Chapter 3 - Combining single-cell transcriptomics with whole genome sequencing and bulk RNA-sequencing to improve cell-type annotations

3.1 Overview

A primary goal of this study is to generate a single-cell atlas of the maternal-fetal interface by comprehensively and robustly profiling the identities and potential functions of its constituent cell types and cell states. Beyond annotating cells by their gene expression patterns as detected by scRNA-seq, additional analyses utilizing other sequencing approaches are important for validating and further describing the identified cell populations. In this chapter, I first present our approaches for inferring the maternal or fetal genetic origin of single cells from decidua and placenta scRNA-seq data. I then discuss our use of bulk RNA-sequencing data to assess whether the detected gene expression signature of a specific cell population from scRNA-seq is a potential artifact attributable to single-cell sequencing protocols rather than a reflection of true biological variability.

3.2 Maternal-fetal single-cell genotyping

3.2.1 Droplet-based scRNA-seq and whole genome sequencing data

Previously, 37239 cells derived from four decidual and two placental samples were sequenced using droplet-based scRNA-seq (10x Genomics Chromium[110]) (sample information in **Table 1**). Unsupervised graph-based clustering was applied to the data, and marker genes from Appendix 2 were used to annotate clusters as specific cell populations (**Figure 1**). Overall, 28 cell states were predicted by this approach, including populations of decidual fibroblasts, glandular epithelial cells, endothelial cells, dNK cells, decidual myeloid cells, and decidual T cells, as well as subsets of trophoblasts, fetal fibroblasts, and Hofbauer cells (fetal macrophages).

Table 1. Droplet-based scRNA-seq and whole genome sequencing (WGS) sample information.

Donor	Post-conceptional weeks	Pregnancy	10x Chromium - Decidua	10x Chromium - Placenta	WGS - Tissues	Fetus sex
D1	9 to 11	2nd	6514 cells		Fetal placenta / Maternal blood	female
D2	6 to 8	1st	6546 cells		Fetal placenta / Maternal blood	female
D3	5 to 7	6th	8070 cells	5510 cells	Fetal placenta / Maternal decidua	male
D4	9 to 11	3rd	8860 cells	2029 cells	Fetal placenta / Maternal decidua	male

For each of the decidual and placental samples, genomic DNA from the corresponding mother and fetus were sequenced at 30x coverage (see Methods). Information about the genomic DNA samples

and fetus sex, as determined by filtered SNP calls on the sex chromosomes, are also described in **Table 1**.

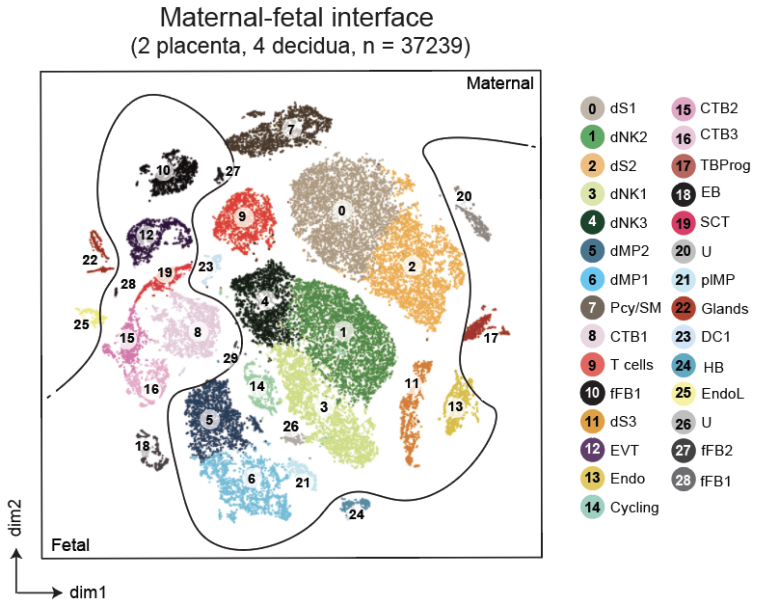

Figure 1. t-SNE visualization of cell populations at the maternal-fetal interface. Cells are colored by predicted cell type or cell state, as inferred by unbiased clustering of decidua and placenta droplet-based scRNA-seq data and annotation based on known markers (from Vento-Tormo R, Efremova M et al., under submission). Legend for cell population abbreviations: CTB = cytotrophoblast cells; EVT = extravillous trophoblast cells; SCT = syncytiotrophoblast cells; TBProg = trophoblast progenitor cells; dS = decidual stromal cells; DC = dendritic cells; dMP = decidual mononuclear phagocytes; plMP = placental mononuclear phagocytes; EndoL = lymphatic endothelial cells; Endo = endothelial cells; Tcell = T cells; dNK = decidual NK cells; ILC = innate lymphoid cells; fFB = fetal fibroblast cells; Glands = decidual glandular cells; HB = Hofbauer cells; Pcy/SM = pericytes/vascular smooth muscle cells.

