



NUI Galway
OÉ Gaillimh



Characterising transcriptional variation between tissue-resident macrophage subsets using single-cell RNA-seq

A thesis submitted

by

Sarah Ennis

to

The Discipline of Bioinformatics,
School of Mathematics, Statistics & Applied Mathematics
National University *of* Ireland, Galway

in partial fulfilment of the requirements for the degree of

M.Sc. in Biomedical Genomics

19th July 2019

Thesis Supervisors:

Dr. Pilib Ó Broin, Dr. Daniel Gaffney, Dr. Sarah Teichmann

Original Paper

Characterising transcriptional variation between tissue-resident macrophage subsets using single-cell RNA-seq

Sarah Ennis^{1,*}, Tomás Gomes², Sarah A. Teichmann², Daniel Gaffney² and Pilib Ó Broin¹,

¹School of Mathematics, Statistics and Applied Mathematics, National University of Ireland, Galway, H91 TK33, Ireland and

²Department of Cellular Genetics, Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, United Kingdom.

*To whom correspondence should be addressed.

Abstract

Motivation: Tissue-resident macrophages are important cells of the innate immune system that display marked heterogeneity in their cellular phenotype and functional roles within tissues. This heterogeneity is partly a consequence of divergent tissue-specific transcriptional regulation and is likely reflected in the gene expression profile of cells from different tissues. Here we compiled single-cell RNA-seq data from human studies in more than 15 tissues and compared the transcriptomes of resident macrophage populations from each tissue.

Results: In total we identified 72,860 macrophages in 15 different tissues. These macrophages showed differential expression of 7,189 genes including previously known markers such as CX3CR1 for brain-resident macrophages and CCL18 for lung-resident macrophages. Differentially expressed genes showed enrichment for immune processes and reflect how resident macrophages in different tissues respond to signals in their environment and carry out their immune effector functions.

Supplementary information: All code used to process cells is available on github: www.github.com/Sarah145/MSc_Project. The cross-tissue macrophage dataset can be explored using an interactive web portal: <http://tiny.cc/cross-tissue-macrophages>.

Contact: s.ennis6@nuigalway.ie

1 Introduction

Tissue-resident macrophages are cells of the innate immune system that perform tissue-specific functions. These cells play an important role in homeostasis, immune surveillance and defence against invading pathogens and display marked heterogeneity in their morphology, functions and overall cell phenotype. The source of the variation seen among these cells cannot be explained by simplified models of macrophage activation and is more likely a consequence of unknown, divergent developmental processes (Davies *et al.*, 2013). Originally, tissue-resident macrophages were thought to be derived from monocytes in the blood that migrate into tissues and mature into tissue-resident cells but more recently it has been shown that the majority of resident-macrophage populations are established early during development from progenitors in the yolk sac and fetal liver (Mass *et al.*, 2016). The fact that these populations have been subject to tissue-specific transcriptional regulation from an early stage means that there is likely

a large amount of variation in the transcriptomes of cells from different tissues.

Importantly dysregulation of tissue-resident macrophages underpins many human diseases which means they represent potential therapeutic targets and require systematic and unbiased characterisation. Unfortunately, the fact that these cells are embedded in tissue makes them particularly difficult to access and, as a consequence, they are poorly characterised. However, advances in single-cell sequencing technology have recently made it possible to characterise variation within and among the diverse cells of the innate immune system on an unprecedented scale (Svensson *et al.*, 2018). These studies have led to fascinating insights, including the discovery of new subtypes of cells, and shed light on the source of the vast heterogeneity present among immune cells (Vento-Tormo *et al.*, 2018). The goal of this project was to assemble a cross-tissue collection of macrophages using single-cell transcriptome data from existing published studies and unpublished data generated as part of the Human Cell Atlas initiative. Table 1 details the datasets included in this study.

Table 1. Human single-cell RNA-seq datasets used for the analysis.

Reference	Tissue	Protocol	# of Cells
Popescu <i>et al.</i> , 2019	Fetal Liver, Fetal Skin	10X	273,856
Zheng <i>et al.</i> , 2017	Blood	10X	163,234
Henry <i>et al.</i> , 2018	Prostate	10X	109,061
Vento-Tormo <i>et al.</i> , 2018	Blood, Decidua, Placenta	10X, SS2	70,325
Young <i>et al.</i> , 2018	Kidney	10X	48,347
Sohni <i>et al.</i> , 2019	Testis	10X	34,729
Unpublished	Colon, mLN	10X	32,228
Vieira Braga <i>et al.</i> , 2019	Lung Parenchyma, Upper Airway	10X	26,013
Guo <i>et al.</i> , 2018	Testis	10X	12,985
Zhang <i>et al.</i> , 2018	Blood, Colon, Tumour	SS2	11,138
Menon <i>et al.</i> , 2018	Kidney	DropSeq	9,840
Baron <i>et al.</i> , 2016	Pancreas	inDrop	8,569
Gierahn <i>et al.</i> , 2017	Blood	SeqWell	5,584
Masuda <i>et al.</i> , 2019	Brain (Microglia)	CEL-Seq2	4,408
Nowakowski <i>et al.</i> , 2017	Brain	C1	4,261
La Manno <i>et al.</i> , 2016	ES Cells, Brain	C1	3,712
Segerstolpe <i>et al.</i> , 2016	Pancreas	SS2	3,363
Muraro <i>et al.</i> , 2016	Pancreas	CEL-Seq2	2,126
Li <i>et al.</i> , 2019	Intestine	10X	1,886
Miragaia <i>et al.</i> , 2019	Blood, Colon, Skin	SS2	1,168
Wang <i>et al.</i> , 2016	Pancreas	SMARTer	635
Total			827,448

mLN, Mesenteric lymph node; ES cells, Embryonic stem cells; 10X, 10X Genomics; SS2, SmartSeq2; C1, Fluidigm C1

2 Methods

2.1 Preparation of macrophage dataset

To compile this cross-tissue collection of macrophages, the guidelines outlined by Luecken and Theis (2019), in a recent benchmark study of best practices in single-cell RNA-seq (scRNA-seq) data analysis were followed. Given the size of the dataset (>820,000 cells) and the efficiency of Python in processing large datasets, the majority of the analyses were implemented with SCANPY v1.4.3, a Python-based toolkit for analyzing single-cell gene expression data (Wolf *et al.*, 2018).

2.1.1 Quality Control (QC)

To begin with, the raw dataset of 827,448 cells was separated into 21 subsets depending on which study they originated from and QC was performed on these 21 datasets individually to filter out low-quality cells. Cells were filtered on three different QC metrics: number of counts per cell, number of genes per cell and percentage of reads mapping to the mitochondrial genome. Cells with relatively high numbers of counts or genes were removed to avoid including doublets and cells expressing a relatively high percentage of mitochondrial genes were removed to avoid including apoptotic cells. Where possible, we adhered to the thresholds set for these metrics by the authors of the original studies. If this information was not available from the original publication, thresholds for number of counts and number of genes per cell were defined as the top and bottom percentile for these metrics in each dataset and the threshold for percentage of mitochondrial genes was defined as the top percentile for all cells in each dataset. Cells exceeding these thresholds were removed. It was necessary to perform QC on each dataset individually in this way because each dataset had a unique distribution of QC metrics owing to the different protocols used in each study and technical variation among batches. After QC 808,034 high quality cells remained in the filtered dataset.

2.1.2 Identification of macrophages in each tissue

The filtered dataset was then separated into 19 subsets based on the tissue of origin for each cell. These datasets were analysed individually to find macrophage populations in each tissue. This involved performing the following processing steps on each tissue dataset:

- Filtering out genes that were expressed in less than 3 cells to reduce the size of the dataset.
- Normalisation of expression data to counts per million (CPM) using SCANPY's `normalize_total` function and transforming to log scale ($\ln x + 1$).
- Identification of top 3,000 highly variable genes.
- Computing UMAP coordinates using the top 30 principal components to reduce the dimensionality of the data (McInnes *et al.*, 2018).
- Using the Leiden algorithm (resolution = 0.5) to perform network clustering and label closely connected cells as belonging to the same cluster (Traag *et al.*, 2018).
- Visualising data and identifying clusters expressing known macrophage markers (e.g. CD68, CD14, CD163).

The macrophages from each tissue were compiled into a final macrophage dataset that included 72,860 cells from 15 different tissues. A schematic overview of the preparation and composition of the final macrophage dataset is depicted in Fig 1.

2.2 Comparison of tissue-resident macrophage transcriptomes

The compiled dataset of macrophages was processed according to the steps mentioned in Section 2.1.2 above (normalisation, identification of highly variable genes, dimensionality reduction and clustering).

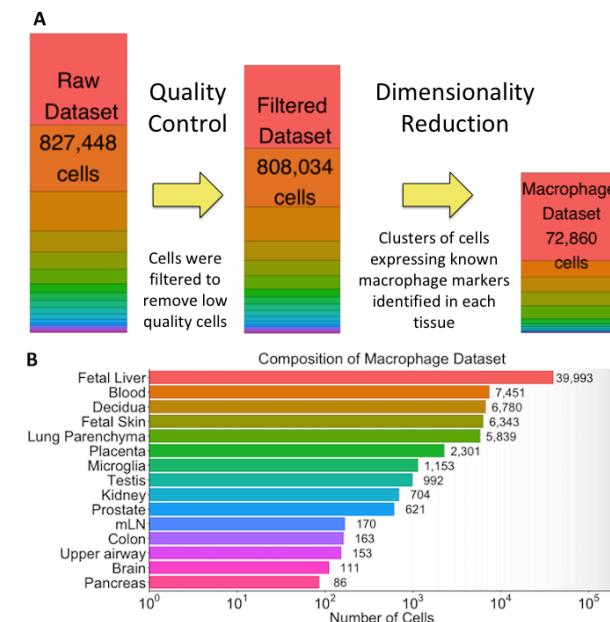


Fig. 1. (A) Schematic representation of the preparation of the macrophage dataset. Where possible cells were filtered using the same QC criteria used in the original publications. Cells in each tissue were visualised with UMAP plots to find clusters of macrophage populations. (B) The contribution of each tissue type to the composition of the final macrophage dataset. The exact contribution of each of the individual studies to the macrophage dataset can be found in Supplementary Table 1. mLN, Mesenteric lymph node.

