Chapter 6

Concluding remarks

In this PhD, I optimised and conducted large-scale single-cell RNA sequencing experiments to study cellular variation in the type I interferon response in fibroblasts of 70 healthy human individuals. Using this dataset, I first studied heterogeneity in the unstimulated state, comparing to *ex vivo* skin data to confirm the relative homogeneity of the *in vitro* cultured fibroblasts used. Using matched whole exome sequencing data, somatic mutations in sub-populations of cells within each donor were detected, and clonal populations identified. Applying cardelino to 32 of the HipSci fibroblast lines identified hundreds of differentially expressed genes between cells from different somatic clones, with cell cycle and proliferation pathways frequently enriched.

Returning to innate immunity, I performed analyses into the variability in the innate immune response across mammalian species, showing a link with evolutionary divergence. Within the human dataset I generated, I characterised the innate immune response at single cell resolution, elucidating the dynamics of the response across donors and defining discrete gene modules. Harnessing the scRNA-seq data, I defined several phenotypes to capture variability in this response. Applying quantitative trait loci approaches to study the genetic basis of this heterogeneity in innate immunity, I

identified 391 response genes with a QTL from either bulk, pseudobulk, or single-cell expression traits.

Moving into the future, experimental work will be required for functional validation of genetic variants altering the innate immune response. One approach is the use of knock down or knock out experiments (for example, through transfection with siRNAs) prior to innate immune stimulation with poly(I:C)/IFN-β. This would allow elucidation of the role of individual genes in the type I interferon response, and could be applied to genes identified through temporal analysis (Chapter 4) or genetic analysis (Chapter 5) in order to investigate regulation within the system. For the validation of specific genetic variants, such as those described in Chapter 5, a CRISPR approach could be used. With this, cell lines could be engineered to contain the alternative genotype at the specific site of interest, prior to monitoring the effect on response. Where variants are suspected to affect binding of transcription factors, this could be confirmed by ChIP-seq experiments. Furthermore, stimulation experiments may be extended to understand how genetic variants relate to susceptibility phenotypes in particular individuals. *In vitro* infection with specific viruses, rather than poly(I:C) and IFN-β, could allow a more focused look at the role of variation in infectious diseases.

As described in Chapter 3, fibroblasts form just one element of the skin millieu. To place this component of the innate immune response within the tissue environment, transcription can be measured spatially. Single-molecule fluorescence in situ hybridisation (smFISH) [209] provides a method to detect individual mRNA molecules of tissue sections, but is limited in the number of transcripts assayable. Recent developments in multiplexing, for example multiplex error-robust FISH (MERFISH) [210] and sequential FISH (seqFISH) [211] have addressed this bottleneck. Applying these methods to innate immune stimulation in the skin would deepen our understanding of the spatial

nature of type I interferon signalling, which may shed light on the heterogeneity in this response.

While this work has focused on variability in transcription, technologies to profile single cells at different molecular levels have vastly evolved over recent years. For example, there are several techniques to capture epigenetic regulation, such as single cell reduced representation bisulfite sequencing (scRRBS) [212, 213] and single-cell methylome and transcriptome sequencing (scMT-seq) [214], which profile DNA methylation. Single cell proteomic assays are developing rapidly, however they are still limited in the number of proteins that can be studied within an experiment. Proteomic methods, such as fluorescence-activated cell sorting (FACS) and cytometry by time of flight (CyTOF) do allow higher throughput of cells than sequencing-based technologies. To gain a more complete picture of the heterogeneity in innate immune response, it will be necessary to utilise and integrate these assays.

This work has highlighted the role of scRNA-sequencing technology in understanding variability within healthy donors, both in the unstimulated (Chapter 3) and activated (Chapters 4 and 5) states. We are currently at the boundary of throughput for the use of single cell sequencing in population genetics. However, the increasing scale of these technologies will soon allow this approach to become more commonplace, allowing application to many biological processes.

Looking further into the future, it is intriguing to speculate on the ability to use our understanding of variability in the innate immune response in a translational context. We are at a point of technological advance in two directions: an increasing characterisation of genetic variability, with initiative such as the 100,000 Genomes Project, and a rapid increase in the resolution and methodologies with which we can profile individual cells. This will need to be accompanied by development of sophisticated computational methods to handle such large -omics data. However, with

increased data availability, we may be able to link molecular phenotypes to physiological responses, incorporating information such as infection history. This will pave the way for translating an understanding of the molecular basis and impact of variability in innate immune response to personalised therapies.