3.2.1 Limitations in unbiased assignment of maternal/fetal origin to cells without genotype data

In order to evaluate the distribution of maternal and fetal cell populations in the decidua and placenta, we sought to independently determine the maternal or fetal genetic origin of all cells sequenced from these tissues. We first assessed whether it was possible to distinguish maternal-derived and fetal-derived cells based solely on variant calling from decidual and placental scRNA-seq data. Due to runtime constraints, we only called known SNPs from the reads in each barcoded cell that were mapped to the 1000 most highly expressed genes in placenta and endometrium, as determined from Human Protein Atlas[127,128] tissue-level gene expression data (see Methods). We then performed a probabilistic PCA (PPCA) on the called SNPs among all decidual and placental cells to assess whether cells originating from the mother would be distinguishable from cells originating from the fetus. While PPCA yielded separation between a fraction of the decidual cells, which are generally maternal in origin, and

the placental cells, which are generally fetal in origin, it failed to separate a majority of the remaining cells by tissue or expected individual of origin (**Figure 2a**).

We reasoned that the inherent sparsity of scRNA-seq data[148,149], the limited sequence information extractable from 3' end sequencing protocols such as the 10x Chromium droplet-based scRNA-seq[110], and the intuition that different cell types often exhibit divergent sets of expressed genes, all impede the ability to call a sufficient number of SNPs that are both consistently detectable across many analyzed cells and informative for genetically differentiating mother from fetus. Indeed, we observed that a significantly higher proportion of the genes sampled for variant calling were expressed in decidual non-immune cells and placental cells relative to decidual immune cells ($p < 2 \times 10^{-16}$) (**Figure 2b**). This suggests that fewer informative SNPs were callable in decidual immune cells, thus hindering the ability to distinguish maternal or fetal genotypes in these cells. We therefore then performed PPCA on a subset of non-immune cells which were previously inferred by single-cell transcriptome-based analyses to be maternal stromal cells, fetal trophoblast cells, and an unknown cell population (cluster 21) found in both decidua and placenta (**Figure 2c**). The PPCA reveals apparent separation between the maternal and fetal cells and places the unknown population closer to the fetal trophoblast population, suggesting that this approach is able to detect interindividual genetic heterogeneity among single cells belonging to specific cell types. However, we show that if the PPCA is repeated using a decidual immune cell population, such as the maternal macrophages, in place of maternal stromal cells, the qualitative separation between maternal and fetal cells is no longer apparent (**Figure 2d**). Overall, our implementation of an scRNA-seq-only variant calling and single-cell genotyping pipeline appears to be able to qualitatively distinguish certain cell types at the maternal-fetal interface by individual of origin but requires further development to be able to fully resolve maternal and fetal identities among all decidual and placental cells in an unbiased, cell type-independent manner.