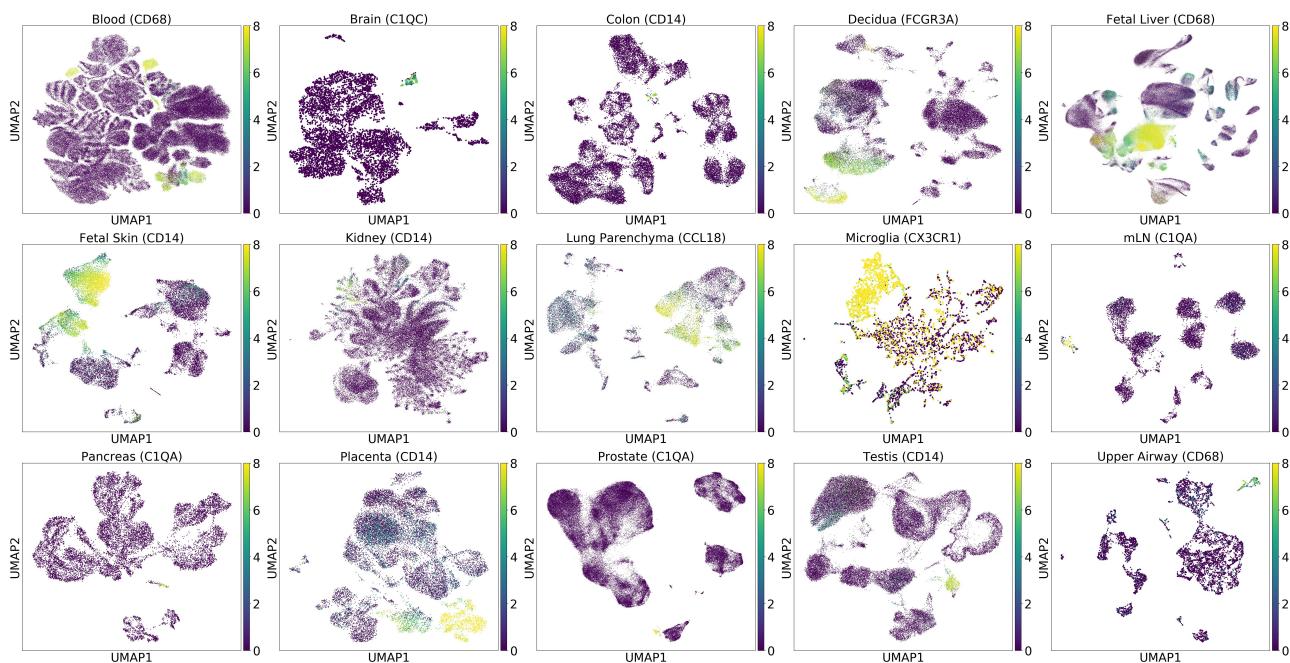


Fig. 2. UMAP plots of cells from each tissue with the macrophage population highlighted. Cells are coloured based on their expression levels (in log(CPM)) for macrophage marker genes. The title of each plot depicts which tissue the cells are from and also the gene used to colour the plot.

2.2.1 Batch Integration

Given that the dataset comprised of cells from several different studies that were conducted using different protocols, correcting batch effects due to technical variation was a challenge. Because only one dataset was available for many tissues, correcting for variation due to dataset would also remove variation due to tissue of origin which we wanted to avoid. Also it has been shown that using traditional batch correction methods to integrate data from different scRNA-seq studies can lead to overcorrection especially when integrating collections of datasets with considerable differences in cellular composition (Hie *et al.*, 2019). For these reasons, we chose to use BBKNN v1.3.4 (Batch-Balanced K-Nearest-Neighbours, Park *et al.* (2018)) to align batches when computing the connectivities of the neighbourhood graph, using ‘Protocol’ as the batch key, where ‘Protocol’ corresponds to the sequencing protocol used in the original study (e.g. 10X Genomics, SmartSeq2 etc.). By postponing batch correction to the neighbour graph inference step in this way, BBKN avoids overcorrecting the data and is not as computationally intensive as other batch correction algorithms.

2.2.2 Differential Expression Analysis

To perform differential expression (DE) analysis the macrophage dataset was loaded into R using Seurat v3.0.0 (Stuart *et al.*, 2019). A recent large-scale comparison study of DE analysis found that MAST (Finak *et al.*, 2015) was the best-performing single-cell DE testing method (Soneson and Robinson, 2018), therefore we chose to use Seurat’s implementation of MAST to find genes that were differentially expressed in each tissue compared to the rest and also genes that were differentially expressed in each of the predicted clusters compared to the rest. The min.pct parameter was set to 0.3 to only consider genes that were expressed by at least 30% of cells in that tissue/cluster.

2.2.3 Functional Enrichment

To interpret the DE results, genes that were found to be significantly upregulated ($\log FC \geq 1$, FDR adjusted p-value ≤ 0.05) in each tissue and each cluster were annotated based on involvement in common

biological processes. The g:Profiler Python module was used to find Gene Ontology Biological Process (GO:BP) annotations and Reactome pathways that were enriched in each list of upregulated genes (Raudvere *et al.*, 2019).

3 Results

3.1 A cross-tissue collection of macrophages

The final macrophage dataset contained expression levels for 22,298 genes in 72,860 cells across 15 tissues from 18 different studies. Fig 2 displays a UMAP of cells in 15 different tissues with the macrophage population highlighted. These macrophages were identified based on the expression levels of commonly used macrophage markers. Using canonical markers to identify cell types in single-cell data is a common practice but can prove challenging at times as cell surface markers that have typically been used for identifying a specific cell type in traditional molecular biology techniques such as fluorescence-activated cell sorting (FACS), don’t always have the same power to distinguish cell types in single-cell data. Fig 3 shows the expression levels of different marker genes in

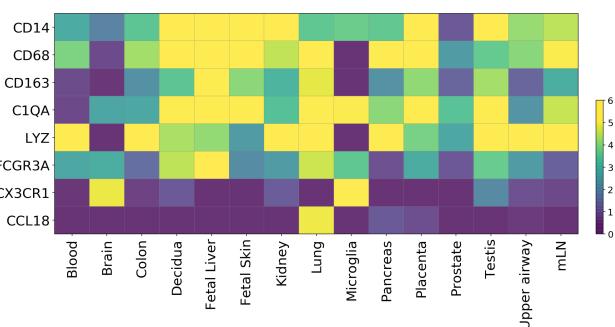


Fig. 3. Matrixplot showing the expression levels (in log(CPM)) of commonly used macrophage marker genes in macrophage populations from different tissues.

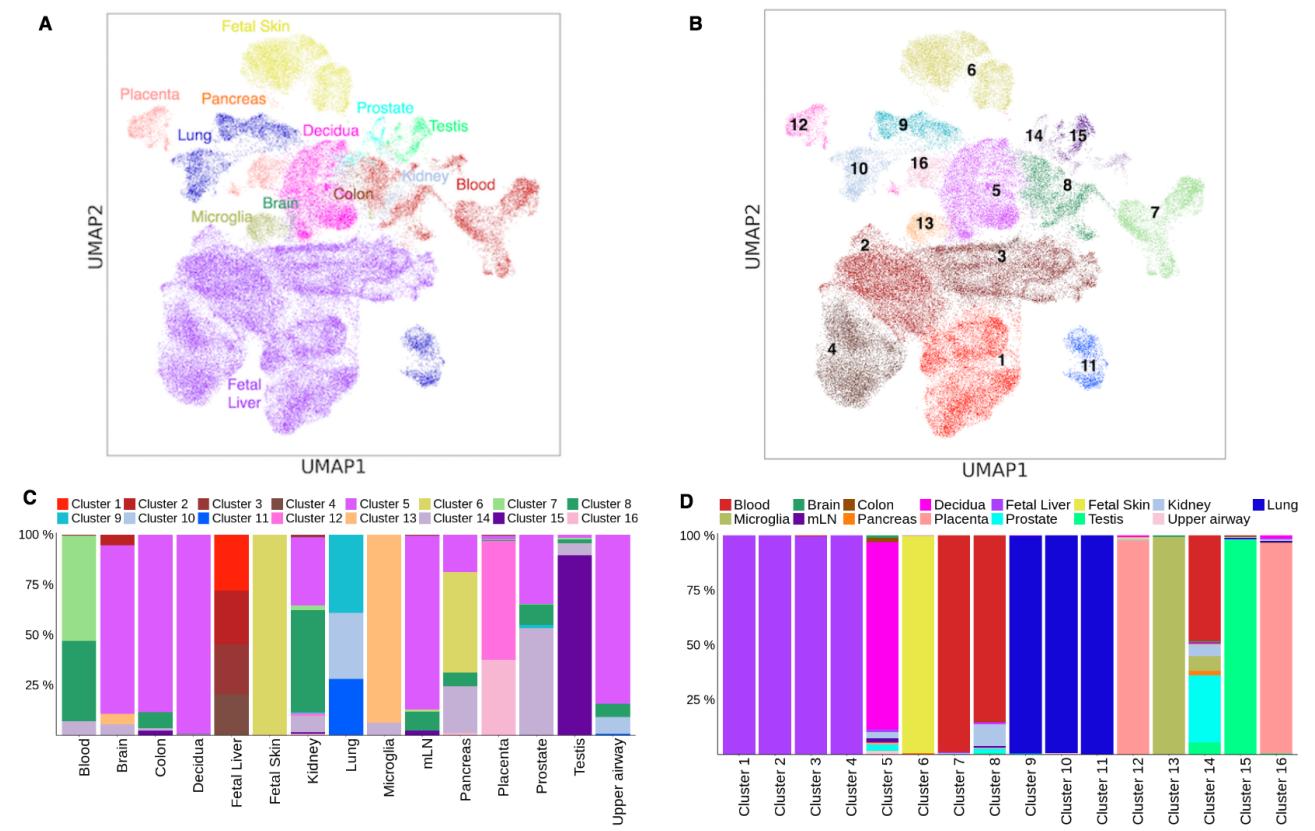


Fig. 4. (A) UMAP plot of the integrated macrophage dataset coloured by tissue of origin for each cell. (B) UMAP plot of the integrated macrophage dataset coloured by predicted cluster for each cell. (C) Cluster composition of cells grouped by tissue. (D) Tissue composition of cells grouped by cluster.

macrophages from each tissue. CD14 and C1QA performed most reliably for identifying macrophage populations in the most tissues while some markers worked particularly well in one specific tissue (e.g. CX3CR1 in brain, CCL18 in lung).

The integrated macrophage dataset is shown in Fig 4 (a view of the data before and after batch integration with BBKNN is displayed in Supplementary Fig 1). The Leiden algorithm used to predict clusters in this dataset placed cells in 16 different clusters which correlated reasonably well with the tissue of origin for each cell. Some tissues formed more than one cluster (Fetal Liver, Lung, Blood) potentially reflecting different subtypes of macrophages in one tissue and some clusters contained cells from more than one tissue (Cluster 5, Cluster 8, Cluster 14) which could indicate macrophage subtypes that are shared among tissues. These cells could potentially represent monocyte-derived macrophages that extravasate into tissues from the blood as opposed to tissue-resident macrophages that colonise tissues during early development.

Fig 5 displays the correlation between the gene expression profiles of macrophages from different tissues and demonstrates how certain groups of tissues are more similar than others. For example, macrophages from kidney, upper airway, colon and mLN have highly correlated transcriptomes, as do macrophages from microglia and brain, potentially reflecting a shared developmental lineage for macrophages in these tissues.

