بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ

In God's Name, Most Loving, Most Compassionate

Using single nucleus RNA-seq to determine the effect of immunosuppression on the transcriptional activity of resident CNS cells in a mouse model of CNS-TB

Compiled by Dr. Ilyaas Amien

Master's thesis submitted in fulfilment of the requirements for degree of Master of Medical Science specialising in Medical Cell Biology

Neuroscience Institute

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Primary supervisor: Dr. Dorit Hockman
Secondary supervisors: Prof. Muazzam Jacobs, Prof. Joseph Raimondo

Department of Human Biology
Faculty of Health Sciences
University of Cape Town
South Africa
November 2024

University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

Plagiarism declaration:

1) I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

(2) I have used the Nature convention for citation and referencing. Each contribution to and quotation in this assignment from the work(s) of other people have been acknowledged, cited and referenced.

(3) This assignment is my own work.

(4) I have not allowed, and will not allow anyone to copy my work with the intention of passing it off as his or her own work.

(5) I acknowledge that copying someone else's assignment or essay, or part of it, is wrong, and declare that this is my own work.

(6) This thesis has been submitted to the Turnitin module.

Date: 17/11/2024

Signed: Dr. Ilyaas Amien

Signed by candidate

"Tuberculosis is as old as humanity itself. It has afflicted kings and queens, poets and politicians, revolutionaries and writers, activists and actors.

Most of its victims, however, are poor, marginalised or malnourished, and the out-of-pocket costs associated with treating TB expose them to financial hardship or drive them further into poverty.

TB is the definitive disease of deprivation. Turning the tide on TB means screening and treatment for those it strikes, and preventing it by addressing its drivers and developing a new vaccine.

Only by working together can we turn the tide against this ancient killer."

Dr Tedros Adhanom Ghebreyesus
Director-General World Health Organization

Table of Contents

LIST OF FIGURES **5**

LIST OF TABLES **7**

SUPPLEMENTARY DATA **8**

CODE AVAILABILITY **8**

ABBREVIATIONS **9**

ACKNOWLEDGEMENTS **12**

ABSTRACT **13**

CHAPTER 1: INTRODUCTION **14**

1.1 TUBERCULOSIS: AN ANCIENT AND PERSISTENT DISEASE **14**
1.1.1 TB PATHOGENESIS 14
1.1.2 CENTRAL NERVOUS SYSTEM TB (CNS-TB) 15
1.2 TRANSCRIPTOMIC STUDIES **18**
1.3 SINGLE NUCLEUS RNA SEQUENCING (SNRNA-SEQ) **19**
1.3.1 HOST-DIRECTED THERAPY 20
1.3.2 SNRNA-SEQ AND TB 20
1.3.3 SNRNA-SEQ AND CNS-TB 21
1.4 AIMS AND OBJECTIVES **22**
1.4.1 AIMS 22
1.4.2 OBJECTIVES 23
1.5 RESEARCH HYPOTHESES **23**

CHAPTER 2: RESEARCH METHODOLOGY **24**

2.1 EXPERIMENTAL DESIGN **24**
2.2 CREATING AN ANIMAL MODEL OF CNS-TB (PERFORMED BY DR NAI-JEN HSU) **25**
2.3 TISSUE COLLECTION AND NUCLEI ISOLATION (PERFORMED BY DR'S NAI-JEN TSU AND EMILY TANGIE) **26**
2.4 SINGLE-NUCLEUS RNA SEQUENCING DATASET GENERATION **28**
2.5 SNRNA-SEQ BIOINFORMATICS ANALYSIS **30**
2.5.1 READ ALIGNMENT AND QUANTIFICATION (CELL RANGER) 30
2.5.2 REMOVING AMBIENT RNA (CELL BENDER – PERFORMED BY DR VUYANI MOSES) 30
2.5.3 DEMULTIPLEXING REPLICATES (SCSPLIT – PERFORMED BY DR VUYANI MOSES) 31
2.5.4 SEURAT SNRNA ANALYSIS PIPELINE 32
2.5.5 DIFFERENTIAL EXPRESSION ANALYSIS 33
2.5.6 VISUALIZING DIFFERENTIAL EXPRESSION DATA 35
2.5.7 GENE ONTOLOGY ANALYSIS 36

CHAPTER 3: RESULTS 37

3.1 snRNA-seq library preparation and initial quality control measures 37
3.2 Pre-processing and quality control of snRNA-seq datasets 40
3.3 Data integration, clustering, and annotation 43
3.4 Differential expression analysis 49
3.4.1 Principal component analysis (PCA) 49
3.4.2 Preliminary conditions tested with DESeq2 51
3.4.3 DEGs in the general TB response 52
3.4.4 DEGs in the TB response specifically in immunocompromised mice ($TNF^{-/-}$) 55
3.4.5 DEGs in TB-infected mice over time 57
3.5 Gene ontology analysis 59
3.5.1 GO analysis of the general TB response 59
3.5.2 GO analysis of the TB response specifically in immunocompromised mice ($TNF^{-/-}$) 65
3.5.3 GO analysis in TB-infected mice over time 68

CHAPTER 4: DISCUSSION 71

4.1 Generating replicates to facilitate differential gene expression analysis 71
4.2 Cell type annotation 72
4.3 Selecting cell types for downstream analysis 73
4.4 Selecting comparisons for further analysis 73
4.5 Relationship between sample nuclei and number of differentially expressed genes 74
4.6 Differentially expressed genes: Candidate gene analysis 74
4.6.1 Candidate genes in the general TB response 74
4.6.2 Candidate genes in the immunocompromised TB response 77
4.6.3 Candidate genes in TB-infected mice over time 80
4.7 Differentially expressed genes: Pathway analysis 81
4.7.1 Pathway analysis in the general TB response 81
4.7.2 Pathway analysis in the immunocompromised TB response 84
4.7.3 Pathway analysis in TB-infected mice over time 85
4.8 Synthesizing candidate gene and pathway analyses 86
4.9 Limitations and future work 86
4.9.1 Limitations related to experimental design 86
4.9.2 Limitations in sample preparation and data processing 87
4.9.3 Limitations in data analysis 88
4.9.4 Limitations of using transcription to understand brain function 88
4.10 Conclusion 89

REFERENCES: 90

List of figures

Figure 2.1: A schematic of the experimental design.

Figure 2.2: A schematic of the mice brain including the site of injection with either saline or Mtb.

Figure 2.3: Graphical representation of the nuclei isolation protocol.

Figure 2.4: snRNA-seq library generation.

Figure 2.5: A schematic of the snRNA-seq bioinformatics pipeline used in this study.

Figure 2.6: A summary of the steps in the ScSplit protocol.

Figure 2.7: A summary of the steps in the DESeq2 protocol.

Figure 2.8: Visualization of the interaction term.

Figure 3.1: Tapestation reports (Day 9).

Figure 3.2: Tapestation reports (Day 17).

Figure 3.3: Graphical summary of quality control metrics post Seurat filtering.

Figure 3.4: Bar graph showing the number of nuclei per sample before and after Seurat filtering.

Figure 3.5: Annotation of nuclei identifies 27 subtypes across the 24 datasets.

Figure 3.6: Stacked bar plot noting the proportion of nuclei (y axis) with respect to each cell type per sample (x axis). Violin plots showing the distribution of the number of transcripts (nUMI) and the number of genes (nGene) per cell type across all the datasets.

Figure 3.7: Dot plot showing the expression of known cell type-specific markers (y axis) with respect to each annotated cell type (x axis).

Figure 3.8: PCA plots for each cell type in the 2D plane showing the first two principal components on the x and y-axis, respectively for the TB vs saline comparison.

Figure 3.9: PCA plots for each cell type in the 2D plane showing the first two principal components on the x and y-axis, respectively for the Day 17 vs Day 9 comparison.

Figure 3.10: Volcano plots showing differentially expressed genes in the TB vs saline condition for each cell type. UpSet plot showing the number of genes intersecting across each cell type in the TB vs saline condition.

Figure 3.11: Volcano plots showing differentially expressed genes in the immunocompromised TB response (with respect to Wildtype). UpSet plot showing the number of genes intersecting across each cell type in the immunocompromised TB response (with respect to Wildtype).

Figure 3.12: Volcano plots showing differentially expressed genes in TB-infected mice over time (Day 17 vs Day 9) amongst the neuronal cell types. UpSet plot showing the number of genes intersecting across each neuronal cell type in TB-infected mice over time.

Figure 3.13: G:Profiler outputs showing the GO driver terms in oligodendrocytes in the general TB response.

List of tables

Table 3.1: Table summarizing Qubit concentrations before and after library preparation during the 10X Genomics protocol.

Table 3.2: Table with quality control metrics generated from Cell Ranger pipeline.

Table 3.3: Table showing the distribution of nuclei per condition after the ScSplit protocol was conducted.

Table 3.4: Table summarising the quality control metrics for each nucleus per sample post Seurat filtering.

Table 3.5 Table showing the number of nuclei per cell type for each sample studied.

Table 3.6: Table showing the number of upregulated and downregulated differentially expressed genes per condition for each chosen cell type

Supplementary Data

Link to supplementary data

Code availability

Link to all scripts

Script 1: Cell Ranger pipeline

Script 2: Cell Bender pipeline

Script 3: ScSplit pipeline

Script 4: Matching ScSplit output with Seurat input requirements (Python)

Script 5: Seurat - Quality control

Script 6: Seurat - Clustering and integration

Script 7: Seurat - Label transfer using Allen Brain Institute dataset

Script 8: DESeq2 pipeline and Volcano plots

Script 9: Upset plots

Abbreviations

Listed in alphabetical order

AIM2 – Absent in melanoma 2
Aqp4 – Aquaporin 4
APOL – Apolipoprotein L
Astro – Astrocytes
AD – Alzheimer's disease
AUC – Area Under the Curve
BAM – Binary Alignment Map
BATF2 – Basic leucine zipper transcription factor 2
BBB – Blood-Brain Barrier
BCG – Bacillus Calmette-Guérin
BCL-2 – B-cell lymphoma 2
BCSFB – Blood – cerebrospinal fluid barrier
BDDN – Brain Damage-Dependent Neurogenesis
BSA – Bovine Serum Albumin
CA2-IG-FC – Cornu-ammonis 2-industrium griseum-fasciola cinerea
CCL – C-C motif chemokine ligand
CFU – Colony Forming Unit
Chodl – Chondrolectin
CNS – Central Nervous System
CNS-TB – Central Nervous System tuberculosis
CPEC – Choroid Plexus Epithelial Cells
CR – Cajal Retzius
CRP – C-Reactive Protein
CSF – Cerebrospinal Fluid
CSPG4 – Chondroitin sulfate proteoglycan 4
CXCL – C-X-C motif chemokine ligand
DCs – Dendritic cells
DEG – Differentially expressed gene
EBV – Epstein-Barr Virus
Endo – Endothelial cells
FC – Fold change
FU – Fluorescent Units
GABA – Gamma-aminobutyric acid
GAD – Glutamic acid decarboxylase
GADD45α – Growth Arrest and DNA Damage 45a
G-CSF – Granulocyte colony-stimulating factor
GDP – Guanosine diphosphate
GEM – Gel Bead in Emulsion
GM-CSF – Granulocyte-macrophage colony-stimulating factor
GO – Gene Ontology
Gp120 – Glycoprotein 120
GPB – Guanylate binding protein
GTP – Guanosine triphosphate
HD – Huntington's disease
HDG – Hippocampal Dentate Gyrus

HDT – Host-Directed Therapy
HIV – Human immunodeficiency virus
IFN – Interferon
IL – Interleukin
ISC – Injection Site Cortex
KEGG – Kyoto Encyclopedia of Genes and Genomes
Lamp5 – LAMP family member 5
LCMV CL-13 - Lymphocytic choriomeningitis virus Clone 13
LEF1 – Lymphoid enhancer-binding factor 1
L2/3 IT CTX – Cortex intratelencephalic layer 2/3
L2/3 IT PPP – Parasubiculum, postsubiculum, presubiculum intratelencephalic layer 2/3
L4 RSP-ACA – Retrosplenial, anterior cingulate layer 4
L4/5 IT CTX – Cortex intratelencephalic layer 4/5
L5 IT CTX – Cortex intratelencephalic layer 5
L5 PT CTX – Cortex pyramidal tract layer 5
L5/L6 IT TPE-ENT – Temporal association/perirhinal/ectorhinal intratelencephalic layer 5/6
L5/6 NP CTX – Cortex near-projecting layer 5/6
L6 CT CTX – Cortex corticothalamic layer 6
L6 IT CTX – Cortex intratelencephalic layer 6
L6b/CT ENT – Entorhinal cortex layer 6b
L6b CTX – Cortex layer 6b
MCL-1 – Myeloid-cell-leukemia sequence-1
MDA – Malondialdehyde
MMP – Matrix metalloproteinases
MPTP – 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine
MS – Multiple-sclerosis
Mtb – Mycobacterium tuberculosis
Microglia-PVM – Microglia and peripheral vascular macrophages
Nef – Negative Regulatory Factor
NETs – Neutrophil extracellular traps
NF-κβ – Nuclear factor kappa-light-chain-enhancer of activated B-cells
NK – Natural Killer
NLRC – Nucleotide-binding oligomerization domain-like receptor caspase recruitment domain
NLRP3 – Nucleotide-binding oligomerization domain-like receptor family pyrin domain containing 3
NMDA – N-methyl-D-aspartate
NPs – Nanoparticles
NPCs – Neural Progenitor Cells
NSB – Nuclei Suspension Buffer
NSCs – Neural Stem Cells
OADC – Oleic Albumin Dextrose and Catalase
OASL1 – 2'–5' oligoadenylate synthetase-like 1
Oligo – Oligodendrocytes
OPCs – Oligodendrocyte Precursor Cells
PBMCs – Peripheral Blood Mononuclear Cells
PBS – Phosphate Buffered Saline
PCA – Principal Component Analysis
PCR – Polymerase Chain Reaction
Pip-Seq – Particle-templated Instant Partition Sequencing
PLGA – Poly(lactic-co-glycolic acid)

PRR – Pattern recognition receptors
Pvalb – Parvalbumin
RNA – Ribonucleic Acid
Rorb – Retinoic acid receptor-related orphan receptor B
ROS – Reactive oxygen species
RT Reagent B – Reverse Transcriptase Reagent B
SA – South Africa
ScRNA-seq – Single-Cell RNA Sequencing
SERPING1 – Serpin1 peptidase inhibitor, clade G, member 1
SIMS – Scalable, Interpretable Machine Learning for Single Cell
SMC-Peri – Smooth muscle cell- pericytes
SN – Substantia nigra
snATAC-seq – Single-Nucleus Assay for Transposase-Accessible Chromatin with high-throughput sequencing
Sncg – Synuclein gamma
SNP – Single Nucleotide Polymorphism
SnRNA-seq – Single Nucleus RNA Sequencing
Sst – Somatostatin
SVZ – Subventricular Zone
Tat – Trans-Activator of Transcription
TB – Tuberculosis
TBI – Traumatic brain injury
TBM – Tuberculous meningitis
Th1 – T-helper 1 cells
TJ – Tight Junction
TMEV – Theiler's murine encephalomyelitis virus
TNF – Tumor Necrosis Factor Alpha
TNF$^{-/-}$ – Tumor Necrosis Factor Alpha knockout
UMAP – Uniform Manifold Approximation and Projection
UMI – Unique Molecular Identifier
VEGF – Vascular endothelial growth factor
Vip – Vasoactive intestinal peptide
Vipr2 – Vasoactive Intestinal Peptide Receptor 2
Vpr – Viral Protein R
VLMC – Vascular/leptomeningeal cell
WHO – World Health Organization
WT – Wildtype
Zeb1 – Zinc Finger E-Box Binding Homeobox 1

Acknowledgements

The submission of this Master's dissertation was only possible because of the incredible love, patience and support shown to me from my family, friends and colleagues.

To my incredible wife, Wassila. Due to working full time there were many days and weekends where I had to sacrifice time with you and yet you always remained supportive. There were times when I did not believe I would complete this thesis, but you were always encouraging. I love you! Thank you to my parents for always being willing to create an optimal environment for me to focus on my dissertation.

To Dorit, my amazing supervisor. Thank you for your constant support and patience with me! There were many occasions when my progress was slow and despite this you were always understanding. Your detailed feedback on my thesis drafts, despite being in the advanced stages of your pregnancy, was incredibly valued.

Vuyani, thank you for the countless hours helping me navigate the complexities of R and bioinformatics analyses. Teresa, thank you for all the advice and code that you shared to assist me whilst completing your own Master's dissertation. A special acknowledgement to Joe and Muazzam for being supportive co-supervisors. To Nai-Jen, thank you for always being available to answer all my mice-related queries.

Lastly, but most importantly, my Creator, without whom nothing in this life is ever achieved.

Abstract

Tuberculosis (TB) is the leading infectious cause of death worldwide and is particularly prevalent in South Africa. Central Nervous System TB (CNS-TB), specifically TB Meningitis (TBM) is the deadliest complication of the disease. Given the high co-prevalence of HIV and CNS-TB, many transcriptional studies have examined the immunocompromised response to CNS-TB, but none have done so at single cell resolution.

Single nucleus RNA Sequencing (snRNA-seq) was used to determine the effect of immunosuppression on the transcriptional activity of resident CNS cells in a mouse model of CNS-TB. The model consisted of two mouse genotypes, namely Wildtype and TNF knockout ($TNF^{-/-}$). *Mycobacterium Tuberculosis* was injected into the somatosensory cortex of the treatment groups, with control samples receiving a saline injection. Brains were harvested at 9- and 17-days post inoculation.

Following library generation, removal of ambient RNA and data processing using Seurat, differential gene expression analysis was used to identify differentially expressed genes (DEGs). Oligodendrocytes had the most DEGs in the general TB response, whereas excitatory neurons had the highest DEGs in the immunocompromised TB response and with disease progression. Potential mechanisms by which these cell types contribute to the pathogenesis of CNS-TB were explored, such as excitotoxicity, demyelination and neurodegeneration. Pathways related to neurogenesis were identified in the immunocompromised TB response. The study adds to the body of knowledge by identifying cell type-specific resident gene expression responses in the context of CNS-TB.

Chapter 1: Introduction

1.1 Tuberculosis: an ancient and persistent disease

Tuberculosis (TB) is a respiratory disease caused by *Mycobacterium tuberculosis* (*Mtb*) and in 2023 (following the COVID-19 pandemic) returned to being the world's leading cause of death from a single agent (1). An ancient illness, *Mtb* has existed for over 70 000 years (2), and despite various treatment strategies, remains a major public health concern with nearly 11 million new cases registered in 2023 (1).

TB is particularly common in South Africa (SA) (3). According to the World Health Organization (WHO), the 2023 incidence of TB in SA was 427 active cases per 100 000 people (4). In addition, an estimated 55 000 South Africans died from the disease that same year, equating to a case fatality ratio of 22% (4). TB is also often described as a social disease, due to its association with malnutrition, homelessness and poor ventilation, all of which are prevalent in SA, underscoring the importance of primary prevention strategies (5).

1.1.1 TB pathogenesis

Mtb is transmitted via aerosol droplets, which are spread by TB patients when they cough, sneeze or talk (5). After entering the infected host, *Mtb* travels through the respiratory tract and into the lungs. As few as three *Mtb* bacilli are required in the alveolar space to cause infection and activate the innate immune response (6). On a molecular level, the innate response to TB leads to the secretion of numerous cytokines such as interleukin (IL)-1 alpha (α) and beta (β), tumour necrosis factor α (TNF), IL-6 and IL-12. In addition, antimicrobrial peptides along with lipoxins and prostaglandins are also released (7). Lipoxins are thought to induce necrosis in TB and contribute to host survival, whereas prostaglandins may induce apoptosis (7).

The secretion of IL-12 and chemokines such as CCL-19 and CCL-21 stimulate resident dendritic cells (DCs) to migrate to local lymph nodes and promote the differentiation of T-helper 1 (Th1) cells. These cells drive the release of interferon gamma (IFN-γ) and TNF at the site of infection (7), which together activate macrophages and DCs to initiate the formation of a granuloma, a collection of immune cells designed to contain the TB infection. In some patients, *Mtb* may be eradicated completely whereas in others the granuloma persists, causing patients to enter a latent state of infection (8, 9, 10). In other patients, especially those with compromised immune systems, such as in Human Immunodeficiency Virus (HIV) and uncontrolled diabetes, TB infection will quickly progress to active disease and manifest with symptoms, such as cough, weight loss, night sweats and fever (11). The ongoing immune response in these patients may lead to tissue destruction in the lung and the dissemination of *Mtb* to other parts of the body, such as the lymph nodes, pericardium, genitourinary tract and the central nervous system (12, 13, 14).

1.1.2 Central nervous system TB (CNS-TB)

In the central nervous system (CNS), TB can manifest in three ways, namely TB meningitis (TBM), intracranial tuberculoma or a brain abscess (15). These disease processes can also be interrelated since tuberculomas can rupture, resulting in *Mtb* entering the subarachnoid space and ventricles causing TBM. However, not all TBM cases are thought to originate from tuberculomas (16, 17).

TBM is the most severe and deadly complication of TB, with neurological complications such as hydrocephalus, cognitive impairment and epilepsy diagnosed in up to half of all affected patients (18, 19). Expectedly, CNS-TB is also more prevalent in patients with compromised immune systems (20) and is thought to result from haematogenous spread of the TB bacilli from the lung. Less frequently, TB dissemination to the brain can also occur secondary to reactivation of latent TB from another region in the body (21).

To enter the CNS, *Mtb* must bypass the blood brain barrier (BBB), a selectively permeable layer that separates brain parenchyma from the systemic circulation. Several theories have been proposed to account for TB dissemination into the brain. Firstly, *Mtb* is thought to traverse the BBB via infected macrophages and neutrophils, known as the 'Trojan horse' mechanism (22). Secondly, TB bacilli may invade the brain endothelium, likely through mediation by the *Mtb* Protein Kinase D (pknD) (Rv0931c), a serine/threonine kinase which also plays a role in signal transduction and cell wall biosynthesis (23). Other mediators that have been implicated in BBB dysfunction include matrix metalloproteases (MMP), vascular endothelial growth factor (VEGF), caspases, and cytokines such as IFN-γ, IL-1*β*, TNF, IL-8 and IL-6, but these mechanisms are currently not well understood (24). Intriguingly, one 2022 study found that TB induces disruption of the blood – cerebrospinal fluid barrier (BCSFB), and not the BBB following an intratracheal inoculation in mice with N15, a virulent strain of *Mtb* (25). Therefore, more work is needed to understand the precise mechanisms by which *Mtb* reaches the CNS.

1.1.2.1 Resident CNS cells involved in CNS-TB

Once in the brain, TB proceeds to predominantly infect microglia, the resident immune cells of the CNS (26). Microglia recognise *Mtb* via pattern recognition receptors (PRR) and subsequently secrete a host of inflammatory cytokines, such as *IL-1α, IL-1β, IL-6, IL-10, IL-12p40, IL-18, TNF, Granulocyte colony-stimulating factor (G-CSF)*, *Granulocyte-macrophage colony-stimulating factor (GM-CSF*), C-C motif chemokine ligand 2 (CCL2*), CCL5*, *C-X-C motif chemokine ligand 9 (CXCL9*) and *CXCL10* (26).

Whilst these responses are important to contain the infection, they may also contribute to excessive inflammation and immune disruption (22, 27). In support of this, one 2024 study showed that subsequent to infection, cerebrospinal fluid (CSF)-enriched microglia (Macro_C01) are involved in activation of the complement system, resulting in excessive inflammation (28).

In addition to microglia, other CNS cells have been implicated in CNS-TB, although less described. For example, *Mtb* has been found to be incorporated by astrocytes, with one human culture noting an average of 1.3 bacilli per astrocyte in a model of CNS-TB (27). Neurons have also been identified as hosT-cells for *Mtb*. One South African study showed how *Mtb* can be incorporated by neurons using

both *in vitro* and *in vivo* conditions. However, the study argued that the number of bacilli is less abundant compared to microglia (29). The exact mechanism by which *Mtb* enters neurons is unclear. However, one proposed route may involve phagocytosis of the bacterium by neurons (30).

In addition to the presence of *Mtb* in neurons, neuronal destruction has also been described as a complication of CNS-TB. One study suggested a link between mild traumatic brain injury (TBI) and CNS-TB, arguing that the chronic neuro-inflammatory processes in both pathologies may result in neurodegeneration (31). In spite of the above, there remains a paucity of research examining the role of resident CNS cells in CNS-TB.

1.1.2.2 Innate immune response in CNS-TB

The activation of microglia described above also initiates the development of inflammasomes, a large multiprotein complex consisting of two pathways - canonical and non-canonical. The canonical pathway results in the production of specific inflammasomes, namely NOD-like receptor family pyrin domain containing 3 (NLRP3), Absent in melanoma 2 (AIM2) and NLR family CARD domain containing 4 (NLRC4). NLRP3 is activated upstream by caspase-1 and subsequently generates IL-1β and IL-18. AIM2 is released in response to intracellular nuclei acids and NLRC4 is activated by gram-negative secretion pathways (32). In contrast, the non-canonical inflammasome refers to the caspase 4/5-dependent pathway in humans, and the caspase 11 pathway in mice. The outcome of both inflammasome pathways is the activation of pyroptosis-mediated cell death and the release of pro-inflammatory markers (32).

Apart from inflammasomes, neutrophils also have been shown to phagocytose *Mtb* and can release cytotoxic proteins such as myeloperoxidase and neutrophil elastase to kill bacteria (33). This process, referred to as neutrophil extracellular traps (NETs) has been largely described as neuroprotective (34). NETs also drive platelet activation as well as the complement pathway (34).

Another component of the innate immune response in CNS-TB is TNF. In addition to its role in maintaining granuloma integrity in the lung, TNF is produced by various CNS cells such as microglia, macrophages, neutrophils, T-cells and DCs. In the CNS, TNF increases the permeability of the BBB, allowing for increased cytokine release and peripheral immune cell recruitment (35). TNF, particularly myeloid and lymphoid derived TNF, is considered essential in the immune response against CNS-TB.

The crucial role of TNF in both TB and CNS-TB has led some studies to explore the impact of TNF knockout ($TNF^{-/-}$) in TB models. First described in 1996, $TNF^{-/-}$ was achieved by manipulation of embryonic stem cells along with homologous recombination techniques to disrupt the TNF gene (36). One study, using a murine model of $TNF^{-/-}$ showed that intracerebral inoculation of *Mtb* resulted in uncontrolled bacterial replication and subsequent mortality (37). Interestingly, the same study found that neuron-specific $TNF^{-/-}$ had no significant influence on morbidity and mortality (37). This underlies the importance of other sources of TNF, such as myeloid and lymphoid cells, which are collectively required in the immune response to CNS-TB (38). Global TNF deficiency was also associated with a decreased number of activated innate immune cells and was shown to disrupt cytokine and chemokine synthesis in the brain (37). Other studies have shown how infliximab, a TNF-neutralizing

drug, contributes to increased host susceptibility to TB (39). Taken together, these studies highlight the critical role played by TNF in CNS-TB and how $TNF^{-/-}$ attenuates the innate immune response and increases susceptibility towards severe disease. Thus, $TNF^{-/-}$ sufficiently models the immunocompromised brain.

1.1.2.3 Adaptive immune response in CNS-TB

The dissemination of *Mtb* into the subarachnoid space also initiates a local T-cells response (40). This response is partly mediated by the release of CD137L by activated microglia (41). In addition, neutrophils have also been shown to 'prime' CD4+ T-cells in TB by mediating DC movement and antigen presentation (42). T-cells, specifically Th1, are critical to sustain microglia activation through the release of IFN-γ (43).

T-cells impairment is considered a hallmark of both TB and CNS-TB (44), with one paediatric study demonstrating downregulation of T-cells development and function over the disease course using blood bulk Ribonucleic Acid (RNA) sequencing of HIV negative patients with CNS-TB (45). This was supported by an additional study, which found that children with CNS-TB have downregulated T-cells signalling in blood transcripts in comparison to controls (31). Collectively, the immune response in CNS-TB has been shown to involve innate immune activation coupled with attenuated T-cell functioning.

1.1.2.4 Excitotoxicity and CNS-TB

Beyond the classical immune system, other factors have been shown to play a role in the immune response to CNS-TB, such as excitotoxicity. Excitotoxicity is defined as the ability of excitatory amino acids (particularly glutamate) to contribute towards neuronal death as a result of excessive exposure (46). Glutamate, the main excitatory neurotransmitter of the brain, is known to act on post-synaptic neurons and neuroglia (47). It has been suggested that the infarction often associated with CNS-TB can lead to stimulation of glutamate release (48). Glutamate binds to neuronal receptors, driving calcium influx into the cell which then causes the release of various pro-injury substances, such as calpains, caspases, endonucleases and phospholipases. The subsequent cellular damage and membrane breakdown ultimately results in necrosis or apoptosis (49, 50).

