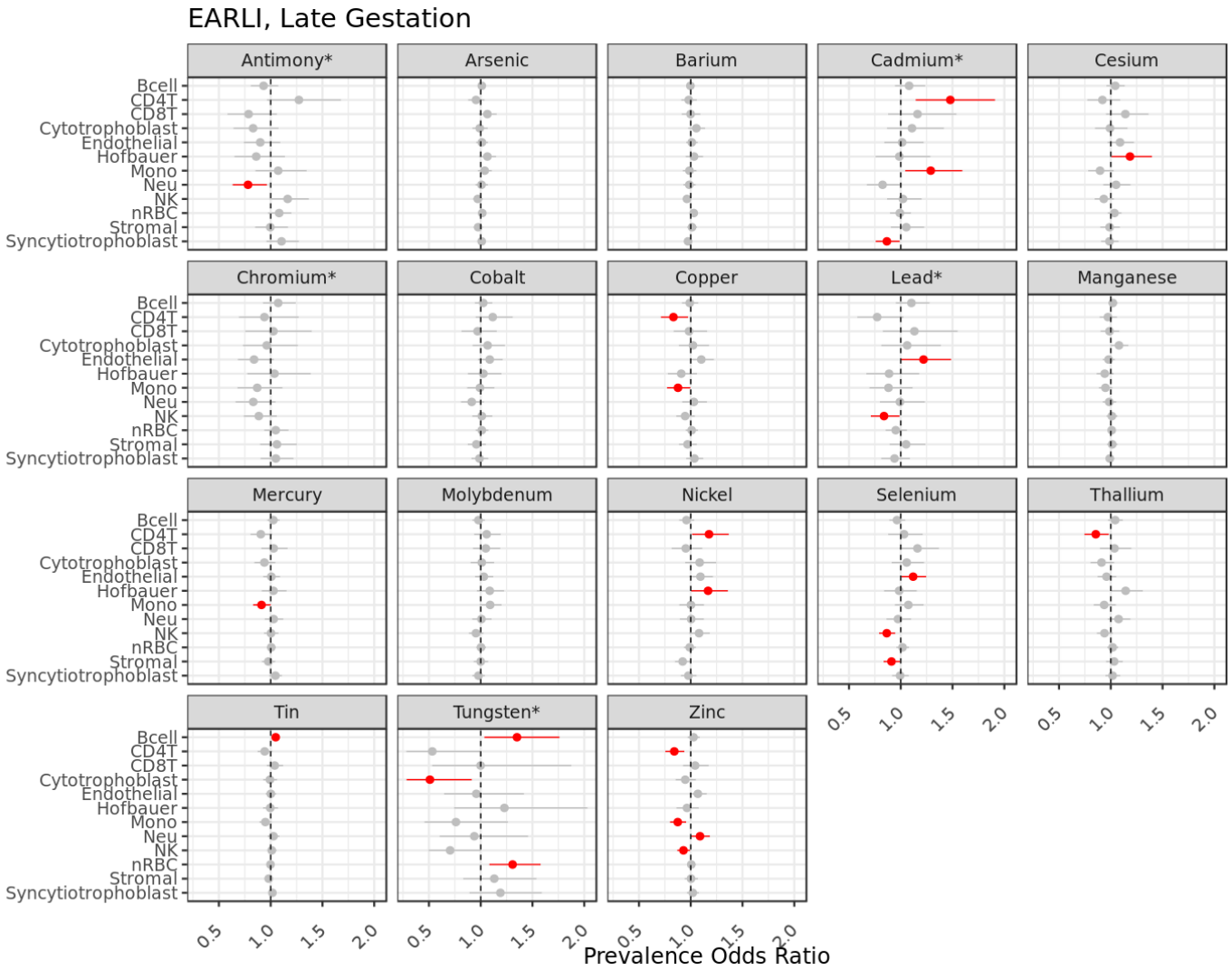


Figure 4.10 Forest plot of prevalence odds ratio estimates of cell composition changes in EARLI for a doubling in late gestation exposure concentration or exposed versus unexposed for the metals marked with an asterisk. Line segments represent 95% confidence intervals. Red color denotes statistical significance. Abbreviations: monocytes (Mono), natural killer cells (NK), neutrophils (Neu), and nucleated red blood cells (nRBC).

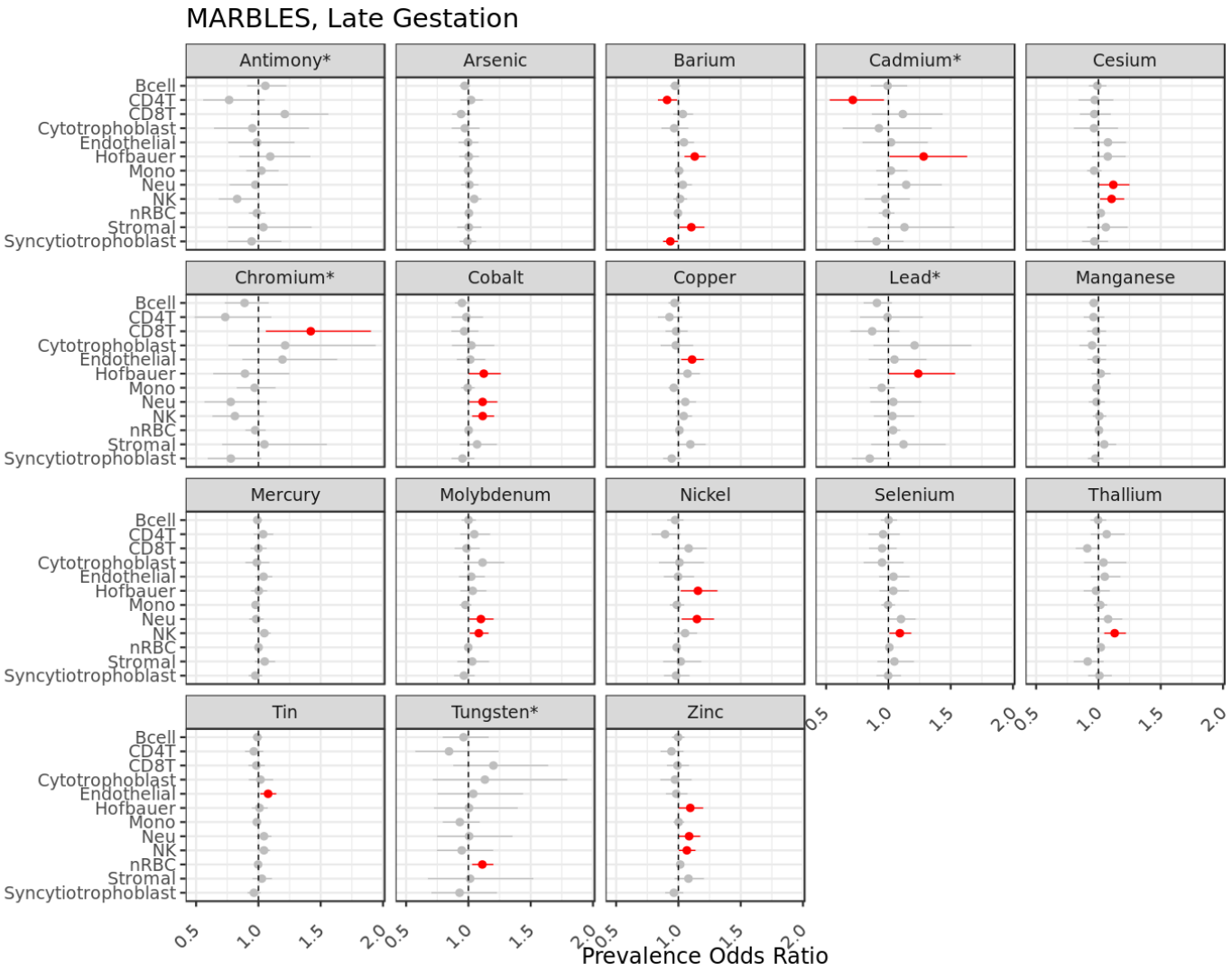


Figure 4.11 Forest plot of prevalence odds ratio estimates of cell composition changes in MARBLES for a doubling in late gestation exposure concentration or exposed versus unexposed for the metals marked with an asterisk. Line segments represent 95% confidence intervals. Red color denotes statistical significance. Abbreviations: monocytes (Mono), natural killer cells (NK), neutrophils (Neu), and nucleated red blood cells (nRBC).

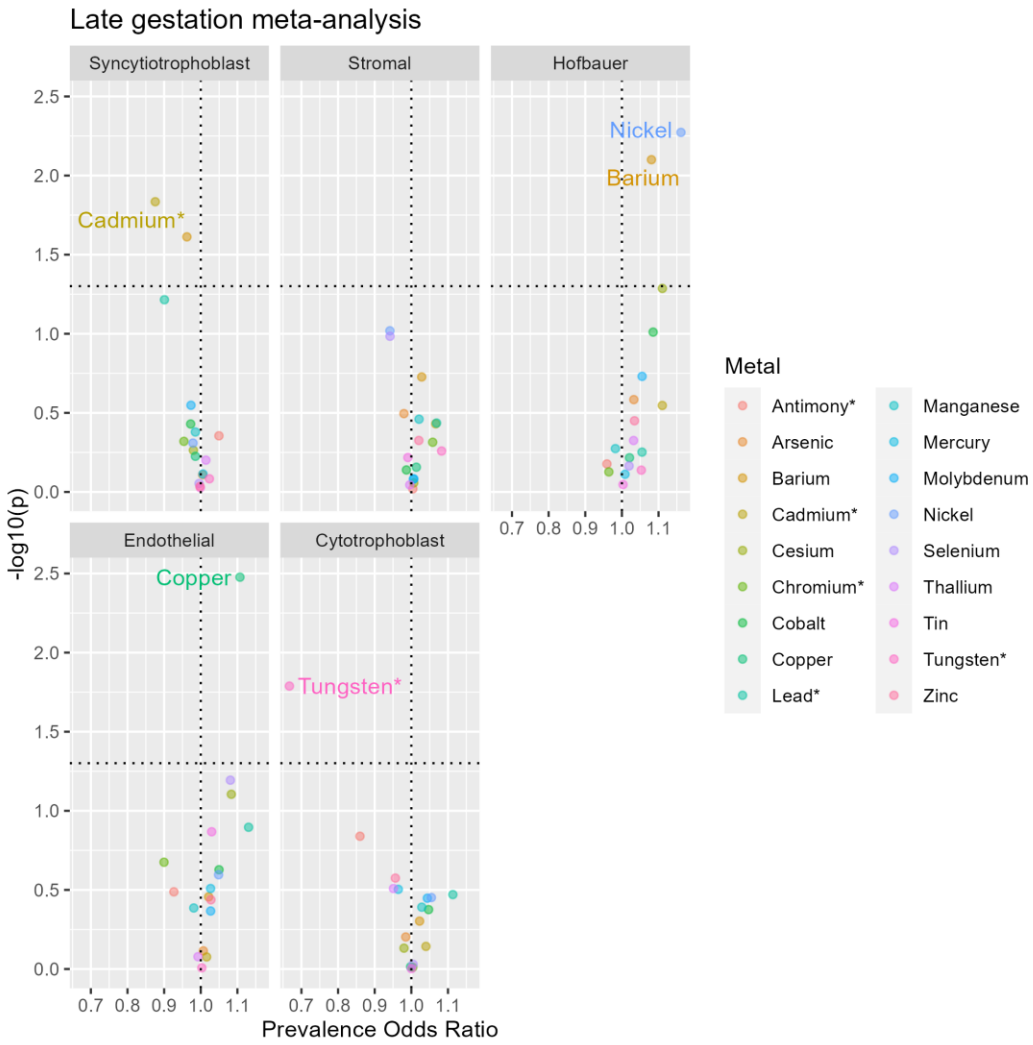


Figure 4.12 Volcano plot of inverse variance weighted meta-analyze prevalence odds ratio estimates of cell composition changes for a doubling in early gestation exposure concentration or exposed versus unexposed for the metals marked with an asterisk. Horizontal dashed line indicates nominal p-value of 0.05. Vertical dashed line indicates a null effect estimate of 1. Labelled points indicate a top hit with false discovery rate-adjusted q-value < 0.33.