3.2 Tissue-resident macrophages possess unique gene expression profiles

The DE analysis found a total of 7,189 genes that were significantly ($\log_{2} \text{FC} \geq 1$ or ≤ -1 , FDR adjusted $p\text{-value} \leq 0.05$) differentially expressed among

macrophages from different tissues and 6,280 genes that were differentially expressed among macrophages in different clusters. DE genes for each cluster were largely similar to the DE genes for the tissue that each cluster represented. These results confirm that macrophages from different tissues possess unique transcriptional profiles. MA plots showing the average expression and fold change for DE genes in each tissue and cluster can be found in Supplementary Fig 2 & 3 and the top 10 DE genes for each tissue can be found in Supplementary Table 2.

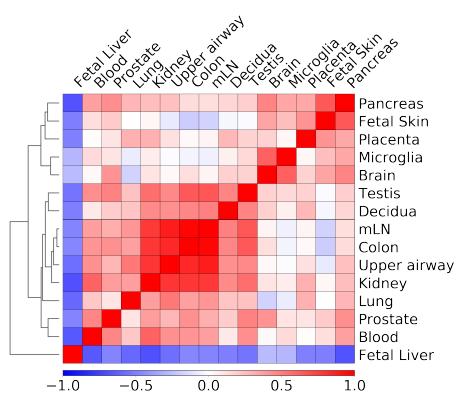


Fig. 5. Correlation plot displaying the Pearson correlation coefficient for average expression of genes in each tissue.

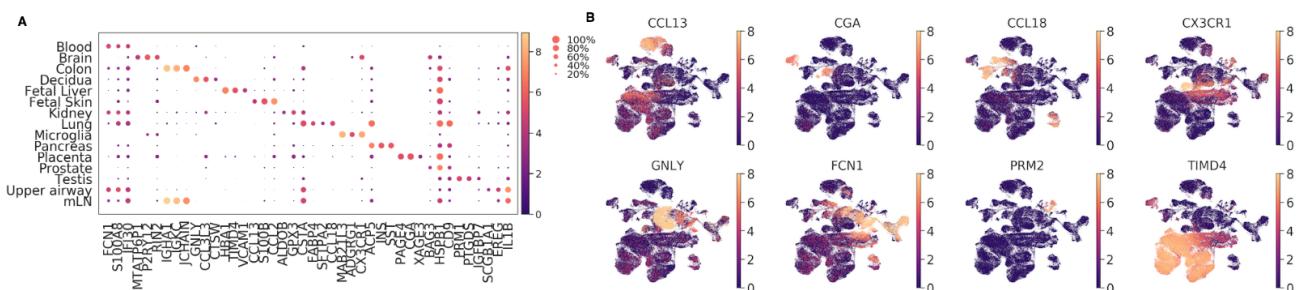


Fig. 6. (A) Dotplot displaying the top 3 DE genes for each tissue. The colour of each dot represents the expression level (in log(CPM)) of a given gene in a given tissue and the size of each dot represents the percentage of cells expressing a given gene in a given tissue. (B) UMAP feature plots of the integrated macrophage dataset coloured by expression levels of selected marker genes.

DE genes that displayed high fold changes in one tissue compared to all others were identified as marker genes for that tissue. Figure 5 shows the expression levels of marker genes in the macrophage dataset. These results confirm many previously known markers such as CX3CR1 and the more recently identified P2RY12 in the brain (Bennett *et al.*, 2016) and CCL18 in the lungs and identify potentially novel markers in some tissues such as CCL13 in fetal skin macrophages, CCL3L3 in decidual macrophages and TIMD4 in fetal liver macrophages.

In many cases the marker genes predicted for macrophages in a given tissue made sense from a biological perspective. For example, CGA was found to be upregulated ($\log FC = 5.85$) in macrophages in the placenta compared to other tissues. This gene encodes the alpha subunit of the four human glycoprotein hormones (chorionic gonadotrophin, luteinizing hormone, follicle stimulating hormone, thyroid stimulating hormone) and is expressed exclusively by the placenta and thyroid gland (Uhlén *et al.*, 2015). Similarly, PRM1 and PRM2 which encode protamines that are involved in the packaging of DNA within sperm cells were found to be significantly upregulated ($\log FC = 6.82$ and 6.62 respectively) in testis macrophages and genes which encode hemoglobin subunits (HBA1, HBA2, HBB, HBG1, HBG2, HBM, HBZ) were found to be upregulated in the fetal liver where one of the main roles of macrophages is recycling hemoglobin from red blood cells.

Interestingly, TIMD4 was identified as significantly upregulated ($\log FC = 4.79$) in macrophages from the fetal liver compared to other tissues. TIMD4 encodes a receptor that detects ‘eat-me’ signals on apoptotic and damaged red blood cells and has been proposed as a marker for distinguishing tissue-resident macrophages from monocyte-derived macrophages in several tissues including the liver, the gut and the heart (Scott *et al.*, 2016; Shaw *et al.*, 2018; Dick *et al.*, 2019). However, a recent study in which a fate-mapping reporter system was developed to track the lineage of macrophages in a murine system found that while TIMD4 performed reasonably well in distinguishing monocyte-derived macrophages from embryonic-derived macrophages early on in development, after some time monocyte-derived macrophages also start expressing TIMD4 (Liu *et al.*, 2019). Therefore, we hypothesised that TIMD4 is highly expressed by macrophages in fetal tissues compared to adult tissues and to test this we separated the macrophage data according to ‘Donor Type’ where ‘Donor Type’ indicates whether the cells came from fetal tissue ($n = 46,424$), biopsied adult tissue ($n = 25,396$) or deceased tissue ($n = 1,040$) and compared the expression of TIMD4. The results (shown in Supplementary Fig 4) confirm that TIMD4 expression is significantly ($p\text{-value} = 2e-16$) upregulated in macrophages from fetal tissues.

Surprisingly, the gene that displayed the highest fold change of all DE genes was MAB21L3 in microglia ($\log FC = 8.89$). MAB21L3 is not a

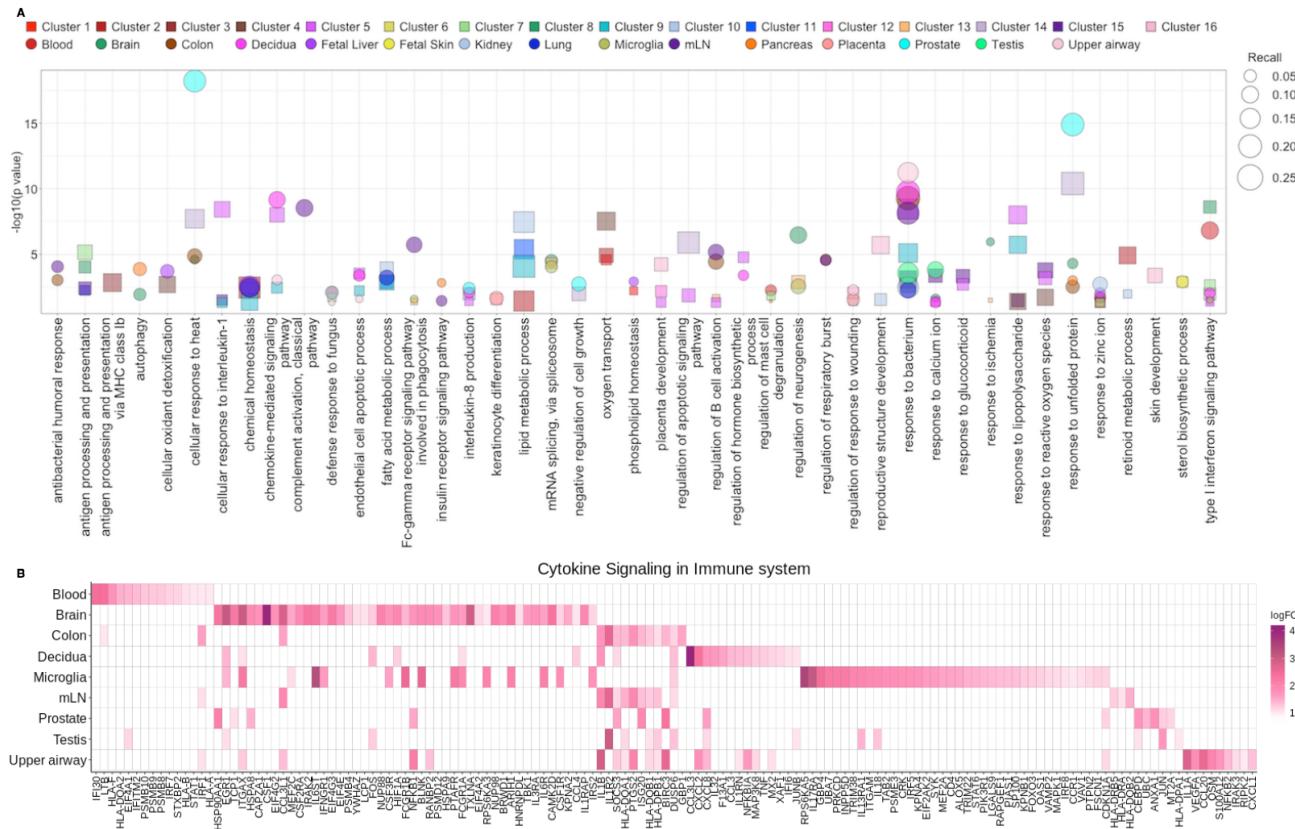
particularly well characterised gene in humans but it is implicated in cell-fate decisions (Takahashi *et al.*, 2015) and shares considerable sequence homology with cGAS, which is a receptor expressed on innate immune cells that senses cytosolic DNA (de Oliveira Mann *et al.*, 2016). These results suggest a potential role for MAB21L3 in regulating macrophage function in the brain.

3.3 Immune processes are enriched among upregulated genes

To investigate how the transcriptional diversity among tissue-resident macrophage populations may relate to their behaviour, we performed functional enrichment analysis to find functional categories (Gene Ontology: Biological Process (GO:BP) terms and Reactome pathways) that were overrepresented among upregulated genes in each tissue and each cluster. Similar to the DE analysis, functional categories that were enriched for each cluster largely overlapped with functional categories that were enriched for the tissue that cluster represented. This returned a total of 3,100 significantly enriched ($\text{adj. } p\text{-value} \leq 0.05$) GO terms and 402 significantly enriched Reactome pathways. Because a lot of these terms were redundant (e.g. ‘response to stimulus’, ‘cellular response to stimulus’ and ‘cellular response to chemical stimulus’) the most significant results for each tissue were filtered manually and are presented in Fig 7A.