In addition to glutamate, recent studies have implicated tryptophan levels with excitotoxicity, with one study noting higher lumbar CSF tryptophan concentrations in CNS-TB. The study also found that higher tryptophan levels in CNS-TB were associated with increased mortality (51). Taken together, these studies emphasise how the immunopathogenesis of CNS-TB involves an array of pathways that are related, but not confined to, the classical immune system.

1.1.2.5 HIV and CNS-TB

Another important factor in the immune response to CNS-TB is HIV infection. HIV/TB co-infection is remarkably common in SA. Of the 450 000 South Africans who are diagnosed with TB every year, 270 000 are already known with HIV (52). Therefore the study of TB (and CNS-TB) in the

immunocompromised setting is critical. TB patients who are co-infected with HIV are known to be at an increased risk of CNS-TB, with the risk inversely proportional to the CD4 count (53). Moreover, HIV positive patients concurrently infected with CNS-TB are known to have higher bacterial loads and CNS neutrophil counts (54). In addition, mortality in CNS-TB is three times greater in HIV patients in comparison to those not infected (55). Therefore, since HIV infection is a major risk factor for both morbidity and mortality in CNS-TB, it is essential to understand the ways in which HIV contributes to CNS-TB pathogenesis.

From a mechanistic perspective, HIV is thought to facilitate *Mtb* entry into the brain with some studies showing how HIV-related proteins can disrupt the integrity of the BBB. Some of these proteins include glycoprotein 120 (gp120), a potent neurotoxin detected in the serum of HIV patients (56), which is known to increase the permeability of the endothelial layer and facilitate monocyte migration in *in vitro* models (57). Secondly, HIV Trans-Activator of Transcription (Tat) Protein, which is necessary for viral replication (58), has been shown to regulate BBB integrity through oxidative stress and modulation of endothelial tight junction (TJ) protein expression (59, 60). Thirdly, HIV Negative Regulatory Factor (NEF) protein, which downregulates CD4 expression in T-cells (57), had been linked with BBB disruption likely through the induction of MMPs (61). Lastly, HIV Viral Protein R (Vpr), which is an important component of the HIV cell cycle (62), has been shown to alter astrocyte metabolism, which in turn can affect BBB integrity (63). In spite of the above, the precise mechanisms of BBB disruption are still being investigated and even less is known about the role of resident CNS cells in HIV patients with CNS-TB.

1.2 Transcriptomic studies

Transcriptomics refers to the study of all the RNA transcripts produced by the genome under specific conditions or in a specific cell type. Transcriptomics involves calculating the levels of RNA, most commonly messenger RNA (mRNA) in samples, which can be useful to measure gene expression and identify cellular processes (64).

Transcriptomics has been used to study a wide range of infectious diseases. Prior studies have used transcriptomics to identify 'transcriptional signatures' to distinguish between disease processes. For example, one study identified genes that could reliably delineate bacterial and viral infections (65). Moreover, a 2007 publication used transcriptomics to discriminate between acute infections and non-infectious illnesses (66). Thirdly, a 2013 study demonstrated that apoptotic pathways were upregulated in patients with rapid disease progression in HIV, in comparison to those with slower disease progression (67). Furthermore, another study discovered transcriptional differences in the host gene expression response to influenza and pneumococcal vaccines. The study showed how influenza vaccines evoked anti-viral pathways, whereas inflammatory signatures were more prevalent in response to the pneumococcal vaccine (68).

Transcriptomic studies have also been used in the context of TB. One area in which transcriptomics has been applied is in the differentiation between latent and active TB, where one study developed a

seven blood gene signature that was able to distinguish the two disease states with an area under the curve (AUC) of 0,88, suggesting strong predictive power (69).

Other studies have attempted to use transcriptomics to delineate TB from other infectious diseases, which may have similar phenotypes. For example, sarcoidosis and TB both have similar clinical presentations. A comparison of the whole blood genome-wide transcriptional profiles demonstrated enrichment of IFN-inducible responses in TB, whereas eukaryotic initiation factor 2 signalling was prominent in sarcoidosis (70). The study highlights the utility of transcriptomics in the distinguishment of similar clinical diseases. Another study used whole blood RNA sequencing to prospectively identify patients at risk of developing active TB (71), which if widely available could serve as useful screening tools.

Transcriptomic studies have also been utilized to investigate CNS-TB. A bulk RNA analysis studied the transcriptional activity between the ventricular and lumbar compartments in paediatric patients with CNS-TB (31). CNS-TB was shown to have different transcriptional signatures in the two compartments, with the ventricular profile characteristic of brain injury and decreased neutrophil markers. This was in contrast to the lumbar profile which showed increased cytokine signalling (31). The same study also identified markers of inflammasome activation in the peripheral nervous system that were not observed to the same extent in the central nervous system (31). Therefore, bulk RNA sequencing can be useful in providing more clarity on the nature of immune responses in different areas of the body.

Bulk sequencing has also been used to develop gene signatures in CNS-TB. One study proposed a four gene biomarker (*MCEMP1*, *NELL2*, *ZNF354C* and *CD4*), which was shown to confidently predict mortality in CNS-TB patients, with or without HIV infection (HIV negative AUC: 0,88, HIV positive AUC: 0,86)(72). Another 2024 study using whole blood showed that *guanylate-binding protein 3* (*GBP3)* was able to delineate CNS-TB from other neurological infections (73). The use of transcriptional signatures in CNS-TB holds much promise, especially in consideration that CNS-TB patients present with non-specific symptoms, making diagnosis challenging (40).

Despite the benefits mentioned above, one of the challenges of RNA sequencing is that the gene expression data are usually analysed in bulk, meaning that the data sequenced would represent an average of the genes transcribed from the collected cells. As a result, relevant differences between specific cell types may be masked (74). Another disadvantage is that rare cell subpopulations would be more difficult to identify (74). In view of these limitations, RNA sequencing tools with single cell/nucleus resolution have become increasingly utilised in the literature.

1.3 Single nucleus RNA sequencing (snRNA-seq)

snRNA-seq emerged as an exciting technique that allows for the transcriptome of single nuclei to be quantified and analysed in high resolution. It differs from single cell RNA sequencing (scRNA-seq) as only the nuclei are collected (75). One of the commonly used tools to profile gene expression on a single nuclear level is 10x Genomics snRNA-seq. This particular workflow involves capturing isolated nuclei in microfluidic droplets along with beads coated in uniquely barcoded primers. RNA is released

from the nuclei to be captured by the primers and then undergoes reverse transcription to generate barcoded cDNA for sequencing, library generation and further downstream analysis (76).

Used widely in cellular biology, snRNA-seq provides insight into heterogenous cell populations, new cell types and regulatory pathways. One study identified 47 cell type subclasses in the somatosensory cortex and hippocampal CA1 region of a mouse brain (77). In addition, scRNA-seq was used to study the cells of the visual cortex in mice and noted 49 different cell types (78). The technique was also used to identify a specific subset of microglia associated with Alzheimer's disease (AD) (79). Furthermore, snRNA-seq allows for the identification of developmental hierarchies and lineage relationships along with the profiling of healthy and diseased tissue, which is useful when studying pathogenesis (80).

1.3.1 Host-directed therapy

snRNA-seq also enables the analysis of cell type-specific host responses to infection. The characterisation of host cell responses opens the possibilities to host-directed treatment. Host-directed therapy (HDT) is a treatment approach that seeks to manipulate the hosts' immune system to enhance the response to infection, in comparison to directly targeting the pathogen. For example, imiquimod, a toll-like receptor agonist, is used topically to treat external genital and perianal warts (81). Host-directed therapy has also been explored in oncology. One scRNA-seq study showed how S100A4, a calcium binding protein, regulates immunosuppressive T and myeloid cells in glioblastoma patients. Intriguingly, a deletion of S100A4 in non-cancer cells was found to promote survival (82).

HDTs have been increasingly studied in the context of TB, in part due to the prolonged treatment duration (6 months in drug-sensitive TB) as well as the rise of antibiotic resistance to standard TB therapy (83). The goals of HDT in TB involve modulating immune mechanisms and reducing excess inflammation, thereby preventing organ damage and preserving lung function (83). The most commonly used HDT in TB is corticosteroids, which are thought to reduce overall inflammation (84). Curiously, a 2022 study identified a role for nanoparticles (NPs) as HDT against TB (85). Poly(lactic-co-glycolic acid) (PLGA) was used to produce immunotherapeutic NPs that were able to activate *Mtb*-infected macrophages. The NPs were also shown to reduce bacterial burden over 72 hours without producing excess host cell toxicity (85). Therefore, HDT represents a promising area of future research into novel TB treatment strategies.

1.3.2 snRNA-seq and TB

Multiple studies have used sn/scRNA-seq to gain more insight into the pathogenesis and subsequent immune response to TB. One study used scRNA-seq technology to explore the functional heterogeneity between macrophages in a murine model of pulmonary TB. Important functional differences were identified between alveolar and interstitial macrophages following *Mtb* infection (86). The study was able to identify four discernible subgroups in both populations and showed that both alveolar and interstitial macrophages have permissive and controller subtypes. For example, one interstitial macrophage (IM) subtype was found to be present in the non-infectious state and expresses *ZEB2*, which is involved with tissue maintenance. In response to infection, the subtype

responds by upregulating genes related to resistance towards TB progression (86). Another 2017 study used Seq-well, a microfluidic-based scRNA-seq system, to profile human macrophages infected with *Mtb* (87). Baseline functional heterogeneity was demonstrated amongst the different macrophage subclusters, which were associated with cell growth, hypoxia and metabolism respectively (87). These studies collectively highlight the role of sn/scRNA-seq in determining functional heterogeneity between cellular subtypes.

snRNA-seq has also been used to add insight into the function of granulomas. Using a zebrafish-marine mycobacteria model along with the analysis of *Mtb*-infected macaques, one 2021 study showed how type 2 immune responses are also involved in granuloma integrity (88). Specifically, it showed how type 2 signals were associated with macrophage epithelization, the process whereby macrophages take on functional characteristics of epithelial cells and form barriers to contain the spread of infection (88). A second study used scRNA-seq to analyse the granulomas of TB-infected macaques and demonstrated extensive heterogeneity in granuloma composition (89). The study showed how granulomas with greater bacterial clearance contained higher levels of IFN-γ and TNF. This was in contrast to other granulomas in which *Mtb* persisted, which were found to have increased levels of regulatory T-cells (89). These studies demonstrate how sn/scRNA-seq can be used to obtain a more detailed understanding of the immune processes within a cellular environment, such as granulomas.

snRNA-seq has also been used to delineate between active and latent TB through an analysis of cellular subtypes. A 2020 study isolated Peripheral Blood Mononuclear Cells (PBMCs) from healthy, latent TB and active TB patients before conducting scRNA-seq using 10x Genomics. Curiously, the group observed a loss of a natural killer (NK) cell subset (CD3-CD7+GZMB+) in active TB compared to the other patient groups, with healthy controls showing the largest number of cells. Interestingly, the group observed an increase in the frequency of the NK subset following TB treatment (90). Therefore, this NK subset has been suggested as a novel marker of active TB.

Lastly, scRNA-seq has been used to examine the transcriptomic changes associated with disease progression and reinfection in TB. One study found higher proportions of inflammatory immune cells and lower fractions of lymphocytes in patients with severe TB, suggesting that severe pulmonary TB is associated with lymphopaenia (91). The study also noted an upregulation of apoptotic pathways in patients with severe disease and the loss of lymphocytes was suggested as a contributor to this finding (91). A second study used scRNA-seq to study the role of $CD4^+$ T-cells in TB reinfection (92). Using antibody-based depletion of $CD4^+$ T-cells prior to TB reinfection of macaques, the study showed how $CD4^+$ T-cells play a protective role by reducing inflammation and neutrophilia along with decreasing disease severity in TB reinfection (92). Both of the above studies highlight the importance of the adaptive immune system in TB severity and therefore the modulation of this system serves as a promising area of future research.

1.3.3 snRNA-seq and CNS-TB

Although there are many studies using sn/scRNA-seq to investigate pulmonary TB, the technology has been scarcely applied in the context of CNS-TB, with only two studies identified at the time of writing. The first study injected attenuated Bacille-Calmette-Guerin (BCG) vaccine in the cerebellum of C57BL6

mice to model CNS-TB, with saline injected into the control group (93). The infected mice were found to have an expanded subarachnoid and ventricular space along with an increased density of infiltrating cells, all of which are common observations in CNS-TB (93). SnRNA-seq was performed and lead to the identification of 15 cell types. Notably, ependymal cells were the most represented cell type followed by microglia. The study found that upregulated genes in the CNS-TB group were associated with antigen presentation and chemokine signalling, suggestive of a hyperactive immune response. Interestingly, cell type-specific responses were also noted, with *CXCL1* upregulated in choroid plexus epithelial cells (CPEC) and *CCL3* overexpressed in microglia (93). In addition to the above findings, genes related to oxidative phosphorylation was found to be downregulated in the neuronal population of CNS-TB mice. The authors inferred from this that in the context of CNS-TB, neuronal metabolism activities may be dysregulated (93). Remarkably, gene ontology analysis in the same study found an association between CNS-TB and neurodegenerative disease (93).

The second study used scRNA-seq to sequence PBMCs and CSF cells from children with CNS-TB using 10x Genomics (28). Children with viral or cryptococcal meningitis were used as controls and six distinct cell types were identified, namely CD4, CD8, NK, Myeloid, B-cells and proliferating T-cells (28). In addition, the study analysed the subtypes of each cell and demonstrated functional heterogeneity between cell types (28). For example, CD4_C04 cells were enriched for glycolysis and lipid metabolism, whereas CD4_C05 were associated with typical T regulatory cell activity and higher cytokine expression. Amongst the CNS-TB patients, the study also identified higher levels of complement protein C1Q, C-Reactive Protein (CRP), TNF and IL-6 amongst microglia in the CSF compared to blood. The authors concluded from these findings that complement activation, mediated by microglia may be a key driver of excessive inflammation in the CSF of patients with CNS-TB (28).

Even though the two aforementioned studies have provided insight into the immunopathogenesis of CNS-TB with single cell resolution, significant gaps remain in the literature. For example, the first study used the BCG vaccine (and not live *Mtb)* to model CNS-TB. This is important because BCG lacks important TB virulence factors such as ESAT6 and CFP10, and therefore may not be fully representative of *Mtb* infection (93). In addition, the second study only examined CNS-TB transcriptional activity using blood and CSF samples (and not resident CNS cells). Thirdly, no studies to our knowledge have used snRNA-seq to investigate CNS-TB in the immunocompromised setting, which is important given the high co-prevalence between HIV and TB (52).

1.4 Aims and Objectives

1.4.1 Aims

In view of the above, the aim of this study was to use snRNA-seq to determine the transcriptional activity of resident CNS cells in a general and immunocompromised setting using a mouse model of CNS-TB.

1.4.2 Objectives

1. To develop a murine model of CNS-TB in both immunocompetent and immunocompromised mice, using $TNF^{-/-}$ as a proxy of suppressed immunity.
2. To perform tissue collection, nuclei isolation and dataset generation of samples followed by the snRNA-seq bioinformatics pipeline.
3. To identify differentially expressed genes and thereafter conduct candidate gene and pathway analyses.

1.5 Research hypotheses

Whilst the explorative nature of this study meant that hypotheses were generated and not tested, some preliminary hypotheses were made. Firstly, I expect to see significant transcriptional differences in TB-infected mice (both genotypes) compared to saline-injected mice (both genotypes) in all cell types. I anticipate that differentially expressed genes in the general TB response (compared to saline) will be associated with neuroinflammation that will worsen with disease progression. I also expect to identify cell type-specific responses between CNS cells, such as excitotoxicity amongst neurons of TB-infected mice compared to saline controls. I also anticipate seeing increased markers of cytotoxicity in TB-infected immunocompromised mice with respect to immunocompetent controls.

Chapter 2: Research Methodology

2.1 Experimental design

All procedures received ethical approval for animal research from the UCT Animal Ethics Committee (AEC018/017, AEC018/028, AEC024/002). The experiment was designed to test eight conditions (Figure 2.1). Wildtype C57B6 (WT) and Tumour Necrosis Factor knockout mice ($TNF^{-/-}$) were used to model immunocompetent and immunocompromised mice respectively (94, 95). In order to obtain homozygous $TNF^{-/-}$, Polymerase Chain Reaction (PCR) genotyping was performed on each litter by Dr Nai-Jen Hsu until sufficient numbers of homozygous mice were bred for the experiment. The age of these mice ranged between 6 and 12 weeks prior to culling. Half of the mice were infected with *Mtb* and the other half were injected with normal saline. In addition to serving as a control for TB infection, the normal saline samples also represented controls for the injury sustained from the creation of the burr hole and subsequent injection. In addition, half the mice were culled at day 9 and the other half at day 17 respectively post saline injection or TB infection. The separate time points allowed for monitoring of disease progression. A previous study showed that $TNF^{-/-}$ mice succumbed within three weeks after TB infection (37). Hence, the day 17 time point was chosen with the presumption that sufficient brain architecture would be present to ensure viability of cells for nuclei isolation.

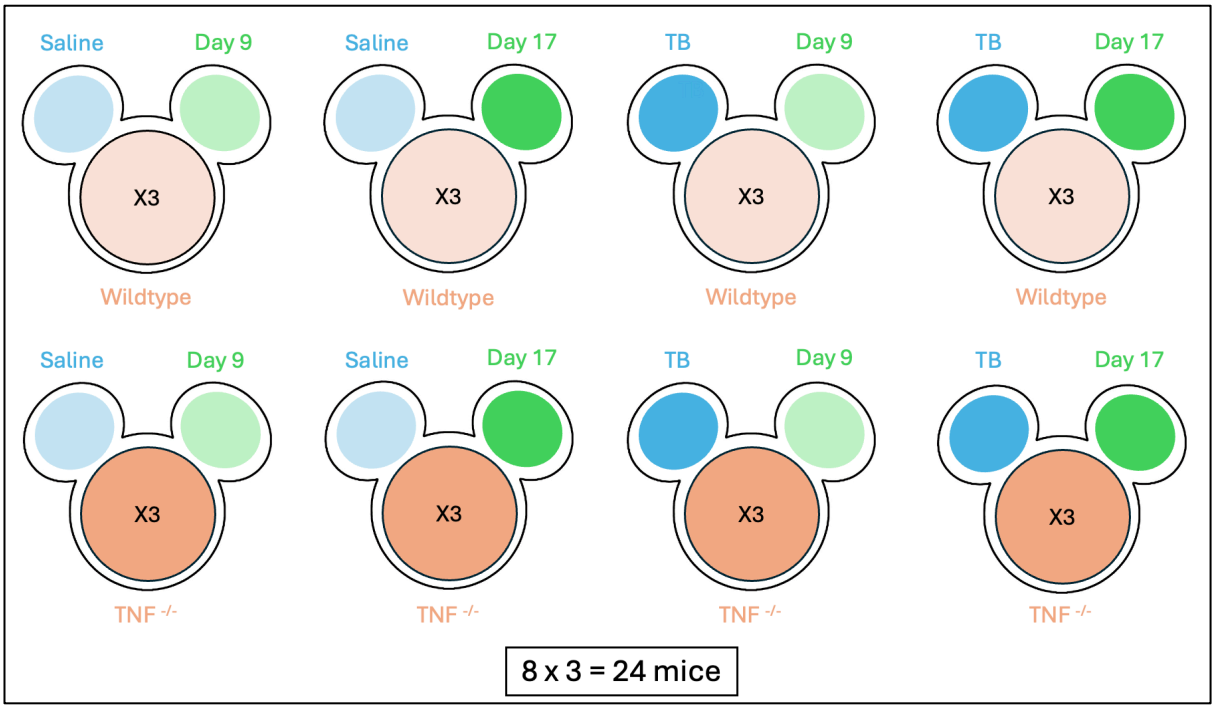


Figure 2.1: **A schematic of the experimental design**. The mice were stratified into three main categories: immune status (immunocompetent vs immunocompromised), infection status (saline treated vs TB-infected) and day of brain collection (day 9 vs day 17 post saline injection/TB infection) – yielding eight sample conditions. Three mice were utilised per condition resulting in 24 mice being used in the whole experiment.

Three mice were used per sample condition (i.e. 3 x 8 = 24 mice), however the dissected tissue from each of the three mice was pooled for nuclei extraction and subsequent snRNA-seq library preparation. In addition to a reduction in cost, the above approach ensured that replicates were placed under the same experimental conditions during library preparation, which would theoretically minimise technical variability between the samples. The data from individual mice was retrieved using ScSplit (96), a bioinformatics tool that aims to restore replicates within a pooled sample (See Section 2.5.3).

2.2 Creating an animal model of CNS-TB (performed by Dr Nai-Jen Hsu)

Mtb strain H37Rv was grown in Middlebrook 7H9 broth at 37°C, in a solution with 10% Oleic Albumin Dextrose and Catalase (OADC) growth supplement and 0,5% Tween-80 until log phase was reached. The resulting sample was aliquoted and stored at -80°C. Thereafter, the sample was thawed, passed 30 times through 29 gauge needles before being diluted in sterile saline.

All mice were anesthetized with halothane and then positioned in a stereotactic frame. The heads were cut with scissors and a small burr hole was created inside the skull over the somatosensory cortex, anterior to the bregma and to the left of the midline (Figure 2.2). Subsequently, the mice were intracerebrally injected with either *Mtb* strain H37Rv or 0.9% normal saline in double distilled water (97). The mice injected with TB were inoculated with $1x10^4$ – $1x10^5$ colony forming units (CFUs) using a Hamilton syringe (Gastight no. 1701, Switzerland). After inoculation the burr hole was closed with bone wax and the overlying skin was sutured. Most of the head wounds closed within one week. All mice showed initial mild weight loss (< 6%) following inoculation of saline or *Mtb* before returning to baseline after 9 days. All mice who were not culled at 9 days subsequently gained weight with the exception of TB-infected $TNF^{-/-}$ mice, which showed 10% weight loss at 17 days post inoculation (Supp. Figure 2.1 A).

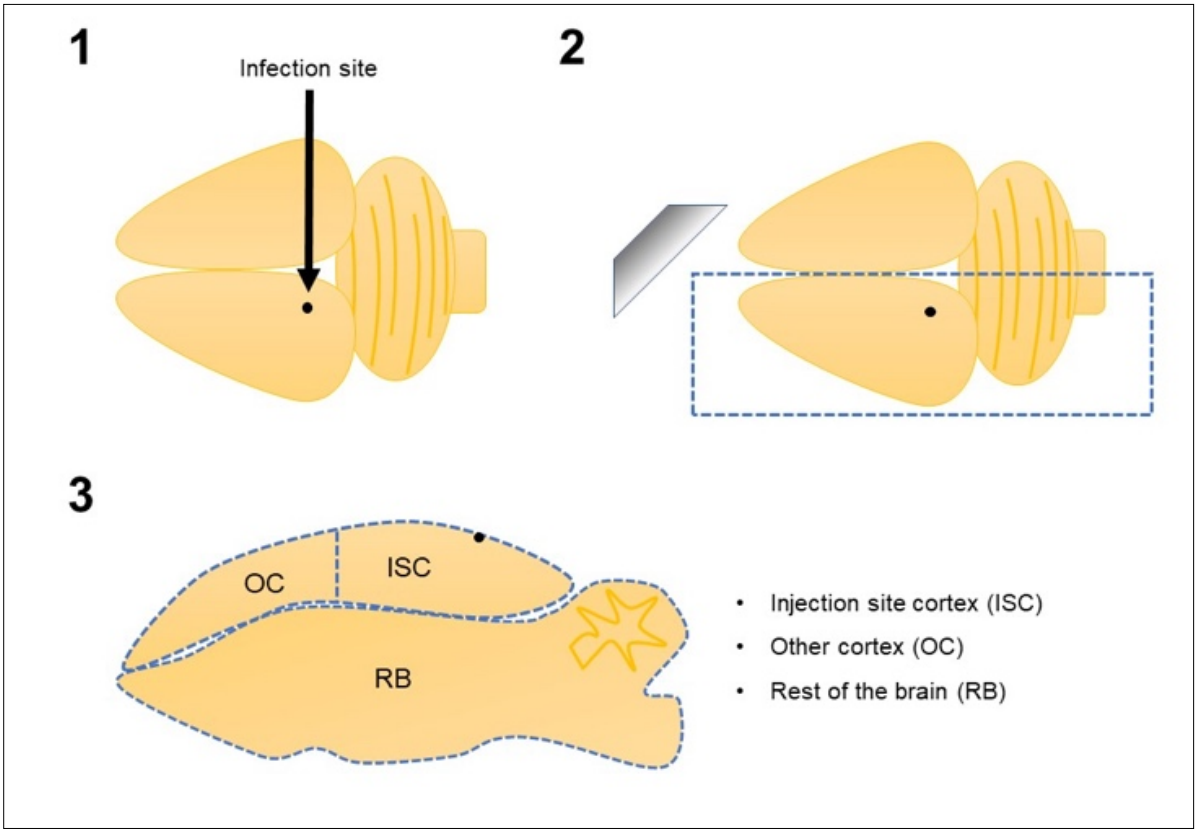


Figure 2.2: A schematic of the mice brain including the site of injection with either saline or Mtb*. 1,2) Dorsal views of mouse brain. Mice were anaesthetized with halothane and positioned in a stereotactic frame. A burr hole was created (labelled with black dot) over the injection site, which was located over the somatosensory cortex, anterior to the bregma and to the left of the midline. 3) Sagittal view of mouse brain with injection site cortex labelled, along with other cortical regions and the rest of the brain tissue. The injection site cortex was subsequently collected for downstream analysis. Picture used with permission by Dr Nai-Jen Hsu.*

2.3 Tissue collection and nuclei isolation (performed by Dr's Nai-Jen Tsu and Emily Tangie)

The mice were anesthetized, decapitated, and their brains harvested on either Day 9 or Day 17 post-saline injection/TB infection. The injection site cortex (ISC) (Figure 2.2) was isolated using a dissection kit. To confirm the presence of TB in TB-inoculated mice, colony forming unit (CFU) enumeration was performed on the brain hemispheres not used for snRNA-seq. Mycobacterial colonies were counted after three weeks of incubation at 37°C and showed a higher bacterial burden in $TNF^{-/-}$ mice compared to WT (Supp. Figure 2.1 B).

A previously established protocol was adapted for isolating nuclei from brain tissue (98). Tissue lysis was achieved by pooling the three samples per condition in a single glass dounce homogenizer on ice, which contained 1 ml of ice-cold nuclei EZ lysis buffer [Sigma-Aldrich, NUC101] (Figure 2.3). The tissue was ground 50 times with pestle A and pestle B respectively, before being transferred to a 15 ml conical tube and then adding 2 ml nuclei EZ lysis buffer to incubate for 5 minutes.

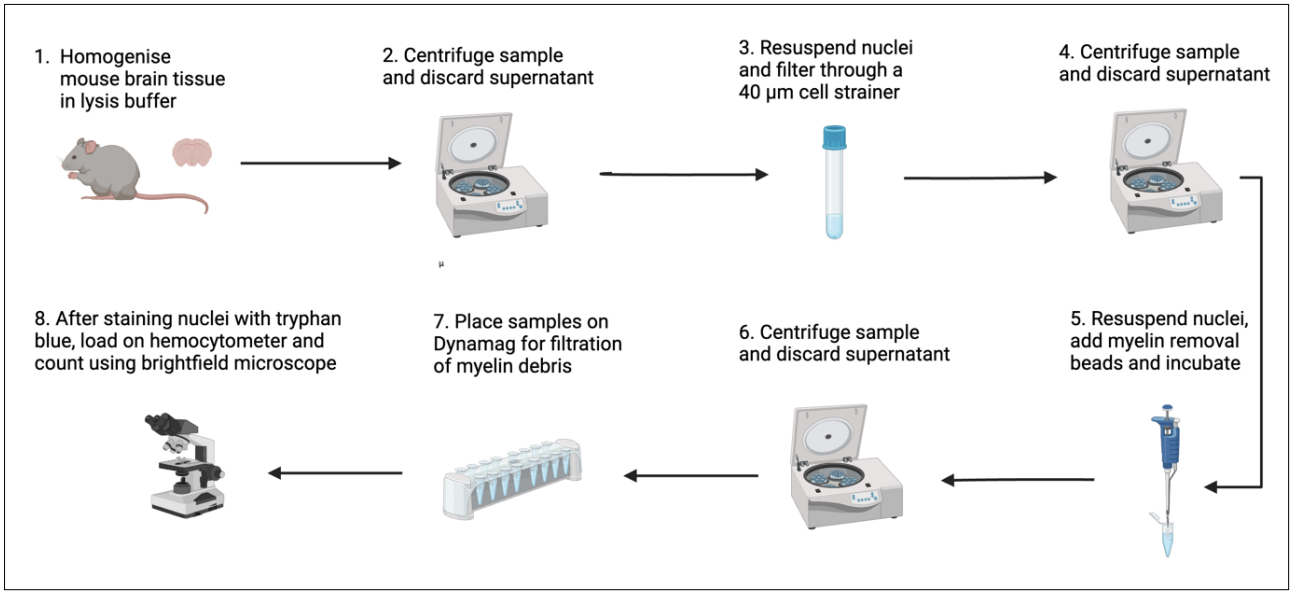


Figure 2.3: Graphical representation of the nuclei isolation protocol*. 1,2) Mouse brain tissue was homogenized in lysis buffer before being centrifuged after which the supernatant was discarded. 3,4) Nuclei were then resuspended and filtered through a cell strainer before a second centrifugation. 5,6) Nuclei were resuspended before myelin removal beads were added followed by a third centrifugation. 7) Samples were placed in Dynamag to filter myelin debris. 8) Nuclei were then stained with tryphan blue before being loaded on a hemocytometer and counted using a brightfield microscope. Figure compiled with BioRender.*

The resulting sample was centrifuged 500 x g for 5 minutes at 4 °C, after which the supernatant was discarded and the pellet resuspended in 3 ml of ice-cold nuclei suspension buffer (NSB) (1x phosphate-buffered saline [Sigma-Aldrich, P4417-60TAB,US], 0.01% bovine serum albumin [Sigma-Aldrich, A2153-10G, US], 0.2 U/µl RNAsin Plus RNase inhibitor [Promega, N2615, US]). A 40 *µm* filter was placed on the 50 ml tube and the resuspended nuclei were passed through the filter before being centrifuged at 500 x g for 10 minutes at 4 °C to pellet the nuclei. A 50 *µL* solution remained above the pellet after the supernatant was discarded.