Table 4.1 Sample descriptive statistics for EARLI and MARBLES samples. Statistical testing compares the distribution of demographic variables across cohorts.

Characteristic	**EARLI**, N = 94[1]	**MARBLES**, N = 83[1]	**p-value**[2]
Child Sex			0.3
Female	40 (43%)	29 (35%)	
Male	54 (57%)	54 (65%)	
Gestational Age (wks)	39.40 (39.00, 40.00)	39.14 (38.86, 39.86)	0.3
Maternal Age	34.0 (31.0, 37.0)	34.8 (30.8, 39.0)	0.5
Maternal Race/ethnicity			0.5
White	51 (54%)	41 (49%)	
Non-white	43 (46%)	42 (51%)	
Maternal Education			0.3
Less than Bachelor's Degree	38 (40%)	40 (48%)	
Bachelor's or higher	56 (60%)	43 (52%)	

Characteristic	EARLI, N = 94[1]	MARBLES, N = 83[1]	p-value[2]
Specific Gravity	1.016 (1.011, 1.022)	1.013 (1.009, 1.018)	0.004
Illumina Array Type			
450k	94 (100%)	0 (0%)	
EPIC	0 (0%)	83 (100%)	

[1] n (%); Median (IQR)

[2] Pearson's Chi-squared test; Wilcoxon rank sum test

Table 4.2 Urinary metals distributions, median (Q1, Q3) for the EARLI cohort sample stratified by Early vs. Late Gestation timepoints.

	EARLI (µg/L)		
Metal	Early, N = 94[1]	Late, N = 84[1]	p-value[2]
Antimony	0.04 (0.03, 0.07)	0.05 (0.03, 0.07)	>0.9
Arsenic	7.36 (4.53, 17.99)	6.25 (3.78, 16.49)	0.3
Barium	2.37 (1.24, 4.09)	2.75 (1.34, 4.61)	0.4
Cadmium	0.10 (0.05, 0.21)	0.08 (0.06, 0.17)	0.3
Cesium	3.79 (2.68, 5.66)	3.19 (2.05, 5.19)	0.066

	EARLI (µg/L)		
Metal	**Early**, N = 94[1]	**Late**, N = 84[1]	**p-value**[2]
Chromium	0.57 (0.15, 0.84)	0.25 (0.11, 0.77)	0.045
Cobalt	0.71 (0.42, 1.20)	1.05 (0.67, 1.64)	<0.001
Copper	8.48 (6.31, 12.18)	9.91 (6.46, 16.04)	0.082
Lead	0.28 (0.14, 0.41)	0.24 (0.08, 0.42)	0.4
Manganese	0.25 (0.14, 0.35)	0.23 (0.11, 0.42)	>0.9
Mercury	0.24 (0.15, 0.40)	0.16 (0.10, 0.28)	<0.001
Molybdenum	57.03 (37.54, 106.35)	49.71 (28.71, 80.05)	0.062
Nickel	3.72 (2.01, 6.90)	4.02 (2.61, 6.14)	0.4
Selenium	40.23 (29.14, 67.54)	37.25 (22.24, 67.28)	0.2
Thallium	0.15 (0.10, 0.20)	0.12 (0.07, 0.17)	0.024
Tin	0.40 (0.17, 0.81)	0.33 (0.14, 0.72)	0.2
Tungsten	0.11 (0.06, 0.20)	0.10 (0.03, 0.16)	0.067
Zinc	252.85 (135.49, 436.60)	290.53 (141.46, 578.01)	0.2

[1] Median (IQR)

	EARLI (µg/L)		
Metal	**Early**, N = 94[1]	**Late**, N = 84[1]	**p-value**[2]

[2] Wilcoxon rank sum test

Table 4.3 Urinary metals distributions, median (Q1, Q3) for the MARBLES cohort sample stratified by Early vs. Late Gestation timepoints.

	MARBLES (µg/L)		
Metal	**Early**, N = 59[1]	**Late**, N = 79[1]	**p-value**[2]
Antimony	0.05 (0.04, 0.06)	0.06 (0.04, 0.07)	0.4
Arsenic	5.67 (4.00, 12.81)	6.01 (3.71, 12.65)	>0.9
Barium	1.77 (0.90, 3.01)	2.73 (1.28, 5.09)	0.009
Cadmium	0.08 (0.05, 0.12)	0.08 (0.04, 0.12)	0.3
Cesium	4.03 (2.69, 5.61)	4.07 (2.21, 5.37)	0.6
Chromium	0.49 (0.12, 0.66)	0.41 (0.11, 0.62)	0.4
Cobalt	0.75 (0.52, 0.97)	1.33 (0.73, 2.03)	<0.001
Copper	6.36 (4.98, 10.32)	11.01 (6.97, 16.01)	<0.001
Lead	0.14 (0.07, 0.20)	0.14 (0.08, 0.33)	0.3

	MARBLES (µg/L)		
Metal	**Early**, N = 59[1]	**Late**, N = 79[1]	**p-value**[2]
Manganese	0.22 (0.18, 0.29)	0.25 (0.16, 0.38)	0.4
Mercury	0.17 (0.11, 0.33)	0.15 (0.07, 0.27)	0.2
Molybdenum	42.66 (26.65, 72.72)	44.80 (29.48, 82.98)	0.7
Nickel	4.41 (3.55, 6.73)	5.79 (3.94, 7.91)	0.062
Selenium	31.74 (20.04, 52.28)	28.96 (18.14, 48.34)	0.4
Thallium	0.10 (0.08, 0.14)	0.09 (0.06, 0.15)	0.2
Tin	0.26 (0.18, 0.49)	0.28 (0.17, 0.62)	0.6
Tungsten	0.14 (0.07, 0.24)	0.16 (0.07, 0.28)	0.4
Zinc	241.65 (125.32, 382.00)	295.66 (145.77, 531.61)	0.060

[1] Median (IQR)
[2] Wilcoxon rank sum test

Table 4.4 Distribution of placental cell composition deconvolution estimates, stratified by cohort. Abbreviations: monocytes (Mono), natural killer cells (NK), neutrophils (Neu), and nucleated red blood cells (nRBC).

Cell proportion (%)	**EARLI**, N = 94[1]	**MARBLES**, N = 83[1]	**p-value**[2]
Bcell	0.66 (0.43, 0.90)	0.43 (0.25, 0.65)	<0.001
CD4T	0.92 (0.30, 1.82)	0.55 (0.00, 1.30)	0.010
CD8T	0.00 (0.00, 0.02)	0.00 (0.00, 0.00)	0.2
Cytotrophoblast	7.77 (5.90, 10.72)	5.55 (2.95, 9.22)	0.003
Endothelial	5.59 (4.35, 7.59)	5.90 (4.13, 7.63)	0.8
Hofbauer	1.16 (0.29, 2.32)	2.94 (2.16, 4.43)	<0.001
Mono	0.00 (0.00, 0.21)	0.00 (0.00, 0.00)	0.009
Neu	0.00 (0.00, 0.27)	0.89 (0.41, 1.18)	<0.001
NK	0.73 (0.41, 1.11)	0.93 (0.60, 1.39)	0.011
nRBC	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	>0.9
Stromal	17.06 (15.12, 19.40)	18.60 (13.29, 22.25)	0.3

Cell proportion (%)	**EARLI**, N = 94[1]	**MARBLES**, N = 83[1]	**p-value**[2]
Syncytiotrophoblast	63.20 (58.96, 68.00)	60.07 (54.57, 68.03)	0.11

[1] Median (IQR)
[2] Wilcoxon rank sum test

Chapter 5 Addressing Cellular Heterogeneity in Molecular Epidemiology and Future Directions

This chapter has been adapted in part from a manuscript published in *Current Environmental Health Reports* (2020) [1].

Authors: Kyle A. Campbell[1], Justin A. Colacino[2], Sung Kyun Park[1,2], and Kelly M. Bakulski[1]
Affiliations:
[1]Department of Epidemiology, University of Michigan School of Public Health, University of Michigan, Ann Arbor, Michigan,
[2]Department of Environmental Health Sciences, University of Michigan School of Public Health, University of Michigan, Ann Arbor, Michigan

5.1 Brief summary of the dissertation research chapters

In **Chapter 2**, I generated a cell type-specific placental deconvolution gene expression reference with novel and previously published single-cell RNA-sequencing data. data to create the largest deconvolution reference of 19 fetal and 8 maternal cell types from placental villous tissue (n=9 biological replicates) at term (n=40,494 cells). I deconvoluted eight published microarray case-control studies of preeclampsia (n=173 controls, 157 cases). Preeclampsia was associated with excess extravillous trophoblasts (POR = 1.94, 95% CI [1.61, 2.34]) and fewer mesenchymal (POR = 0.79, 95% CI [0.73, 0.85]) and Hofbauer cells (POR = 0.67, 95% CI [0.59, 0.77]). Adjustment for cellular composition reduced preeclampsia-associated differentially expressed genes ($\log_2$ fold-change cutoff=0.1, FDR<0.05) from 1,154 to 0, whereas downregulation of mitochondrial biogenesis, aerobic respiration, and ribosome biogenesis were robust to cell type adjustment, suggesting direct changes to these pathways. Cellular composition

mediated a substantial proportion of the association between preeclampsia and *FLT1* (37.8%, 95% CI [27.5%, 48.8%]), *LEP* (34.5%, 95% CI [26.0%, 44.9%]), and *ENG* (34.5%, 95% CI [25.0%, 45.3%]) overexpression. Our findings indicate substantial placental cellular heterogeneity in preeclampsia contributes to previously observed bulk gene expression differences. This novel deconvolution reference lays the groundwork for cellular heterogeneity-aware investigation into placental dysfunction and adverse birth outcomes.

In **Chapter 3**, I generated a cell type-specific placental deconvolution DNA methylation reference with novel and previously published sorted placental cell type DNA methylation profiles. I analyzed 192 new and previously published DNA methylation profiles from 12 cell types, including cytotrophoblasts (n=32), endothelial cells (n=19), Hofbauer cells (n=26), stromal cells (n=29), syncytiotrophoblasts (n=4), six types of adult lymphocytes (n=36), and nucleated red blood cells (n=11), as well as 35 bulk placental tissue samples. Consistent with placental biology, bulk placental tissue cell type proportion estimates (mean ± standard deviation) from methylation measures were predominately syncytiotrophoblast (57.8% ± 8.3%), stromal (20.6% ± 5.9%), cytotrophoblast (11.0% ± 4.1%), endothelial (7.5% ± 2.2%), Hofbauer cells (1.5% ± 1.2%), and CD4+ T cells (0.89% ± 0.84%). Other cell types had mean estimates less than 0.5%. This innovative cell type DNA methylation reference panel can robustly estimate cell composition from placental DNA methylation data in epidemiological studies to detect unexpected non-placental cell types, reveal biological mechanisms, and improve casual inference.