As was the case with the DE genes, many of the enriched functional categories reflected the different biological roles of macrophages in different tissues. For example, ‘lipid metabolism’ was found to be enriched among upregulated genes for the 3 clusters of macrophages from lung tissue where one of the main roles of macrophages is to metabolise and clear surfactant. Similarly, ‘oxygen transport’ was found to be enriched among upregulated genes for 3 of the 4 clusters of macrophages in the liver, where one of the main functions of macrophages is to recycle hemoglobin (carrying oxygen) from red blood cells. Another GO term that demonstrated tissue-specificity is ‘regulation of neurogenesis’ which was enriched among upregulated genes from microglia, brain and Cluster 13 - which comprised mostly of cells from microglia and brain (Fig 4D). These results also indicate that macrophages in different tissues may be equipped for dealing with different types of pathogens. For example, ‘type I interferon signaling pathway’ which is involved in defence against viruses was found to be enriched for macrophages in the blood - one cluster of blood macrophages in particular (Cluster 8) - whereas ‘response to bacterium’ was enriched in other tissues - especially upper airway.

Many of these terms and pathways exhibited shared enrichment in several tissues despite the fact that a different set of upregulated genes was being analysed for each tissue. One such pathway, ‘Cytokine signaling in Immune System’, was enriched among upregulated genes from 9 out of 15 tissues. This is a very large pathway (more than 1,000 associated



4 Discussion

Macrophages are an integral part of all organs and play an instrumental role in regulating tissue homeostasis. Although they were originally thought to be short-lived bone marrow-derived cells, recent work has shown that tissue-resident macrophages have a mixed ontogeny - with the majority of cells originating in the developing embryo - and the ability to self-renew throughout the lifespan (Hoeffel and Ginhoux, 2018). These cells are likely subject to tissue-specific transcriptional regulation from an early developmental stage which means macrophages from different tissues will have unique gene expression profiles dictated by their tissue of origin. Here we present the first large-scale integration of tissue-resident macrophage

gene expression data across 15 different tissues and a comparison of these cells at the single-cell transcriptome level. The results confirm that the heterogeneity seen among tissue-resident macrophages is reflected in their diverse gene expression profiles.

The vast number of DE genes identified in this study demonstrates the extent of the heterogeneity among tissue-resident macrophages. Unsurprisingly, microglia and brain were the tissues that had the largest amounts of DE genes (1,364 and 1,358 genes respectively). Given that brain-resident macrophages make up the majority of immune cells in the brain and play a crucial role in brain function and development, these cells need to be equipped to respond to a broad range of signals that are likely very different from the signals encountered by macrophages in other tissues (Tanabe and Yamashita, 2018).

Genes that showed particularly drastic fold changes (such as MAB21L3 in microglia and INS in pancreas), along with functional annotations that showed high tissue-specific enrichment warrant experimental validation to determine whether or not they are genuine results or artefacts from technical variation. For example, the annotations ‘cellular response to heat’ and ‘response to unfolded protein’ displayed highly significant enrichment (adj. p-value = 5.77e-19 and 1.23e-15 respectively) among upregulated genes in prostate macrophages. This is likely due to the fact that many of these genes encode heat-shock proteins and could indicate that these cells were stressed prior to sequencing due to the experimental protocol used, although the percentage of reads mapping to the mitochondrial genome (a commonly used measure of cellular stress in scRNA-seq data) was relatively low for cells in this dataset. Results like this highlight how this study could be much improved by the inclusion of additional datasets for each tissue, as only one replicate was available for many of the included tissues. Additional replicates would make it easier to tease apart how much batch effects are contributing to results, as would better tools for integrating data from multiple scRNA-seq studies in different tissues as most of the currently available integration tools assume that cell types are shared among batches. Given the rapid pace at which the field of single-cell genomics is moving, these limitations could hopefully be overcome in the near future and these results could be further built upon.

Overall, this study represents a valuable resource for understanding transcriptional variation between macrophages in different tissues, which is essential when contemplating the fundamental role of macrophages in many disease conditions. The DE genes found in this study could potentially be used to inform the development of more accurate *in vitro* macrophage models - such as macrophage cell lines derived from induced pluripotent stem cells (iPSCs) - for investigating how macrophages can contribute to - and also have the potential to treat - many diseases and conditions including obesity (Zhang *et al.*, 2017), cancer (Cheng *et al.*, 2016) and infectious diseases (Hong *et al.*, 2018).

In keeping with the strong commitment to data-sharing exemplified by the Human Cell Atlas Consortium, we used the *cellxgene* software (www.github.com/chanzuckerberg/cellxgene) developed by the Chan-Zuckerberg Initiative to create an interactive web application where this cross-tissue collection of macrophages can be explored by members of the scientific community (<http://tiny.cc/cross-tissue-macrophages>). We hope the results presented here can inform the pursuit of experimental avenues that will further illuminate how tissue-resident macrophages carry out their diverse functions and manage to maintain homeostasis in highly heterogeneous environments.

Acknowledgements

We acknowledge with gratitude the Human Cell Atlas Consortium and the Teichmann group at the Sanger Institute in particular, for generating and sharing the single-cell RNA-seq data that made this analysis possible.

References

- Baron, M., Veres, A., Wolock, S. L., Faust, A. L., Gaujoux, R., Vetere, A., Ryu, J. H., Wagner, B. K., Shen-Orr, S. S., Klein, A. M., Melton, D. A., and Yanai, I. (2016). A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell systems*, **3**, 346–360.e4.
- Bennett, M. L., Bennett, F. C., Liddelow, S. A., Ajami, B., Zamanian, J. L., Fernhoff, N. B., Mulinyawe, S. B., Bohlen, C. J., Adil, A., Tucker, A., Weissman, I. L., Chang, E. F., Li, G., Grant, G. A., Hayden Gephart, M. G., and Barres, B. A. (2016). New tools for studying microglia in the mouse and human CNS. *Proceedings of the National Academy of Sciences*, **113**(12), E1738–E1746.
- Cheng, W. Y., Huynh, H., Chen, P., Peña Llopis, S., and Wan, Y. (2016). Macrophage parp1 inhibits grpr132 to mediate the anti-tumor effects of rosiglitazone. *eLife*, **5**(27692066), e18501.
- Davies, L. C., Jenkins, S. J., Allen, J. E., and Taylor, P. R. (2013). Tissue-resident macrophages. *Nature immunology*, **14**, 986–995.
- de Oliveira Mann, C. C., Kiefersauer, R., Witte, G., and Hopfner, K.-P. (2016). Structural and biochemical characterization of the cell fate determining nucleotidyltransferase fold protein mab2111. *Scientific reports*, **6**(27271801), 27498–27498.
- Dick, S. A., Macklin, J. A., Nejat, S., Momen, A., Clemente-Casares, X., Althagafi, M. G., Chen, J., Kantores, C., Hosseinzadeh, S., Aronoff, L., Wong, A., Zaman, R., Barbu, I., Besla, R., Lavine, K. J., Razani, B., Ginhoux, F., Husain, M., Cybulsky, M. I., Robbins, C. S., and Epelman, S. (2019). Self-renewing resident cardiac macrophages limit adverse remodeling following myocardial infarction. *Nature immunology*, **20**, 29–39.
- Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A. K., Slichter, C. K., Miller, H. W., McElrath, M. J., Prlic, M., Linsley, P. S., and Gottardo, R. (2015). Mast: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome biology*, **16**(26653891), 278–278.
- Gierahn, T. M., Wadsworth II, M. H., Hughes, T. K., Bryson, B. D., Butler, A., Satija, R., Fortune, S., Love, J. C., and Shalek, A. K. (2017). Seq-well: portable, low-cost RNA sequencing of single cells at high throughput. *Nature Methods*, **14**, 395.
- Guo, J., Grow, E. J., Micochova, H., Maher, G. J., Lindskog, C., Nie, X., Guo, Y., Takei, Y., Yun, J., Cai, L., Kim, R., Carrell, D. T., Goriely, A., Hotaling, J. M., and Cairns, B. R. (2018). The adult human testis transcriptional cell atlas. *Cell Research*, **28**(12), 1141–1157.
- Henry, G. H., Malewski, A., Joseph, D. B., Malladi, V. S., Lee, J., Torrealba, J., Mauck, R. J., Gahan, J. C., Raj, G. V., Roehrborn, C. G., Hon, G. C., MacConnara, M. P., Reese, J. C., Hutchinson, R. C., Vezina, C. M., and Strand, D. W. (2018). A cellular anatomy of the normal adult human prostate and prostatic urethra. *Cell Reports*, **25**(12), 3530 – 3542.e5.
- Hie, B., Bryson, B., and Berger, B. (2019). Efficient integration of heterogeneous single-cell transcriptomes using scanorama. *Nature Biotechnology*, **37**(6), 685–691.
- Hoeffel, G. and Ginhoux, F. (2018). Fetal monocytes and the origins of tissue-resident macrophages. *Cellular Immunology*, **330**, 5–15.
- Hong, D., Ding, J., Li, O., He, Q., Ke, M., Zhu, M., Liu, L., Ou, W.-B., He, Y., and Wu, Y. (2018). Human-induced pluripotent stem cell-derived macrophages and their immunological function in response to tuberculosis infection. *Stem cell research & therapy*, **9**(29482598), 49–49.
- La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., Borm, L. E., Stott, S. R. W., Toledo, E. M., Villaescusa, J. C., Lonnerberg, P., Ryge, J., Barker, R. A., Arenas, E., and Linnarsson, S. (2016). Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell*, **167**, 566–580.e19.
- Li, N., van Unen, V., Abdelaal, T., Guo, N., Kasatkaya, S. A., Ladell, K., McLaren, J. E., Egorov, E. S., Izraelson, M., Chuva de Sousa Lopes, S. M., Höllt, T., Britanova, O. V., Eggerton, J., de Miranda, N. F. C. C., Chudakov, D. M., Price, D. A., Lelieveldt, B. P. F., and Koning, F. (2019). Memory CD4+ T cells are generated in the human fetal intestine. *Nature Immunology*, **20**(3), 301–312.
- Liu, Z., Gu, Y., Chakarov, S., Bleriot, C., Chen, X., Shin, A., Huang, W., Dress, R. J., Dutertre, C.-A., Schlitzer, A., Chen, J., Wang, H., Liu, Z., Su, B., and Ginhoux, F. (2019). Fate mapping via ms4a3 expression history traces monocyte-derived cells. *bioRxiv*.
- Luecken, M. D. and Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol*, **15**(6), e8746.
- Mass, E., Ballesteros, I., Farlik, M., Halbritter, F., Günther, P., Crozet, L., Jacome-Galarza, C. E., Händler, K., Klughammer, J., Kobayashi, Y., Gomez-Perdiguer, E., Schultz, J. L., Beyer, M., Bock, C., and Geissmann, F. (2016). Specification of tissue-resident macrophages during organogenesis. *Science*, **353**(6304).
- Masuda, T., Sankowski, R., Staszewski, O., Böttcher, C., Amann, L., Sagar, Scheiwe, C., Nessler, S., Kunz, P., van Loo, G., Coenen, V. A., Reinacher, P. C., Michel, A., Sure, U., Gold, R., Grün, D., Priller, J., Stadelmann, C., and Prinz, M. (2019). Spatial and temporal heterogeneity of mouse and human microglia at single-cell resolution. *Nature*, **566**(7744), 388–392.
- McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.
- Menon, R., Otto, E. A., Kokurada, A., Zhou, J., Zhang, Z., Yoon, E., Chen, Y.-C., Troyanskaya, O., Spence, J. R., Kretzler, M., and Cebrian, C. (2018). Single-cell analysis of progenitor cell dynamics and lineage specification in the human fetal kidney. *Development (Cambridge, England)*, **145**.
- Miragaia, R. J., Gomes, T., Chomka, A., Jardine, L., Riedel, A., Hegazy, A. N., Whibley, N., Tucci, A., Chen, X., Lindeman, I., Emerton, G., Krausgruber,

