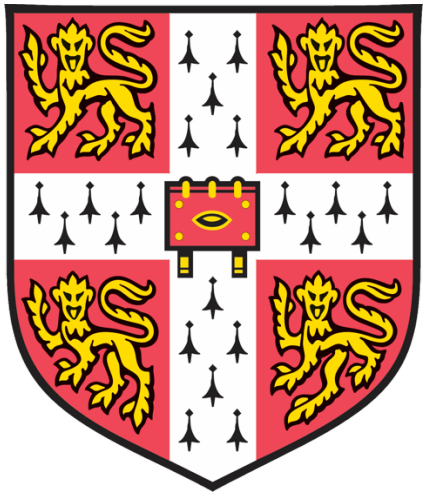
Integrating single-cell transcriptomics with orthogonal experimental approaches and datasets to enhance characterization of the maternal-fetal interface



Angela Elizabeth Zou

Wellcome Trust Sanger Institute
Churchill College
University of Cambridge

This dissertation is submitted for the degree of Master of Philosophy.
20 April 2018

Abstract

During early pregnancy, the maternal immune system shapes the uterine environment to facilitate the implantation and development of an antigenically foreign fetus. Critical to orchestrating this process are unique cell populations within the uterine lining (maternal decidua) and the fetal placenta, which together comprise the maternal-fetal interface. In this study, we sought to comprehensively profile specific cell types and their respective functions at the maternal-fetal interface at single-cell resolution, through analytical approaches combining single-cell RNA-sequencing (scRNA-seq) with other experimental datasets.

scRNA-seq data derived from early pregnancy (6-13 weeks) decidua, placenta, and maternal peripheral blood were analyzed in tandem with other sequencing data, flow cytometric data, and publicly available gene signature and gene expression datasets. We first employed cutting-edge methods to determine via scRNA-seq the maternal or fetal origin of cell populations in the decidua and placenta, both with and without incorporation of genotype data. We then evaluated the utility of bulk RNA-sequencing as quality control for scRNA-seq by analyzing the expression of scRNA-seq-upregulated genes associated with dissociation-induced effects in decidual bulk RNA-sequencing data. Our analysis identified a small number of heat shock protein genes (HSPs) as scRNA-seq-associated technical artifacts. Next, through paired analyses of scRNA-seq with FACS index sorting and mass cytometry data, we further characterized novel subpopulations of decidual NK cells, T cells, and mononuclear phagocytes at both the transcriptome and protein level. Additional comparison of marker genes distinguishing the maternal mononuclear phagocyte populations with known macrophage gene signatures and with tissue-level gene expression data from the Human Protein Atlas showed that these cell subsets do not strictly exhibit classical M1/M2 polarization and express placenta- and endometrium-enriched genes, suggesting their heavily tissue-specific function. Finally, we curated genes with established links to pregnancy complications and fertility and used scRNA-seq to study their cell-type specific expression patterns at the maternal-fetal interface. Our results highlight, among other insights, the specialized role of decidual stromal cell subsets in mediating insulin-like growth factor (IGF) signaling, an important regulator of fetal growth. Cumulatively, these analyses have enhanced the potential of scRNA-seq to robustly unravel novel biology at the maternal-fetal interface.

Declaration of originality

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution. Finally, my dissertation does not exceed the word limit prescribed by the Degree Committee for the Faculty of Biology (it is 19,921 words in length, excluding abstract, figures, tables, references, and appendices).

Preface

The work presented in this dissertation was part of a study conceived by Sarah Teichmann and Roser Vento-Tormo. Tissue dissection and processing, fluorescence-activated cell sorting (FACS), droplet- and plate-based single-cell RNA-sequencing, and mass cytometry (CyTOF) experiments were carried out by Roser Vento-Tormo with help from Margherita Turco, Rachel Botting, Jongeun Park, and Rebecca Payne. DNA/RNA library preparation, whole genome sequencing, and bulk RNA-sequencing were carried out by the Sanger Institute sequencing facility. Mirjana Efremova and Roser Vento-Tormo performed the initial quality control, analysis, interpretation, and annotation of cell populations in the single-cell RNA-sequencing data. I performed the single-cell genotyping analyses, bulk RNA-sequencing evaluation of heat shock proteins, CyTOF data analysis, public gene signature-based analysis and functional annotation of maternal mononuclear phagocyte populations, and cell type-level analysis of genes associated with complications of pregnancy and fertility-related conditions, which are presented in this thesis. Davis McCarthy, Raghd Rostom, and Krzysztof Polanski provided helpful advice and scripts facilitating our approach for genotyping single cells without whole genome sequencing data.

Acknowledgements

First and foremost, I wish to thank my supervisor, Sarah Teichmann, for opening her lab to me over this past year and providing invaluable mentorship and direction for my MPhil research. Sarah's continuous support and thoughtful insights have greatly shaped the development of my work and of this thesis. I am also tremendously grateful to Roser Vento-Tormo and Mirjana Efremova, for their immeasurable help and guidance over the course of this project. Apart from giving me the chance to play around with some truly fascinating, one-of-a-kind datasets, Roser and Mirjana have taught me so much about maternal-fetal biology and single-cell data analysis, and they have provided endless doses of feedback and encouragement during my research and the writing of this thesis.

I am also very indebted to the members of my thesis committee, Jacqui Shields and Martin Hemberg, for taking the time to offer constructive advice about my work throughout this year. A big thank you as well to Davis McCarthy, Raghd Rostom, and Krzysztof Polanski, for the discussions and pieces of code they shared in helping me set up a pipeline to genotype single cells based on single-cell RNA-seq variant calling. And I wish to thank the entire Teichmann group for hosting my stay and for all of the conversations – both enlightening and entertaining – that we have shared at group meetings, in Murray's, or around the water cooler.

I am very lucky to have had my year in Cambridge supported by a Churchill Scholarship from the Winston Churchill Foundation of the United States. Churchill College and the Wellcome Trust Sanger Institute have been a wonderful home for me this year, and my friends on both sides of the Atlantic, both new and old, have been the best source of strength, inspiration, and laughter that I could have asked for. Finally, I want to thank my parents and boyfriend for their unwavering love, and for believing in me, supporting me, and cheering me on through thick and thin.

Table of Contents