In **Chapter 4**, I applied the DNA methylation deconvolution reference developed in **Chapter 3** to test the association between prenatal urinary metals concentrations and placental cell composition at term in samples from two prospective pregnancy cohort studies, the Markers of Autism Risk in Babies and Learning Early Signs (MARBLES n=83) and Early Autism Risk Longitudinal Investigation (EARLI) (n=94). At the mean cytotrophoblast cell type proportion across both cohorts of 7.8%, a doubling in early gestation zinc concentration was associated with a 0.8% decrement to 7.0% (POR = 0.89, 95% CI [0.82, 0.97], q=0.26). At the mean Hofbauer cell type proportion across both cohorts of 2.3%, a doubling in late gestation barium concentration was associated with a 0.2% increment to 2.5% (POR = 1.08, 95% CI [1.02, 0.14], q=0.25). Similarly, a doubling in late gestation nickel concentration was associated with a 0.4% increment in expected Hofbauer cell proportion to 2.7% (POR = 1.16, 95% CI [1.05, 1.27], q=0.25). At the mean syncytiotrophoblast cell type proportion across both cohorts of 62.5%, exposure to cadmium concentrations above the limit of detection was associated with a 3.0% decrement to 59.5% (POR = 0.88, 95% CI [0.78, 0.98], q=0.31). At the mean endothelial cell type proportion across both cohorts of 6.1%, a doubling in late gestation copper concentration was associated with a 0.7% increment to 6.8% (POR = 1.11, 95% CI [1.04, 1.18], q=0.24). These results implicated the divalent metal cations zinc, cadmium, barium, and nickel which may directly interact with endothelial, Hofbauer, and syncytiotrophoblast cell types via divalent metal transporter proteins and affect placental structure. This chapter provides evidence that prenatal urinary metals concentrations are associated with term placental cell composition, which may have implications for

increased disease risk related to placental dysfunction as well as for molecular studies of the placenta that fail to account for potential cell composition differences.

Overall, my literature review demonstrates that placental cell composition has been an overlooked factor in perinatal molecular epidemiology studies, and these results show how future work can unlock new biologic insights by incorporating placental cell composition. This dissertation not only highlights these limitations of prior research, with a review of the state of the literature but contextualizes prior insufficiencies in the context of a causal inference framework. Furthermore, I generated novel and integrated placenta cell type-specific single-cell gene expression and sorted cell type DNA methylation microarray profiles to bioinformatically estimate cell composition from bulk placenta samples. Estimation of cell composition from bulk placenta samples allows investigators to model cell composition in a carefully considered causal inference framework to reveal mechanistic insights and eliminate systematic sources of bias. Finally, I applied this deconvolution approach in causal inference-informed models to reveal that cell composition differences in preeclampsia drive previously observed gene expression differences and that prenatal metals urinary metals concentrations are associated with placental cell composition itself. These applications demonstrate the extent cell composition can drive apparent tissue-level gene expression differences or how cell composition can be studied as an outcome of environmental exposures. Future placental investigations must similarly consider cell composition, in a casual inference framework, as a critical factor in understanding the effects of environmental exposures and disease on perinatal health outcomes. The

deconvolution approaches developed in this dissertation and their application provide a template to do so.

5.2 Utility of cellular heterogeneity in research questions and epidemiologic modeling

The biomedical molecular revolution in measuring gene expression, DNA methylation, and other biomolecules has provided unparalleled resolution in connecting exposures, outcomes, and biological mechanisms. As the associated technologies have become more cost-effective and scalable, investigators have increasingly bio-banked, collected, or biopsied tissues for clinical or research purposes. However, these bulk molecular measures represent averages across all the different cell types contained in the target tissue. Epigenetic mechanisms allow each cell type to manifest a specialized, differentiated phenotype despite each cell in the tissue containing a nearly identical genome. Failure to account for the inherent heterogeneity of cell types across any given molecular measure obfuscates the underlying causal relationships between exposure, outcome, and molecular measure. Careful consideration of cell type heterogeneity in a causal inference framework can inform study design and interpretation of results.

Once cell type heterogeneity has been accounted for, one can ask critical questions about DNA methylation, exposures, and disease, though this reasoning applies equally to any bulk molecular measure. The appropriate approach depends on the study sampling framework, timing of measures, and hypothesized relationships between exposures, cellular heterogeneity, DNA methylation, and disease. Causal diagrams [253] are frequently employed in epidemiologic studies to evaluate and communicate the relationships between key variables and identify appropriate

approaches to address bias [254]. Below, I use causal diagrams to describe five study hypotheses involving an exposure, a disease, tissue-level DNA methylation, and cell type-specific DNA methylation epigenotypes, though, again, this framing could equally be applied to other omics, including gene expression. These five hypotheses are well-studied in epidemiologic frameworks: mediation, confounding, biomarker of disease, biomarker of exposure, and precision variables. By directly measuring epigenetics within sorted cell types (cell type-specific epigenotypes), researchers simplify casual diagrams and associated statistical models. Simpler causal diagrams require fewer assumptions, use simpler analytic methods, minimize sources of bias, and improve interpretability of study results [253, 254]. In each setting, I demonstrate that studying cell type-specific epigenotypes simplifies the causal diagram and reduces sources of potential bias.

DNA methylation dysregulation is a candidate to mediate early-life environmental exposures and later life health outcomes [255, 256]. Mediation refers to the indirect effect an exposure has on an outcome by acting through an intervening variable [257]. Though perhaps the most biologically compelling, mediation studies were the among rarest study designs according to recent DNA methylation mediation reviews [258, 259]. Testing mediation requires assumptions under different analytic frameworks, but all models rely on faithfully capturing exposure-mediator and mediator-outcome relationships [260] (**Figure 5.1—Tissue Epigenotype**). These relationships must be identifiable, unconfounded, and free of bias to establish evidence of a causal relationship [261]. Mediation testing in epigenetic perturbations is difficult due to a lack of cell type-specific studies in tissues relevant to the disease process [262] and a lack of

observational studies that span the perinatal period until the end of life [263]. One example of a recent mediation-based study design was a case-control study of rheumatoid arthritis and genetic risk that controlled for cell type heterogeneity and removed DNA methylation signatures due to arthritis onset. Ten differentially methylated regions were identified in a mediation analysis [264]. Recently, a cell type deconvolution algorithm for DNA methylation demonstrated a quantitation of the mediation of phenotypic associations with DNA methylation by cellular heterogeneity in 23 DNA methylation microarray datasets across 13 studies [70]. In **Chapter 2**, I successfully applied a causal mediation framework to demonstrate cell composition changes, particularly an overabundance of extravillous trophoblasts, mediated a large fraction of the association between preeclampsia and placental gene expression. This work also highlights the possibility that placental cell composition may represent a response to exposure or disease state. For example, the elevation of extravillous trophoblasts observed may represent an immature development state or a compensatory attempt to better regulate blood supply to the placenta. In **Chapter 4**, I identified an association between prenatal urinary metals concentrations and placental cell composition, establishing the exposure-mediator relationship that's essential for mediation analysis. The observed elevation of Hofbauer cells and endothelial cells and decreased abundance of syncytiotrophoblasts may similarly represent a compensatory response to direct divalent cation toxicity or to increase sequestration of such cations. Cell type-specific assessment of DNA methylation avoids mediation by cell type heterogeneity altogether (**Figure 5.1—Cell Type Epigenotypes**). Well-designed DNA methylation

mediation studies that account for cell type heterogeneity may identify the mechanisms by which environmental exposures affect DNA methylation directly and disease etiology.

Now, I focus on modeling the tissue epigenotype as the outcome to understand the role cellular heterogeneity plays in the relationship between exposure and tissue epigenotype in the mediation framework. These examples Typically, investigators do not adjust for a mediator as it represents a contributor to the total effect of the exposure on the outcome. In epigenetic studies, assessing mediation by cell type heterogeneity is essential to distinguish direct intranuclear changes (**Figure 1.3a-b**) from shifts in cell type heterogeneity (**Figure 1.3c**) [59], each of which offers insights into disease etiology [46]. Note that single-cell or cell type-specific assessment is required to distinguish a global direct effect (**Figure 1.3a**) from a vulnerable cell type scenario **(Figure 1.3b**). When the goal is to identify direct DNA methylation changes, cell type deconvolution and adjustment can block the pathway that is mediated by cell type heterogeneity. For example, a recent epigenome-wide meta-analysis identified associations robust to cell type adjustment between exposure to maternal smoking in pregnancy and over 6,000 newborn blood DNA methylation sites (**Figure 5.1—Tissue Epigenotype, boxed)** [35]. Again, Cell type-specific assessment of DNA methylation circumvents cell type heterogeneity **(Figure 5.1—Cell Type Epigenotypes, boxed)**. Researchers must take care in the design, analysis, and interpretation of epigenetics studies where the exposure is thought to affect cell type heterogeneity, prompting the consideration of a mediation framework.

In a second scenario, epigenetic measures serve as exposure biomarkers. Because DNA methylation is labile to environmental exposures and generally stable

once established [265], DNA methylation can serve as a proxy measure of past exposures [266, 267]. Several studies have linked maternal smoking during pregnancy to changes in newborn or later childhood blood DNA methylation, though the potential health consequences of these changes beyond a biomarker is unclear [35, 268, 269]. The framework is identical in structure to the mediation scenario, except that the tissue epigenotype does not affect disease (**Figure 5.1—Tissue Epigenotype**). For similar reasons, cell type-specific epigenotypes should be prioritized over tissue epigenotypes whenever possible (**Figure 5.1—Cell Type Epigenotypes**). DNA methylation could reduce information bias in epidemiological studies by extending the reach of exposure assessment backward in time and more accurately quantifying an individual's exposure.

In a third scenario, cell type heterogeneity could be a confounder in the relationship between exposure and tissue-level DNA methylation (**Figure 5.1c—Tissue Epigenotype**). Confounding refers to a non-causal association between an exposure and outcome due to a shared common cause (the "confounder") [270]. Most current epigenome-wide association studies implement cell type proportions as adjustment covariates in regression models, with the stated goal to account for potential confounding due to cell type heterogeneity. Cellular heterogeneity may affect the metabolism, storage, or cellular response to an environmental toxicant, impacting biomarker measured toxicant levels. For example, in five cell lines across 20 heavy metal toxicants, an Nrf2-dependent oxidative stress response varied by cell type [271]. Further, we know that cell type heterogeneity predicts tissue-level DNA methylation [272]. The common cause of cell type heterogeneity could therefore distort the measure of association between exposure and tissue-level DNA methylation. As before,

evaluating DNA methylation at the cell type level simplifies the causal diagram as well as eliminates the potential for confounding by cell type heterogeneity (**Figure 5.1—Cell Type Epigenotypes)**.