- T., Shields, J., Haniffa, M., Powrie, F., and Teichmann, S. A. (2019). Single-cell transcriptomics of regulatory t cells reveals trajectories of tissue adaptation. *Immunity*, **50**(2), 493 – 504.e7.
- Muraro, M. J., Dharmadhikari, G., Grun, D., Groen, N., Dielen, T., Jansen, E., van Gurp, L., Engelse, M. A., Carlotti, F., de Koning, E. J. P., and van Oudenaarden, A. (2016). A single-cell transcriptome atlas of the human pancreas. *Cell systems*, **3**, 385–394.e3.
- Nowakowski, T. J., Bhaduri, A., Pollen, A. A., Alvarado, B., Mostajo-Radji, M. A., Di Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S. J., Velmeshv, D., Ounadjela, J. R., Shuga, J., Wang, X., Lim, D. A., West, J. A., Leyrat, A. A., Kent, W. J., and Kriegstein, A. R. (2017). Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science (New York, N.Y.)*, **358**, 1318–1323.
- Park, J.-E., Polański, K., Meyer, K., and Teichmann, S. A. (2018). Fast batch alignment of single cell transcriptomes unifies multiple mouse cell atlases into an integrated landscape. *bioRxiv*.
- Popescu, D.-M., Botting, R. A., Stephenson, E., Green, K., Jardine, L., Calderbank, E. F., Efremova, M., Acres, M., Maunder, D., Vegh, P., Goh, I., Gitton, Y., Park, J., Polanski, K., Vento-Tormo, R., Miao, Z., Rowell, R., McDonald, D., Fletcher, J., Dixon, D., Poyner, E., Reynolds, G., Mather, M., Moldovan, C., Mamanova, L., Greig, F., Young, M., Meyer, K., Lisgo, S., Bacardit, J., Fuller, A., Millar, B., Innes, B., Lindsay, S., Stubbington, M. J. T., Kowalczyk, M. S., Li, B., Ashenbrg, O., Tabaka, M., Dionne, D., Tickle, T. L., Slyper, M., Rozenblatt-Rosen, O., Filby, A., Villani, A.-C., Roy, A., Regev, A., Chedotal, A., Roberts, I., Göttgens, B., Laurenti, E., Behjati, S., Teichmann, S. A., and Haniffa, M. (2019). Decoding the development of the blood and immune systems during human fetal liver haematopoiesis. *bioRxiv*.
- Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*.
- Scott, C. L., Zheng, F., De Baetselier, P., Martens, L., Saeys, Y., De Prijck, S., Lippens, S., Abels, C., Schoonooghe, S., Raes, G., Devogd, N., Lambrecht, B. N., Beschin, A., and Guillemins, M. (2016). Bone marrow-derived monocytes give rise to self-renewing and fully differentiated kupffer cells. *Nature communications*, **7**(26813785), 10321–10321.
- Segerstolpe, A., Palasantza, A., Eliasson, P., Andersson, E.-M., Andreasson, A.-C., Sun, X., Picelli, S., Sabirsh, A., Clausen, M., Bjursell, M. K., Smith, D. M., Kasper, M., Ammala, C., and Sandberg, R. (2016). Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell metabolism*, **24**, 593–607.
- Shaw, T. N., Houston, S. A., Wemyss, K., Bridgeman, H. M., Barbera, T. A., Zangerle-Murray, T., Strangward, P., Ridley, A. J. L., Wang, P., Tamoutounour, S., Allen, J. E., Konkel, J. E., and Grainger, J. R. (2018). Tissue-resident macrophages in the intestine are long lived and defined by tim-4 and cd4 expression. *The Journal of experimental medicine*, **215**, 1507–1518.
- Sohni, A., Tan, K., Song, H.-W., Burow, D., de Rooij, D. G., Laurent, L., Hsieh, T.-C., Rabah, R., Hammoud, S. S., Vicini, E., and Wilkinson, M. F. (2019). The neonatal and adult human testis defined at the single-cell level. *Cell reports*, **26**, 1501–1517.e4.
- Soneson, C. and Robinson, M. D. (2018). Bias, robustness and scalability in single-cell differential expression analysis. *Nature methods*, **15**, 255–261.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell*, **177**(7), 1888–1902.e21.
- Svensson, V., Vento-Tormo, R., and Teichmann, S. A. (2018). Exponential scaling of single-cell rna-seq in the past decade. *Nature protocols*, **13**(4), 599.
- Takahashi, C., Kusakabe, M., Suzuki, T., Miyatake, K., and Nishida, E. (2015). mab21-13 regulates cell fate specification of multiciliate cells and ionocytes. *Nature communications*, **6**, 6017.
- Tanabe, S. and Yamashita, T. (2018). The role of immune cells in brain development and neurodevelopmental diseases. *International Immunology*, **30**(10), 437–444.
- Traag, V. A., Waltman, L., and van Eck, N. J. (2018). From louvain to leiden: guaranteeing well-connected communities. *CoRR*, **abs/1810.08473**.
- Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, r., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szligarto, C. A.-K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hoher, S., Alm, T., Edqvist, P.-H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J. M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J., and Pontén, F. (2015). Tissue-based map of the human proteome. *Science*, **347**(6220), 1260419.
- Vento-Tormo, R., Efremova, M., Botting, R. A., Turco, M. Y., Vento-Tormo, M., Meyer, K. B., Park, J.-E., Stephenson, E., Polański, K., Goncalves, A., Gardner, L., Holmqvist, S., Henriksson, J., Zou, A., Sharkey, A. M., Millar, B., Innes, B., Wood, L., Wilbrey-Clark, A., Payne, R. P., Ivarsson, M. A., Lisgo, S., Filby, A., Rowitch, D. H., Bulmer, J. N., Wright, G. J., Stubbington, M. J. T., Haniffa, M., Moffett, A., and Teichmann, S. A. (2018). Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature*, **563**(7731), 347–353.
- Vieira Braga, F. A., Kar, G., Berg, M., Carpaíj, O. A., Polanski, K., Simon, L. M., Brouwer, S., Gomes, T., Hesse, L., Jiang, J., Fasouli, E. S., Efremova, M., Vento-Tormo, R., Talavera-López, C., Jonker, M. R., Affleck, K., Palit, S., Strzelecka, P. M., Firth, H. V., Mahbubani, K. T., Cvejic, A., Meyer, K. B., Saeb-Parsy, K., Luinge, M., Brandsma, C.-A., Timens, W., Angelidis, I., Strunz, M., Koppelman, G. H., van Oosterhout, A. J., Schiller, H. B., Theis, F. J., van den Berge, M., Nawijn, M. C., and Teichmann, S. A. (2019). A cellular census of human lungs identifies novel cell states in health and in asthma. *Nature Medicine*.
- Wang, Y. J., Schug, J., Won, K.-J., Liu, C., Naji, A., Avrahami, D., Golson, M. L., and Kaestner, K. H. (2016). Single-cell transcriptomics of the human endocrine pancreas. *Diabetes*, **65**, 3028–38.
- Wolf, F. A., Angerer, P., and Theis, F. J. (2018). Scanpy: large-scale single-cell gene expression data analysis. *Genome Biology*, **19**(1), 15.
- Young, M. D., Mitchell, T. J., Vieira Braga, F. A., Tran, M. G. B., Stewart, B. J., Ferdinand, J. R., Collord, G., Botting, R. A., Popescu, D.-M., Loudon, K. W., Vento-Tormo, R., Stephenson, E., Cagan, A., Farndon, S. J., Del Castillo Velasco-Herrera, M., Guzzo, C., Richoz, N., Mamanova, L., Aho, T., Armitage, J. N., Riddick, A. C. P., Mushtaq, I., Farrell, S., Rampling, D., Nicholson, J., Filby, A., Burge, J., Lisgo, S., Maxwell, P. H., Lindsay, S., Warren, A. Y., Stewart, G. D., Sebire, N., Coleman, N., Haniffa, M., Teichmann, S. A., Clatworthy, M., and Behjati, S. (2018). Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science (New York, N.Y.)*, **361**, 594–599.
- Zhang, L., Yu, X., Zheng, L., Zhang, Y., Li, Y., Fang, Q., Gao, R., Kang, B., Zhang, Q., Huang, J. Y., Konno, H., Guo, X., Ye, Y., Gao, S., Wang, S., Hu, X., Ren, X., Shen, Z., Ouyang, W., and Zhang, Z. (2018). Lineage tracking reveals dynamic relationships of t cells in colorectal cancer. *Nature*, **564**(7735), 268–272.
- Zhang, Y., Rao, E., Zeng, J., Hao, J., Sun, Y., Liu, S., Sauter, E. R., Bernlohr, D. A., Cleary, M. P., Suttlies, J., and Li, B. (2017). Adipose fatty acid binding protein promotes saturated fatty acid-induced macrophage cell death through enhancing ceramide production. *Journal of immunology (Baltimore, Md. : 1950)*, **198**(27920274), 798–807.
- Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., Bharadwaj, R., Wong, A., Ness, K. D., Beppu, L. W., Deeg, H. J., McFarland, C., Loeb, K. R., Valente, W. J., Ericson, N. G., Stevens, E. A., Radich, J. P., Mikkelsen, T. S., Hindson, B. J., and Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, **8**, 14049.

Supplementary Materials

Supplementary Tables

Supplementary Table 1

Contribution of each of the individual datasets to the raw, filtered and macrophage datasets.