In order to remove myelin, 3 ml blocking buffer (1x Phsophate Buffered Saline (PBS); 1% Bovine Serum Albumin (BSA); 0.2 U/µl RNAsin Plus RNase inhibitor) was added to pelleted nuclei and then resuspended until no tissue accumulation was visible. 90 µ*L* myelin removal beads [Miltenyi Biotec, 130-096-733, US] were added and the solution was mixed gently 5 times before being incubated for 15 minutes at 4°C. 3 ml blocking buffer was then added and centrifuged at 300 x g for 5 minutes. The supernatant was removed and 2 ml of clean blocking buffer was added, which was gently resuspended 5 times. The nuclei were then moved to a 2 ml Eppendorf tube, which was placed on a Dynamag magnet to be stored for 15 minutes in a refrigerator at 4 °C.

The resulting supernatant was transferred into a new tube. 10 *uL* of nuclei suspension was stained with trypan blue and loaded onto a hemocytometer, which was counted under a microscope. The nuclei were centrifuged once more at 500 x g for 5 minutes and resuspended in an NSB to achieve a targeted concentration of 700-1000 nuclei*/uL*, which was confirmed by counting again using a hemocytometer. The final concentration was used to determine the volume of nuclei suspension to use in the following snRNA-seq library creation steps.

2.4 Single-nucleus RNA sequencing dataset generation

SnRNA-seq library preparation was performed using the 10x Genomics Chromium Next GEM Single Cell 3' Reagent Kit (v3.1), according to the protocols stipulated by the manufacturer (76). The volume of the nuclei loaded was determined by using the Cell Suspension Volume Calculator Table in the 10x Genomics user manual to achieve approximately 10 000 nuclei per sample.

Sample dilutions were performed with nuclease-water, and the nuclear suspension was combined with the necessary volume of the master mix prepared on ice according to the protocol (59.1% Reverse transcriptase (RT) Reagent B, 7.5% Template Switch Oligo, 6.3% Reducing Agent B, and 27.4% RT Enzyme C).

The samples, comprising the nuclear suspension in master mix, were loaded into the row labelled 1 on the 10X Genomics Single cell 3' chip G. Gel beads were loaded into the used wells corresponding to samples in the row labeled 2, and partitioning oil was loaded into the corresponding sample wells in the row labeled 3. A solution of 50% glycerol was prepared and pipetted into the unused wells of the chip. The chip was loaded into the Chromium Controller. While in the Chromium Controller, individual nuclei were captured in a lipid drop*let* along with a uniquely barcoded gel bead to form gel beads in emulsion (GEMs) (Figure 2.4 A). mRNAs were captured by the barcoded oligonucleotides that were released from the gel bead.

The GEM solution was removed from the chip. A cDNA library was created by reverse transcription of the mRNA, which resulted in transfer of the barcode from the oligonucleotide onto the newly generated cDNA. At step 2.2 in the manufacturer's protocol (76), the cDNA was amplified with 11 cycles of PCR in order to generate a sufficient concentration for library construction. After cDNA clean-up, TapeStation (Agilent) quality control was performed by the Central Analytical Facility (Stellenbosch University) to analyse the size distribution of the cDNA, and to confirm that the cDNA fragments generated were of the appropriate length. Qubit was also performed to measure cDNA concentrations. The subsequent steps of cDNA library preparation included size selection of transcripts, adaptor ligation for sequencing, and the addition of sample-specific indices.

For the majority of the samples, the cDNA was amplified using 11 PCR cycles during the sample-indexing step (step 3.5 in the manufacturer's protocol) (76). However, the cDNA from the Day 17 $TNF^{-/-}$ TB sample was amplified with 15 PCR cycles as the pre-library concentration was much lower than initially anticipated.

The outcome of this process lead to each transcript in the cDNA library being assigned to a unique molecular identifier (UMI), as well as a barcode specific to the captured nucleus and an index for pooling samples. The final libraries were analysed once more by TapeStation and Qubit analysis at the Central Analytical Facility (Stellenbosch University). cDNA libraries were then sequenced on the Illumina NovaSeq platform by Novogene (Singapore).

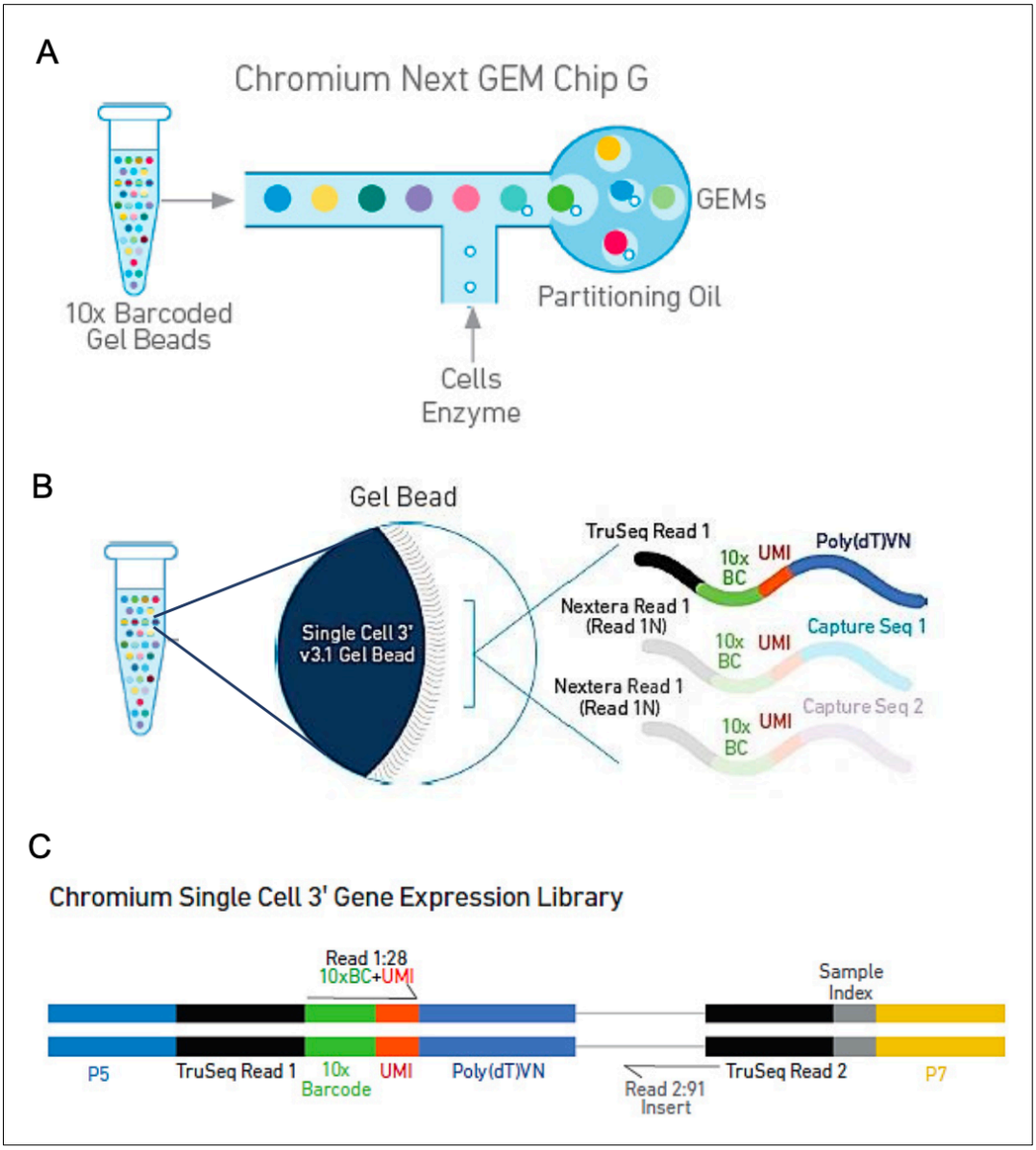


Figure 2.4: snRNA-seq library generation*. A) In 10X Genomics Chromium controller, nuclei/cells are captured in a partitioning oil together with a reverse transcriptase master mix and a gel bead with a specific primer set, resulting in the formation of a gel-bead in emulsion, known as GEM. B) The gel beads are coated with thousands of oligonucleotides. Each oligonucleotide contains a unique barcode, which is used to distinguish the transcripts from each nucleus. In addition, each gel bed contains a unique molecular identifier (UMI) which distinguishes each read from all other reads within a single nucleus. This allows for detection of technical and biological replicates, thereby minimising the effects of PCR amplification bias. Each transcript is captured by the Poly(dT) tail section of the oligonucleotides. C) A representative fragment from a 10X Genomics Chromium 3' gene expression library. In addition to the 10x barcode and UMI, P5 and P7 primers shown are used for cDNA transcription and sequencing. In addition, TruSeq Reads 1 and 2 are used for sequencing with Illumina. Sample indices are used to differentiate reads from different samples. Figure adapted from the Chromium Next GEM Single Cell 3' Reagent Kits v3.1 User Guide.*

2.5 SnRNA-seq bioinformatics analysis

2.5.1 Read alignment and quantification (Cell Ranger)

Demultiplexed, paired-end libraries were received from Novogene in Fastq format. The Cell Ranger pipeline [version 7.2.0] was used to map paired-end sequences to the mouse reference transcriptome (refdata-gex-mm10-2020-A), as well as filtering the data to exclude empty barcodes (99). The inclusion of introns was incorporated in the function script (Code availability, Script 1). Figure 2.5 provides a summarised schematic of the bioinformatics pipeline.

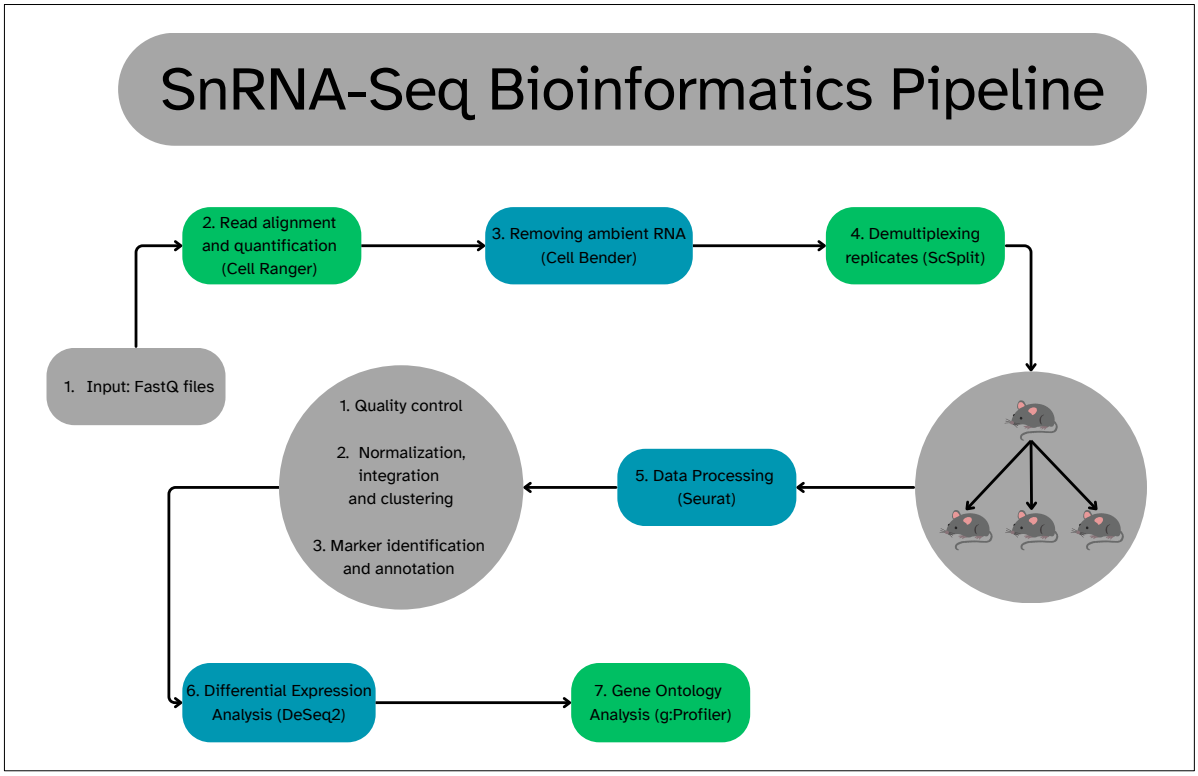


Figure 2.5: A schematic of the snRNA-seq bioinformatics pipeline used in this study*. 1,2) Read alignment and quantification was performed using Cell Ranger with FastQ files as input. 2) Cell Bender removed background contamination in the form of ambient RNA. 3) ScSplit was used to demultiplex the pooled nuclei into three separate samples per condition. 4) The Seurat pipeline was used for further data processing, including quality control, normalisation, integration, clustering along with marker identification and annotation. 4) Differential expression analysis was performed using DESeq2. 5) Gene ontology analysis was conducted using g:Profiler. Diagram made with Canva.*

2.5.2 Removing ambient RNA (Cell Bender – performed by Dr Vuyani Moses)

One of the confounders of snRNA-seq analysis is the presence of RNA surrounding individual nuclei, known in the literature as 'ambient RNA' (100). Ambient RNA can become trapped in GEMs along with the single nucleus, which can lead to an incorrect measurement of the true gene expression profile from the captured nucleus (100). To account for the presence of ambient RNA in the data, Cell Bender was used to model the presence of ambient RNA along with other sources of noise, using the count

matrix from Cell Ranger as input (100). By estimating and removing the influence of ambient RNA on the observed expression levels, Cell Bender provides a more accurate understanding of the true biological activity in each nucleus (Code availability, Script 2).

2.5.3 Demultiplexing replicates (ScSplit – performed by Dr Vuyani Moses)

In order to identify differentially expressed genes between conditions using DESeq2, biological replicates would be required. As each experimental condition only had one pooled snRNA-seq library (Section 2.1), ScSplit was used to separate the pooled nuclei into three original samples per experimental condition (96).

ScSplit uses a hidden state model to demultiplex sc/snRNA-seq data without taking into account information about the reference genome (96). Instead, ScSplit uses genetic differences to map clusters to the samples of origin. Existing bioinformatics tools are also employed by the ScSplit protocol to generate the demultiplexed data. The ScSplit protocol involved various steps (Figure 2.6; Code availability, Script 3). Firstly, a quality control step was performed using samtools, to remove low quality, duplicate and unmapped reads from the count matrix output from Cell Bender (101). Thereafter, the resulting Binary Alignment Map (BAM) file was sorted and indexed. Variant calling using freebayes (102) was used to identify the Single Nucleotide Variation (SNV) for each mice to determine the unique characteristics that would allow for the differentiation between each of the three mice pooled per condition. Once the SNV were identified, Bcftools was then used to filter out SNVs with a quality score of less than 30 (101). In addition, insertion – deletion mutations (indels) and low- quality variants were identified and removed. An allele count matrix was then built for all the chromosomes, which was used to demultiplex cells from each condition into four categories: mouse 1, mouse 2, mouse 3 and doublets.

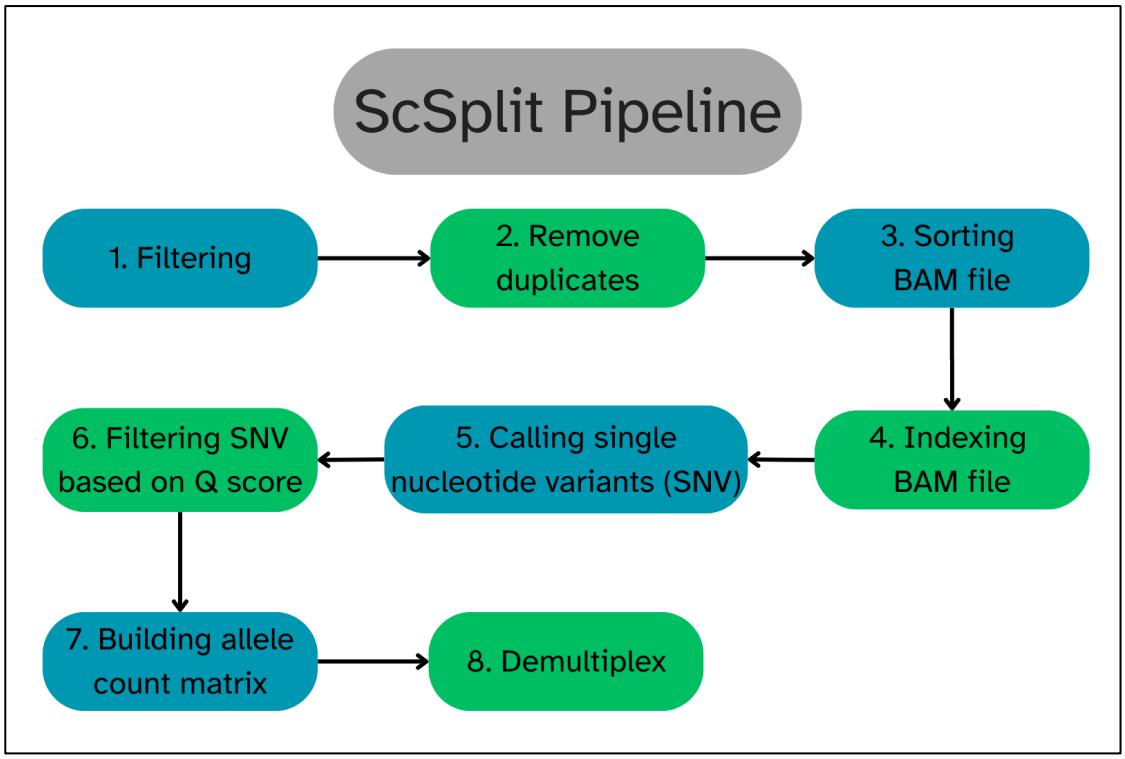


Figure 2.6 – see legend on following page

Figure 2.6: A summary of the steps in the ScSplit protocol*. 1,2) Quality control steps were performed to remove low quality, duplicate and unmapped reads from the count matrix output from Cell Bender. 3,4) The resulting BAM file was sorted and indexed. 5) Variant calling using freebayes was used to identify SNVs and determine the unique characteristics between each mice. 6) Bcftools was then used to filter out SNVs with a quality score of less than 30. In addition, insertion – deletion mutations (indels) and low- quality variants were identified and removed. 7,8) An allele count matrix was then built for all the chromosomes, which was used to demultiplex cells from each condition into four categories: mouse 1, mouse 2, mouse 3 and doublets. Diagram made with Canva.*

Doublets refer to the phenomenon whereby two cells were captured in a single GEM, thus having the same molecular barcode (103). DNA transcripts from the two captured cells would appear to stem from one cell, which could lead to incorrect interpretation of results in downstream analysis. Therefore, the identification and subsequent removal of doublets is crucial in enhancing data quality. With the generation of separate barcodes from each mouse per condition and the removal of said doublets, further data processing could be conducted using the Seurat pipeline.

2.5.4 Seurat SnRNA analysis pipeline

2.5.4.1 Quality control:

The output from the ScSplit protocol was formatted to match the input requirements needed for the Seurat pipeline, the next step in snRNA-seq analysis (Code availability, Script 4). Firstly, a quality control step was performed by removing barcodes with poor quality nuclei or gel beads without nuclei (105)(Figure 2.5; Code availability, Script 5). Barcodes were only retained if they contained more than 500 UMIs or more than 250 expressed genes. Other filtering steps included keeping barcodes where the log10genesperUMI was greater than 0.8 and mitochondrial ratio less than 0.2. Subsequently, mitochondrial genes were also excluded from the dataset as only nuclei were being captured, and hence mitochondrial transcripts were not expected.

2.5.4.2 Data Normalization, integration and clustering:

Principal component analysis (PCA) was used to examine the influence of the mitochondrial ratio and cell cycle phase on the inter-sample variation between nuclei (Code availability, Script 6). Normalization was then conducted with Seurat's SCTransform() function to address variations in sequencing depth across nuclei (105)(Code availability, Script 6). In addition, variance stabilization was performed to ensure that the observed variation was not driven solely by highly expressed genes. The mitochondrial ratio was subsequently regressed out of the analysis as it was a considered to be a non-useful source of variation. The cell cycle phase was also not found to substantially contribute to data variation.

The need for dataset integration was determined using Uniform Manifold Approximation and Projection (UMAP) analysis (104). Following this, the datasets were integrated to align similar cell types across the datasets. The functions required for Seurat integration included SelectIntegrationFeatures, PrepSCTIntegration, FindIntegrationAnchors and IntegrateData (105)(Code availability, Script 6).

Following integration, dimensionality reduction was performed (104)(Code availability, Script 6). The nature of single nuclei analysis involves processing thousands of genes expressed by numerous cell types. The purpose of dimensionality reduction is to summarise the most important features across the dataset for ease of analysis and to present this variation in two-dimensional space. 30 dimensions were specified. The FindClusters function was applied at a resolution of 0.8, thereby yielding 53 clusters.

2.5.4.3 Marker identification and annotation:

After clustering, the next step was to identify cell type-specific markers and annotate each cluster to one of the expected cell types in the brain. There are two major approaches to annotate clusters, namely manual and automatic. The manual approaches involve comparing known cell type-specific markers to the highest expressed genes in each cluster. The markers were the same markers use for the Allen Mouse Brain Atlas (105).

Automatic approaches can be divided into automated annotation or label transfer tools. These tools use reference datasets to label clusters. Label transfer is thought to be the more accurate of the two approaches, because it compares the dataset to a known reference base of the entire genetic expression profile of each cell (106). This contrasts with automated annotation tools, which only compare the top expressed genes in each cluster to the reference database. In this analysis the label transfer was performed by using the Allen Mouse Brain Atlas as a reference (105)(Code availability, Script 7). For computing purposes, the reference atlas was filtered to only include the relevant regions in the mouse brain. These regions included the Anterior Cingulate Area, Entorhinal area, Posterior Parietal Association areas, Primary Somatosensory Area, Retrosplenial area, Temporal Association Areas, Perirhinal areas, Ectorhinal areas, Lateral Visual Area, Primary Visual area (VISp) and the Medial Visual Area (VISm).

2.5.5 Differential expression analysis

After annotation of the clusters, differential expression analysis was performed using the DESeq2 package (107)(Code availability, Script 8). DESeq2 is a statistical tool that is used to analyse RNA-seq data by identifying differentially expressed genes between conditions (Figure 2.7). DESeq2 uses a negative binomial distribution to account for the variability in the count data, since the data produced by RNA-seq is discrete (number of reads mapped to a particular gene). Additionally, DESeq2 employs a pseudobulk strategy for analysing snRNA-seq data by aggregating the read counts of cells within specific cluster (108).

DESeq2 requires an input matrix of pseudobulk read counts and a metadata file which specifies the sample conditions. DESeq2 then normalizes the count data using size factors. This ensures that the comparisons between samples are not biased by differences in sequencing depth and library composition. Thirdly, RNA-Seq data is often over dispersed, where the variability is higher than what would be expected from a normal Poisson distribution. DESeq2 accounts for the dispersion by creating a variance term that models that variance between biological replicates. To estimate dispersion across all genes, DESeq2 fits a curve using the negative binomial model, and then shrinks the gene dispersions towards the fitted curve. This process improves the estimates for genes with low counts or replicates.

DESeq2 then performs statistical testing to determine significant gene expression differences between conditions of interest (107).

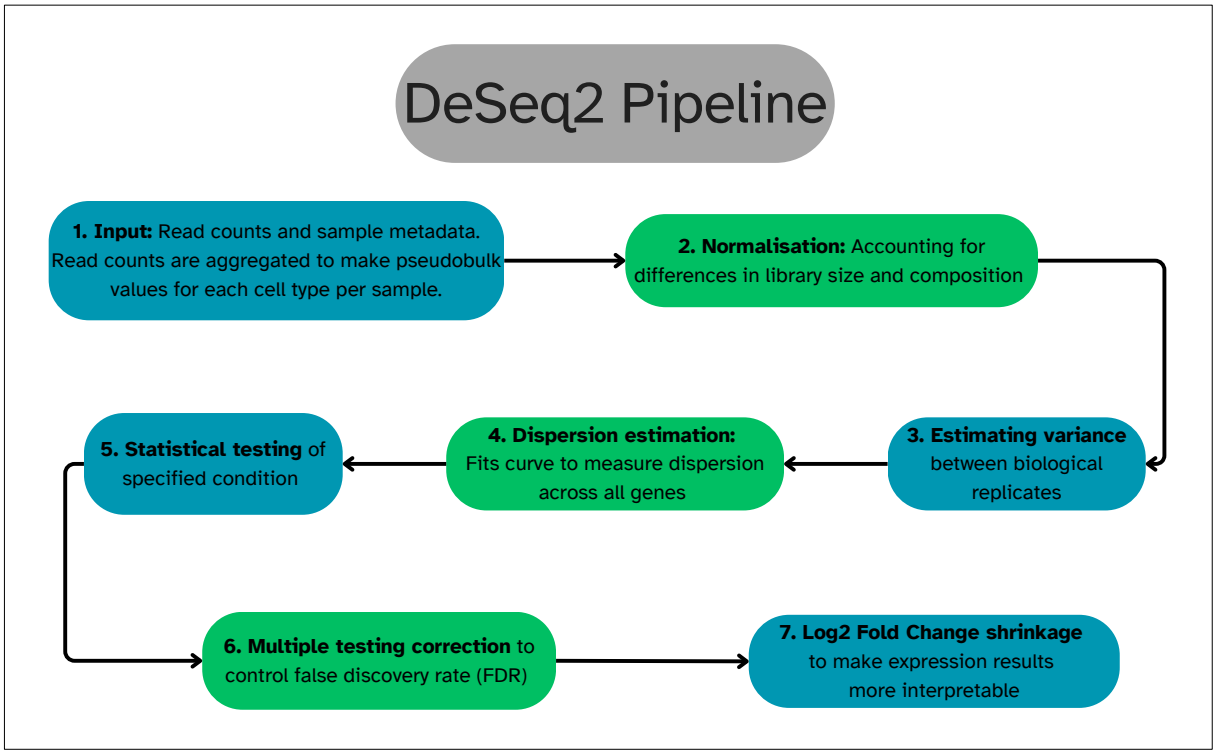


Figure 2.7: A summary of the steps in the DESeq2 protocol*. 1) Raw read counts and metadata are used as input. Raw reads are aggregated to form pseudobulk values for each cell type per sample. 2) Normalisation is used to account for variations in library size and composition. 3) Variance estimation is then performed between biological replicates. 4) Thereafter, dispersions across all genes are measured and a curve is fitted. 5,6) Statistical testing of each desired condition is then performed followed by corrections for multiple testing. 7) Finally,* $\log_2$FC shrinkage is calculated. Diagram made with Canva.

The output of DESeq2 includes a p-value, which evaluates if the observed differences are due to chance. Since numerous genes are being tested simultaneously, DESeq2 adjusts for multiple testing by outputting an adjusted p-value. Another output of the DESeq2 pipeline is the fold change (FC) for each gene, which is defined as the ratio of gene expression between the treatment and control group. Since the FC can vary widely between genes in a sample, DESeq2 generates a $\log_2$FC, which makes the data more interpretable. Therefore, a positive $\log_2$FC (FC > 1) implies that the gene is upregulated – and vice versa.

DESeq2 can accommodate a variety of comparisons between conditions, the most basic of which is the comparison between two groups (such as TB-infected mice compared to saline mice). In addition, DESeq2 can also model how one factor changes depending on the presence of another factor (109). For example, DESeq2 could determine how the TB response changes in $TNF^{-/-}$ compared WT mice. To achieve this, DESeq2 introduces an interaction term, which accounts for scenarios where the observed expression levels of two factors (eg. TB and $TNF^{-/-}$) are not additive.