Fourth, epigenetic measures can be biomarkers of disease. The application of DNA methylation as a disease biomarker can reduce outcome misclassification, serve as a surrogate endpoint, and monitor disease progression, prognosis, and treatment response [273, 274]. DNA methylation is routinely employed as a disease biomarker in cancers and is being investigated for use in psychiatric conditions and chronic diseases such as cardiovascular disease [275–277]. DNA methylation of placentally derived DNA from maternal plasma has been used as a noninvasive biomarker of aneuploidy [278]. A recent meta-analysis revealed an association between neonatal blood DNA methylation and birthweight, although it is unclear if DNA methylation was a mediating cause of birthweight changes or simply a birthweight biomarker [279]. DNA methylation is an auspicious vehicle for the promise of precision medicine and might soon be established as the ‘universal’ disease biomarker, given the vast array of disease-specific DNA methylation profiles that are being uncovered [280]. Identification of DNA methylation perturbations related to negative health outcomes will likely identify at-risk individuals and lead to novel preventive and therapeutic strategies [259]. Notice that assessment of DNA methylation as a biomarker of disease is identical in diagram structure to the mediation scenario presented above (**Figure 5.1d—Tissue Epigenotype**). Therefore, the same recommendation for assessing cell type-specific epigenotypes can be applied to the use of DNA methylation as a biomarker of disease (**Figure 5.1d—Cell Type**

Epigenotypes). The use of DNA methylation as a disease biomarker still requires methods to account for cell type heterogeneity.

Finally, cell type heterogeneity may be an important precision variable in epigenetic studies (**Figure 5.1e—Tissue Epigenotype**). A precision variable is a predictor of the outcome that is unrelated to the exposure. A precision variable increases statistical efficiency when adjusted for in a model [281]. Cell type proportions are strong predictors of DNA methylation, often accounting for the first principal component of variability in DNA methylation data. Tissue DNA methylation studies may account for cell type proportions in regression models to improve precision in estimating DNA methylation associations with other variables. When cell type-specific DNA methylation is measured, cell type heterogeneity may no longer be relevant to estimating the direct effect of the exposure on cellular DNA methylation (**Figure 5.1e—Cell Type Epigenotypes**). However, independence between cell type heterogeneity and cellular DNA methylation may be an unrealistic assumption due to cell-cell interactions in tissues [53]. Even when cell type heterogeneity is unrelated to the exposure, cell type heterogeneity may be an important precision variable. The capability to model cell type proportions is only a first step in addressing the issue of cellular heterogeneity in bulk molecular measures.

At various genomic locations, tissues, and developmental times, DNA methylation is a promising biomarker of exposure and disease, as well as a potential mediator between environmental exposure and health outcomes. Careful attention must be paid to the hypothesized role of cellular heterogeneity in an epidemiological research question and to the assumptions and analytic methods required to faithfully test the

hypothesized relationship between exposure, bulk molecular measure, cell type heterogeneity, and outcome. Subject matter expertise, transparency of model assumptions, and appropriate methods to accommodate and evaluate potential study hypotheses will be required to improve causal inference and interpretability in environmental epigenetics. It is important that investigators clearly state hypotheses and analytic assumptions to generate valid, replicable, and interpretable study results.

5.3 Limitations and future directions

Dissociation bias, which describes the differential resilience of cell types to dissociation or biosample preparation procedures, has continued to pose a challenge to studying cellular heterogeneity in complex tissues [72]. Reference-based approaches require cell type-specific molecular profiles that can be derived from cell sorting or single-cell approaches; both approaches are affected by dissociation bias. The abnormal size of the multi-nucleated placental syncytiotrophoblast especially is unamenable to conventional cell sorting and single-cell approaches. Single-nuclei preparations are an alternative option to conventional single-cell assays. Single nucleus RNA sequencing has been used to characterize a syncytiotrophoblast cell model and may be more appropriate for such cell types [175]. Future studies of the human placenta should consider such a method to better characterize the syncytiotrophoblast beyond the single-cell RNA sequencing approach used here. Deconvolution estimates may be more accurate than cell sorting or single-cell approaches when the effect of dissociation bias is large. For example, in **Chapter 2**, fetal cytotrophoblasts, endothelial cells, and syncytiotrophoblast were underrepresented in raw single cell counts compared to deconvoluted estimates **(Figure 2.15**). Dissociation bias may also be

exacerbated by cryopreservation or similar biobanking procedures commonly employed in epidemiological studies and should be considered at the study design stage [174].

Notably, when I deconvoluted RNA microarray samples (n=330) in **Chapter 2**, the median (Q1, Q3) estimated placental syncytiotrophoblast composition was 23.7% (20.8%, 26.7%) compared to 63.2% (59.0%, 68.0%) in DNA methylation based deconvolution estimates in EARLI (n=94) and 60.1% (54.6%, 68.0%) in MARBLES (n=83) presented in **Chapter 4**. This was the largest discrepancy across cell types assayed between the two methods. Superficially, these results suggest that a gene expression-based approach underestimates syncytiotrophoblast proportions or a DNA methylation approach overestimates syncytiotrophoblast proportions. However, this may be due to the differing absolute contribution of syncytiotrophoblasts to RNA or DNA molecules measured. Compared to cytotrophoblast nuclei, syncytiotrophoblast nuclei exhibit increased heterochromatin formation and transcriptional inactivity, particularly associated with syncytial knots [282, 283]. Transcriptionally inactive nuclei still contribute to the DNA methylation signal but not active gene transcription while transcriptionally active nuclei contribute to both the DNA methylation and gene expression signals, resulting in a larger relative contribution to the DNA methylation as compared to gene expression. The nonlinear contribution of certain cell types to different levels of molecular biology may underlie other complex tissues beyond the placenta and syncytiotrophoblasts. Future directions to address these challenges may involve combinations of matched gene expression and epigenomic characterization of placental tissue, *in situ* based approaches such as spatial transcriptomics or immunohistochemical investigations, matched samples collected from different regions

of the placenta, and the continued use of bulk deconvolution of whole tissues with highly specific single-cell references that are likely less susceptible to dissociation bias such as the approaches used in this investigation.

The ability to estimate cell composition represents a critical advancement in addressing the challenges of cellular heterogeneity in complex tissues by providing a way to consider cell composition directly in epidemiologic study designs [1, 46]. Deconvolution represents a promising and scalable approach to meet these challenges. Recent statistical approaches have implemented an interaction term between a DNA methylation signal and deconvoluted cell type proportions to detect cell type-specific effects in epigenome-wide association studies without the need for single cell assays [284]. In **5.1**, I reviewed just 5 common scenarios of modelling cell composition in causal inference with direct acyclic graphs. Future work will not only need to consider the technical and biological challenges to measuring cell composition, but also how to incorporate cell composition in causal inference to obtain valid causal estimates in epidemiological research questions.

5.4 Strengths

There were several strengths to this dissertation research. Firstly, the deconvolution references integrated both primary and secondary datasets from multiple sources with leading computational approaches and advanced, complementary laboratory methods. When possible, I also conducted *in silico* deconvolution simulations to validate deconvolution performance. These steps provide confidence in the generalizability and robustness of the reference panels' performance. Further, I verified that cell type and not batch effect drove the variation in molecular data. Next, both

deconvolution references incorporate information from unexpected or cryptic cell types such as comingled maternal cell types or particular cell subtypes that could be a previously unobserved source of bias. Finally, I explicitly considered cell composition as a feature of interest for use in causal inference models in the research application of these methodological advancements. Such an application can serve as a template to elevating traditional analyses that too commonly merely controlled for cell composition as a putative confounder without careful study design considerations. Undoubtably, deployment of the cost-effective and scalable research products of this dissertation will reduce systematic bias and improve mechanistic insight of perinatal epidemiology investigations.

5.5 Public health significance

This dissertation features several novel developments of public health and scientific significance. First, I advance the goal of identifying the molecular underpinnings of adverse birth outcomes, which pose a serious public health threat. In addition, mounting evidence for the developmental origins of health and disease hypothesis suggests the mitigation of adverse birth outcomes may prevent numerous adverse sequelae.

Second, this research addresses cell composition not only as a source of systematic bias but as an outcome itself. Additionally, this work will broadly improve rigor and reproducibility in any field that studies placental tissue and lessons learned here can be applied more broadly to any bulk tissue investigation. Technological advancements have provided unparalleled resolution and cost savings in studying human health. Such advancements come with bioinformatic challenges, including bulk

tissue convolution. This work fills critical gaps by executing labor-intensive cell sorting and analysis to generate and integrate cell type-specific references that will be made widely available for deconvolution. This allows researchers to leverage commonly collected and already-collected bulk placental tissue data to improve precision, reduce potential sources of bias, and improve interpretability of study results.

Third, this dissertation is highly interdisciplinary and serves as a model for broadly integrating epidemiological causal inference techniques into bioinformatic biomedical studies. To carry execute this novel approach, I developed or integrated advanced methods in primary tissue collection, dissociation, and cell type sorting, single-cell gene expression characterization and analysis, bulk gene expression characterization and analysis, bulk DNA methylation characterization and analysis, creation and validation of gene expression and DNA methylation deconvolution references, application of deconvolution to epidemiologically model cell composition, general bioinformatics, bioinformatic data integration, meta-analysis, molecular epidemiology, and causal inference.

This dissertation fills a key lacuna in applying causal inference techniques to various research questions and study designs in molecular epidemiology (**Figure 5.1**). In summary, I provide additional or novel cell type-specific gene expression or DNA methylation profiles for public use in deconvolution. In matched samples, I integrate complementary gene expression technologies single-cell RNA sequencing, bulk RNA sequencing of cell types and apply both RNA-based and DNA methylation-based cell type estimation. I contribute to the understanding of the molecular underpinnings of

preeclampsia and effects of early-life metals exposure on placental structure commonly experienced by humans.