Reference	Raw Dataset	Filtered Dataset	Macrophage Dataset
Baron <i>et. al.</i> , 2016	8,569	8,272	63
Gierahn <i>et.al.</i> , 2017	5,584	4,027	1,542
Guo <i>et.al.</i> , 2018	12,985	12,645	375
Henry <i>et.al.</i> , 2018	109,061	103,057	628
LaManno <i>et.al.</i> , 2016	3,712	3,712	29
Li <i>et.al.</i> , 2019	1,886	1,886	0
Masuda <i>et.al.</i> , 2019	4,408	4,247	1,153
Menon <i>et.al.</i> , 2018	9,840	9,649	6
Miragaia <i>et.al.</i> , 2019	1,168	1,164	0
Muraro <i>et.al.</i> , 2016	2,126	2067	11
Nowakowski <i>et.al.</i> , 2017	4,261	4,219	82
Popescu <i>et.al.</i> , 2019	273,856	271,113	46,336
Segerstolpe <i>et.al.</i> , 2016	3,363	3,132	12
Sohni <i>et.al.</i> , 2019	34,729	34,021	617
Vento-Tormo <i>et.al.</i> , 2018	70,325	70,325	10,765
Vieira Braga <i>et.al.</i> , 2019	26,013	125,848	5,992
Wang <i>et.al.</i> , 2016	635	613	0
Young <i>et.al.</i> , 2018	48,347	48,333	698
Zhang <i>et.al.</i> , 2018	11,138	10,771	0
Zheng <i>et.al.</i> , 2017	163,234	157,784	4,225
Unpublished	32,228	31,169	333
Total	827,448	808,034	72,860

Supplementary Table 2

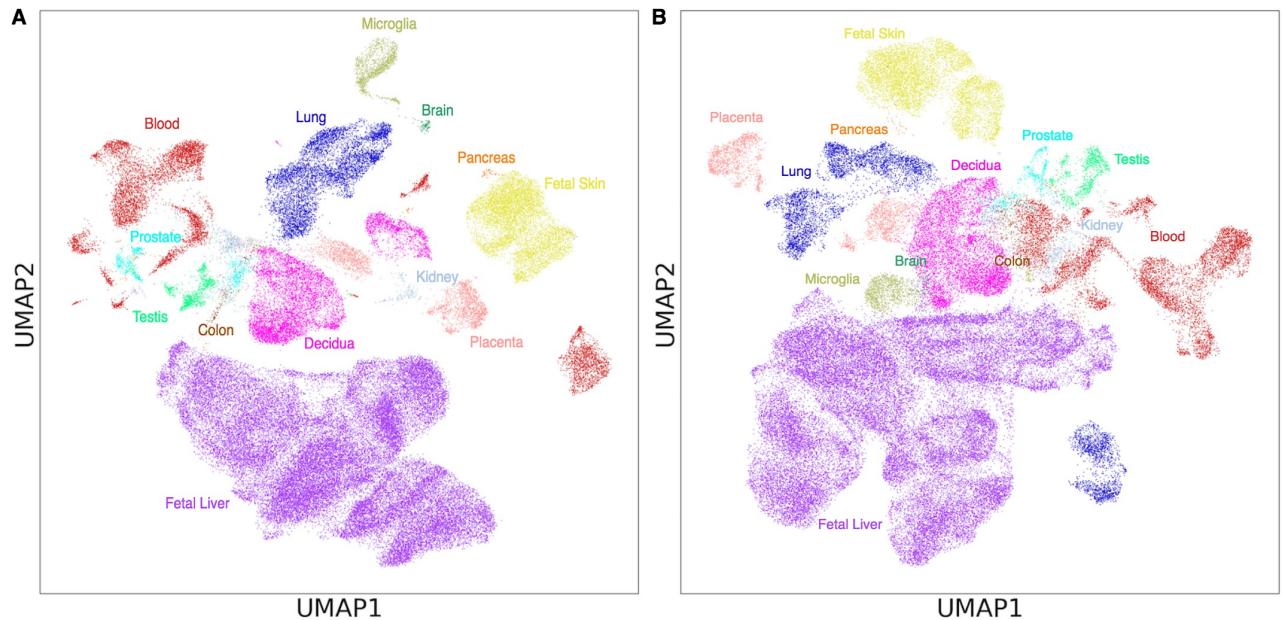
Gene symbol, name, logFC and adj. p-value for the top 10 DE genes in each tissue. An extended version of this table, which includes the UniProt description of each gene is available at this github link.

Tissue	Gene Symbol	Gene Name	Avg. logFC	Adj. p-value
Blood	MTRNR2L12	MT-RNR2 Like 12	3.91	0
Blood	MTRNR2L8	MT-RNR2 Like 8	3.77	0
Blood	VCAN	Versican	3.23	0
Blood	FCN1	Ficolin 1	2.94	0
Blood	S100A8	S100 Calcium Binding Protein A8	2.54	0
Blood	IFI30	IFI30 Lysosomal Thiol Reductase	2.5	0
Blood	LTB	Lymphotoxin Beta	2.44	0
Blood	POLD4	DNA Polymerase Delta 4, Accessory Subunit	2.2	0
Blood	LYZ	Lysozyme	2.12	0
Blood	ICAM3	Intercellular Adhesion Molecule 3	2.03	0
Brain	RPL13AP5	Ribosomal Protein L13a Pseudogene 5	5.58	3.04E-238
Brain	MTATP6P1	MT-ATP6 Pseudogene 1	6.25	1.06E-231
Brain	EEF1A1P5	Eukaryotic Translation Elongation Factor 1 Alpha 1 Pseudogene 5	4.91	3.23E-217
Brain	FTLP3	Ferritin Light Chain Pseudogene 3	3.5	1.13E-215
Brain	H3F3AP4	H3 Histone, Family 3A, Pseudogene 4	5.58	7.25E-212
Brain	RPL3P4	Ribosomal Protein L3 Pseudogene 4	3.58	2.30E-211
Brain	RPL21P75	Ribosomal Protein L21 Pseudogene 75	4.46	5.53E-207
Brain	MTND2P28	MT-ND2 Pseudogene 28	3.91	6.10E-205
Brain	RPL21P28	Ribosomal Protein L21 Pseudogene 28	4.64	1.18E-184
Brain	HMGB1P5	High Mobility Group Box 1 Pseudogene 5	4.1	3.11E-181
Colon	IGHA1	Immunoglobulin Heavy Constant Alpha 1	4.72	0
Colon	IGLC2	Immunoglobulin Lambda Constant 2	4.52	0
Colon	IGKC	Immunoglobulin Kappa Constant	3.62	0
Colon	JCHAIN	Joining Chain Of Multimeric IgA And IgM	3.52	6.77E-292
Colon	IGLC3	Immunoglobulin Lambda Constant 3 (Kern-Oz+ Marker)	2.81	6.01E-276
Colon	IGHA2	Immunoglobulin Heavy Constant Alpha 2 (A2m Marker)	4.03	1.48E-258
Colon	IGLC7	Immunoglobulin Lambda Constant 7	3.82	7.72E-108
Colon	LYZ	Lysozyme	1.6	3.37E-82
Colon	CD69	CD69 Molecule	2.17	7.70E-60
Colon	TNFAIP3	TNF Alpha Induced Protein 3	1.93	7.99E-60
Decidua	GNLY	Granulysin	4.9	0
Decidua	CCL3L3	C-C Motif Chemokine Ligand 3 Like 3	4.09	0
Decidua	XCL2	X-C Motif Chemokine Ligand 2	3.85	0
Decidua	TRDC	T Cell Receptor Delta Constant	3.47	0
Decidua	GZMA	Granzyme A	3.17	0
Decidua	CTSW	Cathepsin W	3.11	0
Decidua	HAMP	Hepcidin Antimicrobial Peptide	2.82	0
Decidua	MIR24-2	MicroRNA 24-2	2.64	0
Decidua	SNORD3D	Small Nucleolar RNA, C/D Box 3D	2.54	0
Decidua	CXCL2	C-X-C Motif Chemokine Ligand 2	2.39	0
Fetal Liver	HBG2	Hemoglobin Subunit Gamma 2	6.68	0
Fetal Liver	HBA1	Hemoglobin Subunit Alpha 1	5.96	0
Fetal Liver	CETP	Cholesteryl Ester Transfer Protein	5.8	0
Fetal Liver	HBA2	Hemoglobin Subunit Alpha 2	5.34	0
Fetal Liver	HBM	Hemoglobin Subunit Mu	5.15	0
Fetal Liver	VCAM1	Vascular Cell Adhesion Molecule 1	4.83	0
Fetal Liver	TIMD4	T Cell Immunoglobulin And Mucin Domain Containing 4	4.79	0
Fetal Liver	AHSP	Alpha Hemoglobin Stabilizing Protein	4.7	0
Fetal Liver	ALB	Albumin	4.63	0
Fetal Liver	APOA2	Apolipoprotein A2	4.56	0
Fetal Skin	DDIT4L	DNA Damage Inducible Transcript 4 Like	4.55	0
Fetal Skin	CCL13	C-C Motif Chemokine Ligand 13	3.36	0
Fetal Skin	GAPLINC	Gastric Adenocarcinoma Associated, Positive CD44 Regulator, Long Intergenic Non-Coding RNA	3.22	0
Fetal Skin	NT5DC2	5'-Nucleotidase Domain Containing 2	3.21	0
Fetal Skin	LINC01480	Long Intergenic Non-Protein Coding RNA 1480	3.01	0
Fetal Skin	SLC12A5	Solute Carrier Family 12 Member 5	3	0
Fetal Skin	S100B	S100 Calcium Binding Protein B	2.95	0
Fetal Skin	CCL2	C-C Motif Chemokine Ligand 2	2.79	0
Fetal Skin	SPARC	Secreted Protein Acidic And Cysteine Rich	2.71	0
Fetal Skin	ATP6V0D2	ATPase H+ Transporting V0 Subunit D2	2.67	0
Kidney	MIOX	Myo-Inositol Oxygenase	5.6	0
Kidney	PDZK1IP1	PDZK1 Interacting Protein 1	5.31	0
Kidney	NAT8	N-Acetyltransferase 8 (Putative)	4.89	0
Kidney	ALDOB	Aldolase, Fructose-Bisphosphate B	4.59	0
Kidney	RBP5	Retinol Binding Protein 5	4.28	0
Kidney	GPX3	Glutathione Peroxidase 3	3.58	0
Kidney	PRAP1	Proline Rich Acidic Protein 1	4.02	1.04E-316