Figure 2.8 shows how the observed expression levels of Gene X in $TNF^{-/-}$ TB mice are not necessarily equal to the summation of the individual background effects of $TNF^{-/-}$ and TB, when each is compared to the control condition. This is why an interaction term was included to model the unique effects on

gene expression when TNF$^{-/-}$ is combined with TB. If the interaction term has a positive log_2FC, it implies that the observed gene expression is more than what would be expected from summation of the background effects of TB and TNF$^{-/-}$ compared to controls i.e. upregulation. Consequently, if the interaction term has a negative log_2FC, it implies that the observed gene expression is less than what would be expected from summation of the background effects i.e. downregulation. A non-significant interaction term implies that the observed gene expression is not significantly different from the summation of background effects. Therefore, the expression of this gene is not influenced by the interaction between TNF$^{-/-}$ and TB. In summary, DESeq2 is a powerful tool that was used to determine significant differences in cell type-specific gene expression between multiple conditions.

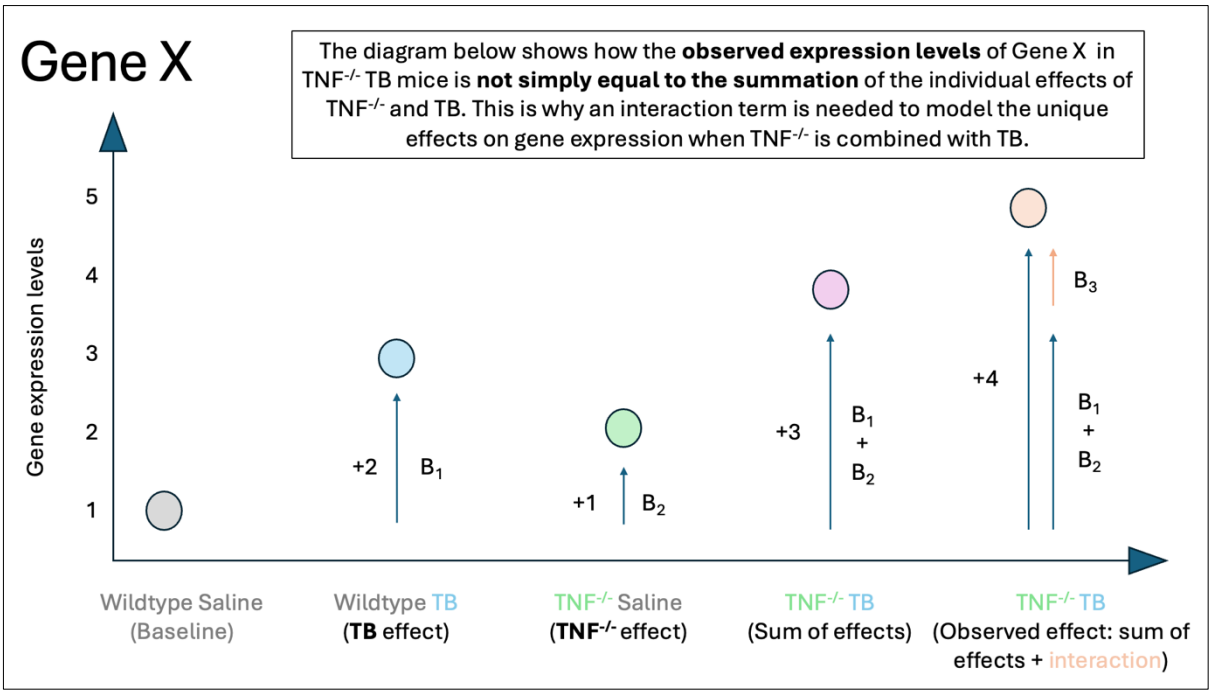


Figure 2.8: Visualization of the interaction term*. Diagram showing how the observed expression levels of Gene X in TNF$^{-/-}$ TB mice is not necessarily equal to the sum of the individual background effects of TNF$^{-/-}$ and TB when compared to the controls conditions. This is why an interaction term is needed to model the unique effects on gene expression when TNF$^{-/-}$ is combined with TB.*

2.5.6 Visualizing differential expression data

Volcano plots were used to visually represent significantly differentially expressed genes (DEGs) for each cell type per condition (Code availability, Script 8). The x-axis represented the log_2FC and the y-axis showed the negative logarithm of the p-adjusted value. For example, highly upregulated DEGs with large statistical significance would be in the upper right quadrant of the volcano plot. UpSet plots were used to visualize the number of shared and uniquely expressed DEGs per cell type for each condition (110)(Code availability, Script 9).

2.5.7 Gene Ontology analysis

Gene Ontology (GO) analysis is a tool used to understand the biological significance of gene expression data. GO categorizes DEGs into functional groups within three major categories: molecular functions, biological processes and cellular components. Molecular functions refer to cellular activities occurring at the molecular level, such as protein binding. The combination of different molecular functions forms the foundation for biological processes, such as synapse organization. Cellular components refer to the locations within a cell where a gene may be active, such as the cytoplasm.

G:Profiler was used to conduct gene ontology analysis using the output data from DESeq2 (111). G:Profiler compares the specified DEG list to a background selection of all genes for a particular organism (*eg. Mus Musculus*). The resultant GO terms were then categorized into functional groups. G:Profiler also highlights 'driver terms', which are specific GO terms that contribute most significantly to the overall enrichment score (111). Driver terms are useful because they assist researchers in identifying the most relevant pathways or functions in the context of the data. The output of g:Profiler can also be visualized in the form of a Manhattan-like plot and flow diagrams showing the relationships between GO terms. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database, part of the g:Profiler output, was used to identify disease associations with the DEGs (111).

Chapter 3: Results

3.1 snRNA-seq library preparation and initial quality control measures

A total of 8 pooled snRNA-seq libraries were generated from TB- and saline injected mouse brain samples (Table 3.1). To determine the quality of the cDNA libraries, Agilent Tape Station was used to measure cDNA fragment sizes, before and after completion of library preparation (Figure 3.1 and 3.2). The peak fragment size before and after library preparation was 1051bp and 444bp respectively (excluding the Day 17 TB $TNF^{-/-}$ library). This is in keeping with the expected fragment sizes pre- and post-library preparation. It should be noted that the Day 17 TB $TNF^{-/-}$ library had a very low concentration and as mentioned in the methods section, 15 PCR cycles were used during sample index PCR for this sample compared to 11 cycles for the other samples. Qubit concentrations were also recorded to quantify cDNA concentrations both before and after the completion of the library preparation procedure (Table 3.1). All samples were deemed to have sufficient concentrations for sequencing.

Table 3.1 *Qubit concentrations before and after library preparation during the 10X Genomics protocol.*

Sample name	Qubit concentration: Pre- library prep (ng/uL)	Qubit concentration: Post- library prep (ng/uL)
Day 9 WT Saline	39	39
Day 9 $TNF^{-/-}$ Saline	35,2	37,4
Day 9 WT TB	39,2	45,6
Day 9 $TNF^{-/-}$ TB	37,4	43,6
Day 17 WT Saline	27,8	28,4
Day 17 $TNF^{-/-}$ Saline	21,4	23
Day 17 WT TB	25,8	23,2
Day 17 $TNF^{-/-}$ TB	0,362	13,3

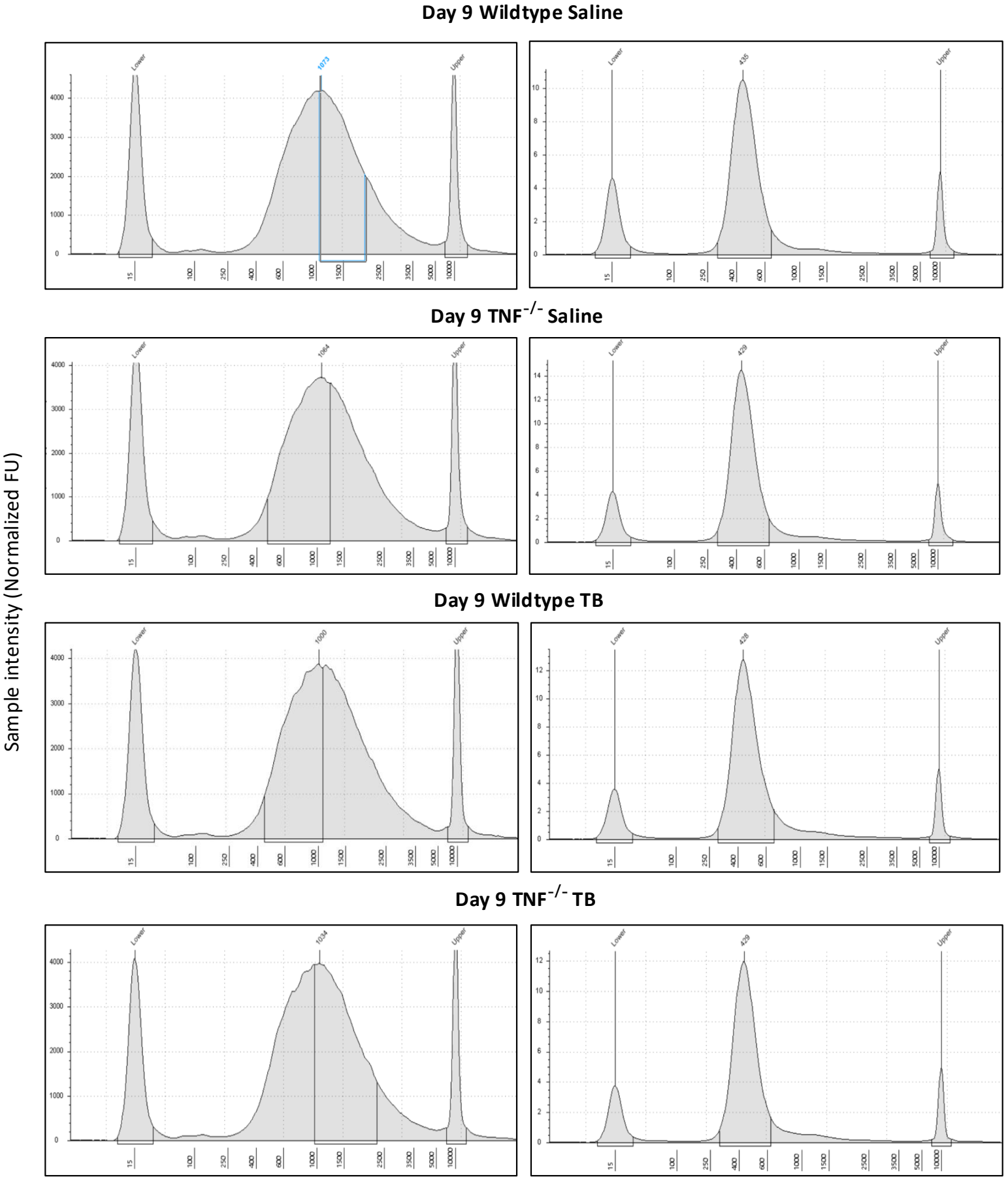


Figure 3.1: Tapestation reports (Day 9). *Graphical reports showing cDNA fragment sizes for each of the snRNA-seq Day 9 libraries, before (left) and after (right) library preparation. The number above the middle vertical line indicates the peak fragment size. (FU, fluorescence units)*

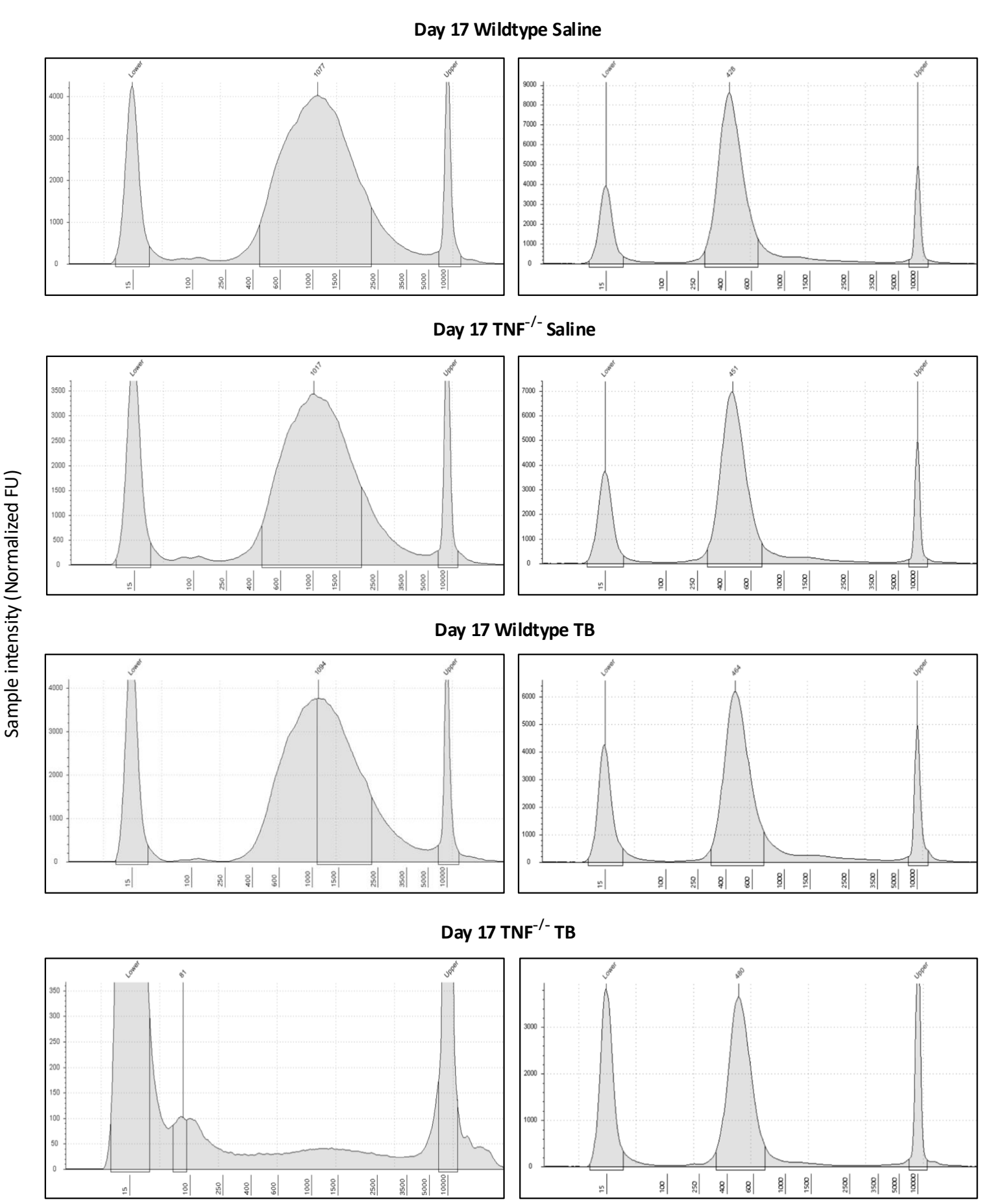


Figure 3.2: Tapestation reports (Day 17). *Graphical reports showing cDNA fragment sizes for each of the snRNA-seq Day 17 libraries, before (left) and after (right) library preparation. The number above the middle vertical line indicates the peak fragment size. (FU, fluorescence units)*

3.2 Pre-processing and quality control of snRNA-seq datasets

The Cell Ranger pipeline [version 7.2.0] was used to map the sequenced snRNA-seq libraries to the mouse transcriptome and perform initial quality checks of the resulting datasets. According to the Cell Ranger Web Summary outputs, the estimated number of nuclei per condition ranged from 8 648 (Day 9 $TNF^{-/-}$ saline) to 11 175 (Day 17 $TNF^{-/-}$ saline) with an average of 10,533 nuclei (Table 3.2), in keeping with the target of 10 000 nuclei per library. The mean reads per nucleus (total number of sequenced reads divided by the number of nuclei) averaged at 18 064, which is slightly below the recommended minimum of 20 000. The sequencing saturation, defined as the fraction of library complexity sequenced in a given experiment, ranged from 26.6% (Day 9 $TNF^{-/-}$ TB) to 54% (Day 17 $TNF^{-/-}$ TB), with an average of 31.7%. The fraction of reads mapped to the genome was also measured with an average of 93.2%. In view of the above findings, the quality of the datasets were assessed as acceptable.

Table 3.2: Quality control metrics generated from Cell Ranger pipeline*. These metrics include the estimated number of nuclei per sample as well as the mean and median reads per nucleus. In addition, the sequencing saturation and percentage of reads mapped to the genome have also been provided*.

Sample	Estimated number of nuclei	Mean reads per nucleus	Median genes per nucleus	Sequencing saturation (%)	Reads mapped to genome (%)
Day 9 WT Saline	10 866	14 543	1 360	32%	93.2%
Day 9 $TNF^{-/-}$ Saline	8 648	15 246	1 377	27.1%	93.4%
Day 9 WT TB	10 563	13 414	1 280	27.1%	93.4%
Day 9 $TNF^{-/-}$ TB	11 594	10 132	1 118	26.6%	93.7%
Day 17 WT Saline	10 248	16 587	1 453	27.6%	92.7%
Day 17 $TNF^{-/-}$ Saline	11 175	22 007	1 647	30.1%	92.9%
Day 17 WT TB	8 740	28 356	1 984	29.3%	93.4%
Day 17 $TNF^{-/-}$ TB	10 547	24 230	1 173	54%	92.8%
Average	**10533**	**18064**	**1424**	**31.7%**	**93.2%**

Following initial processing by the Cell Ranger pipeline and removal of ambient RNA by Cell Bender, the ScSplit tool was used to separate the datasets into their constituent samples (i.e. three samples per condition), thereby yielding 24 separate snRNA-seq datasets. Table 3.3 summarizes the distribution of nuclei across each condition between the three samples for each condition. It also states the number of doublets identified using the ScSplit protocol. The number of doublets per sample varied between 2 029 (Day 17 WT TB) and 2 739 (Day 9 $TNF^{-/-}$ TB), with an average doublet count of 2 324. The average number of nuclei per sample ranged from 1 977 (Day 17 WT TB) to 2 817 (Day 9 $TNF^{-/-}$ TB).

***Table 3.3**: **Distribution of nuclei per condition after the ScSplit protocol was conducted**. The number of doublets per sample condition is also noted along with the average number of nuclei per sample.*

Sample	Mouse 1 Nuclei	Mouse 2 Nuclei	Mouse 3 Nuclei	ScSplit Doublets	Total	Average nuclei per sample
Day 9 WT Saline	2 313	2 229	2 372	2 385	**9 299**	**2 304**
Day 9 TNF$^{-/-}$ Saline	2 193	1 815	2 178	2 298	**8 484**	**2 062**
Day 9 WT TB	2 346	2 419	2 428	2 310	**9 503**	**2 397**
Day 9 TNF$^{-/-}$ TB	2 764	2 843	2 845	2 739	**11 191**	**2 817**
Day 17 WT Saline	2 130	2 183	2 259	2 237	**8 809**	**2 190**
Day 17 TNF$^{-/-}$ Saline	2374	2 524	2 499	2 330	**9 727**	**2 465**
Day 17 WT TB	1 926	1 840	2 166	2 029	**7 961**	**1 977**
Day 17 TNF$^{-/-}$ TB	2 166	2 132	2 144	2 271	**8 713**	**2 147**

After the ScSplit protocol, the Seurat pipeline was used to perform further quality control checks and filtering of the remaining nuclei. The quality control filtering criteria included the average number of transcripts per nucleus (nUMI, should be > 500), the average number of genes within each nucleus (nGene, should be > 250), the log ratio of the number of genes with respect to transcripts, which is indicative of sample complexity (log10 Genes per UMI, should be > 0.8) and finally the ratio of mitochondrial genes to total genes within each nucleus (mitoRatio, should be < 0.2), which is a measure of the proportion of uncontaminated nuclei. Following this quality control step, the average nUMI per nucleus was 3 616, the average nGene was 1 717, the average log10GenesPerUMI was 0.93 and the average mitoratio was 0.0006 (Figure 3.3; Table 3.4). The number of nuclei per sample before and after Seurat filtering is displayed in Figure 3.4. The average number of nuclei removed from the Day 9 samples was 4, in contrast to 33 nuclei that were removed on average from the Day 17 samples.

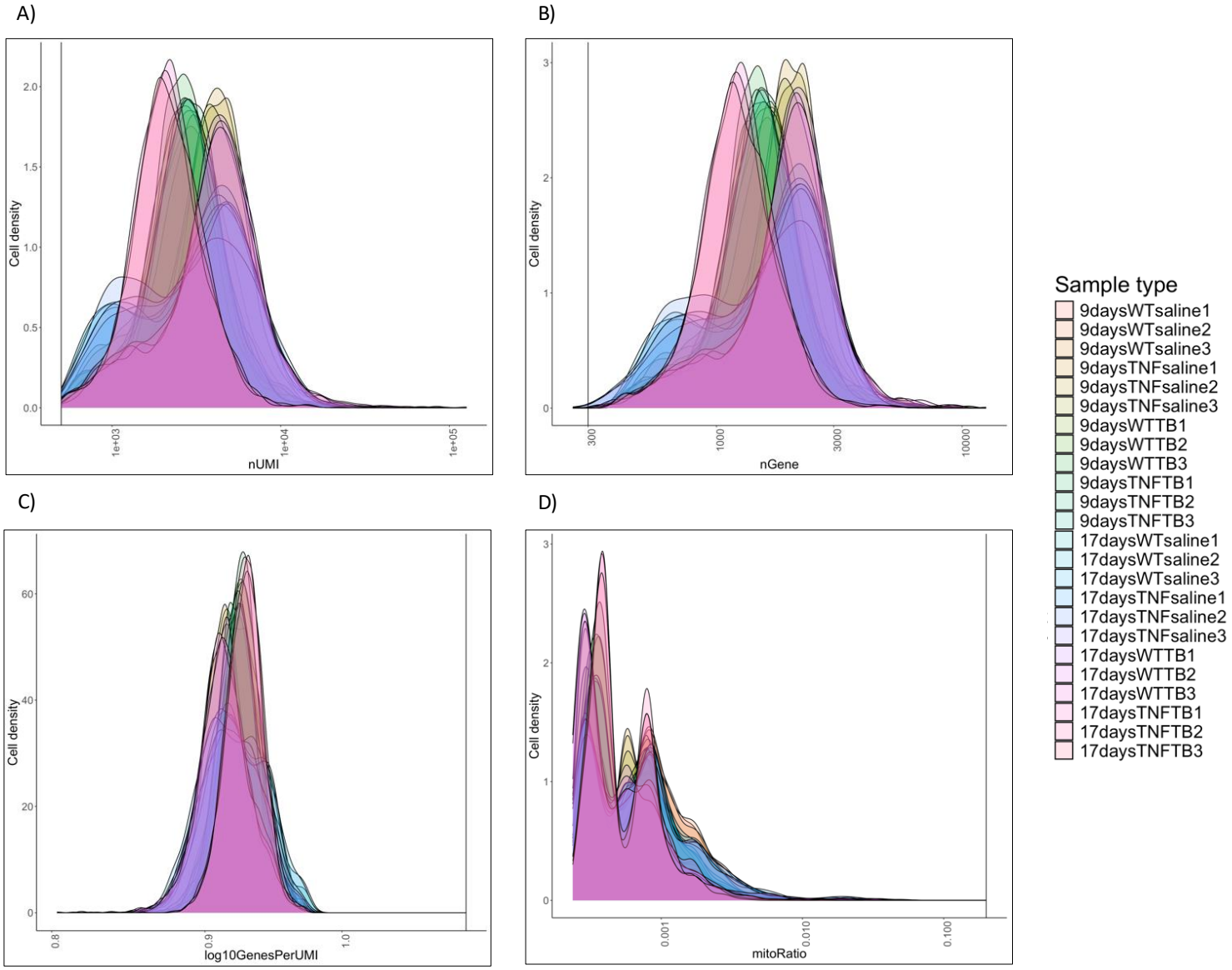


Figure 3.3: Graphical summary of quality control metrics post Seurat filtering. *Metrics included A) Average number of transcripts per nucleus (nUMI), B) Average number of genes within each nucleus (nGene), C) Logratio of the number of genes with respect to transcripts (log10GenesPerUMI) and D) Ratio of mitochondrial genes to total genes within each nucleus (mitoRatio).*

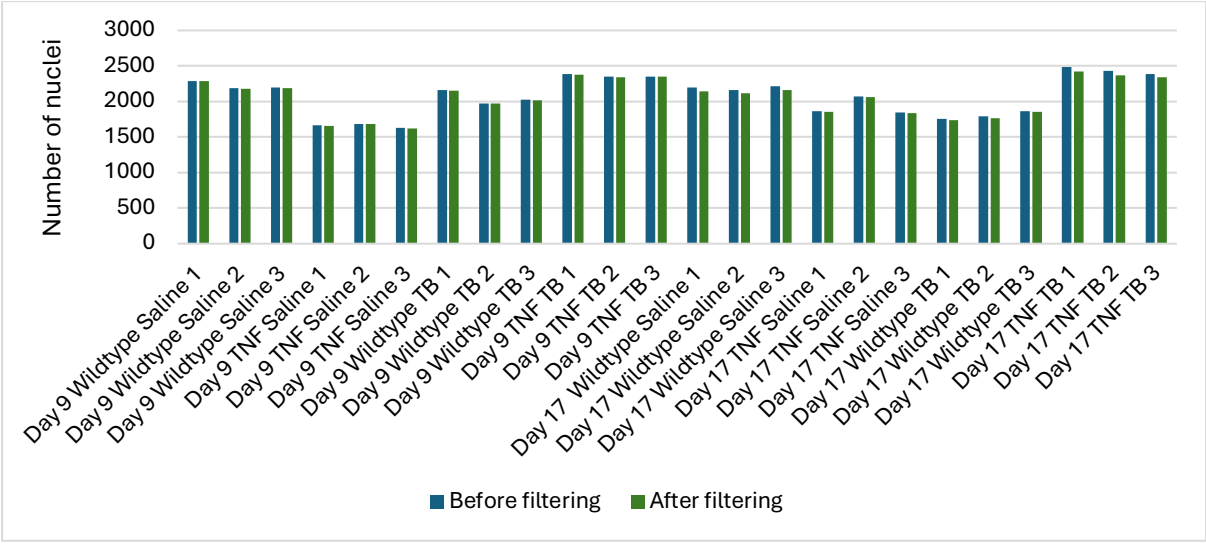


Figure 3.4: Bar graph showing the number of nuclei per sample before and after Seurat filtering.

***Table 3.4: Summary of quality control metrics for each nucleus per sample post Seurat filtering**. Metrics included the nUMI, nGene, log10 Genes per UMI, and mitoRatio.*

Sample	nUMI	nGene	log10GenesPerUMI	mitoRatio
Day 9 WT Saline 1	3 032	1 563	0,93	0,0009
Day 9 WT Saline 2	2 999	1 549	0,93	0,0009
Day 9 WT Saline 3	3 421	1 693	0,93	0,0008
Day 9 TNF$^{-/-}$ Saline 1	4 724	2 058	0,92	0,0005
Day 9 TNF$^{-/-}$ Saline 2	4 732	2 082	0,92	0,0005
Day 9 TNF$^{-/-}$ Saline 3	4 176	1 908	0,92	0,0006
Day 9 WT TB 1	3 181	1 614	0,93	0,0005
Day 9 WT TB 2	3 159	1 599	0,93	0,0004
Day 9 WT TB 3	2 960	1 533	0,93	0,0004
Day 9 TNF$^{-/-}$ TB 1	3 360	1 653	0,93	0,0006
Day 9 TNF$^{-/-}$ TB 2	3 163	1 592	0,93	0,0006
Day 9 TNF$^{-/-}$ TB 3	3 230	1 598	0,93	0,0006
Day 17 WT Saline 1	3 864	1 807	0,93	0,0007
Day 17 WT Saline 2	3 770	1 770	0,93	0,0007
Day 17 WT Saline 3	3 822	1 780	0,93	0,0007
Day 17 TNF$^{-/-}$ Saline 1	4 092	1 806	0,92	0,0008
Day 17 TNF$^{-/-}$ Saline 2	3 705	1 671	0,92	0,0008
Day 17 TNF$^{-/-}$ Saline 3	3 986	1 771	0,92	0,0008
Day 17 WT TB 1	4 688	2 080	0,92	0,0005
Day 17 WT TB 2	4 757	2 087	0,92	0,0005
Day 17 WT TB 3	4 710	2 067	0,92	0,0005
Day 17 TNF$^{-/-}$ TB 1	2 435	1 309	0,93	0,0003
Day 17 TNF$^{-/-}$ TB 2	2 389	1 293	0,93	0,0003
Day 17 TNF$^{-/-}$ TB 3	2 421	1 309	0,93	0,0003
Average	**3 616**	**1 717**	**0,93**	**0,0006**

3.3 Data integration, clustering, and annotation

After the quality control steps were completed, the respective gene counts of each dataset underwent normalization and scaling using the Seurat pipeline. It was also found that cell cycle cycle contributed minimally to the variation observed between the samples. Subsequent to the above steps, data integration was use to align similar cell types across the datasets (Figure 3.5 A). This resulted in greater alignment of the samples compared to the non- integrated data (Figure 3.5 B).

