Chapter 5: Discussion

In this thesis I have used multiple RNA-sequencing technologies to generate a transcriptional map of human adult primary microglia and to compare these cells to available *in-vitro* model systems. I have demonstrated that microglia are constantly responding to the CNS environment. In the brain they react to trauma or disease to respond in a disorder-specific manner and it is the complex CNS environment that appears to give rise to the unique transcriptional signature of the primary cells.

5.1 Sequencing primary human microglia

In the second chapter of this thesis, I described the analysis of the largest RNA-sequencing dataset of fresh, adult primary microglial cells to date and demonstrated that microglia display pathology specific activation patterns, particularly following traumatic brain injury. The scale of this study also allowed for comparisons across a variety of clinical factors and demonstrated only a small impact of age or sex on microglial transcriptomes.

Data described in Chapter 2 of this thesis identified potential pathology driven activation patterns in microglial cells through single cell RNA-sequencing. Identification of marker genes for these subpopulations of cells will allow researchers to understand how different microglial phenotypes impact disease outcome or how the activated microglia may play differing roles in microglial responses to trauma or disease. One limitation of this work is that we have not conducted functional validation to verify potential marker genes or to map the functional consequences for each of the populations. Spatial transcriptomics provides a method to combine transcriptional data with *in-situ* hybridization and allows for the identification of cells expressing specific gene markers within a tissue[270,271]. If brain tissue slices could be collected from patients with particular pathologies, such as traumatic brain injury, spatial transcriptomics could be used to not only verify the marker gene sets identified but also see how particular cell populations are distributed within a brain region.

Transcriptomic studies of any cell come with multiple experimental caveats and challenges. The largest challenge is balancing sample access and control of experimental or technical factors that may unknowingly impact microglial transcriptomes. For instance, certain microglial transcriptomes can never be captured using fresh samples. In Chapter 2, we collected “control” patients but it's important to note that these were unlikely to be truly healthy samples. Additionally, tissue samples from certain disease pathologies, such as Alzheimer’s disease, cannot be collected fresh. In order to sequence microglia from these specific cohorts, they must be collected from post-mortem brain tissue. It is not clear how post-mortem delay may impact microglial transcriptomes, especially as data in this thesis has demonstrated that an active CNS environment is vital for the maintenance of the microglial transcriptional signature. While collecting fresh surgery samples removes the potential impact of post-mortem delay on the transcriptome, there are still stages of the single cell sequencing process, such as tissue dissociation, that might introduce transcriptional changes or cell biases. Single-nucleus sequencing may provide a method to overcome some of the technical biases introduced in single cell sequencing, but these technologies are even more costly.

As mentioned above, single cell and single nucleus sequencing technologies are expensive in comparison to bulk RNA-sequencing. Deconvolution techniques allow for the identification of cell types from within bulk data[272]. This means that single cell maps such as the one generated in Chapter 2, could in future be used to deconvolute even larger collections of whole brain tissue samples to identify microglial populations. Increasing sample size within RNA-sequencing studies would allow for more complex genetic association studies, such as subtype specific eQTL studies that could identify specific cell populations that may be involved in disease. Importantly, deconvolution of bulk whole tissue samples also allows the removal of two major steps required for processing of single cell microglial samples, tissue dissociation and cell sorting, which could potentially have an unknown impact on microglial transcriptomes. However, deconvolution does not come without limitations, particularly when identifying rare populations of cells within tissues such as microglia in the brain.

5.2 Modelling primary microglia *in-vitro*

Studies such as the ones described in Chapter 3 and 4 highlight the need for transcriptional comparisons of *in-vitro* model systems to their primary counterpart in order to identify potential limitations of the culture systems. For instance, monoculture iPSC-derived microglia were shown to lack the specialised CNS-linked transcriptional signature seen in primary microglia and, therefore, some of the CNS connected cell functions may also be lacking in these systems. Organoid cultures can provide certain CNS stimuli and single cell trajectory analysis suggested that a population of organoid derived microglia cells moved further along a differentiation pathway. However, gene set enrichment analysis still suggested that certain specialised CNS linked functions were missing in the model systems, such as oligodendrocyte differentiation and myelination. Even more complex brain organoid models are being developed, such as systems with a developing vasculature network[273] or *in-vitro* systems that mimic the BBB[274]. These extensive models may begin to capture more brain functions and lead to further development of specialised cellular phenotypes such as those seen in primary microglia.

However, these complex systems also come with caveats that have to be considered when deciding which model should be used experimentally. They are time consuming to generate, require expensive equipment and reagents and can be more complicated to assay than monoculture systems. Many of these factors mean that brain organoids cannot be used at scale. Large scale genetics studies, such as quantitative trait loci (QTL) experiments, require experimental data from hundreds of samples across varying genetic backgrounds and, therefore, standard organoid differentiation pipelines would not be a feasible experimental tool for these studies. Single cell sequencing has provided a potential way to overcome this issue; it allows for the deconvolution of pools of iPSC lines from within one sample[275] and can attribute single cells back to their original donors. Pooling of iPSC lines allows for the differentiation of multiple donors within one experimental study. This not only reduces the number of required differentiations but also removes some of the batch effects that can arise from comparing different differentiation experiments across different lines.

While iPSC pooling can increase the scalability of organoid differentiations, the protocols remain expensive and complex and so it is important to understand where using these more extensive model systems is necessary. For instance, monoculture iPSC differentiated cells appear to capture some of the transcriptional profile of primary microglia and studies have shown they have comparable behavioural and morphological features of the primary cell type[197–201]. In many cases it may, therefore, be suitable to study certain aspects of microglia function with the more simple monoculture model systems. However, the monoculture models cannot accurately capture how the cells interact with neurons or how they may respond to environmental changes. In these situations more complex models may be required for studying changes in microglial function.