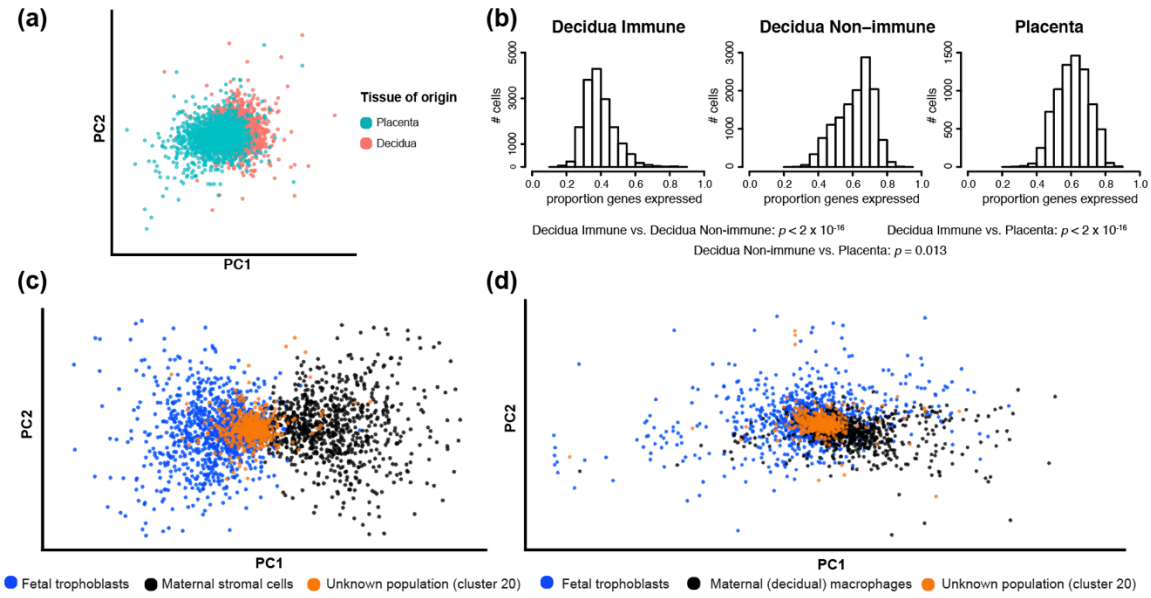

Figure 2. Variant calling from scRNA-seq data alone shows limitations in inferring the maternal or fetal genetic origin of cells from the decidua and placenta. (a) Probabilistic PCA (PPCA) on 19343 SNPs called in cells from Donor 4 yielded substantial overlap between maternal decidua-derived cells and fetal placenta-derived cells, suggesting limited ability to discriminate cells by tissue/individual of origin. **(b)** The 1000 most abundant genes in placenta and endometrium, which contain the SNP sites analyzed by the variant calling pipeline, are expressed in significantly different proportions between decidual immune cells, decidual non-immune cells, and placental cells (pairwise Student's t-test, Benjamini-Hochberg adjusted *p*-values). **(c)** PPCA of SNP calls in cells from three non-immune ($CD45^-$) subsets yields separation between maternal stromal and fetal trophoblast cells and suggests that the unknown cell population is fetal in origin. **(d)** PPCA of SNP calls in cells from both immune and non-immune populations is unable to qualitatively separate fetal trophoblast cells, maternal macrophages, or the unknown cells.

3.2.2 Comprehensive and unbiased assignment of single-cell genetic origin, leveraging demuxlet algorithm and WGS of maternal and fetal genomic DNA

Next, we assessed whether the incorporation of maternal and fetal WGS data would enable us to determine the genetic identity of each cell from the scRNA-seq data in an unbiased fashion. We used the recently published demuxlet tool[125], which leverages inter-individual genetic variation, as determined from provided genotype data, to infer the most likely origin of each cell in scRNA-seq experiments containing cells derived from distinct individuals. First, we performed alignment and variant calling on maternal and fetal WGS datasets corresponding to the decidua and placenta droplet-based scRNA-seq samples (see Methods; **Table 1**). We then used demuxlet to analyze the scRNA-seq data in conjunction with the maternal and fetal genotypes, in order to correlate detectable SNPs overlapping reads from decidual and placental cells with SNPs from maternal and fetal genomic DNA and probabilistically assign each cell as being maternal or fetal in origin. To visualize the results, we overlaid the maternal/fetal assignments from demuxlet onto a t-SNE representation of the decidual and placental

single-cell transcriptome data, facilitating direct comparison between annotated cell populations (**Figure 3a**), tissue of origin (**Figure 3b**), and demuxlet-inferred genetic origin (**Figure 3c**). Demuxlet assignments for each cluster or cell type are largely consistent with the known biology of these cells. For example, dNK cell populations (clusters 1, 3, 4), which are known to make up the largest maternal immune population at the maternal-fetal interface[2], were inferred to be maternal cells, and the trophoblast cells (clusters 8, 12, 15-17, 19), which comprise the growing placenta, were identified to be fetal cells. Additionally, as expected, cells originating from the decidua were largely determined to be of maternal origin, and cells from the placenta of fetal origin. One notable exception is the small fraction of decidua-derived EVT (cluster 12) identified to be fetal cells, which aligns with the observation that they invade the pregnant endometrium[150]. Meanwhile, the unknown cell population found in both decidual and placental samples (cluster 20) was determined to be fetally derived, consistent with what was qualitatively suggested by our DNA-free genotyping pipeline results (**Figure 2c**). Interestingly, a myeloid subset found exclusively in both placenta samples (cluster 21) and initially hypothesized to be a fetal macrophage population was inferred by demuxlet to be maternal in origin. These cells potentially correspond to patrolling maternal monocytes attached to the placental villi as previously described[151].