5.6 Figures and Tables

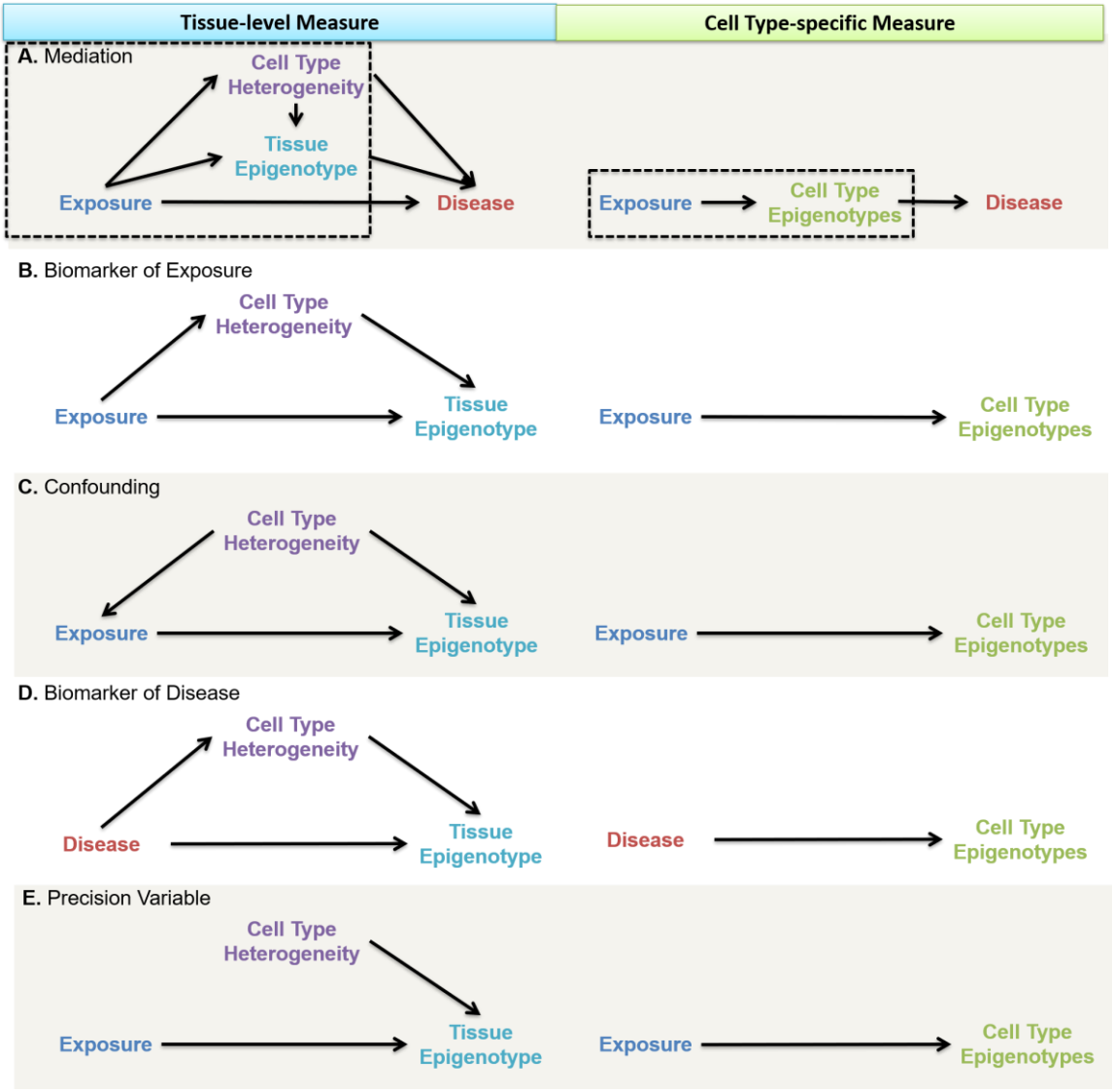


Figure 5.1 Based on hypothesized relationships between exposure, disease, and DNA methylation, multiple study design scenarios are possible. Measures of DNA methylation can be implemented in directed acyclic graphs for identifying model assumptions and analytic strategies for causal inference. The left column represents a DNA methylation epigenotype aggregated over multiple cell types (e.g., tissue) and the right column represents cell type-specific epigenotypes measures. When measured on the cell type-specific level, the causal link between composition heterogeneity and epigenotype is broken and omitted from the cell type-specific diagram. ***A.*** *Mediation: The exposure affects disease indirectly through the tissue epigenotype (direct DNA methylation effects across all cells (Figure 1.3A) or in vulnerable cell types (Figure 1.3B) or through cell type heterogeneity (Figure 1.3C). The exposure may also affect the*

*disease directly. We focus on modeling the tissue epigenotype as the outcome, a subset of the overall causal diagram (**Boxed**). **B.** Biomarker of Exposure: The exposure affects the DNA methylation epigenotype directly and indirectly through cell type composition heterogeneity. C. Confounding: Cell type composition affects the level of exposure and directly affects the epigenotype through heterogeneity. **D.** Biomarker of Disease: The disease state affects the DNA methylation epigenotype directly and indirectly through cell type composition heterogeneity. **E.** Precision Variable: Cell type composition heterogeneity is independent of the exposure but is a strong predictor of DNA methylation epigenotype.*

Bibliography

1. 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

2. Goldenberg RL, Culhane JF, Iams JD, Romero R (2008) Epidemiology and causes of preterm birth. The Lancet 371:75–84. https://doi.org/10.1016/S0140-6736(08)60074-4

3. 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

4. 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

5. Guttmacher AE, Maddox YT, Spong CY (2014) The Human Placenta Project: Placental structure, development, and function in real time. Placenta 35:303–304. https://doi.org/10.1016/j.placenta.2014.02.012

6. 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

7. Barker DJP (1995) Fetal origins of coronary heart disease. BMJ 311:171–174

8. 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

9. 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

10. Midgley AR, Pierce GB, Deneau GA, Gosling JRG (1963) Morphogenesis of Syncytiotrophoblast in vivo: An Autoradiographic Demonstration. Science 141:349–350. https://doi.org/10.1126/science.141.3578.349

11. Loegl J, Hiden U, Nussbaumer E, et al (2016) Hofbauer cells of M2a, M2b and M2c polarization may regulate feto-placental angiogenesis. Reproduction 152:447–455. https://doi.org/10.1530/REP-16-0159

12. Seval Y, Korgun ET, Demir R (2007) Hofbauer Cells in Early Human Placenta: Possible Implications in Vasculogenesis and Angiogenesis. Placenta 28:841–845. https://doi.org/10.1016/j.placenta.2007.01.010

13. Travers A, Muskhelishvili G (2015) DNA structure and function. The FEBS Journal 282:2279–2295. https://doi.org/10.1111/febs.13307

14. Ozsolak F, Milos PM (2011) RNA sequencing: advances, challenges and opportunities. Nature Reviews Genetics 12:87–98. https://doi.org/10.1038/nrg2934

15. Saenen Nelly D., Plusquin Michelle, Bijnens Esmée, et al (2015) In Utero Fine Particle Air Pollution and Placental Expression of Genes in the Brain-Derived Neurotrophic Factor Signaling Pathway: An ENVIRONAGE Birth Cohort Study. Environmental Health Perspectives 123:834–840. https://doi.org/10.1289/ehp.1408549

16. Greally JM (2018) A user's guide to the ambiguous word "epigenetics." Nat Rev Mol Cell Biol 19:207–208. https://doi.org/10.1038/nrm.2017.135

17. Deichmann U (2016) Epigenetics: The origins and evolution of a fashionable topic. Developmental Biology 416:249–254. https://doi.org/10.1016/j.ydbio.2016.06.005

18. Goldberg AD, Allis CD, Bernstein E (2007) Epigenetics: A Landscape Takes Shape. Cell 128:635–638. https://doi.org/10.1016/j.cell.2007.02.006

19. Smith ZD, Chan MM, Humm KC, et al (2014) DNA methylation dynamics of the human preimplantation embryo. Nature 511:611–615. https://doi.org/10.1038/nature13581

20. Messerschmidt DM, Knowles BB, Solter D (2014) DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. Genes Dev 28:812–828. https://doi.org/10.1101/gad.234294.113

21. Chaligné R, Heard E (2014) X-chromosome inactivation in development and cancer. FEBS Letters 588:2514–2522. https://doi.org/10.1016/j.febslet.2014.06.023

22. Doi A, Park I-H, Wen B, et al (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat Genet 41:1350–1353. https://doi.org/10.1038/ng.471

23. Reik W, Dean W, Walter J (2001) Epigenetic Reprogramming in Mammalian Development. Science 293:1089–1093. https://doi.org/10.1126/science.1063443

24. Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. Nat Rev Genet 14:204–220. https://doi.org/10.1038/nrg3354

25. Khavari DA, Sen GL, Rinn JL (2010) DNA methylation and epigenetic control of cellular differentiation. Cell Cycle 9:3880–3883. https://doi.org/10.4161/cc.9.19.13385

26. Virani S, Colacino JA, Kim JH, Rozek LS (2012) Cancer Epigenetics: A Brief Review. ILAR J 53:359–369. https://doi.org/10.1093/ilar.53.3-4.359

27. Berson A, Nativio R, Berger SL, Bonini NM (2018) Epigenetic Regulation in Neurodegenerative Diseases. Trends in Neurosciences 41:587–598. https://doi.org/10.1016/j.tins.2018.05.005

28. Ordovás JM, Smith CE (2010) Epigenetics and cardiovascular disease. Nat Rev Cardiol 7:510–519. https://doi.org/10.1038/nrcardio.2010.104

29. Feil R, Fraga MF (2012) Epigenetics and the environment: emerging patterns and implications. Nat Rev Genet 13:97–109. https://doi.org/10.1038/nrg3142

30. Gluckman PD, Hanson MA, Buklijas T, et al (2009) Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. Nat Rev Endocrinol 5:401–408. https://doi.org/10.1038/nrendo.2009.102

31. Bakulski KM, Halladay A, Hu VW, et al (2016) Epigenetic Research in Neuropsychiatric Disorders: the “Tissue Issue.” Curr Behav Neurosci Rep 3:264–274. https://doi.org/10.1007/s40473-016-0083-4

32. Hannon E, Lunnon K, Schalkwyk L, Mill J (2015) Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. Epigenetics 10:1024–1032. https://doi.org/10.1080/15592294.2015.1100786

33. Walton E, Hass J, Liu J, et al (2016) Correspondence of DNA Methylation Between Blood and Brain Tissue and Its Application to Schizophrenia Research. Schizophr Bull 42:406–414. https://doi.org/10.1093/schbul/sbv074

34. Wang T, Pehrsson EC, Purushotham D, et al (2018) The NIEHS TaRGET II Consortium and environmental epigenomics. Nat Biotechnol 36:225–227. https://doi.org/10.1038/nbt.4099

35. Joubert BR, Felix JF, Yousefi P, et al (2016) DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. The American Journal of Human Genetics 98:680–696. https://doi.org/10.1016/j.ajhg.2016.02.019

36. Sikdar S, Joehanes R, Joubert BR, et al (2019) Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from

personal smoking. Epigenomics 11:1487–1500. https://doi.org/10.2217/epi-2019-0066

37. Bakulski KM, Dou J, Lin N, et al (2019) DNA methylation signature of smoking in lung cancer is enriched for exposure signatures in newborn and adult blood. Sci Rep 9:. https://doi.org/10.1038/s41598-019-40963-2

38. Tsang JCH, Vong JSL, Ji L, et al (2017) Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics. European Genome-Phenome Archive EGAS00001002449. https://ega-archive.org/studies/EGAS00001002449

39. Nakahara A, Nair S, Ormazabal V, et al (2020) Circulating Placental Extracellular Vesicles and Their Potential Roles During Pregnancy. Ochsner J 20:439–445. https://doi.org/10.31486/toj.20.0049

40. Schuster J, Cheng S-B, Padbury J, Sharma S (2021) Placental Extracellular Vesicles and Preeclampsia. Am J Reprod Immunol 85:e13297. https://doi.org/10.1111/aji.13297

41. Maltepe E, Bakardjiev AI, Fisher SJ (2010) The placenta: transcriptional, epigenetic, and physiological integration during development. J Clin Invest 120:1016–1025. https://doi.org/10.1172/JCI41211

42. Knöfler M, Pollheimer J (2013) Human placental trophoblast invasion and differentiation: a particular focus on Wnt signaling. Front Genet 4:. https://doi.org/10.3389/fgene.2013.00190

43. Knott JG, Paul S (2014) Transcriptional regulators of the trophoblast lineage in mammals with hemochorial placentation. Reproduction 148:R121–R136. https://doi.org/10.1530/REP-14-0072

44. 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

45. 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

46. 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

47. 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

48. Horvath S (2013) DNA methylation age of human tissues and cell types. Genome Biology 14:3156. https://doi.org/10.1186/gb-2013-14-10-r115

49. Jylhävä J, Pedersen NL, Hägg S (2017) Biological Age Predictors. EBioMedicine 21:29–36. https://doi.org/10.1016/j.ebiom.2017.03.046

50. Bauer M, Fink B, Thürmann L, et al (2016) Tobacco smoking differently influences cell types of the innate and adaptive immune system—indications from CpG site methylation. Clin Epigenetics 8:. https://doi.org/10.1186/s13148-016-0249-7