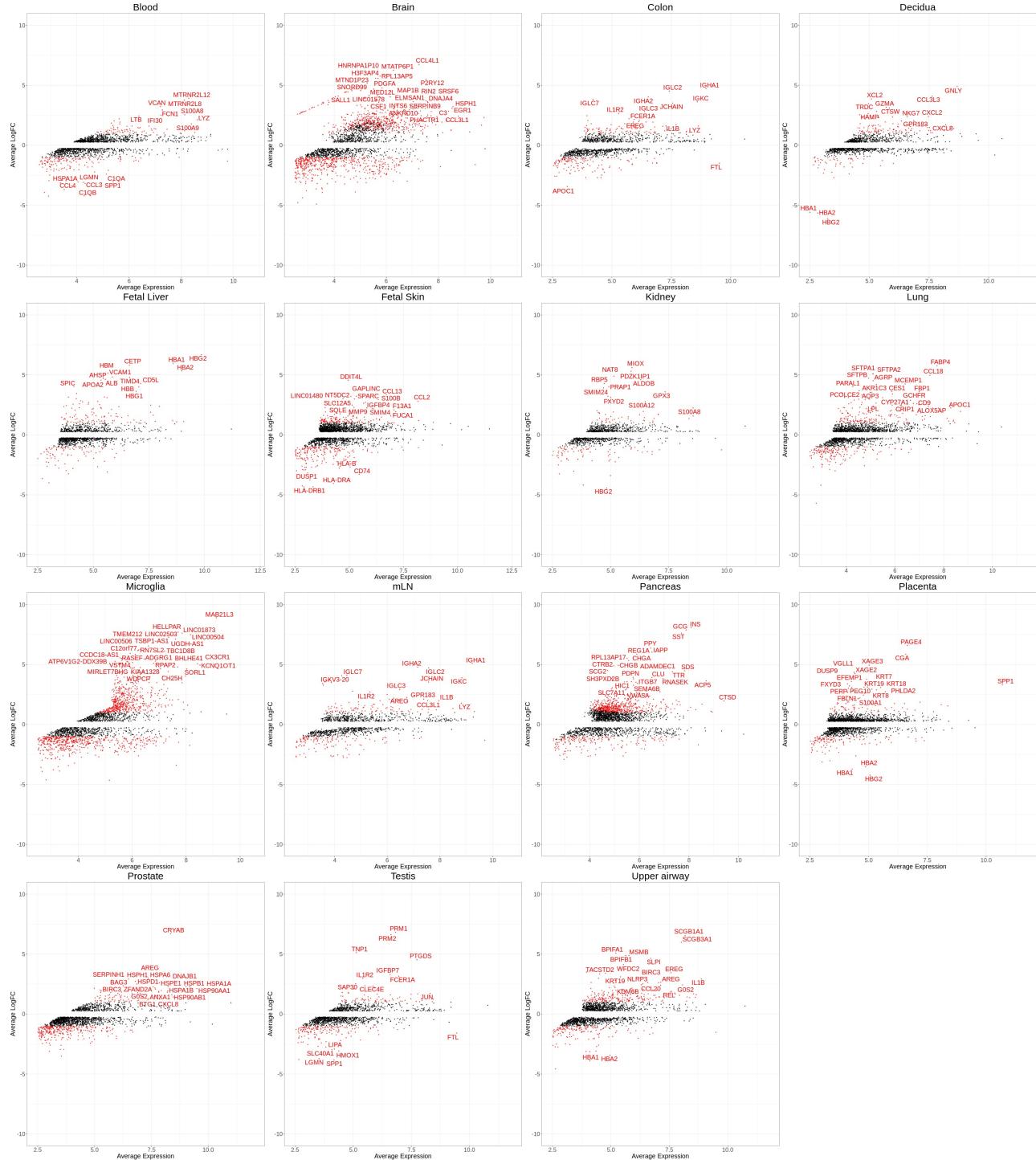
Tissue	Gene Symbol	Gene Name	Avg. logFC	Adj. p-value
Kidney	NBEAL1	Neurobeachin Like 1	1.37	1.69E-296
Kidney	FXYD2	FXYD Domain Containing Ion Transport Regulator 2	3.12	7.12E-296
Kidney	ASS1	Argininosuccinate Synthase 1	3.21	3.83E-290
Lung	FABP4	Fatty Acid Binding Protein 4	5.76	0
Lung	SFTPA2	Surfactant Protein A2	5.09	0
Lung	SFTPA1	Surfactant Protein A1	5.08	0
Lung	CCL18	C-C Motif Chemokine Ligand 18	4.96	0
Lung	SFTPB	Surfactant Protein B	4.72	0
Lung	AGRP	Agouti Related Neuropeptide	4.52	0
Lung	MCEMP1	Mast Cell Expressed Membrane Protein 1	4.24	0
Lung	AKR1C3	Aldo-Keto Reductase Family 1 Member C3	4.2	0
Lung	PARAL1	PPARG Activating RBM14 Associated LncRNA 1	4.02	0
Lung	PCOLCE2	Procollagen C-Endopeptidase Enhancer 2	3.68	0
Microglia	MAB21L3	Mab-21 Like 3	8.89	0
Microglia	HELLPAR	HELLP Associated Long Non-Coding RNA	7.7	0
Microglia	LINC01873	Long Intergenic Non-Protein Coding RNA 1873	7.61	0
Microglia	LINC00504	Long Intergenic Non-Protein Coding RNA 504	7.49	0
Microglia	LINC02503	Long Intergenic Non-Protein Coding RNA 2503	7.26	0
Microglia	UGDH-AS1	UGDH Antisense RNA 1	7.11	0
Microglia	TSBP1-AS1	TSBP1 And BTNL2 Antisense RNA 1	6.55	0
Microglia	TMEM212	Transmembrane Protein 212	6.26	0
Microglia	LINC00506	Long Intergenic Non-Protein Coding RNA 506	6.21	0
Microglia	C12orf77	Chromosome 12 Open Reading Frame 77	6.16	0
mLN	IGHA1	Immunoglobulin Heavy Constant Alpha 1	5.09	0
mLN	IGLC2	Immunoglobulin Lambda Constant 2	4.71	0
mLN	IGKC	Immunoglobulin Kappa Constant	3.89	0
mLN	JCHAIN	Joining Chain Of Multimeric IgA And IgM	3.53	0
mLN	IGLC3	Immunoglobulin Lambda Constant 3 (Kern-Oz+ Marker)	2.91	0
mLN	IGHA2	Immunoglobulin Heavy Constant Alpha 2 (A2m Marker)	4.79	1.10E-319
mLN	IGKV3-20	Immunoglobulin Kappa Variable 3-20	4.06	5.49E-153
mLN	IGLC7	Immunoglobulin Lambda Constant 7	4.11	4.21E-151
mLN	IGHV3-48	Immunoglobulin Heavy Variable 3-48	3.72	1.69E-132
mLN	IGLV3-21	Immunoglobulin Lambda Variable 3-21	3.63	8.26E-121
Pancreas	INS	Insulin	8.09	8.32E-232
Pancreas	SST	Somatostatin	7.61	3.08E-203
Pancreas	GCG	Glucagon	7.87	2.97E-188
Pancreas	RNASEK	Ribonuclease K	3.82	8.00E-140
Pancreas	SDS	Serine Dehydratase	4.48	1.29E-126
Pancreas	REG1A	Regenerating Family Member 1 Alpha	5.85	9.10E-126
Pancreas	ACP5	Acid Phosphatase 5, Tartrate Resistant	3.63	2.09E-121
Pancreas	IAPP	Islet Amyloid Polypeptide	6.02	1.18E-100
Pancreas	RPL13AP17	Ribosomal Protein L13a Pseudogene 17	5.51	4.23E-92
Pancreas	GABARAP	GABA Type A Receptor-Associated Protein	2.1	1.25E-83
Placenta	PAGE4	PAGE Family Member 4	6.59	0
Placenta	CGA	Glycoprotein Hormones, Alpha Polypeptide	5.85	0
Placenta	XAGE3	X Antigen Family Member 3	4.94	0
Placenta	VGLL1	Vestigial Like Family Member 1	4.35	0
Placenta	XAGE2	X Antigen Family Member 2	4.24	0
Placenta	KRT7	Keratin 7	3.78	0
Placenta	EFEMP1	EGF Containing Fibulin Extracellular Matrix Protein 1	3.58	0
Placenta	DUSP9	Dual Specificity Phosphatase 9	3.39	0
Placenta	SPP1	Secreted Phosphoprotein 1	3.32	0
Placenta	FXYD3	FXYD Domain Containing Ion Transport Regulator 3	3.22	0
Prostate	CRYAB	Crystallin Alpha B	6.7	0
Prostate	SERPINH1	Serpin Family H Member 1	2.92	0
Prostate	BAG3	BCL2 Associated Athanogene 3	2.73	0
Prostate	HSPB1	Heat Shock Protein Family B (Small) Member 1	2.29	0
Prostate	DNAJB1	DnaJ Heat Shock Protein Family (Hsp40) Member B1	2.29	0
Prostate	HSPA1A	Heat Shock Protein Family A (Hsp70) Member 1A	2.27	0
Prostate	HSPH1	Heat Shock Protein Family H (Hsp110) Member 1	2.27	0
Prostate	HSPD1	Heat Shock Protein Family D (Hsp60) Member 1	2.18	0
Prostate	HSP90AA1	Heat Shock Protein 90 Alpha Family Class A Member 1	2.12	0
Prostate	HSPE1	Heat Shock Protein Family E (Hsp10) Member 1	1.96	0
Testis	PRM1	Protamine 1	6.82	0
Testis	PRM2	Protamine 2	6.62	0
Testis	TNP1	Transition Protein 1	5.12	0
Testis	PTGDS	Prostaglandin D2 Synthase	4.53	0
Testis	IGFBP7	Insulin Like Growth Factor Binding Protein 7	3.29	0
Testis	FCER1A	Fc Fragment Of IgE Receptor Ia	3.15	0
Testis	SAP30	Sin3A Associated Protein 30	2.53	0
Testis	TCOF1	Treacle Ribosome Biogenesis Factor 1	2.33	0
Testis	FCGR2B	Fc Fragment Of IgG Receptor IIb	2.18	0
Testis	JUN	Jun Proto-Oncogene, AP-1 Transcription Factor Subunit	1.72	0
Upper airway	REL	REL Proto-Oncogene, NF-KB Subunit	2.39	6.48E-175
Upper airway	SCGB1A1	Secretoglobin Family 1A Member 1	6.54	2.37E-131

Tissue	Gene Symbol	Gene Name	Avg. logFC	Adj. p-value
Upper airway	KDM6B	Lysine Demethylase 6B	2.44	1.11E-126
Upper airway	IL1B	Interleukin 1 Beta	2.92	4.87E-126
Upper airway	MSMB	Microseminoprotein Beta	4.88	3.35E-122
Upper airway	BPIFA1	BPI Fold Containing Family A Member 1	5.07	1.52E-121
Upper airway	BPIFB1	BPI Fold Containing Family B Member 1	4.87	1.12E-117
Upper airway	PLAUR	Plasminogen Activator, Urokinase Receptor	1.9	3.24E-114
Upper airway	BIRC3	Baculoviral IAP Repeat Containing 3	2.64	1.74E-113
Upper airway	NLRP3	NLR Family Pyrin Domain Containing 3	2.52	2.88E-96