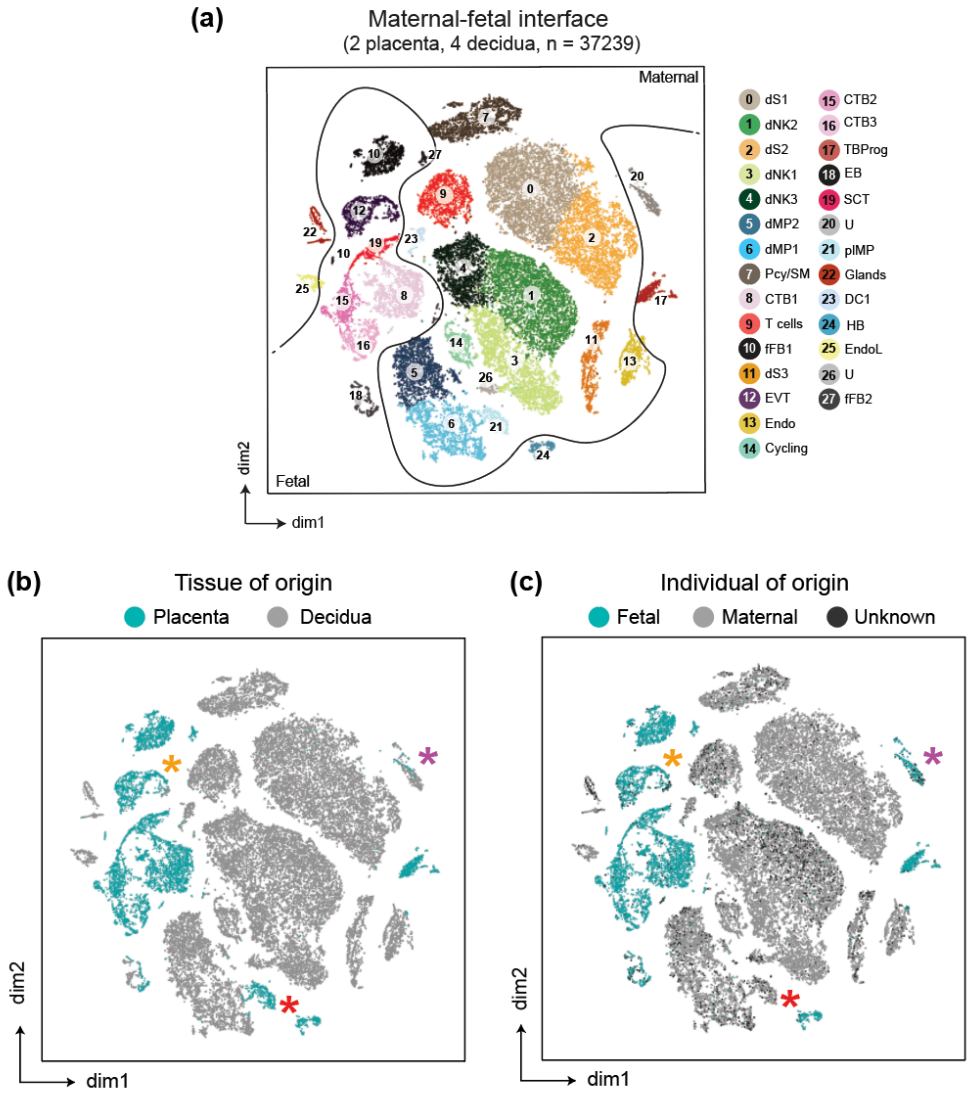

Figure 3. t-SNE visualizations juxtaposing cell type annotations, tissue of origin, and assignment of maternal and fetal identity. (a) Figure 1 reproduced for reference; cells colored by marker-based annotation. **(b)** Cells colored by tissue of origin (decidua or placenta). **(c)** Cells colored by demuxlet inference of origin (maternal, fetal, or unassigned). The orange star labels EVT found in the decidua that were determined to be of fetal origin. The purple star labels the unannotated cell population found in both decidua and placenta determined to be of fetal origin. The red star labels the mononuclear phagocytes derived from the placenta determined to be of maternal origin.

Demuxlet uses a maximum likelihood approach to infer the best-matching genotype and sample identity of each cell[125]. Thus, we further assessed the likelihood differences between assignment to fetal origin and assignment to maternal origin for all analyzed cells (**Figure 4a**). Although likelihood differences vary between individual cells, the 92.8% of cells with conclusively inferred maternal or fetal identity are on average highly confidently assigned. However, we observed that many of the cells with a larger number of quality control (QC)-passed scRNA-seq reads were called with lower confidence than their

counterparts with fewer QC-passed reads. The placental macrophages identified to be maternal in origin were among the cells with fewer QC-passed reads (**Figure 4b**), and we confirmed that confidence in their assignment did not significantly differ from that of other assigned maternal cells (p = 0.2638, Student's t-test).

Next, we evaluated the relative proportions of assigned fetal, maternal, and unknown cells in each of the annotated cell types (**Figure 4c**). Although the percentages of cells with unknown identity varied between cell types, a sufficient number of cells were conclusively identified within each cell type to infer a consensus fetal or maternal origin for each population. The sole exception is the cluster corresponding to erythrocytes and erythroblasts (EB), which contains significant proportions of both maternal and fetal cells, suggesting the presence of both maternal and fetal blood in the tissue samples. As all other documented cell types in the decidua and placenta are known to be exclusively maternal or fetal in origin, our results thus corroborate the single-cell transcriptome-based clustering and cell type annotations at the genetic level. Overall, our results suggest that demuxlet can be applied to confidently infer the genetic identities of individual cells in scRNA-seq samples of mixed maternal and fetal origin. Furthermore, we show that the single-cell genotyping results may in themselves reveal additional biological insights not immediately apparent from purely transcriptome-based analyses.