Thereafter, dimensionality reduction and clustering was performed, which produced 53 clusters (Figure 3.5 C). Annotation was performed using label transfer with a filtered version of the Allen Mouse Brain Atlas as a reference to classify the clusters into 27 different cell subtypes (Figure 3.5 D). These cell subtypes were also manually classified into broad groups, namely excitatory neurons,

inhibitory neurons, astrocytes, microglia-peripheral vascular macrophages (microglia-PVM), oligodendrocytes and non-neural cell types (Figure 3.5 E).

The excitatory neuron group consisted of the following cell subtypes: Cortex intratelecephalic layer 2/3 (L2/3 IT CTX), parasubiculum, postsubiculum, presubiculum intratelencephalic layer 2/3 (L2/3 IT PPP), retrosplenial, anterior cingulate layer 4 (L4 RSP-ACA), cortex intratelencephalic layer 4/5 (L4/5 IT CTX), cortex intratelencephalic layer 5 (L5 IT CTX), cortex pyramidal tract layer 5 (L5 PT CTX), temporal association/perirhinal/ectorhinal intratelencephalic layer 5/6 (L5/L6 IT TPE-ENT), cortex near-projecting layer 5/6 (L5/6 NP CTX), cortex corticothalamic layer 6 (L6 CT CTX), cortex intratelencephalic layer 6 (L6 IT CTX), entorhinal cortex layer 6b (L6b/CT ENT), cortex layer 6b (L6b CTX), cornu-ammonis 2-industrium griseum-fasciola cinerea (CA2-IG-FC) and car3-expressing (Car3) Inhibitory neurons were represented by Cajal Retzius (CR), LAMP family member 5 (lamp5), Synuclein gamma (sncg), Vasoactive intestinal peptide (vip), Somatostain (sst), sst chrondrolectin (chodl) and Parvalbumin (pvalb) neuronal cell sub types. Neuroglial cell types included astrocytes (Astro), microglia-PVM and oligodendrocytes (Oligo). Non-neural cell types consisted of endothelial cells (Endo), smooth muscle cell-pericytes (SMC-Peri) and vascular/leptomeningeal cell(VLMC). The L6b/CT ENT-cells type was removed from further analysis because only one nucleus was mapped to this cell type across all the samples.

Analysis of the proportion of nuclei in each cell type per sample showed an general consistency across the sample replicates generated from the ScSplit protocol (Figure 3.6 A). The Day 17 saline (both WT and $TNF^{-/-}$) samples showed a greater proportion of oligodendrocytes compared to the Day 17 TB samples (both WT and $TNF^{-/-}$). Violin plots were created to represent the number of transcripts (nUMI) and number of genes (nGene) per nucleus for each sample and cell type (Figure 3.6 B-E). The average number of transcripts were noted to be similar between each sample (Figure 3.6 B). Day 17 samples had more expressed genes than Day 9 samples with the exception of the Day 17 $TNF^{-/-}$ TB group, which had fewer expressed genes compared to the other Day 17 samples (Figure 3.6 C). Whilst the average UMIs were consistent between cell types (Figure 3.6 D), a higher number of expressed genes were noted in the neuronal cell types compared to the non- neuronal cell types (Figure 3.6 E).

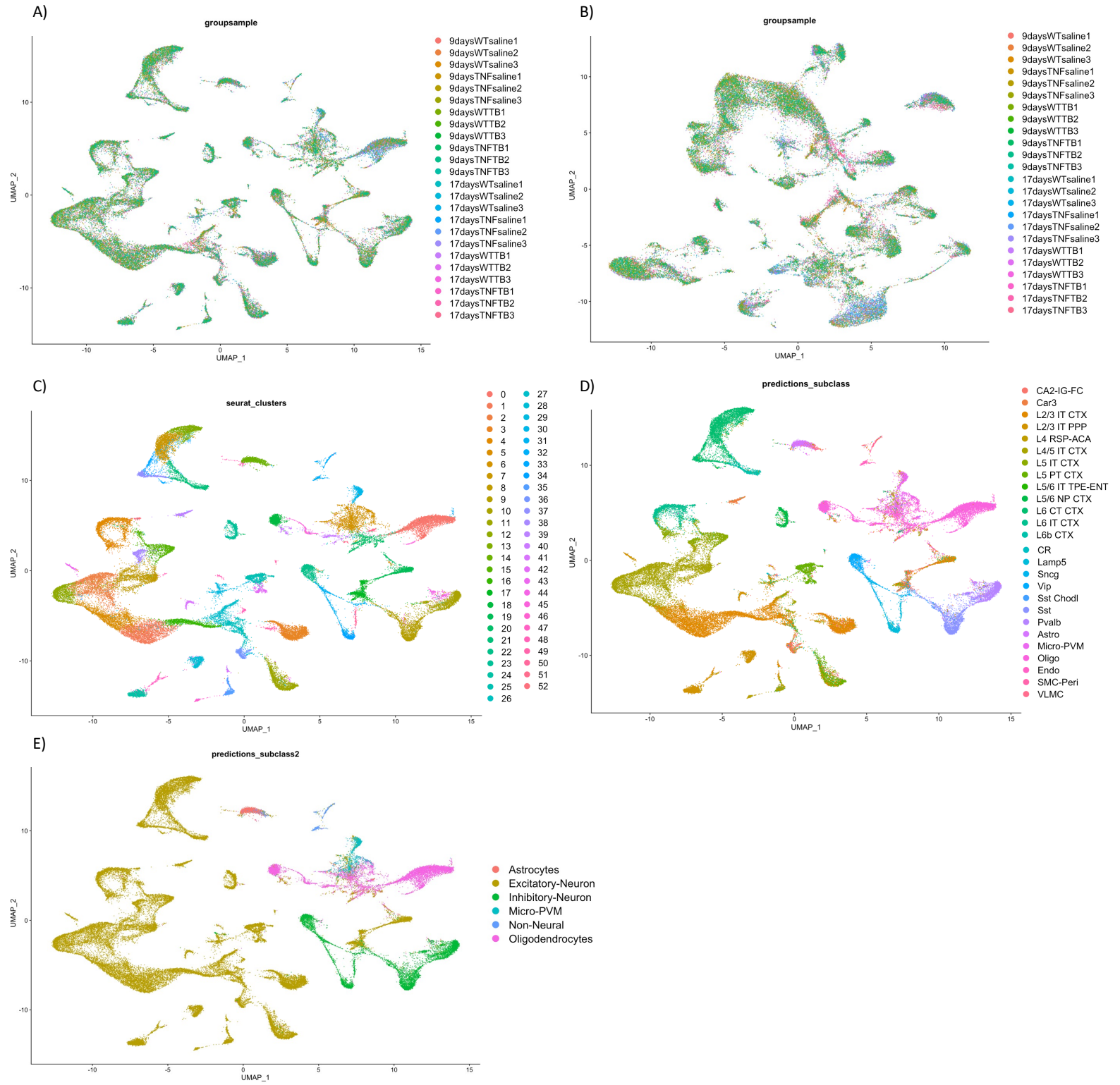


Figure 3.5: Annotation of nuclei identifies 27 cell subtypes across the 24 datasets *A) UMAP plot after integration showing clustering of each sample. B) UMAP plot before integration showing clustering of each sample. B) UMAP plot with clusters coloured by Seurat cluster number. D) Annotated UMAP plot after integration using label transfer showing all cell types. E) Annotated UMAP plot after integration using label transfer showing grouped cell types. L6b/CT ENT was removed from Figure 3.5 D because only one nucleus was mapped to this cell type across all the samples.*

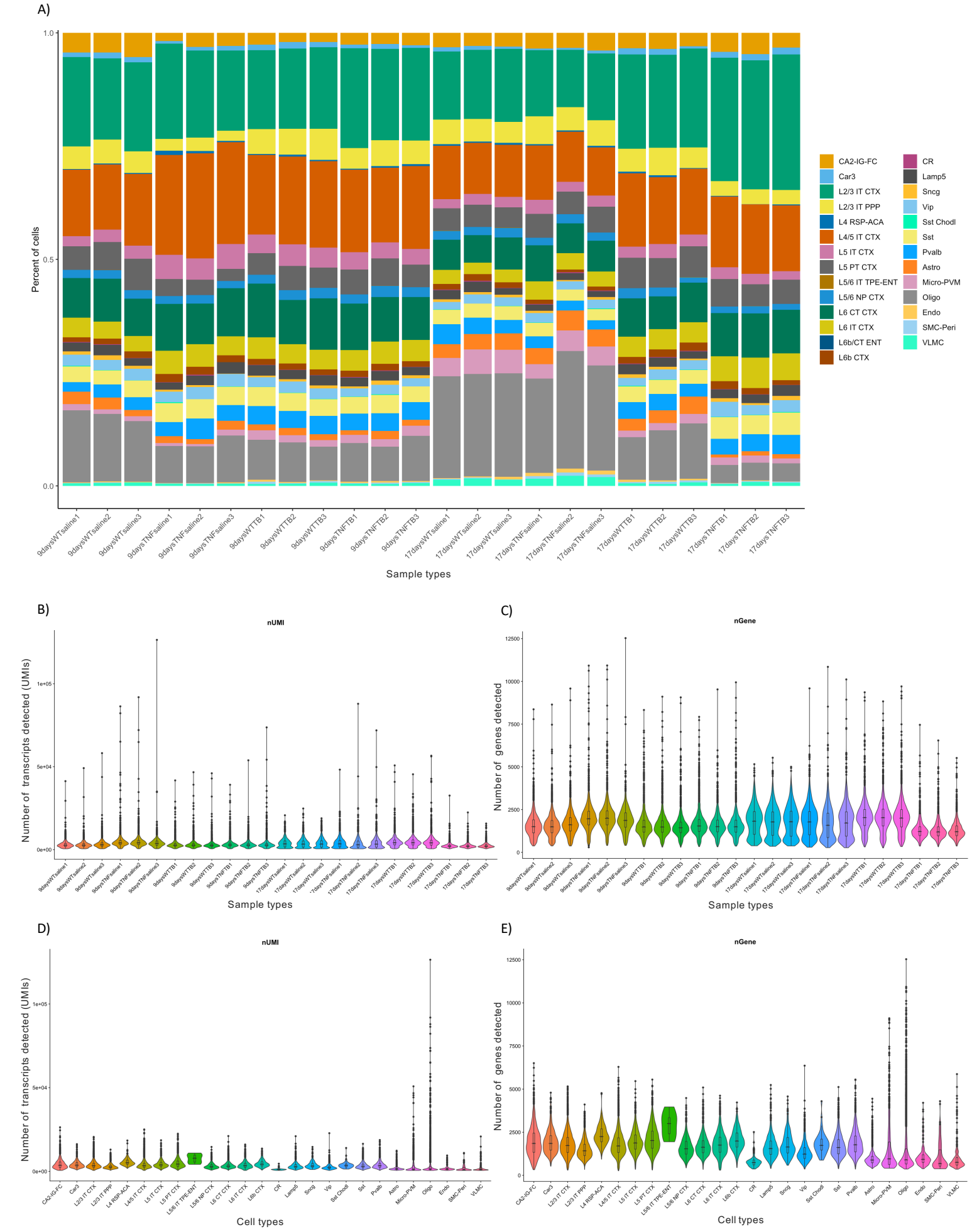


Figure 3.6 – see legend on next page

Figure 3.6: A) Stacked bar plot noting the proportion of nuclei (y axis) with respect to each cell type per sample (x axis). B) Violin plots showing the distribution of the number of transcripts (nUMI) and the number of genes (nGene) for each sample type and cell type. A box and whisker plot is embedded within each violin plot, showing the median, first quartile and third quartile data points. *L6b/CT ENT was excluded from the violin plots.*

The average number of nuclei per cell type per sample was 87 (Table 3.5). The majority of the datasets consisted of neuronal cells, of which the L2/3 IT CTX neuronal subtype was most abundant, followed by L4/5 IT CTX neurons. The oligodendrocytes had the most nuclei on average amongst the non-neuronal cell types. As expected, more excitatory neurons were observed in comparison to inhibitory neurons.

Table 3.5 The number of nuclei per cell type for each sample*. Cell types are grouped into excitatory neurons, inhibitory neurons, neuroglia and non-neural cells. Cell types are coloured by shades of green corresponding to the number of nuclei. Cell types with more nuclei have darker shades of green and vice versa.*

		Day 9												Day 17											
		Wildtype						TNF $^{-/-}$						Wildtype						TNF $^{-/-}$					
		Saline			TB			Saline			TB			Saline			TB			Saline			TB		
	Genes	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Excitatory Neurons	L2/3 IT CTX	386	467	363	441	444	470	448	427	320	632	584	587	379	391	357	355	379	330	317	349	327	649	576	640
	L2/3 IT PPP	113	144	125	146	158	162	62	67	50	149	177	167	125	126	121	100	89	92	132	123	146	83	98	93
	L4 RSP-ACA	6	7	8	7	5	7	21	10	4	6	10	12	10	9	10	5	12	6	10	8	11	4	4	4
	L4/5 IT CTX	307	395	313	441	501	449	482	526	424	554	488	540	295	320	283	242	262	263	253	241	266	338	365	373
	L5 IT CTX	47	78	74	96	100	99	123	102	95	108	104	100	62	56	56	40	56	62	51	65	76	57	52	45
	L5 PT CTX	120	141	133	129	109	113	78	70	40	134	149	131	104	110	112	130	140	118	110	147	146	124	101	123
	L5/6 IT TPE-ENT	2	4	2	0	2	0	0	0	0	0	3	2	3	1	2	1	1	0	0	0	1	4	2	0
	L5/6 NP CTX	36	67	30	58	63	58	37	43	28	55	75	66	53	47	44	22	31	37	30	46	35	44	36	26
	L6 CT CTX	187	230	195	281	267	242	235	202	204	305	289	280	158	170	163	148	149	170	140	168	184	225	203	219
	L6 IT CTX	105	127	95	132	123	147	125	116	87	153	154	147	66	76	90	72	95	87	80	76	99	156	140	136
	L6b/CT ENT	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	L6b CTX	36	36	23	38	42	36	40	36	27	39	47	38	29	29	20	31	18	31	17	14	29	42	35	30
	Car3	21	29	24	40	24	19	14	19	18	24	33	15	23	16	17	18	28	26	10	14	11	31	40	37
	CA2-IG-FC	109	129	129	53	79	71	71	91	64	105	116	117	47	97	82	67	69	72	79	83	80	151	153	115
Inhibitory Neurons	CR	0	4	1	0	0	2	3	1	0	2	1	0	4	2	4	2	0	1	4	2	2	2	1	2
	Lamp5	37	57	43	61	53	39	35	46	45	56	64	52	56	49	32	35	39	31	40	42	34	45	55	45
	Sncg	17	32	15	12	13	8	10	9	4	13	24	19	19	14	12	13	8	9	13	10	8	17	17	19
	Vip	41	56	34	60	51	42	45	54	43	51	54	47	40	51	48	34	35	41	45	43	43	63	50	58
	Sst Chodl	4	1	4	2	3	4	6	2	2	4	6	1	4	4	2	1	1	3	3	2	1	6	6	1
	Sst	77	89	78	112	100	106	95	95	74	107	122	98	74	79	67	67	60	46	65	73	69	107	98	110
	Pvalb	65	81	64	113	82	97	74	103	68	111	116	117	72	112	90	40	79	66	60	48	62	96	102	96
Neuroglia	Astrocytes	40	62	44	22	32	25	34	23	37	32	55	43	85	91	79	74	55	56	111	105	114	20	23	30
	Micro-PVM	20	25	12	28	45	39	8	5	12	36	39	48	134	135	126	38	19	20	119	114	85	27	31	31
	Oligodendrocytes	259	332	240	196	223	173	172	166	178	239	208	260	591	672	585	256	175	160	685	646	530	109	75	89
Non-Neural	Endo	3	5	2	3	6	7	2	2	3	8	9	6	11	11	9	9	8	11	29	22	25	2	5	1
	SMC-Peri	2	2	0	10	14	7	3	2	4	3	6	9	7	7	8	2	6	7	22	24	20	9	3	7
	VLMC	4	8	4	10	8	4	4	7	4	11	8	7	18	33	25	3	7	10	40	43	27	10	8	10

In order to verify the label transfer, the expression of known cell type-specific marker genes was plotted for the annotated cell types (Figure 3.7)(112). For example, inhibitory neurons are known to express high levels of *Glutamic acid decarboxylase 1* and *2(GAD1, GAD2)*, while some subtypes express *VIP* and *SST*. Since the clusters that were annotated as 'inhibitory neurons' express high concentrations of these same markers genes, these annotations were deemed suitable. Similarly, astrocytes are known to express high quantities of *Retinoic acid receptor-related orphan receptor B (RORB)* and Aquaporin 4 (*AQP4)*. The cluster annotated as 'astrocytes' displayed high expression of

those two genes, thereby validating the annotation. The rest of the cells also show a high degree of concordance between cell type-specific marker genes and the annotated cell types.

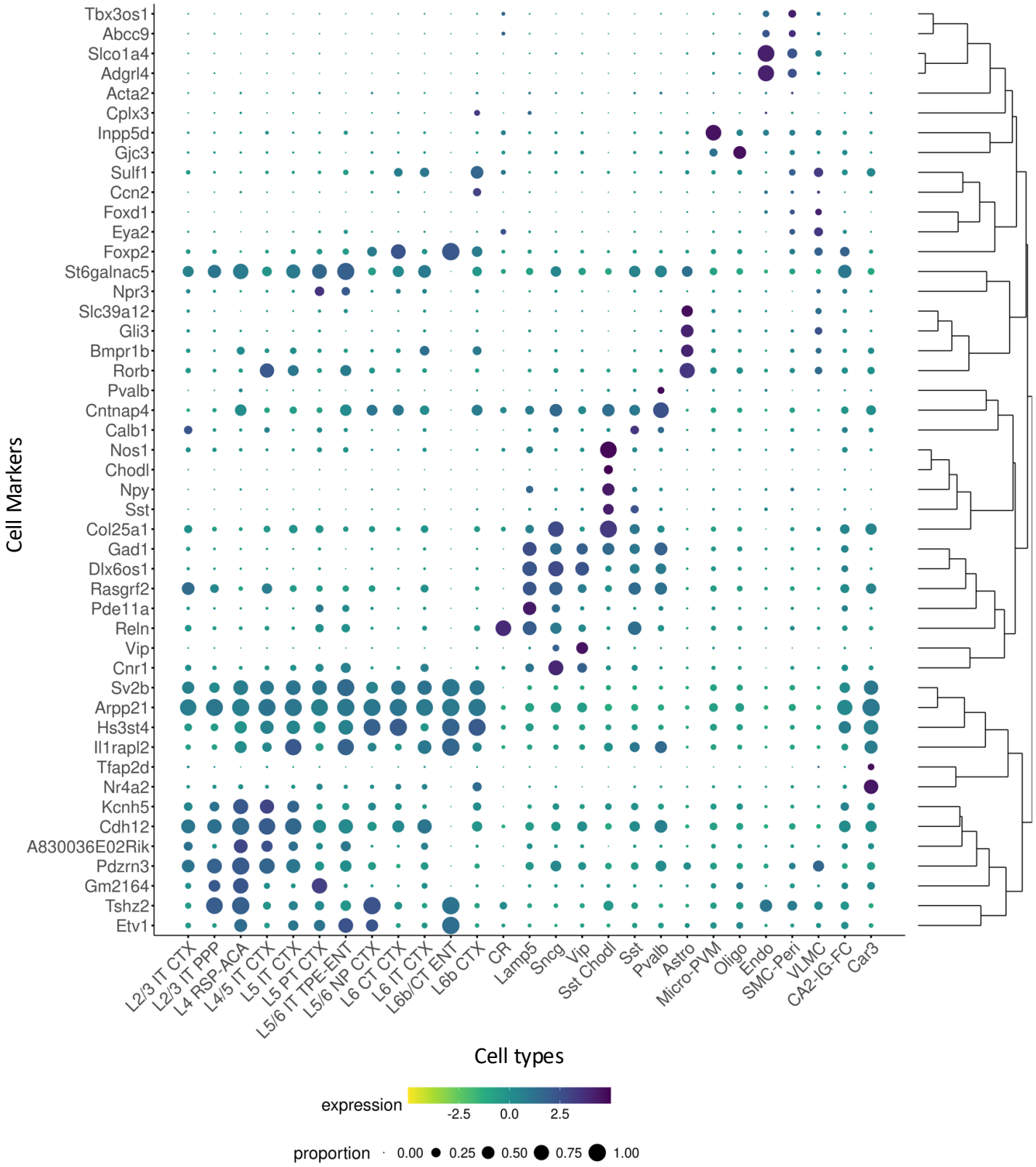


Figure 3.7 – see legend on next page

Figure 3.7: Dot plot showing the expression of known cell type-specific markers (y axis) with respect to each annotated cell type (x axis). *The darkness of the dot denotes the relative expression of the marker in a cell type, whereas the size of the dot indicates the proportional expression of each marker, with respect to other markers in the cell type.*

The analysis thus far has used stringent methods to filter the dataset and obtain sufficient quality nuclei for further analysis. Cell types were annotated with a high degree of confidence in preparation for differential expression analysis across conditions.

3.4 Differential expression analysis

After successfully annotating the clusters, differential gene expression analysis was performed using the DESeq2 pipeline. Analysis was conducted on the broad cell type clusters, namely excitatory neurons, inhibitory neurons, microglia-PVM, astrocytes and oligodendrocytes. Cell types were grouped to address the low number of nuclei per cell subtype and increase the statistical power of the analysis. In addition, excitatory neuronal subtypes L2/3 IT CTX and L4/5 IT CTX were analysed individually since these were the most abundant cell types across all the datasets.

3.4.1 Principal component analysis (PCA)

As part of the DESeq2 pipeline, dispersion estimate plots were created for each cell type, which showed a decrease in dispersion as the mean count increased (Supp. Figure 3.1). Datapoints also approximated the line of best fit, indicating that the model used by the DESeq2 pipeline was appropriate. PCA plots were then generated to determine which variable (day of sample collection [9 vs 17], genotype [WT vs $TNF^{-/-}$] or infection status [saline vs TB infection]) explained most of the variation between the samples (Figure 3.8). For microglia-PVM, astrocytes and oligodendrocytes, most of the variance was seen when delineating by infection status (saline vs TB infection) (Figure 3.8 E-G). However, for the neuronal datasets, a clear variance could be appreciated when stratifying the samples by infection status and day of collection (Day 9 vs Day 17) (Figure 3.8 A-D; Figure 3.9 A-D). In view of the PCA results, the samples were first grouped by infection status for differential expression analysis. The PCA results were then used to guide the comparisons tested using DESeq2.

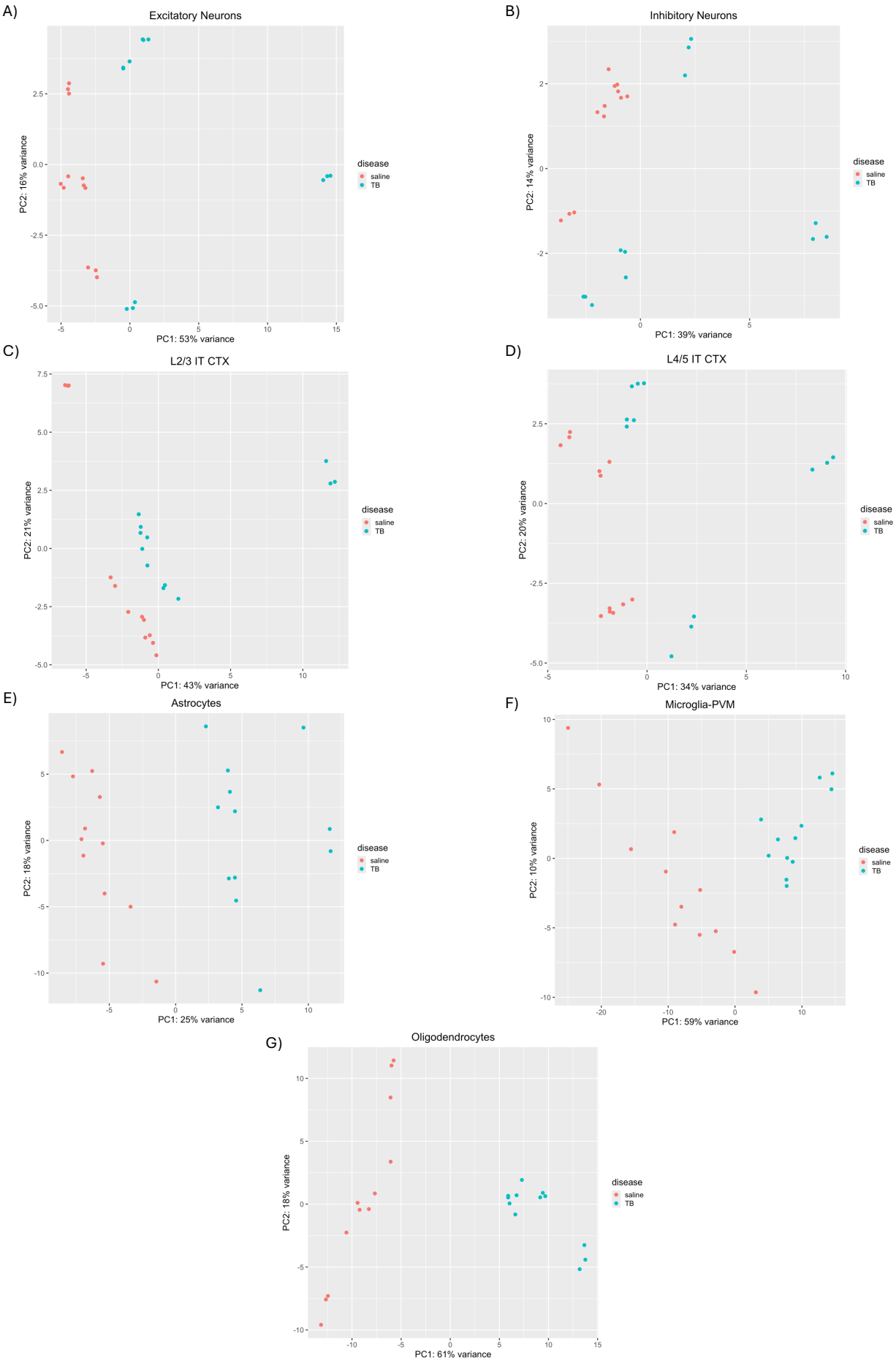


Figure 3.8: PCA plots of TB vs saline comparison. *A-G) PCA plots for each cell type in the 2D plane showing the first two principal components on the x and y-axis, respectively for the TB vs saline comparison. TB datasets are coloured in blue, whereas saline datasets are coloured in red.*

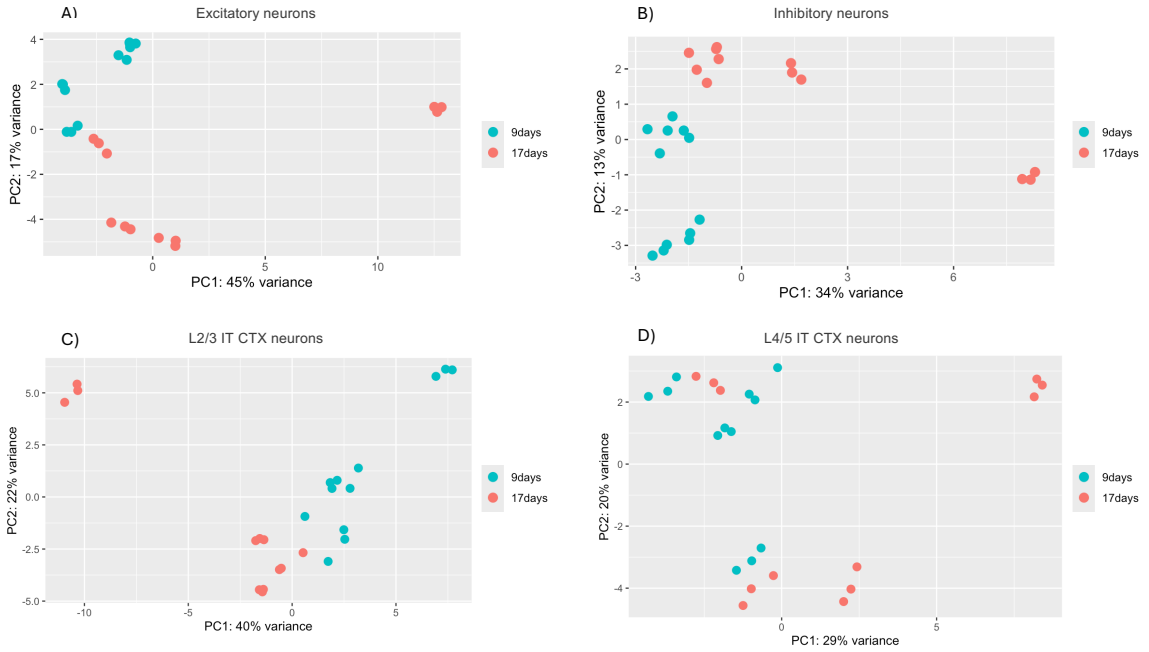


Figure 3.9: PCA plots of Day 17 vs Day 9 comparison. *A-D) PCA plots for each neuronal cell type in the 2D plane showing the first two principal components on the x and y-axis, respectively for the Day 17 vs Day 9 comparison. Day 9 datasets are coloured in blue, whereas Day 17 datasets are coloured in red.*

3.4.2 Preliminary conditions tested with DESeq2

Twelve conditions (Table 3.6) were initially tested using the DESeq2 pipeline. These included 1) the general TB response (irrespective of genotype), 2) the TB response in strictly in immunocompetent mice, 3) the TB response strictly in immunocompromised mice, 4) the effect of genotype on response to injury (i.e. injection of saline), 5) the effect of genotype in response to TB, 6) the TB response specifically in immunocompromised mice after controlling for the main disease and genotype effects, 7) the general TB response over time a focusing on neurons, 8) early TB response (irrespective of genotype) focusing on neurons, 9) late TB response (irrespective of genotype) focusing on neurons, 10) the effect of time on neuronal response to injury (i.e. injection of saline), 11) the effect of time on neuronal response to TB, and lastly 12) the TB response in neurons in infected mice harvested at Day 17, after controlling for the main effects of disease and time. Table 3.6 summarizes the number of upregulated and downregulated genes for each cell type in each of the above conditions. Three conditions (highlighted in bold) were selected for further downstream analysis.