51. Su D, Wang X, Campbell MR, et al (2016) Distinct Epigenetic Effects of Tobacco Smoking in Whole Blood and among Leukocyte Subtypes. PLOS ONE 11:e0166486. https://doi.org/10.1371/journal.pone.0166486

52. Bauer M, Linsel G, Fink B, et al (2015) A varying T cell subtype explains apparent tobacco smoking induced single CpG hypomethylation in whole blood. Clin Epigenet 7:81. https://doi.org/10.1186/s13148-015-0113-1

53. Lappalainen T, Greally JM (2017) Associating cellular epigenetic models with human phenotypes. Nat Rev Genet 18:441–451. https://doi.org/10.1038/nrg.2017.32

54. Herzenberg LA, Parks D, Sahaf B, et al (2002) The History and Future of the Fluorescence Activated Cell Sorter and Flow Cytometry: A View from Stanford. Clinical Chemistry 48:1819–1827

55. Karemaker ID, Vermeulen M (2018) Single-Cell DNA Methylation Profiling: Technologies and Biological Applications. Trends in Biotechnology 36:952–965. https://doi.org/10.1016/j.tibtech.2018.04.002

56. Tanay A, Regev A (2017) Scaling single-cell genomics from phenomenology to mechanism. Nature 541:331–338. https://doi.org/10.1038/nature21350

57. 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

58. 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

59. Houseman EA, Kelsey KT, Wiencke JK, Marsit CJ (2015) Cell-composition effects in the analysis of DNA methylation array data: a mathematical perspective. BMC Bioinformatics 16:95. https://doi.org/10.1186/s12859-015-0527-y

60. Gervin K, Salas LA, Bakulski KM, et al (2019) Systematic evaluation and validation of reference and library selection methods for deconvolution of cord blood

DNA methylation data. Clin Epigenetics 11:. https://doi.org/10.1186/s13148-019-0717-y

61. Gervin K, Page CM, Aass HCD, et al (2016) Cell type specific DNA methylation in cord blood: A 450K-reference data set and cell count-based validation of estimated cell type composition. Epigenetics 11:690–698. https://doi.org/10.1080/15592294.2016.1214782

62. Bakulski KM, Feinberg JI, Andrews SV, et al (2016) DNA methylation of cord blood cell types: Applications for mixed cell birth studies. Epigenetics 11:354–362. https://doi.org/10.1080/15592294.2016.1161875

63. de Goede OM, Lavoie PM, Robinson WP (2016) Characterizing the hypomethylated DNA methylation profile of nucleated red blood cells from cord blood. Epigenomics 8:1481–1494. https://doi.org/10.2217/epi-2016-0069

64. Lin X, Tan JYL, Teh AL, et al (2018) Cell type-specific DNA methylation in neonatal cord tissue and cord blood: a 850K-reference panel and comparison of cell types. Epigenetics 13:941–958. https://doi.org/10.1080/15592294.2018.1522929

65. Salas LA, Koestler DC, Butler RA, et al (2018) An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. Genome Biol 19:1–14. https://doi.org/10.1186/s13059-018-1448-7

66. Guintivano J, Aryee MJ, Kaminsky ZA (2013) A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. Epigenetics 8:290–302. https://doi.org/10.4161/epi.23924

67. Zheng SC, Webster AP, Dong D, et al (2018) A novel cell-type deconvolution algorithm reveals substantial contamination by immune cells in saliva, buccal and cervix. Epigenomics 10:925–940. https://doi.org/10.2217/epi-2018-0037

68. Liang L, Cookson WOC (2014) Grasping nettles: cellular heterogeneity and other confounders in epigenome-wide association studies. Hum Mol Genet 23:R83–R88. https://doi.org/10.1093/hmg/ddu284

69. Leek JT, Storey JD (2007) Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. PLOS Genetics 3:e161. https://doi.org/10.1371/journal.pgen.0030161

70. Houseman EA, Kile ML, Christiani DC, et al (2016) Reference-free deconvolution of DNA methylation data and mediation by cell composition effects. BMC Bioinformatics 17:259. https://doi.org/10.1186/s12859-016-1140-4

71. Zheng SC, Beck S, Jaffe AE, et al (2017) Correcting for cell-type heterogeneity in epigenome-wide association studies: revisiting previous analyses. Nat Methods 14:216–217. https://doi.org/10.1038/nmeth.4187

72. 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

73. Dieckmann L, Cruceanu C, Lahti-Pulkkinen M, et al (2022) Reliability of a novel approach for reference-based cell type estimation in human placental DNA methylation studies. Cell Mol Life Sci 79:115. https://doi.org/10.1007/s00018-021-04091-3

74. Robbins JR, Skrzypczynska KM, Zeldovich VB, et al (2010) Placental syncytiotrophoblast constitutes a major barrier to vertical transmission of Listeria monocytogenes. PLoS Pathog 6:e1000732. https://doi.org/10.1371/journal.ppat.1000732

75. Tomlinson MJ, Tomlinson S, Yang XB, Kirkham J (2012) Cell separation: Terminology and practical considerations: Journal of Tissue Engineering. https://doi.org/10.1177/2041731412472690

76. Akman K, Haaf T, Gravina S, et al (2014) Genome-wide quantitative analysis of DNA methylation from bisulfite sequencing data. Bioinformatics 30:1933–1934. https://doi.org/10.1093/bioinformatics/btu142

77. Schultz MD, He Y, Whitaker JW, et al (2015) Human Body Epigenome Maps Reveal Noncanonical DNA Methylation Variation. Nature 523:212–216. https://doi.org/10.1038/nature14465

78. Welch JD, Kozareva V, Ferreira A, et al (2019) Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. Cell 177:1873-1887.e17. https://doi.org/10.1016/j.cell.2019.05.006

79. Kapourani C-A, Sanguinetti G (2019) Melissa: Bayesian clustering and imputation of single-cell methylomes. Genome Biology 20:61. https://doi.org/10.1186/s13059-019-1665-8

80. Houseman EA, Accomando WP, Koestler DC, et al (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 13:86. https://doi.org/10.1186/1471-2105-13-86

81. Newman AM, Liu CL, Green MR, et al (2015) Robust enumeration of cell subsets from tissue expression profiles. Nature Methods 12:453–457. https://doi.org/10.1038/nmeth.3337

82. Koestler DC, Jones MJ, Usset J, et al (2016) Improving cell mixture deconvolution by identifying optimal DNA methylation libraries (IDOL). BMC Bioinformatics 17:120. https://doi.org/10.1186/s12859-016-0943-7

83. Teschendorff AE, Breeze CE, Zheng SC, Beck S (2017) A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. BMC Bioinformatics 18:105. https://doi.org/10.1186/s12859-017-1511-5

84. Teschendorff AE, Zhuang J, Widschwendter M (2011) Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. Bioinformatics 27:1496–1505. https://doi.org/10.1093/bioinformatics/btr171

85. Gagnon-Bartsch JA, Speed TP (2012) Using control genes to correct for unwanted variation in microarray data. Biostatistics 13:539–552. https://doi.org/10.1093/biostatistics/kxr034

86. Zou J, Lippert C, Heckerman D, et al (2014) Epigenome-wide association studies without the need for cell-type composition. Nat Methods 11:309–311. https://doi.org/10.1038/nmeth.2815

87. Houseman EA, Molitor J, Marsit CJ (2014) Reference-free cell mixture adjustments in analysis of DNA methylation data. Bioinformatics 30:1431–1439. https://doi.org/10.1093/bioinformatics/btu029

88. Rahmani E, Zaitlen N, Baran Y, et al (2016) Sparse PCA corrects for cell type heterogeneity in epigenome-wide association studies. Nature Methods 13:443–445. https://doi.org/10.1038/nmeth.3809

89. Rahmani E, Schweiger R, Shenhav L, et al (2018) BayesCCE: a Bayesian framework for estimating cell-type composition from DNA methylation without the need for methylation reference. Genome Biology 19:141. https://doi.org/10.1186/s13059-018-1513-2

90. Campbell KA, Colacino JA, Puttabyatappa M, et al (2023) Placental cell type deconvolution reveals that cell proportions drive preeclampsia gene expression differences. Commun Biol 6:1–15. https://doi.org/10.1038/s42003-023-04623-6

91. Butwick AJ, Druzin ML, Shaw GM, Guo N (2020) Evaluation of US State–Level Variation in Hypertensive Disorders of Pregnancy. JAMA Netw Open 3:. https://doi.org/10.1001/jamanetworkopen.2020.18741

92. 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

93. 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

94. 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

95. 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

96. 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

97. 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

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

99. 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

100. 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

101. 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

102. 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

103. 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

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

105. 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

106. 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

107. 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

108. 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

109. 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

110. 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

111. 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

112. Pique-Regi R, Romero R, Tarca AL, et al (2019) Single cell transcriptional signatures of the human placenta in term and preterm parturition. Database of Genotypes and Phenotypes phs001886.v1.p1. https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001886.v1.p1

113. 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

114. Zhang F, Yan Y, Kang HM (2021) popscle. https://github.com/statgen/popscle

115. 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

116. 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

117. Amezquita R, Lun A, Hicks S, Gottardo R (2021) Chapter 1 Correcting batch effects | Multi-Sample Single-Cell Analyses with Bioconductor. In: Orchestrating Single-Cell Analysis

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

119. 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

120. Satija R, Butler A, Hoffman P, Stuart T (2020) SeuratWrappers: Community-Provided Methods and Extensions for the Seurat Object

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

122. 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

123. Pierre-Luc (2021) plger/scDblFinder

124. Aran D, Looney AP, Liu L, et al (2019) Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol 20:163–172. https://doi.org/10.1038/s41590-018-0276-y

125. 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

126. Avila Cobos F, Alquicira-Hernandez J, Powell JE, et al (2020) Benchmarking of cell type deconvolution pipelines for transcriptomics data. Nat Commun 11:5650. https://doi.org/10.1038/s41467-020-19015-1

127. 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

128. 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

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

130. 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

131. 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

132. 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

133. 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

134. 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

135. Cervar-Zivkovic M, Stern C (2011) Trophoblast Isolation and Culture. In: Kay HH, Nelson DM, Wang Y (eds) The Placenta. John Wiley & Sons, Ltd, pp 153–162

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

137. 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

138. 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

139. 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

140. 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

141. 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

142. 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

143. 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

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

145. 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

146. 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

147. 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

148. 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

149. Zhou GQ, Baranov V, Zimmermann W, et al (1997) Highly specific monoclonal antibody demonstrates that pregnancy-specific glycoprotein (PSG) is limited to syncytiotrophoblast in human early and term placenta. Placenta 18:491–501. https://doi.org/10.1016/0143-4004(77)90002-9

150. 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

151. Dejana E (2004) Endothelial cell–cell junctions: happy together. Nat Rev Mol Cell Biol 5:261–270. https://doi.org/10.1038/nrm1357

152. van Noesel CJ, van Lier RA, Cordell JL, et al (1991) The membrane IgM-associated heterodimer on human B cells is a newly defined B cell antigen that contains the protein product of the mb-1 gene. J Immunol 146:3881–3888

153. Shapiro-Shelef M, Calame K (2005) Regulation of plasma-cell development. Nat Rev Immunol 5:230–242. https://doi.org/10.1038/nri1572

154. Geissmann F, Jung S, Littman DR (2003) Blood Monocytes Consist of Two Principal Subsets with Distinct Migratory Properties. Immunity 19:71–82. https://doi.org/10.1016/S1074-7613(03)00174-2