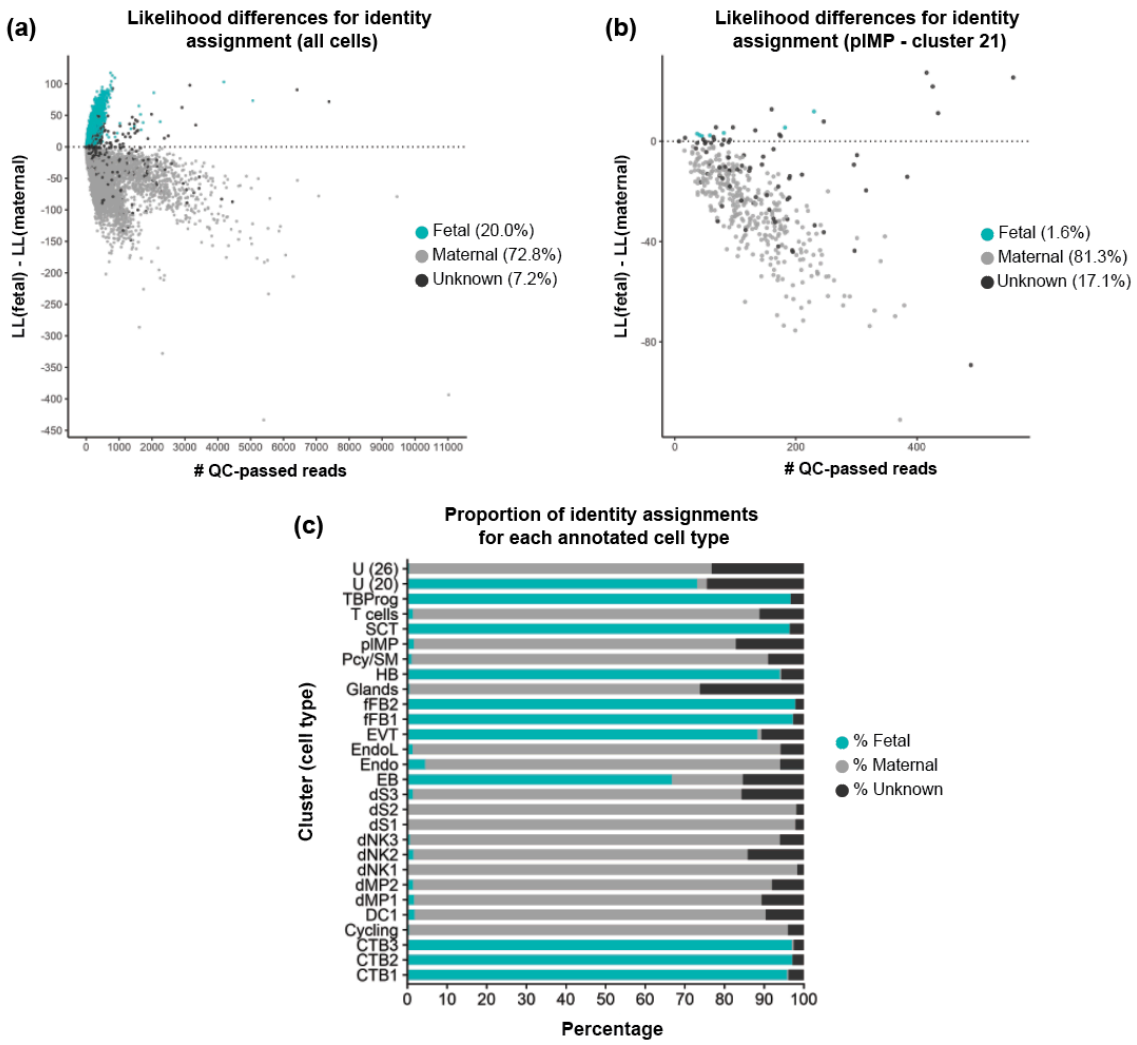

Figure 4. Assessment of demuxlet performance in genotyping all cell populations at the maternal-fetal interface. (a) Log-likelihood differences between assignment to fetal origin or assignment to maternal origin for all cells based on demuxlet SNP calling from the droplet-based scRNA-seq data. Cells are colored by their final origins as determined by demuxlet. **(b)** Log-likelihood differences between assignment to fetal or maternal origin for the maternal mononuclear phagocytes (pIMP) from the placenta. pIMP likelihood differences do not significantly differ from the likelihood differences associated with all other demuxlet-assigned maternal cells (p = 0.2638, Student's t-test). **(c)** Percentages of cells inferred to be of fetal, maternal, or unassigned origin for each annotated cell population.

3.3 Bulk RNA-sequencing evaluation of putative dissociation protocol-associated genes

Initial clustering and annotation of all plate-based scRNA-seq of maternal decidua cells identified a putative decidual T cell population (T3) with many of the most highly upregulated genes belonging to various heat shock protein (HSP) families[152] (**Figure 5a-b**, **Table 2**).