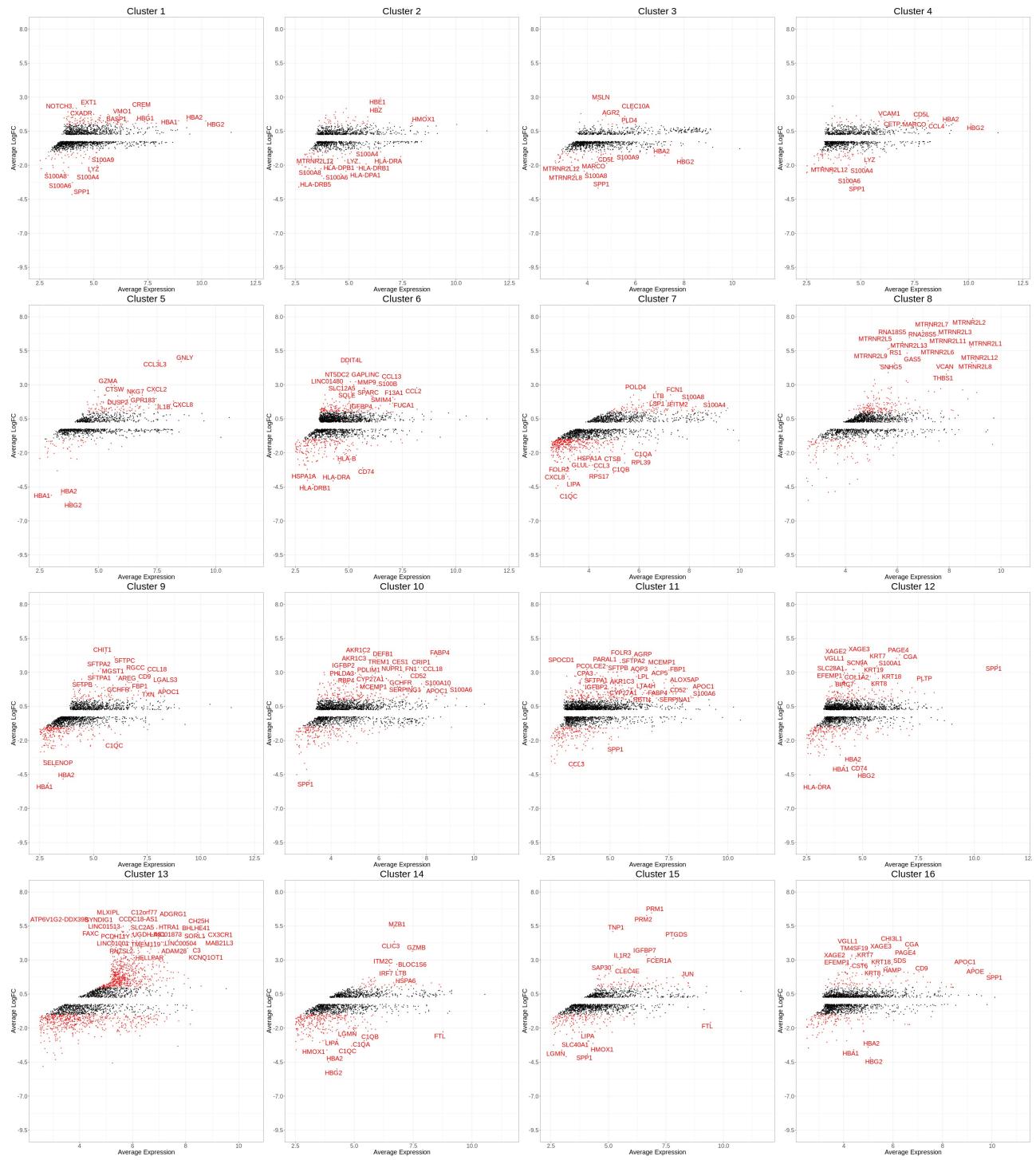
Supplementary Figures



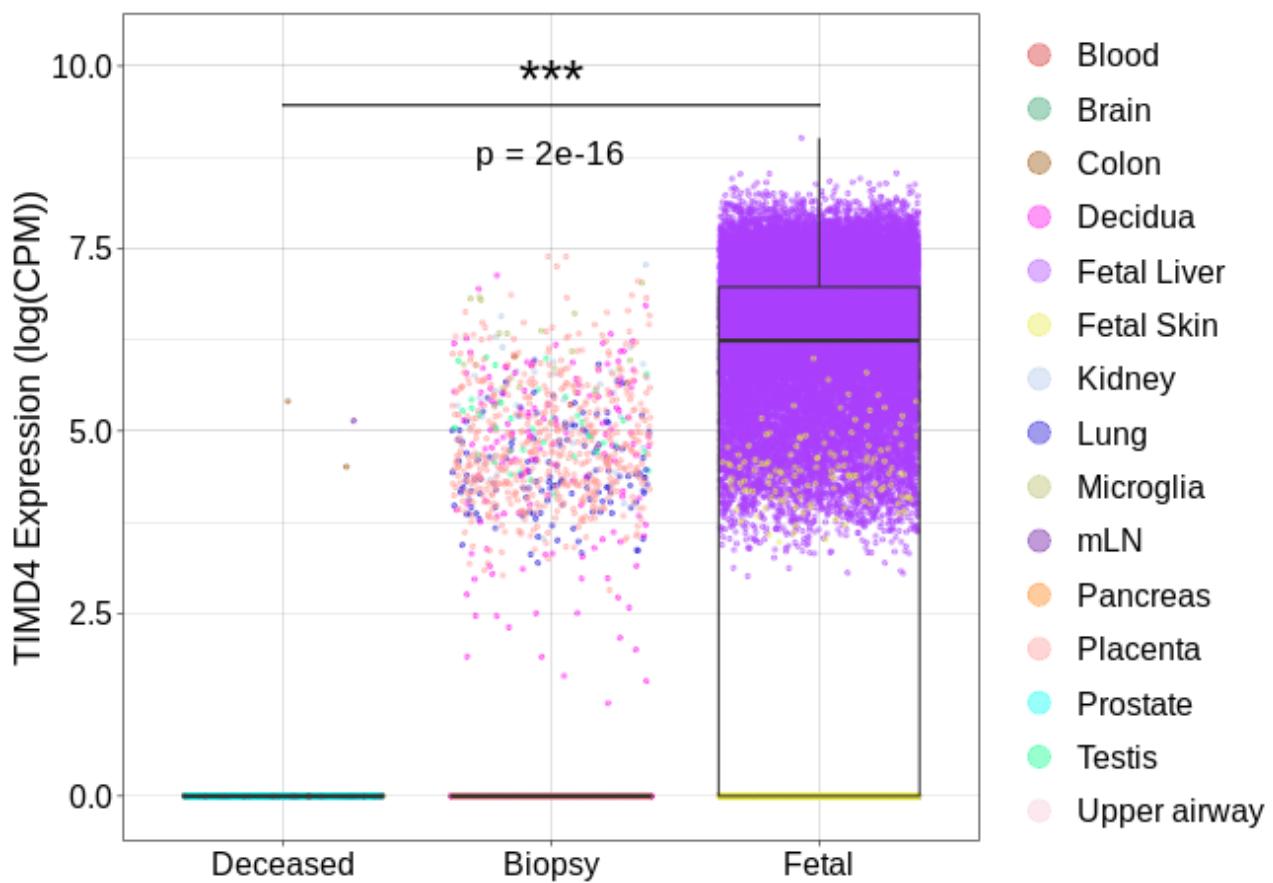
Supplementary Fig 1. (A) A view of the macrophage dataset before batch integration with BBKNN. Here we just used `SCANPY`'s `neighbors` function to compute the connectivities of the UMAP graph. The data is quite spread out and there are multiple clusters for cells in each tissue (blood, decidua, prostate), which don't necessarily represent distinct cell types but more likely represent cells sequenced with different protocols (e.g. 10X Genomics, SmartSeq2, etc.). (B) After using BBKNN to compute the batch-integrated connectivities for the UMAP graph the data are less sparsely dispersed. Cells from the decidua merge into a single group, as do prostate cells, hopefully overcoming batch effects.



Supplementary Fig 2. MA plots showing the average expression levels and \log_2 fold changes for differentially expressed genes in each tissue. The colour of each point represents whether or not it was significantly (adj. p-value ≤ 0.05) differentially expressed (red = significant, black = not significant).



Supplementary Fig 3. MA plots showing the average expression levels and \log_2 fold changes for differentially expressed genes in each cluster. The colour of each point represents whether or not it was significantly (adj. p-value ≤ 0.05) differentially expressed (red = significant, black = not significant).



Supplementary Figure 4. Expression of TIMD4 in macrophages grouped by donor type. TIMD4 was found to be significantly upregulated in fetal cells ($n = 46,424$) compared to biopsied adult cells ($n = 25,396$) and cells from deceased donors ($n = 1,040$). The majority of biopsied cells that expressed TIMD4 were cells from decidual and placental tissue and could represent cells that have migrated in from fetal tissue.

Response to the reviewers

I would like to thank the reviewer for their critical and fair assessment of my work. In the following I address their concerns point by point.

Reviewer 1

Reviewer Point 1 — The reviewer made several comments on the brevity of the manuscript: “Is it possible to include some more content in the Results subsection?”, “Nicely written but brief.”, “Good but too brief.”

Reply: I appreciate this comment from the reviewer and have made several efforts to include more detail in every section of the manuscript - especially figure legends and the discussion section, which were particularly lacking in detail.

Reviewer Point 2 — The reviewer observed that figures could use some improvement: “Re Figures, make sure that the text embedded in each figure is legible and that each figure has a detailed legend.”

Reply: I have addressed this concern by increasing the font size for all text embedded in figures and - as mentioned above - I have expanded on the figure legends for all figures in the manuscript.

Reviewer Point 3 — The reviewer made several important points about the discussion section: “It’s descriptive of the study but has too little interpretation. Good science should be both interesting and useful. You are highlighting some of the interesting aspects of this data analysis but are not identifying reasons why this rich resource may be useful for other studies. You could also critique the study. Highlight its strengths but also its weaknesses, and what are possible future directions? What might be a mechanism to share these data with the scientific community? How does it compare to other resources of single cell RNA-seq data? How have those resources been used by others?”

Reply: I completely agree with the reviewer that “good science should be both interesting and useful” and agree that I may have undersold the potential value of this research in my first draft. I have addressed this by mentioning the various ways in which this study could be useful to others (e.g. for informing the development of more accurate *in vitro* macrophage models). I have also included a paragraph on the limitations of this study (need for more replicates and technical variation from batch effects) and how this work could be improved upon in the future. To address the comments on how this resource could be shared with the scientific community and how it compares to other resources of single-cell RNA-seq data, I created an interactive web portal (<http://tiny.cc/cross-tissue-macrophages>) using the same software that is typically used by the Human Cell Atlas Consortium to share single-cell RNA-seq results. Using this web app, users can explore all data in the final macrophage dataset, including expression levels for 4,000 genes. Hopefully this will allow others in the scientific community to utilise the results from this study and gain further insights.

Reviewer Point 4 — The reviewer made also had a few minor comments: “Manuscript needs a thorough proof-read to identify all typographical and grammatical errors. Make sure that all acronyms have been detailed/explained at first use within the manuscript.”

Reply: I have made my best effort to correct all typos and grammatical errors and make sure that all acronyms are explained.

Again, I would like to sincerely thank the reviewer for their much-appreciated feedback as I feel that these adjustments have really improved the overall quality of my manuscript.