155. 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

156. Musumeci A, Lutz K, Winheim E, Krug AB (2019) What Makes a pDC: Recent Advances in Understanding Plasmacytoid DC Development and Heterogeneity. Frontiers in Immunology 10:

157. 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

158. Yang C, Siebert JR, Burns R, et al (2019) Heterogeneity of human bone marrow and blood natural killer cells defined by single-cell transcriptome. Nat Commun 10:3931. https://doi.org/10.1038/s41467-019-11947-7

159. Santana MA, Esquivel-Guadarrama F (2006) Cell Biology of T Cell Activation and Differentiation. In: International Review of Cytology. Academic Press, pp 217–274

160. 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

161. 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

162. 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

163. 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

164. 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

165. 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

166. 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

167. 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

168. 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

169. 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

170. 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

171. 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

172. 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

173. 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

174. Denisenko E, Guo BB, Jones M, et al (2020) Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. Genome Biol 21:130. https://doi.org/10.1186/s13059-020-02048-6

175. Khan T, Seetharam AS, Zhou J, et al (2021) Single Nucleus RNA Sequence (snRNAseq) Analysis of the Spectrum of Trophoblast Lineages Generated From Human Pluripotent Stem Cells in vitro. Front Cell Dev Biol 9:695248. https://doi.org/10.3389/fcell.2021.695248

176. 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

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

178. 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

179. Suryawanshi H, Max K, Bogardus KA, et al (2022) Dynamic genome-wide gene expression and immune cell composition in the developing human placenta. Journal of Reproductive Immunology 151:103624. https://doi.org/10.1016/j.jri.2022.103624

180. Smith AN, Wang X, Thomas DG, et al (2021) The Role of Mitochondrial Dysfunction in Preeclampsia: Causative Factor or Collateral Damage? Am J Hypertens 34:442–452. https://doi.org/10.1093/ajh/hpab003

181. 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

182. 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

183. Burton GJ, Yung H-W (2011) Endoplasmic reticulum stress in the pathogenesis of early-onset pre-eclampsia. Pregnancy Hypertens 1:72–78. https://doi.org/10.1016/j.preghy.2010.12.002

184. Burton GJ, Yung H-W, Cindrova-Davies T, Charnock-Jones DS (2009) Placental endoplasmic reticulum stress and oxidative stress in the pathophysiology of unexplained intrauterine growth restriction and early onset preeclampsia. Placenta 30 Suppl A:S43-48. https://doi.org/10.1016/j.placenta.2008.11.003

185. Shibuya M (2006) Vascular endothelial growth factor receptor-1 (VEGFR-1/Flt-1): a dual regulator for angiogenesis. Angiogenesis 9:225–230. https://doi.org/10.1007/s10456-006-9055-8

186. Luttun A, Tjwa M, Carmeliet P (2002) Placental growth factor (PlGF) and its receptor Flt-1 (VEGFR-1): novel therapeutic targets for angiogenic disorders. Ann N Y Acad Sci 979:80–93. https://doi.org/10.1111/j.1749-6632.2002.tb04870.x

187. 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

188. 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

189. 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

190. Zhou W, Wang H, Yang Y, et al (2022) Trophoblast Cell Subtypes and Dysfunction in the Placenta of Individuals with Preeclampsia Revealed by Single-Cell RNA Sequencing. Mol Cells 45:317–328. https://doi.org/10.14348/molcells.2021.0211

191. Phillippe M (2015) Cell-Free Fetal DNA, Telomeres, and the Spontaneous Onset of Parturition. Reprod Sci 22:1186–1201. https://doi.org/10.1177/1933719115592714

192. Sharp AN, Heazell AEP, Crocker IP, Mor G (2010) Placental Apoptosis in Health and Disease. American Journal of Reproductive Immunology 64:159–169. https://doi.org/10.1111/j.1600-0897.2010.00837.x

193. 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

194. 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

195. Lapehn S, Paquette AG (2022) The Placental Epigenome as a Molecular Link Between Prenatal Exposures and Fetal Health Outcomes Through the DOHaD Hypothesis. Curr Envir Health Rpt 9:490–501. https://doi.org/10.1007/s40572-022-00354-8

196. Bakulski KM, Lee H, Feinberg JI, et al (2015) Prenatal mercury concentration is associated with changes in DNA methylation at TCEANC2 in newborns. Int J Epidemiol 44:1249–1262. https://doi.org/10.1093/ije/dyv032

197. Cardenas A, Koestler DC, Houseman EA, et al (2015) Differential DNA methylation in umbilical cord blood of infants exposed to mercury and arsenic in utero. Epigenetics 10:508–515. https://doi.org/10.1080/15592294.2015.1046026

198. Hogg K, Price EM, Robinson WP (2014) Improved reporting of DNA methylation data derived from studies of the human placenta. Epigenetics 9:333–337. https://doi.org/10.4161/epi.27648

199. Kim SY, Romero R, Tarca AL, et al (2012) Methylome of Fetal and Maternal Monocytes and Macrophages at the Feto-Maternal Interface. American Journal of Reproductive Immunology 68:8–27. https://doi.org/10.1111/j.1600-0897.2012.01108.x

200. Hamada H, Okae H, Toh H, et al (2016) Allele-Specific Methylome and Transcriptome Analysis Reveals Widespread Imprinting in the Human Placenta. The American Journal of Human Genetics 99:1045–1058. https://doi.org/10.1016/j.ajhg.2016.08.021

201. Yuan V, Hui D, Yin Y, et al (2021) Cell-specific characterization of the placental methylome. BMC Genomics 22:6. https://doi.org/10.1186/s12864-020-07186-6

202. Haftorn KL, Denault WRP, Lee Y, et al (2023) Nucleated red blood cells explain most of the association between DNA methylation and gestational age. Commun Biol 6:224. https://doi.org/10.1038/s42003-023-04584-w

203. Salas LA, Koestler DC (2021) FlowSorted.Blood.EPIC: Illumina EPIC data on immunomagnetic sorted peripheral adult blood cells

204. Salas LA, Gervin K, Jones MC (2021) FlowSorted.CordBloodCombined.450k: Illumina 450k/EPIC data on FACS and MACS umbilical blood cells

205. Morgan M, Shepherd L (2021) ExperimentHub: Client to access ExperimentHub resources

206. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al (2014) Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30:1363–1369. https://doi.org/10.1093/bioinformatics/btu049

207. Heiss JA, Just AC (2018) Identifying mislabeled and contaminated DNA methylation microarray data: an extended quality control toolset with examples from GEO. Clinical Epigenetics 10:73. https://doi.org/10.1186/s13148-018-0504-1

208. Peters TJ, Buckley MJ, Statham AL, et al (2015) De novo identification of differentially methylated regions in the human genome. Epigenetics & Chromatin 8:6. https://doi.org/10.1186/1756-8935-8-6

209. Teschendorff AE, Marabita F, Lechner M, et al (2013) A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics 29:189–196. https://doi.org/10.1093/bioinformatics/bts680

210. Tian Y, Morris TJ, Webster AP, et al (2017) ChAMP: updated methylation analysis pipeline for Illumina BeadChips. Bioinformatics 33:3982–3984. https://doi.org/10.1093/bioinformatics/btx513

211. Ritchie ME, Phipson B, Wu D, et al (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47. https://doi.org/10.1093/nar/gkv007

212. Phipson B, Maksimovic J, Oshlack A (2016) missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics 32:286–288. https://doi.org/10.1093/bioinformatics/btv560

213. Martin EM, Fry RC (2018) Environmental Influences on the Epigenome: Exposure- Associated DNA Methylation in Human Populations. Annual Review of Public Health 39:309–333. https://doi.org/10.1146/annurev-publhealth-040617-014629

214. Rayman MP (2012) Selenium and human health. The Lancet 379:1256–1268. https://doi.org/10.1016/S0140-6736(11)61452-9

215. Mistry HD, Broughton Pipkin F, Redman CWG, Poston L (2012) Selenium in reproductive health. American Journal of Obstetrics and Gynecology 206:21–30. https://doi.org/10.1016/j.ajog.2011.07.034

216. O'Neal SL, Zheng W (2015) Manganese Toxicity Upon Overexposure: a Decade in Review. Curr Envir Health Rpt 2:315–328. https://doi.org/10.1007/s40572-015-0056-x

217. Maccani JZJ, Koestler DC, Houseman EA, et al (2015) DNA methylation changes in the placenta are associated with fetal manganese exposure. Reproductive Toxicology 57:43–49. https://doi.org/10.1016/j.reprotox.2015.05.002

218. Nakayama SF, Iwai-Shimada M, Oguri T, et al (2019) Blood mercury, lead, cadmium, manganese and selenium levels in pregnant women and their determinants: the Japan Environment and Children's Study (JECS). Journal of Exposure Science & Environmental Epidemiology 29:633–647. https://doi.org/10.1038/s41370-019-0139-0

219. Arbuckle TE, Liang CL, Morisset A-S, et al (2016) Maternal and fetal exposure to cadmium, lead, manganese and mercury: The MIREC study. Chemosphere 163:270–282. https://doi.org/10.1016/j.chemosphere.2016.08.023

220. Callan AC, Hinwood AL, Ramalingam M, et al (2013) Maternal exposure to metals—Concentrations and predictors of exposure. Environmental Research 126:111–117. https://doi.org/10.1016/j.envres.2013.07.004

221. Chung Soo Eun, Cheong Hae-Kwan, Ha Eun-Hee, et al (2015) Maternal Blood Manganese and Early Neurodevelopment: The Mothers and Children's Environmental Health (MOCEH) Study. Environmental Health Perspectives 123:717–722. https://doi.org/10.1289/ehp.1307865

222. Erikson KM, Thompson K, Aschner J, Aschner M (2007) Manganese neurotoxicity: A focus on the neonate. Pharmacology & Therapeutics 113:369–377. https://doi.org/10.1016/j.pharmthera.2006.09.002

223. Nandakumaran M, Al-Sannan B, Al-Sarraf H, Al-Shammari M (2016) Maternal–fetal transport kinetics of manganese in perfused human placental lobule in vitro. The Journal of Maternal-Fetal & Neonatal Medicine 29:274–278. https://doi.org/10.3109/14767058.2014.998193

224. Chen Z, Myers R, Wei T, et al (2014) Placental transfer and concentrations of cadmium, mercury, lead, and selenium in mothers, newborns, and young children. J Expo Sci Environ Epidemiol 24:537–544. https://doi.org/10.1038/jes.2014.26

225. Caserta D, Graziano A, Lo Monte G, et al (2013) Heavy metals and placental fetal-maternal barrier: a mini-review on the major concerns. Eur Rev Med Pharmacol Sci 17:2198–2206

226. Geng H-X, Wang L (2019) Cadmium: Toxic effects on placental and embryonic development. Environmental Toxicology and Pharmacology 67:102–107. https://doi.org/10.1016/j.etap.2019.02.006

227. Phuapittayalert L, Norkaew T, Supanpaiboon W, et al (2013) Increasing of Syncytial Knot and Fibrinoid Deposit in High-Cd Accumulated Human Placentas. International Journal of Morphology 31:1210–1215. https://doi.org/10.4067/S0717-95022013000400011

228. Wilson RL, Leemaqz SY, Goh Z, et al (2017) Zinc is a critical regulator of placental morphogenesis and maternal hemodynamics during pregnancy in mice. Sci Rep 7:15137. https://doi.org/10.1038/s41598-017-15085-2