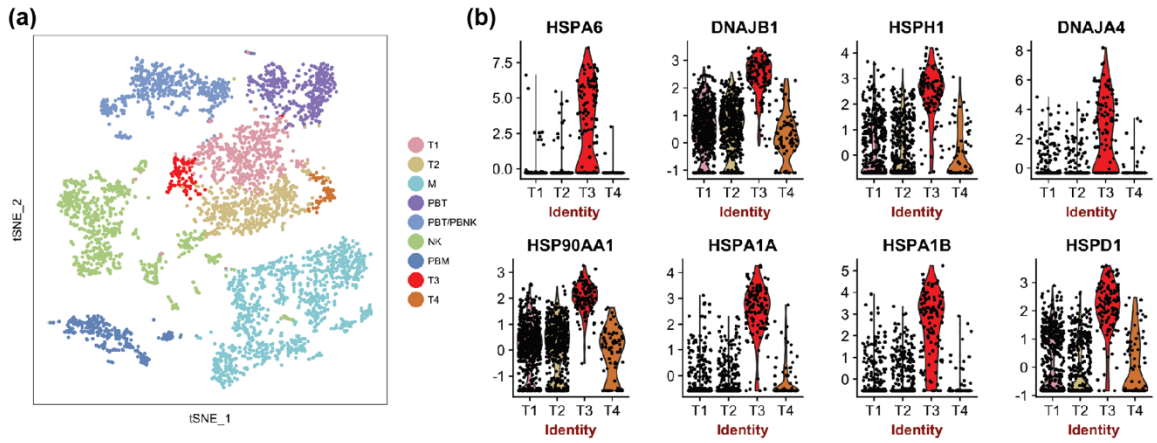

Figure 5. Identification of T cell subpopulation exhibiting significant upregulation of heat shock protein (HSP) genes from decidua plate-based scRNA-seq data. (a) t-SNE mapping of decidual cell populations as inferred from unbiased clustering, with the T cell subpopulation of interest (T3) in red. **(b)** Violin plot comparing the expression of the HSP genes comprising the top 10 most upregulated genes in T3 across all T cell subpopulations.

Table 2. Gene identifiers, fold changes, and *p*-values for the HSP genes of interest in T3. (Genes in blue were previously shown to be induced following tissue dissociation[109].)

Gene name	Ensembl gene ID	p value	ave $\log_2$(FC)	adjusted p value
HSPA6	ENSG00000173110	2.70E-69	3.043	8.17E-65
DNAJB1	ENSG00000132002	6.34E-68	2.910	1.92E-63
HSPH1	ENSG00000120694	1.17E-70	2.377	3.53E-66
DNAJA4	ENSG00000140403	1.08E-56	2.359	3.27E-52
HSP90AA1	ENSG00000080824	3.63E-67	2.286	1.10E-62
HSPA1A	ENSG00000204389	2.39E-78	2.213	7.23E-74
HSPA1B	ENSG00000204388	3.75E-68	2.160	1.14E-63
HSPD1	ENSG00000144381	1.32E-61	2.137	4.00E-57
DNAJA1	ENSG00000086061	6.91E-50	2.051	2.09E-45
HSPA8	ENSG00000109971	2.79E-55	1.991	8.46E-51
HSPB1	ENSG00000106211	1.17E-62	1.108	3.54E-58

A recent study demonstrated that HSPs are among the stress-induced genes in certain subpopulations of cells as a consequence of the tissue dissociation protocols used to extract cells for scRNA-seq[109]. In light of these findings, we evaluated the expression of the upregulated HSP genes in T3 cells in a set of bulk RNA-sequencing data generated on the same decidua samples. As single-cell tissue dissociation is not a necessary component of bulk RNA-sequencing experiments, we reasoned that comparing detectable genes between scRNA-seq and corresponding bulk RNA-seq datasets would enable us to account for the potential effects of these protocols. Upon performing alignment and gene quantification in the decidua bulk RNA-seq samples (see Methods), we specifically examined the expression levels of the HSP genes below, which were either highly upregulated in the T3 subpopulation or implicated in the single-cell dissociation protocol study (**Table 3**).

Table 3. Expression levels of the HSP genes from Table 2 in bulk decidua RNA-seq data (tpm).

Gene name	Ensembl gene ID	Donor 5	Donor 6	Donor 7	Donor 8	Donor 9
HSPA6	ENSG00000173110	0.51	1.43	1.95	1.81	1.04
DNAJB1	ENSG00000132002	13.64	17.67	12.01	15.77	14.66
HSPH1	ENSG00000120694	6.52	6.20	9.23	8.39	6.80
DNAJA4	ENSG00000140403	0.61	1.39	2.32	1.64	2.04
HSP90AA1	ENSG00000080824	39.55	31.69	40.28	49.02	38.08
HSPA1A	ENSG00000204389	0.00	2.32	0.90	3.92	10.80
HSPA1B	ENSG00000204388	0.00	0.00	0.00	0.00	0.04
HSPD1	ENSG00000144381	14.52	13.26	17.50	20.31	16.57
DNAJA1	ENSG00000086061	15.45	9.00	19.87	20.63	18.02
HSPA8	ENSG00000109971	282.95	134.65	289.29	305.98	228.79
HSPB1	ENSG00000106211	118.41	110.06	128.27	155.58	131.52

Overall, the HSP genes are also detectable in the bulk RNA-seq samples, although many are expressed at low levels. Because T cells comprise a small proportion of decidual cells[90], it is conceivable that these genes could be attributed to a particular T cell subpopulation in the decidua. However, we observed that *HSPA1B*, the 7th most highly upregulated gene in T3 decidual T cells, is not significantly expressed in any of the bulk decidua samples, suggesting that the genes found to be most significantly enriched in T3 are likely to be influenced by single-cell dissociation protocols rather than reflect a functionally distinct T cell subpopulation. Thus, we excluded the T3 cells from subsequent single cell-level annotations and analyses. These results suggest that evaluating the expression of a gene of interest in both scRNA-seq and corresponding bulk RNA-seq can serve as a rough heuristic for determining whether it is a potential dissociation protocol-related artifact.