Large scale transcriptional comparisons such as the ones carried out in this thesis could also be used to inform these choices, particularly when studies focus on one specific gene or pathway. Before a model system is chosen, caution should be taken to ensure the gene or pathway of interest is expressed at comparable levels in the model being used to the primary cell type. While this doesn't guarantee comparable responses, it at least provides some evidence that the model system being used has a similar profile to that of primary microglia.

It is also worth noting that all of the studies described in this thesis utilise RNA-sequencing, and therefore, gene expression as a measure of classifying and characterising cell function. However, this does not account for the complicated relationship between gene and protein expression or whether gene/protein expression directly translates to a specific cell function. There are multiple processes following gene transcription that can impact protein expression[276,277] including the translation rate, a protein's half-life and the rate or method by which a protein is transported to its functional location. Variation in any of these stages can lead to a divergence between mRNA levels and protein expression. This is particularly true when cells are transitioning between states and responding to environmental stimuli[277]. This means that the gene expression changes seen in some of the studies described within this thesis may not represent correlated changes in protein levels

and, therefore, functional outputs of the cells. This may be particularly true within the primary microglial single cell dataset where the cells appeared to be dynamically responding to environmental changes.

5.3 Studying microglia in Alzheimer's disease

Microglia are thought to be pathogenic cells in the development and progression of Alzheimer's disease (AD) and therefore each chapter within this thesis has looked at expression of AD linked genes in a variety of contexts. Evidence from the single cell analysis of fresh adult primary microglia in Chapter 1 suggested that microglia respond in a pathology specific manner and studies in both mice and human brain tissue have also demonstrated AD specific activation patterns within microglia[164,166,184]. While some of the genes identified by these studies were expressed across the primary microglia studied in Chapter 1, there was no clear enrichment within a particular cluster which suggested our study did not capture AD specific microglial activation.

It should also be noted that the AD risk gene lists used throughout this thesis were in the most part curated from genes identified in genome wide association studies (GWAS) and these gene lists come with caveats. As described in section 1.5.2 GWAS often identifies a "lead SNP" and associates the SNP to the "nearest gene" despite many of the SNPs falling within the non-coding region of the genome. This may mean that the genes used in this analysis do not represent the true causal risk genes.

Identification of specific gene expression changes that occur in microglia during AD can also highlight genesets and pathways that would need to be mimicked in model systems to accurately capture AD pathology in a dish. Organoid iPSC-based systems have already been used to study AD pathology in a dish, often beginning with iPSC lines containing familial AD mutations to push the cultures towards a disease phenotype[278,279]. With identification of AD specific transcriptional profiles, it may be possible to understand how close *in-vitro* microglia capture the changes seen in

microglia throughout disease progression. One of the major problems with using iPSC differentiated cells to model AD is the maturity of the cultures, age is a major risk factor for neurodegenerative disorders such as AD and capturing that affect in a culture system is challenging as neuronal cultures in particular often more closely represent an immature cell population.

As well as using familial AD mutations within iPSC-derived cultures, it is also possible to engineer late onset AD mutations in iPSC, however there are also caveats with these experiments that should be considered. First, the analysis in this thesis has shown that certain AD risk genes were not expressed at comparable levels in any of the model systems to primary microglia. This means for certain disease genes the effects of risk alleles may not be captured. Even if the expression of the gene of interest is comparable across model systems, the model system chosen is highly dependent on the question and function of interest. For instance, basic microglial functions such as phagocytosis may be well captured by monoculture systems but if the variants are impacting interactions between cell types then more complex models may be required. Unfortunately, for many of the risk alleles associated with AD a clear function has not been identified and so it is difficult to know which model system to choose.

The variants associated with late onset AD risk also tend to have relatively small effect sizes that gradually build throughout life, meaning their effects on individual cell types may be relatively small and not easily seen in cell culture systems. For instance, mutations in the *TREM2* gene in iPSC-derived microglia have been shown to have no impact on cell differentiation, response to stimuli or the ability of microglia to phagocytose compounds[200]. Therefore, it may require the combination of AD risk genes to model AD cell changes in a dish. Polygenic risk scores are statistically based scores that combine genotypes across all risk variants of a disease to predict the likelihood of a person developing a specific trait[280]. Patient-derived cell lines, such as iPSC, could be classified by their polygenic risk scores and differentiated before running functional comparisons across a spectrum of scores. While this would not allow researchers to unpick disease causal mechanisms behind individual genes, it may mean that the subtle impacts of each SNP would combine to generate a more

realistic disease phenotype within cells. Using a spectrum of scores may allow for a greater understanding of how differing levels of disease risk could impact disease progression or development.

5.4 Concluding remarks

In summary, in this thesis I have shown that the microglial transcriptome is constantly reacting to the CNS environment. Initially to develop a unique transcriptional signature and subsequently to respond to disease or trauma. It appears to be signals from the CNS environment that are not well captured by monoculture *in-vitro* model systems. However, more complex systems that culture microglia alongside other neuronal cells and features, such as the BBB, may move the cells closer towards the primary phenotype and the combination of iPSC pooling and single cell sequencing techniques may make large scale studies of these systems more feasible in the future. The potential use of these more complicated and extensive model systems does not always mean they are required. Studies have shown that monoculture *in-vitro* models have certain comparable traits to the primary cell type, such as phagocytosis, whereas other functions of microglia that involve interaction with neuronal signals, like in learning and memory, may only be captured by complex models. It is, therefore, vital to consider the function of interest when identifying an appropriate model system to use for study. This is of particular importance when looking to understand how disease risk genes may modulate cell function. If the model system selected does not accurately capture the linked cellular phenotype then the biological function of a risk gene may be missed.