229. Newschaffer CJ, Croen LA, Fallin MD, et al (2012) Infant siblings and the investigation of autism risk factors. Journal of Neurodevelopmental Disorders 4:7. https://doi.org/10.1186/1866-1955-4-7

230. Hopper JL (2011) Disease-specific prospective family study cohorts enriched for familial risk. Epidemiol Perspect Innov 8:2. https://doi.org/10.1186/1742-5573-8-2

231. Hertz-Picciotto I, Schmidt RJ, Walker CK, et al (2018) A Prospective Study of Environmental Exposures and Early Biomarkers in Autism Spectrum Disorder: Design, Protocols, and Preliminary Data from the MARBLES Study. Environ Health Perspect 126:117004. https://doi.org/10.1289/EHP535

232. Hornung RW, Reed LD (1990) Estimation of Average Concentration in the Presence of Nondetectable Values. Applied Occupational and Environmental Hygiene 5:46–51. https://doi.org/10.1080/1047322X.1990.10389587

233. Bibikova M, Barnes B, Tsan C, et al (2011) High density DNA methylation array with single CpG site resolution. Genomics 98:288–295. https://doi.org/10.1016/j.ygeno.2011.07.007

234. Triche TJ, Weisenberger DJ, Van Den Berg D, et al (2013) Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucleic Acids Res 41:e90. https://doi.org/10.1093/nar/gkt090

235. Chen Y, Lemire M, Choufani S, et al (2013) Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 8:203–209. https://doi.org/10.4161/epi.23470

236. Pidsley R, Zotenko E, Peters TJ, et al (2016) Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol 17:208. https://doi.org/10.1186/s13059-016-1066-1

237. Sjoberg DD, Whiting K, Curry M, et al (2021) Reproducible Summary Tables with the gtsummary Package. The R Journal 13:570–580. https://doi.org/10.32614/RJ-2021-053

238. Wei T, Simko V (2021) R package “corrplot”: Visualization of a Correlation Matrix

239. Cribari-Neto F, Zeileis A (2010) Beta Regression in R. Journal of Statistical Software 34:1–24. https://doi.org/10.18637/jss.v034.i02

240. Balduzzi S, Rücker G, Schwarzer G (2019) How to perform a meta-analysis with R: a practical tutorial. Evidence-Based Mental Health 153–160

241. R Core Team (2022) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria

242. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

243. Knöfler M, Haider S, Saleh L, et al (2019) Human placenta and trophoblast development: key molecular mechanisms and model systems. Cell Mol Life Sci 76:3479–3496. https://doi.org/10.1007/s00018-019-03104-6

244. Reyes L, Golos TG (2018) Hofbauer Cells: Their Role in Healthy and Complicated Pregnancy. Front Immunol 9:2628. https://doi.org/10.3389/fimmu.2018.02628

245. Vimalraj S, Sumantran VN, Chatterjee S (2017) MicroRNAs: Impaired vasculogenesis in metal induced teratogenicity. Reproductive Toxicology 70:30–48. https://doi.org/10.1016/j.reprotox.2017.02.014

246. Garrick MD, Singleton ST, Vargas F, et al (2006) DMT1: which metals does it transport? Biol Res 39:79–85. https://doi.org/10.4067/s0716-97602006000100009

247. Somsuan K, Phuapittayalert L, Srithongchai Y, et al (2019) Increased DMT-1 expression in placentas of women living in high-Cd-contaminated areas of Thailand. Environ Sci Pollut Res Int 26:141–151. https://doi.org/10.1007/s11356-018-3598-2

248. Pi X, Jin L, Li Z, et al (2019) Association between concentrations of barium and aluminum in placental tissues and risk for orofacial clefts. Sci Total Environ 652:406–412. https://doi.org/10.1016/j.scitotenv.2018.10.262

249. Wang C, Pi X, Chen Y, et al (2021) Prenatal exposure to barium and the occurrence of neural tube defects in offspring. Sci Total Environ 764:144245. https://doi.org/10.1016/j.scitotenv.2020.144245

250. Zhang N, Liu Z, Tian X, et al (2018) Barium exposure increases the risk of congenital heart defects occurrence in offspring. Clin Toxicol (Phila) 56:132–139. https://doi.org/10.1080/15563650.2017.1343479

251. A. Gibson E, Nunez Y, Abuawad A, et al (2019) An overview of methods to address distinct research questions on environmental mixtures: an application to persistent organic pollutants and leukocyte telomere length. Environ Health 18:76. https://doi.org/10.1186/s12940-019-0515-1

252. Davis JM, Svendsgaard DJ (1990) U-shaped dose-response curves: their occurrence and implications for risk assessment. J Toxicol Environ Health 30:71–83. https://doi.org/10.1080/15287399009531412

253. Greenland S, Pearl J, Robins J (1999) Causal Diagrams for Epidemiologic Research. Epidemiology 10:37–48

254. Shrier I, Platt RW (2008) Reducing bias through directed acyclic graphs. BMC Medical Research Methodology 8:70. https://doi.org/10.1186/1471-2288-8-70

255. Bianco-Miotto T, Craig JM, Gasser YP, et al (2017) Epigenetics and DOHaD: from basics to birth and beyond. Journal of Developmental Origins of Health and Disease 8:513–519. https://doi.org/10.1017/S2040174417000733

256. Godfrey KM, Lillycrop KA, Burdge GC, et al (2007) Epigenetic Mechanisms and the Mismatch Concept of the Developmental Origins of Health and Disease. Pediatric Research 61:5–10. https://doi.org/10.1203/pdr.0b013e318045bedb

257. Baron RM, Kenny DA (1986) The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology 51:1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

258. Barker ED, Walton E, Cecil CAM (2018) Annual Research Review: DNA methylation as a mediator in the association between risk exposure and child and adolescent psychopathology. Journal of Child Psychology and Psychiatry 59:303–322. https://doi.org/10.1111/jcpp.12782

259. Lin VW, Baccarelli AA, Burris HH (2016) Epigenetics—a potential mediator between air pollution and preterm birth. Environ Epigenet 2:. https://doi.org/10.1093/eep/dvv008

260. Imai K, Keele L, Tingley D (2010) A general approach to causal mediation analysis. Psychological Methods 15:309–334. https://doi.org/10.1037/a0020761

261. VanderWeele TJ (2009) Mediation and mechanism. Eur J Epidemiol 24:217–224. https://doi.org/10.1007/s10654-009-9331-1

262. Bansal A, Simmons RA (2018) Epigenetics and developmental origins of diabetes: correlation or causation? American Journal of Physiology-Endocrinology and Metabolism 315:E15–E28. https://doi.org/10.1152/ajpendo.00424.2017

263. Saffery R (2014) Epigenetic Change as the Major Mediator of Fetal Programming in Humans: Are We There Yet? ANM 64:203–207. https://doi.org/10.1159/000365020

264. Liu Y, Aryee MJ, Padyukov L, et al (2013) Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol 31:142–147. https://doi.org/10.1038/nbt.2487

265. Meehan RR, Thomson JP, Lentini A, et al (2018) DNA methylation as a genomic marker of exposure to chemical and environmental agents. Current Opinion in Chemical Biology 45:48–56. https://doi.org/10.1016/j.cbpa.2018.02.006

266. Shenker N, Ueland P, Polidoro S, et al (2013) DNA Methylation as a Long-term Biomarker of Exposure to Tobacco Smoke. Epidemiology 24:712–716. https://doi.org/10.1097/EDE.0b013e31829d5cb3

267. Guerrero-Preston R, Goldman LR, Brebi-Mieville P, et al (2010) Global DNA hypomethylation is associated with in utero exposure to cotinine and perfluorinated alkyl compounds. Epigenetics 5:539–546. https://doi.org/10.4161/epi.5.6.12378

268. Ladd-Acosta C, Shu C, Lee BK, et al (2016) Presence of an epigenetic signature of prenatal cigarette smoke exposure in childhood. Environmental Research 144:139–148. https://doi.org/10.1016/j.envres.2015.11.014

269. Breton CV, Siegmund KD, Joubert BR, et al (2014) Prenatal Tobacco Smoke Exposure Is Associated with Childhood DNA CpG Methylation. PLOS ONE 9:e99716. https://doi.org/10.1371/journal.pone.0099716

270. Hernán MA, Hernández-Díaz S, Robins JM (2004) A Structural Approach to Selection Bias. Epidemiology 15:615. https://doi.org/10.1097/01.ede.0000135174.63482.43

271. Simmons SO, Fan C-Y, Yeoman K, et al (2011) NRF2 Oxidative Stress Induced by Heavy Metals is Cell Type Dependent. Curr Chem Genomics 5:1–12. https://doi.org/10.2174/1875397301105010001

272. Michels KB, Binder AM, Dedeurwaerder S, et al (2013) Recommendations for the design and analysis of epigenome-wide association studies. Nat Methods 10:949–955. https://doi.org/10.1038/nmeth.2632

273. Schulte PA, Perera FP (2012) Molecular Epidemiology: Principles and Practices. Academic Press

274. Mayeux R (2004) Biomarkers: Potential Uses and Limitations. NeuroRx 1:182–188

275. Udali S, Guarini P, Moruzzi S, et al (2013) Cardiovascular epigenetics: From DNA methylation to microRNAs. Molecular Aspects of Medicine 34:883–901. https://doi.org/10.1016/j.mam.2012.08.001

276. Goud Alladi C, Etain B, Bellivier F, Marie-Claire C (2018) DNA Methylation as a Biomarker of Treatment Response Variability in Serious Mental Illnesses: A Systematic Review Focused on Bipolar Disorder, Schizophrenia, and Major Depressive Disorder. International Journal of Molecular Sciences 19:3026. https://doi.org/10.3390/ijms19103026

277. Mikeska T, Craig JM (2014) DNA Methylation Biomarkers: Cancer and Beyond. Genes 5:821–864. https://doi.org/10.3390/genes5030821

278. Chu T, Burke B, Bunce K, et al (2009) A microarray-based approach for the identification of epigenetic biomarkers for the noninvasive diagnosis of fetal disease. Prenatal Diagnosis 29:1020–1030. https://doi.org/10.1002/pd.2335

279. Küpers LK, Monnereau C, Sharp GC, et al (2019) Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight. Nat Commun 10:1–11. https://doi.org/10.1038/s41467-019-09671-3

280. Levenson VV (2010) DNA methylation as a universal biomarker. Expert Review of Molecular Diagnostics 10:481–488. https://doi.org/10.1586/erm.10.17

281. Schisterman EF, Cole SR, Platt RW (2009) Overadjustment Bias and Unnecessary Adjustment in Epidemiologic Studies. Epidemiology 20:488–495. https://doi.org/10.1097/EDE.0b013e3181a819a1

282. Fogarty NME, Burton GJ, Ferguson-Smith AC (2015) Different epigenetic states define syncytiotrophoblast and cytotrophoblast nuclei in the trophoblast of the human placenta. Placenta 36:796–802. https://doi.org/10.1016/j.placenta.2015.05.006

283. Fogarty NME, Ferguson-Smith AC, Burton GJ (2013) Syncytial knots (Tenney-Parker changes) in the human placenta: evidence of loss of transcriptional activity and oxidative damage. Am J Pathol 183:144–152. https://doi.org/10.1016/j.ajpath.2013.03.016

284. Zheng SC, Breeze CE, Beck S, Teschendorff AE (2018) Identification of differentially methylated cell-types in Epigenome-Wide Association Studies. Nat Methods 15:1059–1066. https://doi.org/10.1038/s41592-018-0213-x