3.4 Discussion

In this chapter, I evaluated different approaches for genetically distinguishing maternal and fetal cells in the decidua and placenta in an unbiased manner. I first presented our approach for inferring the maternal or fetal origin of cells based on variant calling on the scRNA-seq reads alone. This method presented limitations in the unbiased determination of fetal and maternal cells that I overcame by applying an established algorithm, demuxlet[125], on the scRNA-seq data in combination with maternal and fetal genotypes obtained from WGS.

Our results suggest that further optimizations are required to confidently perform maternal and fetal cell assignment based on allele calling from scRNA-seq data alone. One initial barrier to considering a larger pool of callable variants or performing *de novo* variant calling was that the runtime of variant calling using GATK HaplotypeCaller was found to be prohibitive. We therefore limited our analysis to

SNPs overlapping the 1000 genes most highly expressed in placenta and endometrium (see Methods). This approach yielded separation between a fraction of fetal and maternal cells, but failed to distinguish the identities of the remaining majority (**Figure 2a**). To evaluate whether the results are influenced by the cell types analyzed, we repeated the analysis on specific subsets of cells established with high confidence to be maternal and fetal in origin. For example, probabilistic PCA on a combination of maternal stromal cells, fetal trophoblast cells, and cells of unknown origin (cluster 20) from Donor 4 produced qualitative separation between the trophoblast and stromal cells and clustering of the unknown cells with the trophoblasts (**Figure 2c**). This was corroborated by the subsequent demuxlet output and was consistent with our previous knowledge regarding the identities of these cells. However, the same analysis performed on a combination of maternal mononuclear phagocytes, fetal trophoblast cells, and the unknown cells was unable to separate these populations (**Figure 2d**). This suggests that DNA-free single-cell genotyping is limited at least in part by cell type-dependent differences in expressed genes, and in turn, the sites of callable SNPs or other genetic variants. Two cells originating from the same individual may exhibit disparate gene expression profiles if they belong to different cell types, and as a result, they may not share a sufficient number of identifiable SNPs in common to both confidently relate them to each other and differentiate them from cells originating from another individual. This is further supported by the observation that a similar application of this pipeline to calling variants overlapping the 1000 most highly expressed genes in iPSCs performs well at distinguishing iPSCs derived from different human donors that were pooled and profiled by droplet-based scRNA-seq (Teichmann group, unpublished data). In contrast to our study, in which the cells have originated from tissues composed of heterogeneous cell types, the iPSCs were all derived from skin fibroblast cell cultures generated by the Human Induced Pluripotent Stem Cell Initiative (HipSci) resource. Future attempts to improve this pipeline for maternal/fetal identification may evaluate whether increasing the number of called SNPs improves its ability to separate genetically distinct cells, and whether alternative variant callers improve the runtime of this approach. Additionally, best practices for variant calling from RNA-seq data, not to mention scRNA-seq data, are still in development[153]. For example, the filtering parameters applied to called variant alleles and homozygous reference alleles may also merit optimizing.

I next presented our results using the demuxlet algorithm to perform unbiased maternal/fetal assignment of all cells sequenced via droplet encapsulation from the decidua and placenta (**Figure 3**).

I show that demuxlet confidently inferred the identities of ~93% of these cells, and that the assignments were largely consistent with our expectations of the maternal or fetal origins of these cells, given their derivation from either the decidua or placenta or their annotated cell types as determined by unbiased clustering and examination of marker genes from the known literature. For example, demuxlet successfully resolved the small number of invading EVT of fetal origin which were derived from the decidua samples. Demuxlet was also able to clarify or revise our understanding of other cell populations. One such population is cluster 21, corresponding to a mononuclear phagocyte subset localized to the placenta in both of the sequenced samples (**Figure 3b**), which was inferred by demuxlet to be maternal in origin (**Figure 3c**). This represents a particularly interesting finding as previous literature largely reports that the placental myeloid cells are primarily fetal Hofbauer cells[154,155]. A previous study, however, observed the recruitment and attachment of maternal mononuclear phagocytes to placental trophoblast cells[151]. In particular, these cells accumulate on the syncytial surface of the placenta, at sites where the placental layer has withstood damage. This mechanism must be precisely regulated, as increased placental accumulation of macrophages is observed in cases of villitis or other inflammatory conditions[156,157]. These findings suggest that the presence of maternal myeloid cells in the placenta is not a misassignment of single-cell genetic identities or attributable to sample contamination, but rather a bona fide reflection of maternal macrophage biology. Additionally, for cluster 20, a population observed in both the decidua and placenta which was unable to be classified as any known cell type based on transcriptome analysis, demuxlet identified a large majority of these cells to be fetal in origin (**Figure 3c**), which may be useful in guiding further attempts to annotate these cells or determine whether they may represent potential contamination from non-placental fetal tissues. Considering the demuxlet assignments and their likelihoods in aggregate, we noticed that some cells with more QC-passed reads were assigned maternal/fetal identities with lower confidence and that in some cell types, the proportions of cells with unknown identities were much higher than in others (**Figure 4a**). We believe this may be attributed to the fact that in these cells, fewer scRNA-seq reads overlap the variants which are most informative for genetically distinguishing mother and fetus. Since the callable variants in each cell are dependent on which genes the cell expresses and are detectable in the scRNA-seq data, this could result in some cell types being systematically genotyped with lower confidence than others, rather than a more stochastic distribution of confidence levels and unknown assignments across all cell types. Indeed, we observed that the vast majority of cells with high QC-