Table 3.6: The number of upregulated and downregulated differentially expressed genes (DEGs) per condition for each chosen cell type. *Upregulated genes are coloured in green and downregulated genes in red. Colour intensity is proportional to the number of DEGs i.e. darker shades imply more DEGs. Conditions selected for further downstream analysis are highlighted in bold.*

Interpretation	Condition	Excitatory Neurons		Inhibitory Neurons		L2/3 IT CTX		L4/5 IT CTX		Astrocytes		Microglia-PVM		Oligodendrocytes	
		Up	*Down*	*Up*	*Down*	*Up*	*Down*	*Up*	*Down*	*Up*	*Down*	*Up*	*Down*	*Up*	*Down*
General TB response across all conditions	**TB vs saline (combined)**	**799**	**781**	**145**	**42**	**224**	**192**	**301**	**297**	**70**	**7**	**1039**	**949**	**1541**	**1207**
TB response in immunocompetent mice	TB vs saline in wildtype group	521	663	40	20	134	149	169	236	32	3	808	730	1301	1124
TB response in immunocompromised mice	TB vs saline in TNF group	1063	882	98	22	452	425	392	302	56	4	228	226	602	434
Effect of genotype on response to injury	TNF vs WT in saline group	720	557	42	15	377	154	255	188	1	0	80	5	158	112
Effect of genotype on response to TB	TNF vs WT in TB group	457	320	26	11	191	133	119	85	0	0	1	0	40	15
Genes uniquely expressed in TB due to immunosuppression	**Interaction effect (TNF & TB)**	**504**	**620**	**13**	**35**	**245**	**359**	**152**	**208**	0	0	**2**	**9**	**131**	**109**
General TB response over time	Day 17 vs Day 9 combined	1673	1218	374	121	876	502	386	164						
Early TB response (both genotypes)	TB vs saline in Day 9 group	424	549	30	9	119	180	152	177						
Late TB response (both genotypes)	TB vs saline in Day 17 group	1442	1163	236	50	439	258	343	306						
Effect of time on response to injury	Day 17 vs Day 9 in saline group	625	532	17	13	337	234	112	70						
Effect of time on response to TB	**Day 17 vs Day 9 in TB group**	**1946**	**1279**	**453**	**91**	**809**	**355**	**553**	**207**						
Genes uniquely expressed in TB due to time	Interaction effect (Day 17 & TB)	735	297	86	15	274	91	171	86						

3.4.3 DEGs in the general TB response

The first chosen condition measured the global effect of TB (compared to saline) on gene expression irrespective of genotype. In this comparison, the average of number of differentially expressed genes (DEGs) for each cell type was 542. The cell type with the most differentially expressed genes was the oligodendrocytes (1541 upregulated; 1207 downregulated) (Table 3.6). This was followed by the microglia-PVM population (1039 upregulated; 949 downregulated) and then the neuronal cell types (Table 3.6). Astrocytes had the lowest number of differentially expressed genes in the TB vs saline comparison (70 upregulated; 7 downregulated).

For the TB vs saline condition, the most upregulated DEG (denoted by the highest log2FC) in the excitatory neuron population was *GM50237* (Figure 3.10 A; Supp. Table 3.1). The most downregulated gene was *GABRA6* (Figure 3.10 A; Supp. Table 3.1). Amongst the inhibitory neurons, *IIGP1* was most upregulated, and *MAL* was most downregulated (Figure 3.10 B; Supp. Table 3.1). *IIGP1* was also the most upregulated DEG in L4/5 IT CTX neurons, oligodendrocytes, microglia-PVM and astrocytes (Figure 3.10 D-G; Supp. Table 3.1). The most downregulated DEG in these above cell types was *XLR4A, CNPY1, GM48614* and *ZSWIM6* respectively (Figure 3.10 D-G; Supp. Table 3.1). In L2/3 IT CTX neurons, the most upregulated DEG was *GM4951*, with *GM44691* being most downregulated (Figure 3.10 C; Supp. Table 3.1)

An UpSet plot shows the number of DEGs that were unique to or shared between different cell types (Figure 3.10 H). Curiously, 822 downregulated and 652 upregulated DEGs were solely identified in oligodendrocytes. In addition, 309 downregulated and 272 upregulated DEGs were uniquely found in the broad excitatory neuron cluster (and not the neuronal subtypes). Of note is the large number of unique DEGs in microglia-PVM (322 upregulated; 574 downregulated), despite the presence of few nuclei.

It is interesting that even though excitatory neurons had the largest number of nuclei, oligodendrocytes and microglia-PVM had more unique DEGs in the general TB response. Moreover,

357 upregulated and 169 downregulated DEGs were only found in microglia-PVM and oligodendrocytes. In addition, 106 upregulated and 66 downregulated DEGs were only identified amongst excitatory neurons and oligodendrocytes. A small set of DEGs showed opposite responses between cell types. For example, 33 DEGs that were upregulated in oligodendrocytes were also downregulated in microglia-PVM. Similarly, 28 upregulated DEGs in excitatory neurons were downregulated in microglia-PVM (Figure 3.10 H).

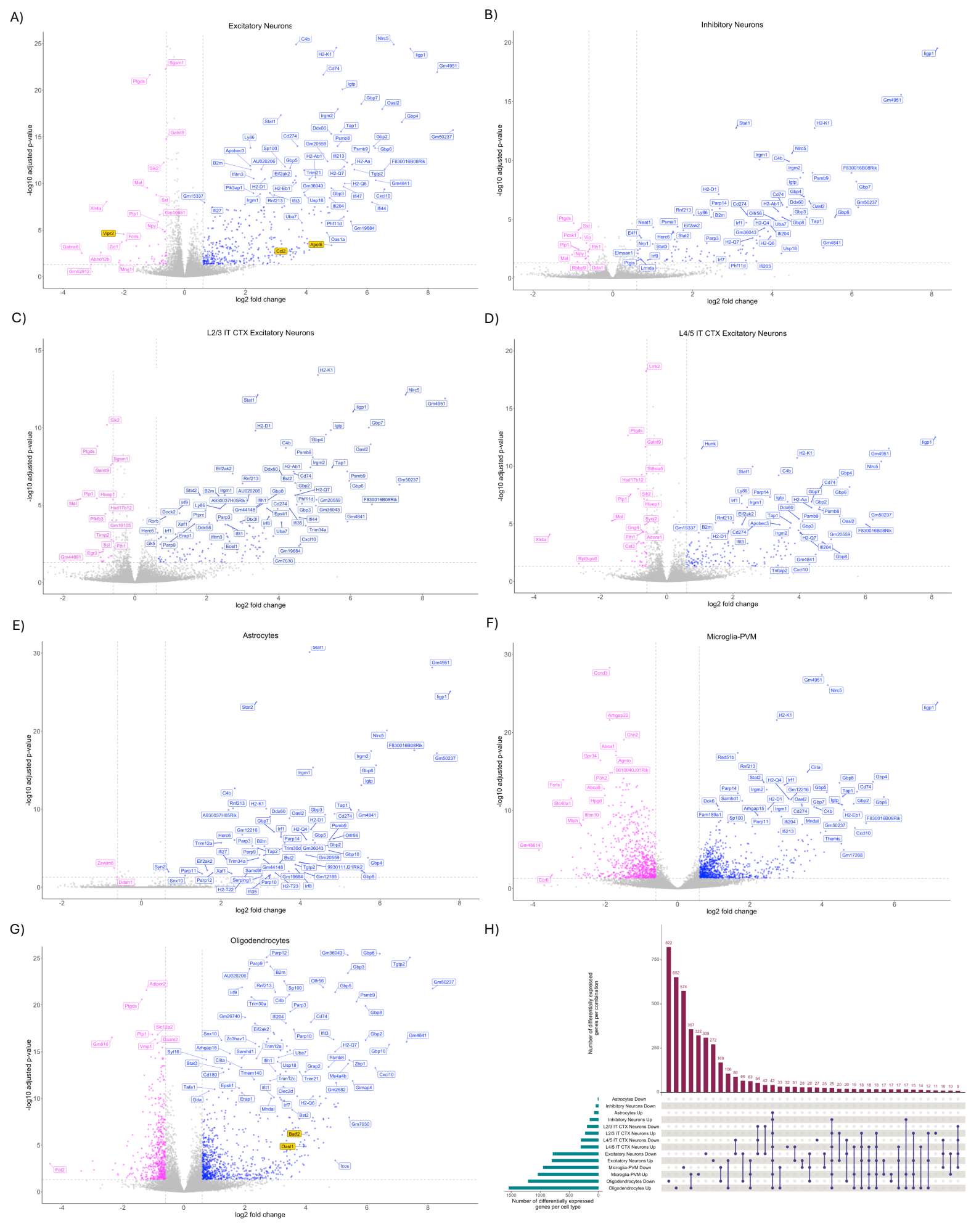


Figure 3.10 – see legend on next page

Figure 3.10: DESeq2 results for the TB vs saline condition. A-G) Volcano plots showing DEGS for each indicated cell type*. The log2 fold change (log_2FC) is shown on the x axes with the – log10 adjusted p-value on the y axes. Upregulated genes ($log_2FC > 0.6$ and adjusted p value < 0.05) are coloured in blue, whereas downregulated genes ($log_2FC < -0.6$ and adjusted p value < 0.05) are coloured in magenta with selected genes labelled*. Genes coloured in yellow were selected for downstream analysis **H) UpSet plot showing the total number of DEGs that were upregulated or downregulated per cell type (blue bars) as well as the number of DEGS intersecting across each cell type (red bars).** Data is only shown for groups with more than 8 DEGs.

3.4.4 DEGs in the TB response specifically in immunocompromised mice ($TNF^{-/-}$)

The second chosen condition identified DEGs in TB-infected $TNF^{-/-}$ mice where the observed expression level was significantly more or less than the background effect of TB and $TNF^{-/-}$ compared to WT controls (Table 3.6). Fewer DEGs on average per cell type (170) were identified compared to the TB vs saline comparison. This was expected since the interaction term has already controlled for the background effects of TB and $TNF^{-/-}$ (109). In this comparison, excitatory neurons had the most differentially expressed genes (504 upregulated; 620 downregulated), followed by L2/3 IT CTX neurons (245 upregulated; 359 downregulated), L4/5 IT CTX neurons (152 upregulated; 208 downregulated) and then oligodendrocytes (131 upregulated; 109 downregulated) (Table 3.6). No differentially expressed genes were identified in the astrocyte population.

The most upregulated DEG in the excitatory neuron population in the above condition was *TTR*, whilst *GM17087* was most downregulated (Figure 3.11 A; Supp. Table 3.2). *TTR* was also the most upregulated DEG in L2/3 IT CTX and L4/5 IT CTX neurons (Figure 3.11 C,D; Supp. Table 3.2). The most downregulated DEG in the above cell types was *GM17089* and *GM20089* respectively (Figure 3.11 C,D; Supp. Table 3.2). Amongst the inhibitory neurons, the most upregulated DEG was *BC1*, with *4921534H16Rik* being the most downregulated (Figure 3.11 B; Supp. Table 3.2). *HK2* was the most upregulated DEG in microglia-PVM, with *NRG3OS* being the most downregulated (Figure 3.11 E; Supp. Table 3.2). Finally, the most upregulated DEG in the oligodendrocytes was *GM3448,* whilst *AC171003.1* was the most downregulated DEG (Figure 3.11 F; Supp. Table 3.2). Curiously, 278 DEGs were upregulated in excitatory neurons that were not differentially expressed in other cells including the two excitatory neuronal subtypes. Similarly, 265 downregulated DEGs were exclusive to excitatory neurons. 89 downregulated and 51 upregulated DEGs were shared between both the broad excitatory neuron population as well as the two excitatory neuronal subtypes (Figure 3.11 G).

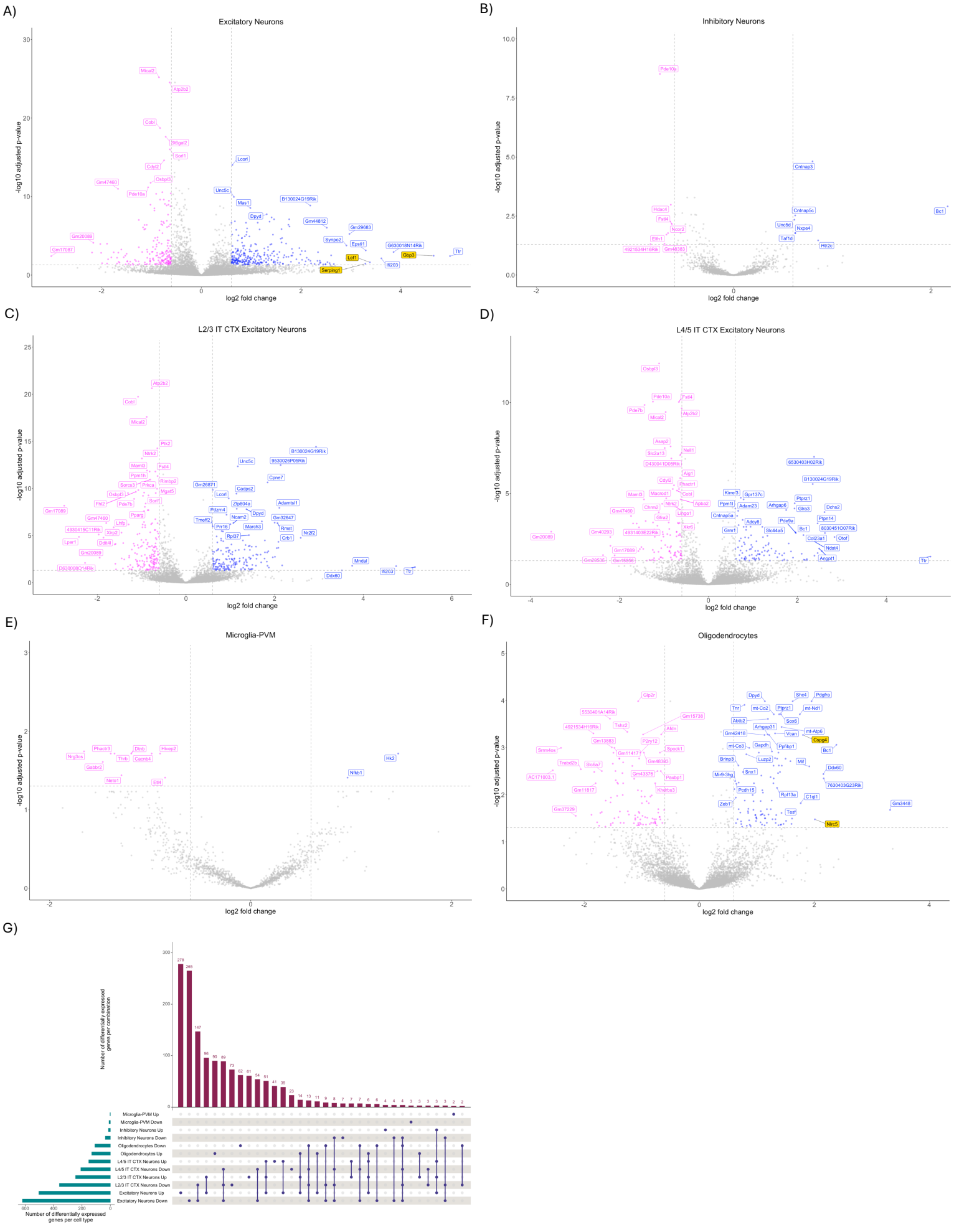


Figure 3.11 – see legend on next page

Figure 3.11: A-F) Volcano plots showing differentially expressed genes in the TB response specifically in immunocompromised mice. *The log_2FC is shown on the x axes with the – log10 adjusted p-value on the y axes. Upregulated genes ($log_2FC > 0.6$ and adjusted p value < 0.05) are coloured in blue, whereas downregulated genes ($log_2FC < -0.6$ and adjusted p value < 0.05) are coloured in magenta with selected genes labelled*. Genes coloured in yellow were selected for downstream analysis***. G) An UpSet plot showing the number of genes (blue bars) intersecting across each cell type (red bars) in the TB response specifically in immunocompromised mice.*** Data is only shown for groups with more than 1 DEG*. Astrocytes were excluded from this figure because no differentially expressed genes were found in this cell type.*

3.4.5 DEGs in TB-infected mice over time

The third chosen condition examined how the TB response changed with time (Day 17 vs Day 9) in TB-infected mice (Table 3.6). Since the PCA plots only showed clear differences across time amongst the neuronal cells, these cell types alone were isolated for further analysis. The average number of DEGs in this condition for the neuronal samples was 711. The cell type with the most DEGs was excitatory neurons (1946 upregulated; 1279 downregulated)(Table 3.6). This was followed by L2/3 IT CTX neurons (809 upregulated; 355 downregulated) and L4/5 IT CTX neurons (553 upregulated; 207 downregulated). Inhibitory neurons had the least number of DEGs (453 upregulated; 91 downregulated).

The most upregulated DEG amongst excitatory neurons in the above comparison was *E330023G01Rik*, whereas *CCN1* was the most downregulated DEG (Figure 3.12 A; Supp. Table 3.3). *NAT8F3* was the most upregulated DEG in inhibitory neurons, whereas *FOS* was most downregulated (Figure 3.12 B; Supp. Table 3.3). In L2/3 IT CTX neurons, the most upregulated DEG was IFI203, with *TTR* being the most downregulated DEG (Figure 3.12 C; Supp. Table 3.3). Amongst L4/5 IT CTX neurons, the most upregulated DEG was *GM20089*. *FOSB* was the most downregulated DEG in this cell type (Figure 3.12 D; Supp. Table 3.3). Curiously, 966 DEGs were exclusively upregulated in excitatory neurons (and not in the excitatory subtypes)(Figure 3.12 E). Similarly, 872 DEGs were exclusively downregulated in excitatory neurons. 227 upregulated and 29 downregulated DEGs were shared across all neuronal cell types. 123 upregulated and 82 downregulated DEGs were only shared between excitatory neurons and the excitatory subtypes (Figure 3.12 E).

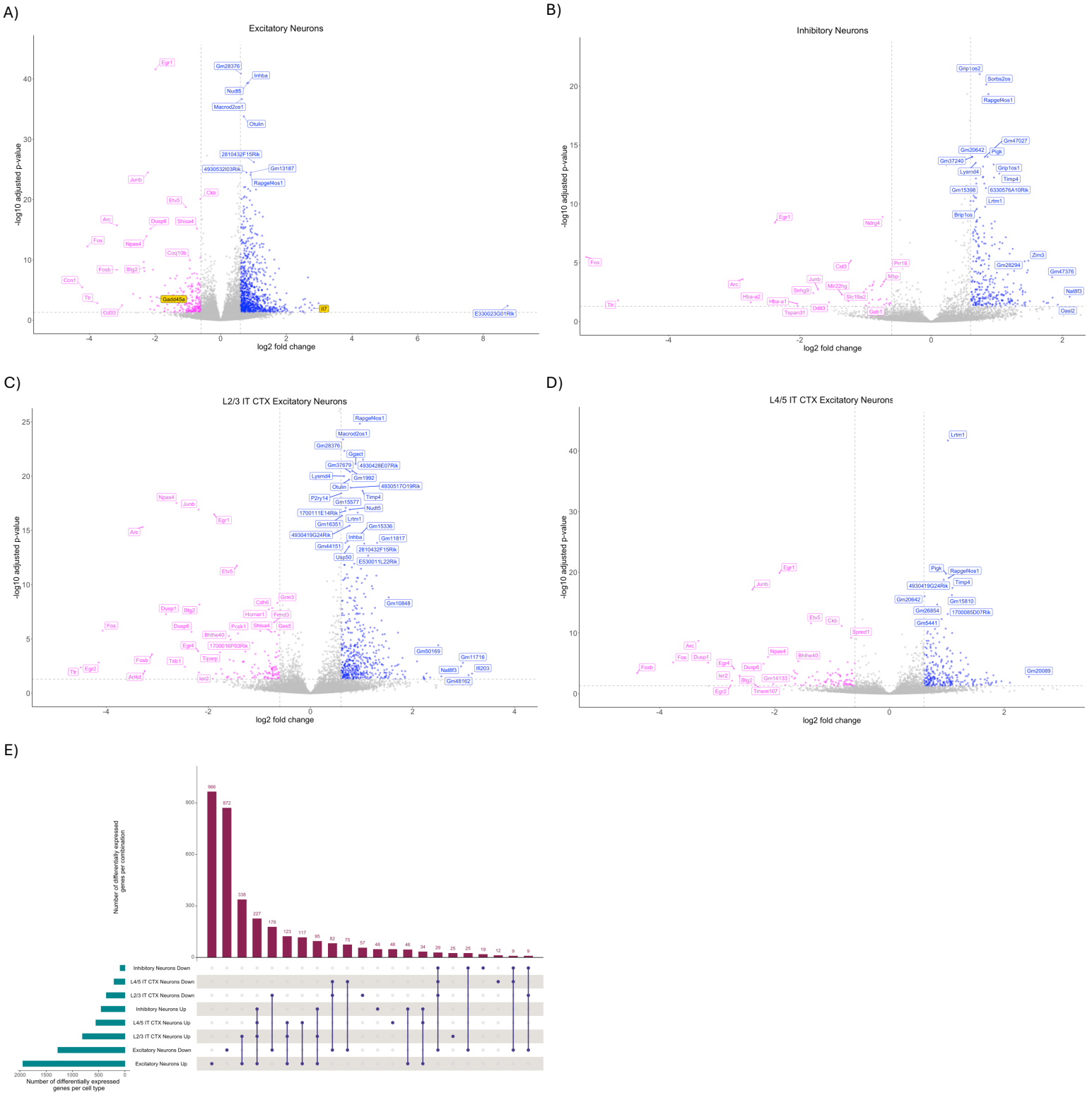


Figure 3.12: A-D) Volcano plots showing differentially expressed genes in TB-infected mice over time (Day 17 vs Day 9) amongst the neuronal cell types. *The* $\log_2$FC *is shown on the x axes with the – log10 adjusted p-value on the y axes. Upregulated genes (*$\log_2$FC *> 0.6 and adjusted p value < 0.05) are coloured in blue, whereas downregulated genes (*$\log_2$FC *< -0.6 and adjusted p value < 0.05) are coloured in magenta with selected genes labelled*. Genes coloured in yellow were selected for downstream analysis. ***E) An UpSet plot showing the number of genes (blue bars) intersecting across each neuronal cell type (red bars) in TB-infected mice over time.*** Data is only shown for groups with more than 8 DEGs.

3.5 Gene ontology analysis

Gene ontology (GO) analysis was performed using g:Profiler to determine if groups of upregulated or downregulated DEGs are implicated in known molecular functions, biological processes, cellular components or diseases (111). The GO terms described below for each cell type were selected amongst the list of g:Profiler-identified "driver terms" from each analysis. Genes associated with each GO term were correlated with the top 10 upregulated and downregulated DEGs per cell type for each condition. In GO terms where the associated genes were amongst the top 10 most DEGs by log_2FC, these genes are highlighted in bold.

3.5.1 GO analysis of the general TB response

3.5.1.1 Oligodendrocyte DEGs

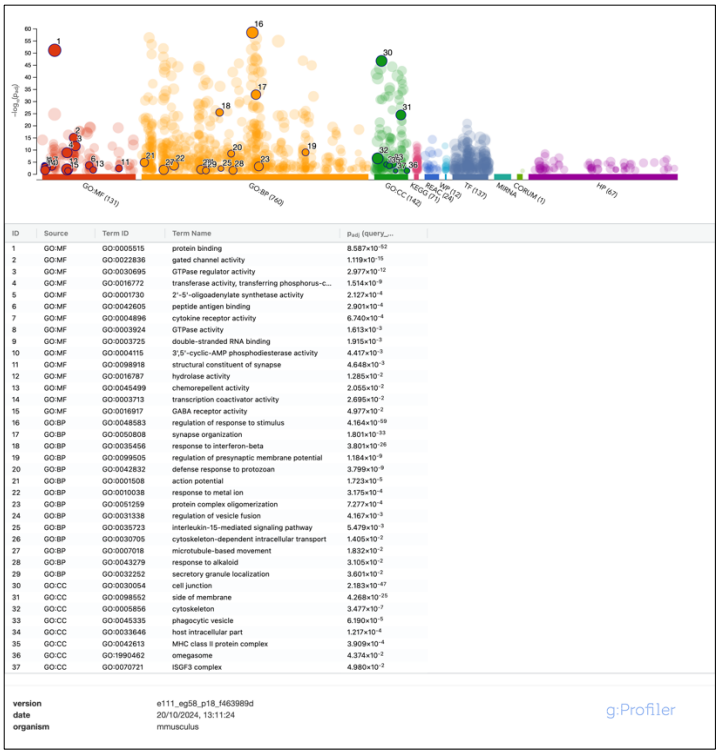
In the oligodendrocyte population, molecular function GO terms associated with DEGs that were upregulated with TB infection included *GTPase* activity (***IIGP1, F830016B08RIK, GBP4, TGTP2, GM4841, IGTP, GBP10, GM4951, GBP6***), hydrolase activity (***IIGP1, F830016B08RIK, GBP4, TGTP2, GM4841, IGTP, GBP10, GM4951, GBP6***), protein binding (***IIGP1, GBP4, IGTP, GBP10***) and Gamma-aminobutyric acid (GABA) receptor activity (*GABBR2, GABRA3, GABRB3, GABRG2, GABRB1, GABRA4, GABRA2*)(Figure 3.13 A; Supp. Table 3.4). Biological processes that were enriched in TB included response to IFN-β (***IIGP1, F830016B08RIK, TGTP2, GM4841, IGTP, GBP10, GM4951***), synapse organization (*SLIT1, GRIA1, CDH9, NEDD9, GPC4, SRGN, C3*) and action potentials (*TRPC4, CACNB2, GRIA1, PKP2, DSP, CD36*)(Figure 3.13 A; Supp. Table 3.4). The TNF signalling pathway was also enriched in TB vs saline samples (*BCL3, FAS ,CCL12, IRF1, IFI47, IL-18R1, CXCL10*)(Supp. Table 3.4).

Downregulated DEGs in TB infection were associated with GO terms such as structural constituents of the myelin sheath (*MAL, PLP1, MBP, MOBP, PLLP, GPM6B*) and ensheathment of neurons (*DLG1, ATRN, GPM6B, EPB41L3, NF1, ANK2*)(Figure 3.13 B; Supp. Table 3.5). This is notable because it suggests that myelination, a core function of oligodendrocytes, is being disrupted as a consequence of CNS-TB. Interestingly, diseases associated with the downregulated genes included Parkinson's disease (PD) (*COX7A2, UBC, PSMD12, PSMC6, ADCY5, KIF5B*) and Huntington's disease (HD) (*PSMD12, CREB3L2, HIP1, PSMC6, CREB1, KIF5B*), both of which are neurodegenerative diseases which could provide insight into the pathogenesis of CNS-TB (Supp. Table 3.5).

GO analysis was also performed on the 1474 DEGs that were uniquely upregulated and downregulated in oligodendrocytes compared to the other cell types. The 652 upregulated DEGs were associated with *GTPase* regulatory activity (*ASAP1, STXBP5, RAPGEF1, FARP1, RALGPS2*)(Figure 3.13 C; Supp. Table 3.6). These DEGs were also associated with the GABAergic synapse (PRKCB, ADCY9, GABBR2, ADCY3) and Th1/Th2 cell differentiation (*JAG2, IL-12RB2, NOTCH3, IL-4RA, ZAP70*)(Figure 3.13 C; Supp. Table 3.6). The 822 DEGs that were uniquely downregulated in oligodendrocytes were associated with catalytic activity (***CLCA4A**, DNAH11, SEPHS2, RAB7B, TLL1, HBA-A2, SCD1*) and processes related to nervous system development (*IQSEC1, ELAVL3, THOC2, KIRREL3*)(Figure 3.12 D; Supp. Table 3.7). As

was noted in the broader gene list, the downregulated DEGs were also associated with PD (*MAPT, ATP5G2, HSPA5, PSMC6, ADCY5, KIF5B*)(Figure 3.12 D; Supp. table 3.7).

A) Oligodendrocytes: Upregulated genes

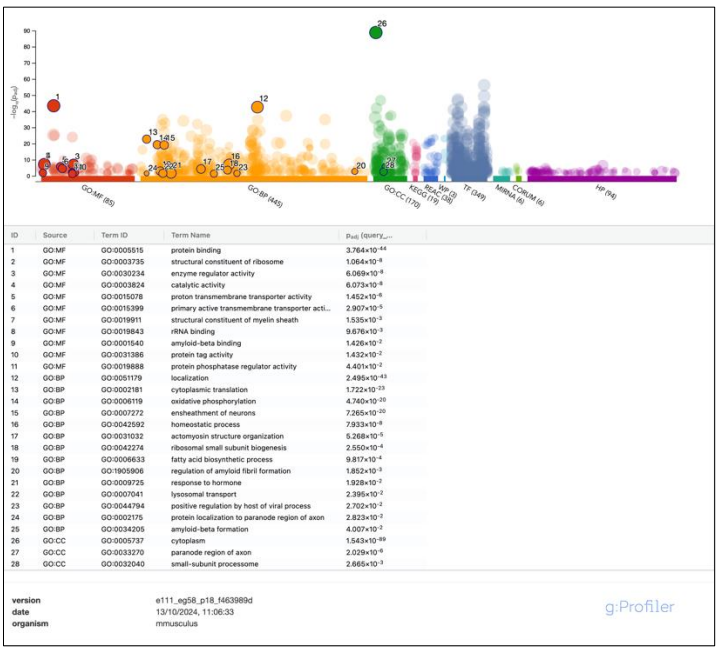


ID	Source	Term ID	Term Name	P_{adj} (query_...
1	GO:MF	GO:0005515	protein binding	8.587×10^{-52}
2	GO:MF	GO:0022836	gated channel activity	1.119×10^{-15}
3	GO:MF	GO:0030695	GTPase regulator activity	2.977×10^{-12}
4	GO:MF	GO:0016772	transferase activity, transferring phosphorus-c...	1.514×10^{-9}
5	GO:MF	GO:0001730	2'-5'-oligoadenylate synthetase activity	2.127×10^{-4}
6	GO:MF	GO:0042605	peptide antigen binding	2.901×10^{-4}
7	GO:MF	GO:0004896	cytokine receptor activity	6.740×10^{-4}
8	GO:MF	GO:0003924	GTPase activity	1.613×10^{-3}
9	GO:MF	GO:0003725	double-stranded RNA binding	1.915×10^{-3}
10	GO:MF	GO:0004115	3',5'-cyclic-AMP phosphodiesterase activity	4.417×10^{-3}
11	GO:MF	GO:0098918	structural constituent of synapse	4.648×10^{-3}
12	GO:MF	GO:0016787	hydrolase activity	1.285×10^{-2}
13	GO:MF	GO:0045499	chemorepellent activity	2.055×10^{-2}
14	GO:MF	GO:0003713	transcription coactivator activity	2.695×10^{-2}
15	GO:MF	GO:0016917	GABA receptor activity	4.977×10^{-2}
16	GO:BP	GO:0048583	regulation of response to stimulus	4.164×10^{-59}
17	GO:BP	GO:0050808	synapse organization	1.801×10^{-33}
18	GO:BP	GO:0035456	response to interferon-beta	3.801×10^{-26}
19	GO:BP	GO:0099505	regulation of presynaptic membrane potential	1.184×10^{-9}
20	GO:BP	GO:0042832	defense response to protozoan	3.799×10^{-9}
21	GO:BP	GO:0001508	action potential	1.723×10^{-5}
22	GO:BP	GO:0010038	response to metal ion	3.175×10^{-4}
23	GO:BP	GO:0051259	protein complex oligomerization	7.277×10^{-4}
24	GO:BP	GO:0031338	regulation of vesicle fusion	4.167×10^{-3}
25	GO:BP	GO:0035723	interleukin-15-mediated signaling pathway	5.479×10^{-3}
26	GO:BP	GO:0030705	cytoskeleton-dependent intracellular transport	1.405×10^{-2}
27	GO:BP	GO:0007018	microtubule-based movement	1.832×10^{-2}
28	GO:BP	GO:0043279	response to alkaloid	3.105×10^{-2}
29	GO:BP	GO:0032252	secretory granule localization	3.601×10^{-2}
30	GO:CC	GO:0030054	cell junction	2.183×10^{-47}
31	GO:CC	GO:0098552	side of membrane	4.268×10^{-25}
32	GO:CC	GO:0005856	cytoskeleton	3.477×10^{-7}
33	GO:CC	GO:0045335	phagocytic vesicle	6.190×10^{-5}
34	GO:CC	GO:0033646	host intracellular part	1.217×10^{-4}
35	GO:CC	GO:0042613	MHC class II protein complex	3.909×10^{-4}
36	GO:CC	GO:1990462	omegasome	4.374×10^{-2}
37	GO:CC	GO:0070721	ISGF3 complex	4.980×10^{-2}

version	e111_eg58_p18_f463989d
date	20/10/2024, 13:11:24
organism	mmusculus

g:Profiler

B) Oligodendrocytes: Downregulated genes

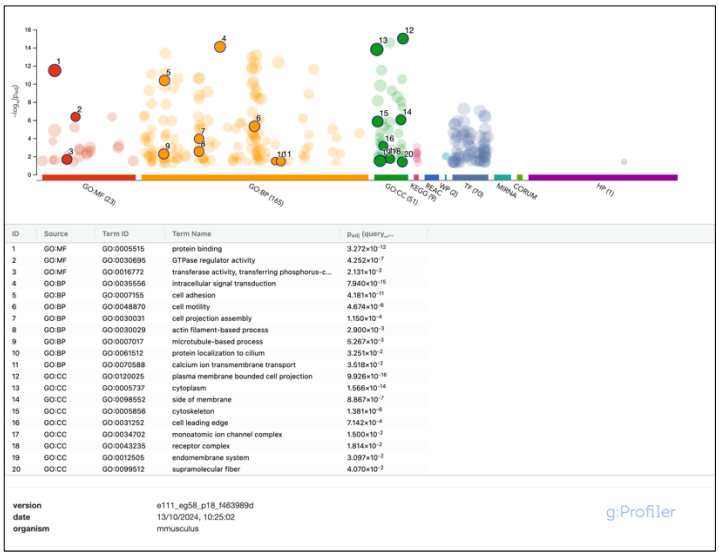


ID	Source	Term ID	Term Name	P_{adj} (query_...
1	GO:MF	GO:0005515	protein binding	3.764×10^{-44}
2	GO:MF	GO:0003735	structural constituent of ribosome	1.064×10^{-8}
3	GO:MF	GO:0030234	enzyme regulator activity	6.069×10^{-6}
4	GO:MF	GO:0003824	catalytic activity	6.073×10^{-6}
5	GO:MF	GO:0015078	proton transmembrane transporter activity	1.452×10^{-6}
6	GO:MF	GO:0015399	primary active transmembrane transporter acti...	2.907×10^{-5}
7	GO:MF	GO:0019911	structural constituent of myelin sheath	1.535×10^{-3}
8	GO:MF	GO:0019843	rRNA binding	9.676×10^{-3}
9	GO:MF	GO:0001540	amyloid-beta binding	1.426×10^{-2}
10	GO:MF	GO:0031386	protein tag activity	1.432×10^{-2}
11	GO:MF	GO:0019888	protein phosphatase regulator activity	4.401×10^{-2}
12	GO:BP	GO:0051179	localization	2.495×10^{-43}
13	GO:BP	GO:0002181	cytoplasmic translation	1.722×10^{-23}
14	GO:BP	GO:0006119	oxidative phosphorylation	4.740×10^{-20}
15	GO:BP	GO:0007272	ensheathment of neurons	7.265×10^{-10}
16	GO:BP	GO:0042592	homeostatic process	7.933×10^{-8}
17	GO:BP	GO:0031032	actomyosin structure organization	5.268×10^{-5}
18	GO:BP	GO:0042274	ribosomal small subunit biogenesis	2.550×10^{-4}
19	GO:BP	GO:0006633	fatty acid biosynthetic process	9.817×10^{-4}
20	GO:BP	GO:1905906	regulation of amyloid fibril formation	1.852×10^{-3}
21	GO:BP	GO:0009725	response to hormone	1.928×10^{-3}
22	GO:BP	GO:0007041	lysosomal transport	2.395×10^{-2}
23	GO:BP	GO:0044794	positive regulation by host of viral process	2.702×10^{-2}
24	GO:BP	GO:0002175	protein localization to paranode region of axon	2.823×10^{-2}
25	GO:BP	GO:0034205	amyloid-beta formation	4.007×10^{-2}
26	GO:CC	GO:0005737	cytoplasm	1.543×10^{-89}
27	GO:CC	GO:0033270	paranode region of axon	2.029×10^{-6}
28	GO:CC	GO:0032040	small-subunit processome	2.665×10^{-3}

version	e111_eg58_p18_f463989d
date	13/10/2024, 11:06:33
organism	mmusculus

g:Profiler

C) Oligodendrocytes: Cell type specific upregulated genes

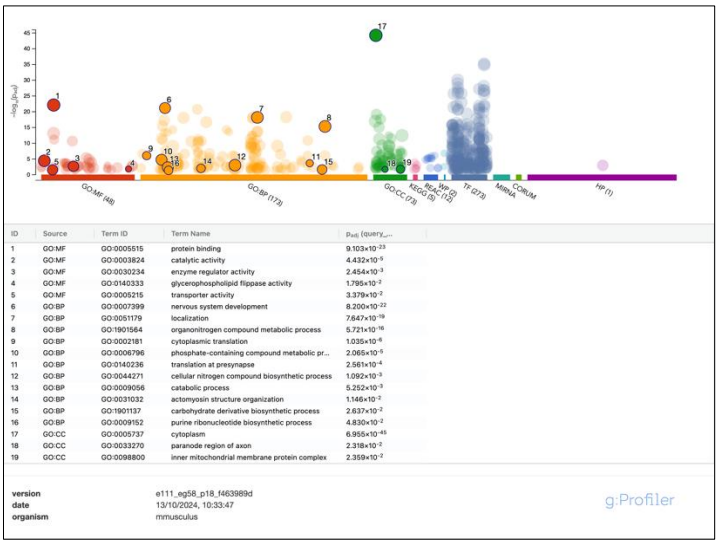


ID	Source	Term ID	Term Name	P_{adj} (query_...
1	GO:MF	GO:0005515	protein binding	3.272×10^{-12}
2	GO:MF	GO:0030695	GTPase regulator activity	4.252×10^{-7}
3	GO:MF	GO:0016772	transferase activity, transferring phosphorus-c...	2.131×10^{-2}
4	GO:BP	GO:0035556	intracellular signal transduction	7.940×10^{-15}
5	GO:BP	GO:0007155	cell adhesion	4.181×10^{-11}
6	GO:BP	GO:0048870	cell motility	4.674×10^{-6}
7	GO:BP	GO:0030031	cell projection assembly	1.150×10^{-4}
8	GO:BP	GO:0030029	actin filament-based process	2.900×10^{-3}
9	GO:BP	GO:0007017	microtubule-based process	5.267×10^{-3}
10	GO:BP	GO:0061512	protein localization to cilium	3.251×10^{-2}
11	GO:BP	GO:0070588	calcium ion transmembrane transport	3.518×10^{-2}
12	GO:CC	GO:0120025	plasma membrane bounded cell projection	9.926×10^{-16}
13	GO:CC	GO:0005737	cytoplasm	1.566×10^{-14}
14	GO:CC	GO:0098552	side of membrane	8.867×10^{-7}
15	GO:CC	GO:0005856	cytoskeleton	1.381×10^{-6}
16	GO:CC	GO:0031252	cell leading edge	7.142×10^{-4}
17	GO:CC	GO:0034702	monoatomic ion channel complex	1.500×10^{-2}
18	GO:CC	GO:0043235	receptor complex	1.814×10^{-2}
19	GO:CC	GO:0012505	endomembrane system	3.097×10^{-2}
20	GO:CC	GO:0099512	supramolecular fiber	4.070×10^{-2}

version	e111_eg58_p18_f463989d
date	13/10/2024, 10:25:02
organism	mmusculus

g:Profiler

D) Oligodendrocytes: Cell type specific downregulated genes



ID	Source	Term ID	Term Name	P_{adj} (query_...
1	GO:MF	GO:0005515	protein binding	9.103×10^{-23}
2	GO:MF	GO:0003824	catalytic activity	4.432×10^{-3}
3	GO:MF	GO:0030234	enzyme regulator activity	2.454×10^{-3}
4	GO:MF	GO:0140333	glycerophospholipid flippase activity	1.795×10^{-2}
5	GO:MF	GO:0005215	transporter activity	3.379×10^{-2}
6	GO:BP	GO:0007399	nervous system development	8.200×10^{-22}
7	GO:BP	GO:0051179	localization	7.647×10^{-19}
8	GO:BP	GO:1901564	organonitrogen compound metabolic process	5.721×10^{-16}
9	GO:BP	GO:0002181	cytoplasmic translation	1.035×10^{-6}
10	GO:BP	GO:0006796	phosphate-containing compound metabolic pr...	2.065×10^{-5}
11	GO:BP	GO:0140236	translation at presynapse	2.561×10^{-4}
12	GO:BP	GO:0044271	cellular nitrogen compound biosynthetic process	1.092×10^{-3}
13	GO:BP	GO:0009056	catabolic process	5.252×10^{-3}
14	GO:BP	GO:0031032	actomyosin structure organization	1.146×10^{-2}
15	GO:BP	GO:1901137	carbohydrate derivative biosynthetic process	2.637×10^{-2}
16	GO:BP	GO:0009152	purine ribonucleotide biosynthetic process	4.830×10^{-2}
17	GO:CC	GO:0005737	cytoplasm	6.955×10^{-45}
18	GO:CC	GO:0033270	paranode region of axon	2.318×10^{-2}
19	GO:CC	GO:0098800	inner mitochondrial membrane protein complex	2.359×10^{-2}

version	e111_eg58_p18_f463989d
date	13/10/2024, 10:33:47
organism	mmusculus

g:Profiler

Figure 3.13. Representative g:Profiler outputs showing the GO driver terms and their adjusted p values (Padj) in oligodendrocytes under the general TB response condition. *GO terms are first ordered by GO category and then by Padj value. Numbered data points are represented by solid-coloured circles in the Manhattan plots. GO driver terms associated with (A) upregulated DEGs, B) downregulated DEGs, C) cell type-specific upregulated DEGs and D) cell type-specific downregulated DEGs are shown.* BP, biological process; CC, cellular component; *MF,* molecular function. *The g:Profiler outputs for the remaining cell types and conditions can be found in the associated supplementary data.*

3.5.1.2 Microglia-PVM DEGs

DEGs upregulated in microglia-PVM were associated with molecular function terms such as *GTPase* regulator activity (***GBP2****, RALGPS1, RASAL2, ARHGAP32, DENND1A, RABGAP1L*), structural constituents of synapses (*ERC1, SPTBN2, CAMK2B, SHANK1, RIMS2*) and IL-15 receptor receptor activity (*IL-15RA, IL-2RB, IL-2RG*)(Supp. Figure 3.2 A; Supp. Table 3.8). Biological processes related to the upregulated DEGs included synapse organization (*NPTN, NEDD4, NBEA, PTK2, MDGA2*), regulation of membrane potential (*PPP3CA, ANK2, NEDD4, ATXN1, NRXN1*) and the innate immune response (***IIGP1, GBP6, GBP4, CXCL10, GBP2, TAP1***)(Supp. Figure 3.2 A; Supp. Table 3.8). Curiously, diseases associated with the upregulated DEGs included Epstein-Barr virus (EBV) infection (***CXCL10, H2-EB1, TAP1***) and Type 1 Diabetes Mellitus (***H2-EB1****, PTPRN2, PTPRN, ICA1*)(Supp. Table 3.8).

Downregulated DEGs were associated with molecular functions such as 1-phosphatidylinositol-4-phosphate 3-kinase activity (*PIK3CG, PIK3CA, PIK3CB, PIK3CD*) and dioxygenase activity (*BCO2, P3H2, PHYHD1, ALOX5AP*)(Supp. Figure 3.2 B; Supp. Table 3.9). Biologically processes such as intracellular signal transduction (***CCR6****, MLPH, CLCF1, HFE*) and B-cell receptor signalling pathway (*CD300A, NFAM1, BLNK, SH2B2*) were also associated with the downregulated DEGs (Supp. Figure 3.2 B; Supp. Table 3.9). As was seen with the oligodendrocytes, diseases related to the downregulated DEGs included PD (*ATP5D, UQCRB, UQCR11, COX7A2*) and HD (*ATP5D, UQCRB, UQCR11, COX7A2*)(Supp. Table 3.9).

The upregulated DEGs exclusively found in microglia-PVM were found to be related to the regulation of resting membrane potential (*NPTN, OXR1, NRXN1, AMPH, GRIK2*), as noted in the wider microglia-PVM group (Supp. Figure 3.2 C; Supp. Table 3.10). Downregulated DEGs solely expressed in microglia-PVM correlated with immune receptor activity (***CCR6****, C5AR2, C3AR1, IL-6RA, CX3CR1, IFNGR1*), B-cell receptor signalling pathway (*CD300A, NFAM1, BLNK, SH2B2, BTK*) as well as negative regulation of B-cell proliferation (*TNFRSF13B, CD300A, BTK, PKN1*)(Supp. Figure 3.2 D; Supp. Table 3.11).

3.5.1.3 Excitatory neuron DEGs

Within the excitatory neuron population, upregulated DEGs were related to molecular function terms including 2'-5'-oligoadenylate synthetase activity (***OASL2****, OAS2, OAS1G, OAS1A*) and pattern recognition receptor activity (*TRIM12A, TRIM30D, TRIM30A, IFIH1*)(Supp. Figure 3.3 A; Supp. Table 3.12). Biological processes associated with the upregulated DEGs included the response to IFN-β (***GM4951, IIGP1, F830016B08RIK, GM4841***), autophagy regulation (***IIGP1****, TRIM12A, TRIM34A, TRIM30D, IRGM1*), IL-1β production (*GBP5, STAT3, PML, NOD1, RIPK2*), T-cell migration (***CXCL10****, ITGA4, ITGAL*) and the TNF signalling pathway (***CXCL10****, IRF1, CSF1, MLKL, CCL12*)(Supp. Figure 3.3 A; Supp. Table 3.12). Diseases that were related to the above DEGs included EBV infection (***CXCL10****, EIF2AK2,H2-T22, STAT2, TAP2, B2M*), HIV infection (*H2-T22, TAP2, TRIM30D, B2M, H2-K1*), Type 1 Diabetes Mellitus (*H2-T22, H2-K1, H2-T23, H2-Q4, H2-D1*) and Hepatitis C (***CXCL10****, EIF2AK2, STAT2, STAT1, STAT3*)(Supp. Table 3.12).

DEGs that were downregulated in excitatory neurons were associated with biological processes such as regulation of *GTPase* activity (*CHN1, RIPOR2, CRK, RAP1GDS1, WNK1*) along with regulation of the

neuron apoptotic process (*SLC1A1, KCNB1, APOE, MT1, CDC42*)(Supp. Figure 3.3 B; Supp. Table 3.13). Neurodegenerative diseases were also shown to be associated with the downregulated DEGs (*UQCRB, MT-ND4, MT-CO3, RPS27A, NDUFA13*)(Supp. Table 3.13). Upregulated DEGs that were solely identified in excitatory neurons were related to kinase activity (*MAPK13, FES, NUAK2, MAPKAPK*)(Supp. Figure 3.3 C; Supp. Table 3.14). No GO terms were identified in DEGs solely downregulated in excitatory neuron DEGs.

3.5.1.4 L2/3 IT CTX neuron DEGs

Upregulated DEGs identified within L2/3 IT CTX neurons were found to be correlated with molecular functions such as *GTPase* activity (***GM4951, F830016B08RIK, GBP7, IIGP1, GBP6, GM4841***) and pattern recognition receptor activity (*TRIM12A, TRIM30D, TRIM30A, IFIH1, TRIM12C*)(Supp. Figure 3.4 A; Supp. Table 3.15). Biological processes that were enriched included the defence response to other organisms (***GM4951, NLRC5, F830016B08RIK, OASL2, GBP7, IIGP1, GBP6, GM4841***), negative T-cell selection (***THEMIS, CD74, DOCK2, PTPRC***) along with positive regulation of lysosome organization and autophagosome maturation (*IRGM1, IGTP, IRGM2*)(Supp. Figure 3.4 A; Supp. Table 3.15). Diseases associated with the above DEGs included Type 1 Diabetes Mellitus (*H2-K1, H2-T23, H2-Q4, H2-D1*), Influenza A (*EIF2AK2, STAT2, STAT1, IRF9*) and Hepatitis C (*EIF2AK2, STAT2, STAT1, IRF9*)(Supp. Table 3.15).

As was seen with oligodendrocytes, DEGs downregulated in L2/3 IT CTX neurons included molecular functions such as the structural constituent of the myelin sheath (***MAL, PLP1****, MBP*)(Supp. Figure 3.4 B; Supp. Table 3.16). Biological processes associated with the downregulated DEGs included the response to organic cyclic compound (***PTGDS, SST****, COMT, HSPA8, MT-ND1, YWHAH*), the response to organonitrogen compound (***SST****, SLC1A1, BAIAP2*) and learning (*SLC1A1, COMT, SYT11, NF1, NTRK2*)(Supp. Figure 3.4 B; Supp. Table 3.16). Degenerative diseases were once again associated with the downregulated DEGs in this cell type (*RPS27A, COX4I1, COX6C, MT-ND1*)(Supp. Table 3.16). Upregulated DEGs that were solely found in L2/3 IT CTX neurons were related to polyribonucleotide nucleotidyltransferase activity (Supp. Figure 3.4 C; Supp. Table 3.17). No GO terms were identified in DEGs solely downregulated in L2/3 IT CTX neurons.

3.5.1.5 L4/5 IT CTX neuron DEGs

In the L4/5 IT CTX neuron population, upregulated DEGs were associated molecular functions relating to *GTPase* activity (***IIGP1, GM4951, F830016B08RIK, GBP6, GBP4, GBP8***) and cytokine receptor activity (*IL-1RAP, IL-15RA, CD74, GHR, IL-10RA, LIFR, ACKR4*)(Supp. Figure 3.5 A; Supp. Table 3.18). Biological processes that were correlated with the upregulated DEGs included the innate immune response (***IIGP1, NLRC5, GBP6, OASL2, GBP4, GBP8***), response to IFN-β (***IIGP1, GM4951, F830016B08RIK***), autophagy regulation (***IIGP1****, TRIM34A, IRGM1, IGTP, IRGM2, IIGP1, TRIM30A, STAT3, TRIM12C*), Nuclear factor kappa-light-chain-enhancer of activated B-cells (NF-κβ) signal transduction (*TRIM34A, GBP7, IRF1, STAT1, IRGM1, IGTP, IRGM2*) and positive regulation of lysosome organization (*IRGM1, IGTP, IRGM2*)(Supp. Figure 3.5 A; Supp. Table 3.18). Diseases associated with the upregulated DEGs included EBV infection (*EIF2AK2, STAT2, TAP2, B2M, H2-K1*), Type 1 Diabetes

mellitus (*H2-K1, H2-T23, H2-Q4, H2-D1, PTPRN2*) and Toxoplasmosis (*STAT1, IRGM1, IGTP, IRGM2, STAT3*)(Supp. Table 3.18).

Amongst the downregulated DEGs in L4/5 IT CTX neurons, GO terms were enriched for receptor localization to synapse (*ARHGAP44, DLG1, CACNG2, STX1B*) and regulation of cell migration (***PLP1****, DMTN, RIPOR2, RAPGEF2, SST, TMSB4X*)(Supp. Figure 3.5 B; Supp. Table 3.19). In DEGs solely upregulated and downregulated in this cell type, no significant GO terms were identified.

3.5.1.6 Inhibitory neuron DEGs

Amongst the inhibitory neurons, molecular functions enriched in upregulated DEGs included *GTPase* activity (***IIGP1, GM4951, GBP7, F830016B08RIK, GBP6, GM4841***), T-cell receptor binding (*H2-T22, H2-K1, H2-T23, H2-Q4*), NKcell lectin-like receptor binding (*H2-T22, H2-K1, H2-T23, H2-Q4*) and CD8 receptor binding (*H2-T22, H2-K1, H2-Q4, H2-D1*)(Supp. Figure 3.6 A; Supp. Table 3.20). Biological processes that were associated with the upregulated DEGS included the innate immune response (***IIGP1, GBP7, GBP6, OASL2***), response to IFN-β (***IIGP1, GM4951, GBP7, F830016B08RIK, GM4841***) along with positive regulation of lysosome organization and autophagosome maturation (*IRGM1, IGTP, IRGM2*)(Supp. Figure 3.6 A; Supp. Table 3.20). Disease processes related to the upregulated DEGs were enriched for EBV infection (*EIF2AK2, H2-T22, STAT2, TAP2, B2M*), Type 1 Diabetes mellitus (*H2-T22, H2-K1, H2-T23, H2-Q4, H2-D1, PTPRN*), Toxoplasmosis (*STAT1, IRGM1, IGTP, IRGM2*) and Hepatitis C (*EIF2AK2, STAT2, STAT1, STAT3*)(Supp. Table 3.20).

Downregulated DEGs in inhibitory neurons were associated with exocytosis (***PTGDS, NPY****, SYT11*) and the myelin sheath (***MAL, PLP1****, STIP1, CALM1, CDC42*)(Supp. Figure 3.6 B; Supp. Table 3.21). No GO terms were identified in DEGs uniquely found in inhibitory neurons.

3.5.1.7 Astrocyte DEGs

In the astrocyte population, upregulated DEGs were associated with molecular functions such as *GTPase* activity (***IIGP1, GM4951, F830016B08RIK, GBP6, IRGM2, GBP4, GBP8, IGTP***), T-cells receptor binding (*H2-T22, H2-K1, H2-T23,H2-Q4, H2-D1*), NK cell lectin-like receptor binding (*H2-T22, H2-K1, H2-T23, H2-Q4, H2-D1*), CD8 receptor binding (*H2-T22, H2-K1, H2-Q4, H2-D1*)(Supp. Figure 3.7 A; Supp. Table 3.22). Biologically processes related to the upregulated DEGs included the defense response to other organisms (***IIGP1, GM4951, F830016B08RIK, NLRC5, GBP6, IRGM2, GBP4, GBP8, IGTP***), NF-κβ signal transduction (***IRGM2, IGTP****, TRIM12A, TRIM34A, TRIM30D*), positive regulation of autophagy (***IRGM2, IGTP****, TRIM12A, TRIM34A*) and positive regulation of lysosome organization and autophagosome maturation (*IRGM2, IGTP, IRGM1*)(Supp. Figure 3.7 A; Supp. Table 3.22). Diseases such as EBV infection (*EIF2AK2, H2-T22, STAT2, TAP2*), Type 1 Diabetes Mellitus (*H2-T22, H2-K1, H2-T23*), HIV infection (*H2-T22, TAP2, TRIM30D, B2M, H2-K1*) and Toxoplasmosis (*IRGM2, IGTP*) were all associated with the upregulated DEGs (Supplementary Table 3.22). No GO terms were identified amongst the set of downregulated DEGs in astrocytes.

3.5.1.8 DEGs that are shared across cell types

The DEGs shared between cell types was used to determine pathways and processes that are prevalent across various cell types, which could give insight into general cellular responses in CNS-TB infection. In the above comparison, 42 DEGs were shared across all the cell types. Curiously, all these genes were upregulated i.e. none were downregulated. GO molecular function terms associated with the upregulated DEGs included *GTPase* activity (*GBP7, GBP5, IRGM1, GBP2, GBP3, GM4841, IGTP, GBP8, GBP4, IRGM2, GBP6, F830016B08RIK, GM4951, IIGP1*), MHC class 1 protein binding (*H2-T22l, TAP2, H2-K1, H2-T23, H2-Q4, H2-D1, TAP1*) and T-cell receptor binding (*H2-T22, H2-K1, H2-T23, H2-Q4, H2-D1*)(Supp. Figure 3.8 A; Supp. Table 3.23). The biological processes that were enriched included the defence response to other organisms (*IGTP, GBP8, GBP4, IRGM2, GBP6, NLRC5, F830016B08RIK, GM4951, IIGP1* amongst others), positive regulation of lysosome organization and autophagosomal maturation (*IRGM1, IGTP, IRGM2*) and NF-κβ signal transduction (*TRIM34A, GBP7, IRF1, STAT1, IRGM1, IGTP, IRGM2*)(Supp. Figure 3.8 A; Supp. Table 3.23). The upregulated DEGs have also been associated with infectious diseases such as HIV (*H2-T22, TAP2, B2M, H2-K1, H2-T23, H2-Q4, H2-D1, TAP1*), EBV (*H2-T22, STAT2, TAP2, B2M, H2-K1, H2-T23, H2-Q4, STAT1, H2-D1, TAP1*) and Herpes Simplex 1 infection (*H2-T22, STAT2, TAP2, B2M, H2-K1, H2-T23, H2-Q4, STAT1, H2-D1, TAP1*). Type 1 Diabetes Mellitus has also been associated with several of the upregulated DEGS (*H2-T22, H2-K1, H2-T23, H2-Q4, H2-D1*)(Supp. Table 3.23).