passed reads and lower confidence of assignment originated from sequencing runs consisting of immune cells isolated from two of the maternal decidua samples. Nevertheless, it was still possible to assign origins for these cells with a certain degree of confidence, given that demuxlet infers single cell identities by using genotypes as an independent, "ground truth" genetic reference. Thus, while the primary envisioned application of demuxlet may be to facilitate the demultiplexing of scRNA-seq experiments deliberately pooling cells from genetically different donors[125], our analysis in the context of the maternal-fetal interface demonstrates that this algorithm can also prove useful in scenarios in which cells from distinct genetic origins biologically coexist in a tissue in different, potentially variable quantities and collectively encompass distinct cell types, and in which the proportions of cells derived from each individual are not known *a priori* and the individuals of origin may themselves be genetically related.

In summary, unbiased genetic inference of the maternal or fetal origin of single cells in the decidua and placenta is associated with several challenges: i) cells exhibiting different transcriptomic profiles must be analyzed in tandem; ii) the mother and fetus share at least 50% identical genetic material; and iii) the cells originating from one individual may be present in very small proportions in a given tissue (e.g. EVT in the decidua). Overall, with the availability of maternal and fetal genotype data, algorithms such as demuxlet remain the most robust means of overcoming these challenges and genetically identifying cells based on single-cell transcriptomes. Ultimately, however, a means of performing accurate, unbiased genetic identification without the necessity of additional genotype or genomic information would be desirable, particularly as the number of samples and tissues to be profiled through scRNA-seq continues to grow.

Lastly, I used bulk RNA-seq to identify potential dissociation protocol-induced artifacts in the scRNA-seq data and prevent the misinterpretation of detected genes in certain cell populations. In this case, we showed how *HSPA1B*, a highly upregulated gene identified by plate-based SmartSeq2 in a T cell population (**Figure 5b**; **Table 2**), was absent in the bulk RNA-seq data from the same samples (**Table 3**), suggesting that the significantly upregulated markers in these cells may be attributable to technical rather than biological noise. At the same time, we still observed expression of other HSP-related genes in the bulk RNA-seq data (**Table 3**), suggesting that bulk RNA-seq sample processing is also capable of imposing stress on cells and that RNA-seq protocols in general are capable of introducing biases in observed gene expression[158]. Nevertheless, comparative analyses between scRNA-seq and matched bulk RNA-seq data still represent a valid strategy to correct for gene expression perturbations

specifically attributable to single-cell tissue dissociation protocols. One approach for detecting scRNA-seq-associated transcriptome alterations in a more comprehensive manner might be to perform a linear regression on the detected gene signatures from the single-cell data to determine which weights best fit the composite, population-level gene expression signals observed in the bulk RNA-seq data, and then determine which genes either from the scRNA-seq data account for the residual or distance in gene expression between these two datasets. Repeating this for different samples and across different tissues might be a potential way to generate a consensus list of genes specifically susceptible to perturbation in scRNA-seq protocols. However, due to the differential enrichment of cell types during tissue processing (e.g. depletion of B cells and epithelial cells from the decidua), a limitation of this approach is that the altered proportions of captured cells in scRNA-seq datasets relative to bulk RNA-seq datasets represents a confounding factor that would make detected residuals more difficult to interpret in an unbiased fashion. Moreover, many of the genes associated through this method with single-cell dissociation might harbor relevant biological functions that would be masked if their upregulation in cells were discounted or fully attributed to protocol-related effects. For example, HSP-family proteins including hsp70 (*HSPA1A/B*) and gp96 are involved in chaperoning peptide complexes for uptake, processing, and presentation by antigen-presenting cells[159]. In addition, HSPs themselves have been shown to modulate T cell responses in chronic inflammatory disease[160].

Conversely, another approach could involve applying computational deconvolution methods to the bulk RNA-seq data to obtain estimates of cell-type specific gene expression profiles, which would be compared in turn with the scRNA-seq datasets. However, several limitations exist here as well. Many existing deconvolution algorithms, including CIBERSORT[161], TIMER[162], and MCP-counter[163], were specifically developed to infer proportions of infiltrating immune cell subsets in tumors. Many currently available methods are also tailored towards deconvolution of microarray data rather than next-generation sequencing data[161,164], or are limited to estimating the enrichment or proportions of cell types in tissue of interest, but not actual cell type-specific gene expression patterns[165,166]. Moreover, these methods may often produce biases in deconvolving mixtures of similar cell types harboring correlated gene expression profiles (e.g. naive vs. activated T cells)[162,166]. For these reasons, these methods were not applied to our decidual and placental bulk RNA-seq data but could nevertheless be evaluated and represent a potential new avenue of study and development.