The upregulated DEGs that were only identified in microglia-PVM and oligodendrocytes were associated with glutamate receptor binding (*NSF, SHANK1, NETO1, GRIN1, SHANK2, RASGRF*), synapse organization (*IL-1RAPL1, EPHA5, STAU2, FRMPD4, CAMK2B, CADM1*), regulation of membrane potential (*ATXN1, OPRM1, CACNA2D1, SLC17A7, KCNQ5, NLGN1*) and nervous system development (*VPS13A, NIN, ATXN1, IL-1RAPL1, MACROD2, EPHA5*)(Supp. Figure 3.8 B; Supp. Table 3.24). Neurodegenerative pathways were also enriched for these shared upregulated DEGs (*ATXN1, CAMK2B, PLCB1, MAPK10, GRIN1*)(Supplementary Table 3.24). Downregulated DEGs only shared between microglia-PVM and oligodendrocytes were related to lymphocyte homeostasis (*SLC40A1, MIF, RPS6, TSC22D3, ABL1, DOCK10*), macroautophagy (*SCOC, MAP1LC3A, CTSD, GABARAP*), and the myelin sheath (*GAPDH, PMP22, UBC,MIF, PPIA, NDRG1*) (Supp. Figure 3.8 C; Supp. Table 3.25). Neurodegenerative diseases such as PD (*ATP5D, UQCR11, COX7A2, NDUFB11*), HD (*ATP5D, UQCR11, COX7A2, NDUFB11*) and AD (*ATP5D, UQCR11, GAPDH, COX7A2)* were also enriched for in the downregulated DEGs (Supplementary Table 3.25).

The upregulated DEGs solely identified in excitatory neurons and oligodendrocytes were associated with negative regulation of IP-10 production (*LIL-RB4A, OAS1G, OAS1A*) and regulation of NF-κβ signal transduction (*TNIP2, ESR1, TRIM25, LIL-RB4A, TRAF1, CARD11, TNFSF10*)(Supp. Figure 3.8 D; Supp. Table 3.26). Downregulated DEGs in the same cell types were correlated with the regulation of amyloid fibril formation (*CRYAB, PFDN2, APP*) and neurodegenerative diseases (*NDUFB9, MT-ND5, NDUFA2, TUBB4A, NDUFS5*)(Supp. Figure 3.8 E; Supp. Table 3.27).

DEGs that were upregulated in oligodendrocytes but downregulated in microglia-PVM were associated with B-cell activation (*CMTM7, CD86, LYN, PLCG2*), regulation of intracellular signal transduction (*SIPA1L2, OPHN1, CD86, LYN, PLCG2, BANK1*) and tolerance induction to self-antigens (*LYN, TGFBR2*)(Supp. Figure 3.9 A; Supp. Table 3.28). Lastly, DEGs that were upregulated in excitatory

neurons but downregulated in microglia-PVM were related to negative regulation of leukocyte activation (*CD84, CD37, INPP5D, MERTK*)(Supp. Figure 3.9 B; Supp. Table 3.29).

3.5.1.9 Correlating GO terms with the top 10 DEGs

The majority of the top 10 upregulated DEGs in each cell type were related to *GTPase* activity (oligodendrocytes, L2/3 IT CTX neurons, L4/5 IT CTX neurons, inhibitory neurons, astrocytes), underscoring the high degree of activity in TB-infected mice. Even though GO driver terms associated with upregulated DEGs in excitatory neurons and microglia-PVM did not include *GTPase* activity, it should be reiterated that *GTPase* activity was associated with the upregulated DEGs shared between all cell types. Other GO terms that showed a strong correlation with the top 10 upregulated DEGs included the response to IFN-β (oligodendrocytes, excitatory neurons, L4/5 IT CTX neurons, inhibitory neurons) and innate immune response (microglia-PVM, L4/5 IT CTX neurons, inhibitory neurons). GO terms that were correlated with the top 10 downregulated DEGs included the structural integrity of the myelin sheath (L2/3 IT CTX neurons and inhibitory neurons).

3.5.2 GO analysis of the TB response specifically in immunocompromised mice ($TNF^{-/-}$)

The second condition identified DEGs in TB-infected $TNF^{-/-}$ mice where the observed expression level was significantly more or less than the sum of the background effects of TB and $TNF^{-/-}$ with respect to WT saline mice. Microglia-PVM and astrocytes were excluded from the analysis due to the low number of DEGs identified.

3.5.2.1 Excitatory neuron DEGs

Upregulated DEGs in excitatory neurons for the above condition were enriched for molecular functions such as glutamate receptor activity (*GRID2, GRIN3A, GRIA3, GRM5, GRIK2*)(Supp. Figure 3.10 A; Supp. Table 3.30). This is interesting because excessive glutamate is one of the markers of excitotoxicity (46, 47). Biological processes associated with the upregulated genes included the regulation of post synaptic membrane potential (*GRID2, GRIN3A, GRIA3, GRM5, PRKN*) and cell surface receptor signaling pathways (***LEF1**, GRID2, GAB2, CDH13, GRIN3A, GIGYF1*) (Supp. Figure 3.10 A; Supp. Table 3.30).

DEGs downregulated in excitatory neurons were associated with molecular function terms such as *GTPase* regulator activity (*IQSEC3, IQSEC1, PLCD4, STARD8, ASAP*) and curiously, glutamate receptor binding (*ATP2B2, CACNG2, AKAP5, GRIN2A, CABP1, SHANK1*)(Supp. Figure 3.10 B; Supp. Table 3.31). Biological processes correlated with the downregulated DEGs included neuron development (*THRB, RUNX1T1, ATP2B2, NTRK2, SPOCK1, AFDN*) and synapse organization (***COLQ**, PTPRT, CACNB4, IQSEC3, ASIC2, ATP2B2*)(Supp. Figure 3.10 B; Supp. Table 3.31). Diseases associated with the downregulated DEGs included long-term depression (*CRHR1, PRKCA, PRKCB, PLCB4*) , nicotine addiction (*GABRA1, SLC17A7, GRIN2A, CACNA1A*) and Cushing syndrome (*CACNA1F, CRHR1, HSD3B3, KCNK2, CACNA1G*) (Supp. Table 31).

The analysis of upregulated DEGs solely identified in excitatory neurons revealed GO terms related to the negative regulation of the innate immune response (***SERPING1****, PRKDC, TRAFD1, GRN, SAMHD1, LGALS9*)(Supp. Figure 3.10 C; Supp. Table 3.32). This is expected since TNF, a key component of the innate immune response (38), is expected to be non-functional in this condition. Neurodegenerative pathways (*PSMA3, CAMK2D, UCHL1, APAF1*) were also associated with upregulated DEGs solely identified in excitatory neurons (Supplementary Table 3.32). Downregulated DEGs only expressed in excitatory neurons were associated with *GTPase* binding (*INF2, FMNL1, VPS9D1, RIC1, ERC1, FNBP1L, RPH3A*)(Supp. Figure 3.10 D; Supp. Table 3.33).

3.5.2.2 L2/3 IT CTX neuron DEGs

In the L2/3 IT CTX neuron population, upregulated DEGs were associated with the regulation of postsynaptic membrane potential (*GRID2, GRIN3A, GRM5, PRKN, GRIK2*)(Supp. Figure 3.11 A; Supp. Table 3.34). Downregulated DEGs were enriched for *GTPase* regulator activity (*IQSEC3, STARD8, ASAP2, DAB2IP, TBC1D*) and synapse organisation (*TNR, IQSEC3, ASIC2, ATP2B2, NTRK2, NOS1AP*)(Supp. Figure 3.11 B; Supp. Table 3.35). Interestingly, neuron development (*TNR, THRB, ATP2B2, NTRK2, PHACTR1, EXT1*) and glutamate receptor signalling pathway (*GRIN2A, KCNB1, KALRN, CAMK2A, PLCB1*) were also associated with the downregulated DEGs (Supp. Figure 3.11 B; Supp. Table 3.35). As with the broader excitatory neuron population, diseases associated with the downregulated DEGs in L2/3 IT CTX neurons included long-term depression (*PRKCA, PRKCB, PLCB4, CACNA1A, IGF1R*), morphine addiction (*GABBR2, PDE10A, PDE7B, PRKCA*) and nicotine addiction (*SLC17A7, GRIN2A, CACNA1A, GABRG3, GRIN2B*)(Supp. Table 3.35).

Downregulated DEGs exclusively identified in L2/3 IT CTX neurons were associated with the regulation of postsynaptic membrane neurotransmitter receptor levels **(*GSG1****, NETO2, CACNG3, CLSTN1, DLG2*)(Supp. Figure 3.11 C; Supp. Table 3.36). No significant GO terms were noted for upregulated DEGs solely identified in L2/3 IT CTX neurons.

3.5.2.3 L4/5 IT CTX neuron DEGs

In keeping with the findings of the other excitatory neuronal groups, upregulated DEGs in L4/5 IT CTX neurons were associated with the regulation of post synaptic membrane potential (*GRIN3A, GRIA3, PRKN, GRIK2, RELN*)(Supp. Figure 3.12 A; Supp. Table 3.37). In addition, upregulated DEGs were curiously enriched for neuron differentiation (*NCAM2, GRIN3A, PTPRZ1, CNTN5, CPEB3*) along with glutamate receptor activity (*GRIN3A, GRIA3, GRIK2, GRM1*) and the positive regulation of long-term synaptic potentiation (*CPEB3, RELN, CHRNA7, ZDHHC2*)(Supp. Figure 3.12 A; Supp. Table 3.37).

Downregulated DEGs in L4/5 IT CTX neurons were associated with *GTPase* activator activity (*STARD8, ASAP2, DAB2IP, ASAP1*) and glutamate receptor binding (*ATP2B2, CACNG2, GRIN2A, SHANK1*)(Supp. Figure 3.12 B; Supp. Table 3.38). In addition, neurodegenerative pathways (*MAPK11, PRKCA, GRIN2A, CAMK2G*) along with atherosclerosis (*MAPK11, HSP90B1, PPARG, PTK2, PRKCA*) were correlated with the downregulated DEGs (Supp. Table 3.38).

When considering the upregulated DEGs exclusively identified in L4/5 IT CTX neurons, GO terms such as the positive regulation of long-term synaptic potentiation (*ZDHHC2, ADCY8, PDE9A*) were noted, as seen in the broader cell group (Supp. Figure 3.12 C; Supp. Table 3.39). No significant GO terms were identified in downregulated DEGs solely identified in L4/5 IT CTX neurons.

3.5.2.4 Oligodendrocyte DEGs

In the oligodendrocyte population, upregulated DEGs were notably associated with glutamatergic synaptic transmission (*NLGN1, GRID2, TNR, GRIN3A, CACNG*)(Supp. Figure 3.13 A; Supp. Table 3.40). In addition, neurodegenerative pathways (*MT-CO3, ITPR2, MT-CO1, UBB, MT-ND4*) along with Diabetic cardiomyopathy (*MT-CO3, MT-CO1, MT-ND4, ATP5G1*) were correlated with the upregulated DEGs (Supp. Table 3.40). Downregulated DEGs were associated with nervous system development (*DYNLT1F, DYNLT1A, DLX6OS1, OPRM1*) (Supp. Figure 3.13 B; Supp. Table 3.41).

In consideration of the upregulated DEGs solely identified in oligodendrocytes, GO analysis revealed an interesting association with neurogenesis pathways (*CSPG4, C1QL1, NLGN1, PTPRK, CTNNA*)(Supp. Figure 3.13 C; Supp. Table 3.42). This is notable because it may indicate how the mouse brain is attempting to respond to TB infection in the immunocompromised state, through the generation of new neurons. Finally, no significant GO terms were noted in downregulated DEGs solely identified in oligodendrocytes.

3.5.2.5 Inhibitory neuron DEGs

Upregulated DEGs in inhibitory neurons, like oligodendrocytes, were also curiously associated with neurogenesis (*CNTNAP5C, CNTNAP3, UNC5D, DYNC1I2*)(Supp. Figure 3.14 A; Supp. Table 3.43). No significant GO terms were identified from the downregulated DEGs in inhibitory neurons.

3.5.2.6 DEGs that are shared across cell types

No DEGs (upregulated or downregulated) were shared across all cell types in this specific condition. Downregulated DEGs that were only identified in excitatory neurons, L2/3 IT CTX neurons and L4/5 IT CTX neurons were associated with glutamate receptor binding (*CACNG2, SHANK1, SHANK2, CAMK2A*) and synapse organization (*DAB2IP, LRFN2, CACNG2, PTK2, RIMBP2*)(Supp. Figure 3.15 A; Supp. Table 3.44). Upregulated DEGs found only in the above cell types did not reveal significant GO terms.

It is curious to note that GO terms related to glutamate activity were both upregulated and downregulated in excitatory neurons. In addition, the TNF signalling pathway was not associated with any of the DEGs for this condition, which is expected since the condition is analysing the TB response specifically in immunocompromised mice (with respect to WT saline controls) where functionality of TNF was removed.

3.5.2.7 Correlating GO terms with the top ten DEGs

There were was not a strong relationship between the GO terms and the top ten upregulated and downregulated DEGs for each cell type in the above condition.

3.5.3 GO analysis in TB-infected mice over time

The third condition analysed how gene expression levels in TB-infected mice changed over two analysed time points. Neuroglia were excluded from this analysis based on the PCA findings described earlier (Figure 3.9).

3.5.3.1 Excitatory neuron DEGs

In the excitatory neuron population, DEGs that were upregulated in TB-infected mice over time were associated with molecular functions such as glutamate receptor binding (*NSF, RASGRF1, GRIP1, DLG4, SHANK2*) and glutamate-gated receptor activity (*GRIN2B, GRID1, GRIN2A, GRIK3, PTK2B*)(Supp. Figure 3.16 A; Supp. Table 3.45). This is interesting because it suggests that glutamate activity increases with disease progression. Biological processes related to the upregulated DEGs included synapse organization (*DAB2IP, LRFN2, CACNG2, PTK2, RIMBP2*)(Supp. Figure 3.16 A; Supp. Table 3.45). Diseases such as dilated cardiomyopathy (*ADCY9, SLC8A3, SLC8A*) and long-term depression (*GNAQ, GNAO1, PRKG1, PRKCG*) were also correlated with the upregulated DEGs (Supplementary Table 3.45).

Downregulated DEGs in excitatory neurons were enriched for neuron projection development (***ARC, EGR2, BTG2***)(Supp. Figure 3.16 B; Supp. Table 3.46). These DEGs were also notably correlated with neurogenerative pathways (*PTGS2, MT-ATP6, ATP5G1, MT-CYTB, MAP1LC3A*) and Diabetic cardiomyopathy (*MT-ATP6, ATP5G1, MT-CYTB, NDUFA13*)(Supplementary Table 3.46).

Amongst the upregulated DEGs solely identified in excitatory neurons, GO analysis revealed an association with the negative regulation of cell population proliferation (*CELF1, PPP2R5C, SMAD3, NEURL1A, KDM2B*)(Supp. Figure 3.16 C; Supp. Table 3.47). No significant GO terms were noted in downregulated DEGs solely identified in excitatory neurons.

3.5.3.2 L2/3 IT CTX neuron DEGs

Upregulated DEGs in L2/3 IT CTX neurons were associated with synapse organization (*SYBU, BAIAP2, SPTBN4, SLC8A3*) along with the regulation of synapse structure (*SYBU, BAIAP2, GRID1, LRRN1*) (Supp. Figure 3.17 A; Supp. Table 3.48). Dilated and Hypertrophic cardiomyopathy (*SLC8A3, SLC8A1, CACNA1D, CACNA1C*) were also related to the upregulated DEGs (Supplementary Table 3.48).

Downregulated DEGs were correlated with cell surface receptor signaling pathways (***FOS, NPAS4, TRIB1***) and neurodegenerative pathways (*MT-ATP6, ATP5G1, MT-CYTB, NDUFA13, MT-CO3)*(Supp. Figure 3.17 B; Supp. Table 3.49). In downregulated DEGs solely identified in L2/3 IT CTX neurons, GO terms such as synapse organization (*SDK1, PDZRN3, CASK, SORBS2, CACNA1A*) were noted (Supp.

Figure 3.17 C; Supp. Table 3.50). No significant GO terms were noted in upregulated DEGs solely found in L2/3 IT CTX neurons.

3.5.3.3 L4/5 IT CTX neuron DEGs

In the L4/5 IT CTX neuron cluster, upregulated DEGs were associated with synapse organization (*SYBU, ANK3, BSN, GABRB3, CADM1, NRXN*) and the regulation of membrane potential (*ANK3, GABRB3, PEX5L, NEDD4L, HCN1*)(Supp. Figure 3.18 A; Supp. Table 3.51). Downregulated DEGs were correlated with synapse organisation (***ARC**, NPAS4, ETV5, HOMER1, APOE, BDN*) and the glutamate receptor signaling pathway (*HOMER1, PLP1, GRM3, CX3CL1, GRIA4, ATP1A3*)(Supp. Figure 3.18 B; Supp. Table 3.52). No significant GO terms were noted in upregulated and downregulated DEGs solely identified in L4/5 IT CTX neurons.

3.5.3.4 Inhibitory neuron DEGs

Amongst the inhibitory neurons, upregulated DEGs were associated with G protein-coupled purinergic nucleotide receptor signalling pathway (*GPR34, P2RY14, P2RY12, GPR171*)(Supp. Figure 3.19 A; Supp. Table 3.53). Downregulated DEGs were correlated with neuron differentiation (***ARC**, ETV5,PLP1, MBP, NDRG4, VEGFA, NR3C1*) and the myelin sheath (*PLP1, MBP, CKB, TUBB4A, SLC25A3, ATP5B*)(Supp. Figure 3.19 B; Supp. Table 3.54). In addition, PD was also related to the downregulated DEGs (**DDIT3**, MT-ATP6, MT-CO2, MT-CO3, TUBB4A)(Supp. Table 3.54).

3.5.3.5 DEGs that are shared across cell types

Upregulated DEGs that were exclusively identified in excitatory neurons, L2/3 IT CTX neurons and L4/5 IT CTX neurons were associated with the regulation of membrane potential (*GRID1, SHTN1, NOS1AP, CACNA1D*)(Supp. Figure 3.20 A; Supp. Table 3.55). Downregulated DEGs in the same cell types were related to synaptic signalling (*EGR2, NPAS4, EGR3, PTGS2, CCK, RGS4*) and negative regulation of the apoptotic process (*DUSP1, BTG2, EGR3, PTGS2, SPRY2, APOE*)(Supp. Figure 3.20 B; Supp. Table 3.56). This is notable because it suggests that apoptotic pathways are more active in excitatory neurons along with the two neuronal subpopulations as the TB infection progresses with time. Downregulated DEGs that were present in all neuronal cells were associated with the structural constituents of the myelin sheath (*PLP1, MBP*), implicating that the myelin sheath integrity becomes disrupted over the course of CNS-TB infection (Supp. Figure 3.20 C; Supp. Table 3.57).

3.5.3.6 Correlating GO terms with the top 10 DEGs

In comparing the GO terms featured in this condition with the top ten downregulated DEGs for each cell type, correlated pathways included neuron projection development (excitatory neurons) and cell surface receptor signalling (L2/3 IT CTX neurons). No significant correlations were noted amongst the upregulated DEGs.

In summary, the gene ontology analysis enabled the identification of several pathways associated with the three conditions of interest. *GTPase* activity, neurodegenerative diseases and EBV were prominent pathways positively associated with the general TB response, whereas myelin sheath integrity was negatively associated. In the immunocompromised TB response, pathways related to neurogenesis were identified along with the negative regulation of the innate immune response in neurons. Interestingly, glutamate activity was both positively and negatively associated with the same condition in neurons. Analysis of the response of TB-infected mice over time revealed similar pathways to the general TB response such as neurodegeneration, excessive glutamate activity and decreased myelin sheath integrity. Pathways in favour of apoptosis were also noted.

Chapter 4: Discussion

The overall aim of the study was to evaluate the cell type-specific gene expression response of resident CNS cells in a general and immunocompromised setting. At the time of writing there have been no studies that have explored this question on a single cell level. Thus, the study aimed to contribute to the field by analyzing DEGs and pathways involved in the disease process along with the impact of factors such as immunosuppression and disease progression.

4.1 Generating replicates to facilitate differential gene expression analysis

In the experimental design, three mouse brain samples were harvested for each of the 8 analysed conditions. The tissue from the three mice for each condition were then pooled together for nuclei extraction. This approach was performed to reduce cost, minimise technical variability during library preparation and detect doublets. Since the DESeq2 pipeline for differential expression analysis required replicates to detect DEGs, ScSplit was used to deconvolute the biological replicates. ScSplit uses a hidden state model to map clusters to the samples of origin, thereby demultiplexing nuclei from each condition and assigning them to one of the three mice (96).

Demultiplexing tools can be divided into two main categories, namely barcode-based and single nucleotide polymorphism (SNP)-based approaches. Barcode-based approaches refer to the use of unique DNA barcodes or 'hashes' that are added to individual cells/ nuclei prior to pooling. Each barcode is usually a short oligonucleotide sequence and are attached to cells using antibodies or through specific chemical tags (113). An example of a cell hashing tool is CITE-seq, which uses oligonucleotide-tagged antibodies to separate cells with different origins before pooling (114). In spite of the above, expenses are a major limitation in cell hashing approaches mainly because of the extra cost of antibodies. In addition, cell hashing may not be compatible with all species, requires sufficient cell numbers and has varying levels of efficacy (115).

The second demultiplexing method involves distinguishing pooled samples based on natural genetic differences, such as SNP's. These SNP-based approaches can be further divided based on whether the technique requires information about the reference genome of each organism used in the experiment. Demuxlet is an example of an SNP-based tool that requires reference genome information prior to pooling (116). However, since many organisms do not have a reference genome, this limits the widespread use of Demuxlet . Thus, SNP-based approaches that do not require a reference genome have been increasingly used in snRNA-seq experiments. ScSplit is an example of such an approach and has been shown to be effective in delineating pooled samples of human data (96). It should be noted that Demuxlet could have been implemented in this study since reference genome data is available for the mouse.

ScSplit has been described to have predictions highly concordant with Demuxlet and other cell-hashing datasets (96). However, some of the limitations of ScSplit is that it has only been thoroughly

tested in human data. The developers of ScSplit acknowledge this but have argued that ScSplit could be applied to other organisms, especially those without available dense genotype chips (96). In spite of this, one 2023 study compared ScSplit with other SNP-based techniques such as souporcell, Vireo and Freemuxlet and suggested that in non-human samples, ScSplit performs less efficaciously than the other methods – giving divergent results for one individual (117). Curiously, the same study also concluded that the other demultiplexing methods had poor doublet detection (117).

Doublet detection is an area of interest in the literature with many tools being developed to increase accuracy. A 2024 study compared the intersectional combinations of different doublet detection methods and concluded that intersectional approaches are more robust in accurately calling doublets (118). One of the limitations of this study is that only one doublet detection method was used, which may have impacted the quality of the data. Even though Figure 3.6 showed similar cell proportions between replicates, it may have been useful to include other doublet detection tools such as Demuxify, which synthesises multiple double detection tools in one analysis (118).

4.2 Cell type annotation

Annotation of the snRNA-seq datasets were performed using label transfer (104), which is a method to assign labels to cells based on a reference dataset of existing annotations. In this case, the Allen Mouse Brain Atlas was used as a reference (105). Label transfer assigns scores to each nucleus based on its similarity to the different cell types within the reference atlas. Through the use of existing data, the method is considered to be a robust way of accurately annotating cells in new datasets (106, 119). Label transfer has also been shown to reduce batch effects, resulting in more accurate interpretations (106). To assist with verifying the annotations in our data, the expression of cell type-specific markers were compared to the expression of each annotated cell type. A high degree of concordance was observed between the annotated cells and the known markers, increasing the confidence in the quality of the annotations.

However, using marker genes to verify cell annotation can pose challenges. Firstly, marker genes may not be uniformly expressed (120). One study showed how OPALIN along with Hyaluronan and Proteoglycan Link Protein 2 (HAPLN2), both considered to be markers of oligodendrocytes in the human brain, have low concentrations amongst subsets of oligodendrocytes at the single cell level (121). Secondly, marker genes are not necessarily consistent in their expression across time. This was supported by a 1995 study which showed how PVALB-positive cortical interneurons are only expressed much later after birth (122). To address these considerations, multiple cell type-specific markers were used to increase the overall confidence of the annotations.

Label transfer methods have also been shown to be vulnerable to preprocessing errors such as batch effects, normalization issues, and misalignments that may limit the accuracy of annotations (123). In addition, the Allen Mouse Brain Atlas is based on healthy tissue. This means that diseased or transitional cell states may not be reliably annotated, especially since label transfer forces nuclei to map onto a reference cell type, even if the predictive score is low (104).

A newer machine-learning annotation approach known as Treearches offers a unique solution to the risk of incorrect annotation by using hierarchical clustering (124). If a cell type cannot be reliably annotated, Treearches may assign it to a broader category within the tree structure without forcing the annotation to an existing cell type. In this way, novel or transitional cells that may have been implicated in disease states can be more easily identified (124). Other methods to improve label transfer methods have recently been developed. For example, a recent 2024 study introduced SIMS (scalable, interpretable machine learning for single cell), which has been shown to handle large and complex datasets with a greater accuracy than standard label transfer methods (125). Future work that uses the data from this study could consider applying these novel tools to more confidently detect diseased cell types.

4.3 Selecting cell types for downstream analysis

As detailed in Chapter 3, 27 cell types were identified following the label transfer annotation. The number of nuclei in each cell type was relatively consistent across the samples, with the exception of the oligodendrocytes, which showed a marked decline in nuclei in TB-infected Day 17 mice (both genotypes) compared to the Day 17 saline condition. This could be a technical finding related to the site of brain tissue harvesting used for nuclei extraction, but may also signify cell death of oligodendrocytes in response to TB infection over time. Future studies could repeat the experiment to assess if this finding is replicable, which would suggest that the observation is less likely due to a technical factor.

Since the average number of nuclei per cell type in each sample was 87, a decision was made to group samples into broader categories, namely excitatory neurons, inhibitory neurons, oligodendrocytes, microglia-PVM and astrocytes. L2/3 IT CTX and L4/5 IT CTX excitatory neurons, which had the highest number of nuclei on average across samples, were also included. The pooling of nuclei strengthened the statistical power of the analysis, whilst the inclusion of the subtypes enabled the possibility of comparisons with the broad excitatory neuron cluster.

4.4 Selecting comparisons for further analysis

Twelve comparisons were initially tested using DESeq2 to identify differentially expressed genes, guided by the PCA findings. Three of these comparisons were chosen for further downstream analysis. Firstly, I chose to study the general TB response, regardless of genotype, to understand how different cell types react to TB infection. I was particularly interested in the neuronal population where there is a significant gap in the literature. Secondly, I opted to explore the TB response specifically in immunocompromised mice compared to WT controls, in part due to the large burden of HIV co-infection but also to determine how the immunocompromised TB response differed from the general TB response. Thirdly, I chose to analyse how the disease process evolved over time, by comparing the gene expression of Day 17 TB-infected mice to Day 9 TB-infected mice. It should be noted that many other comparisons could be performed, such as the TB response strictly in immunocompetent mice

(TB vs saline in WT) and the late TB response (TB vs saline in Day 17 mice), but due to time constraints the analysis was restricted to the three aforementioned conditions.

4.5 Relationship between sample nuclei and number of differentially expressed genes

There was a proportional relationship between the number of nuclei isolated and the number of DEGs per cell type in each condition. This was not unexpected since the larger number of nuclei would theoretically increase the power of the analysis, thus increasing the probability of detecting statistically significant differences between conditions. The main exception to this observation were microglia-PVM, which showed a disproportionately high number of DEGs in the general TB response, despite less than 50 microglia-PVM nuclei being isolated on average per sample (Table 3.5). Barring any technical sources of variation, these findings could indicate a high degree of activity within this cell type in TB-infected mice. The exaggerated number of DEGs in microglia-PVM is not unexpected, given the classical understanding that microglia are the resident immune cells in the brain. As mentioned earlier, one 2024 study found that microglia demonstrated highly activated complement responses in the cerebrospinal fluid (CSF) of a paediatric group with CNS-TB, underpinning the role of microglia in the CNS-TB immune response (28).

4.6 Differentially expressed genes: Candidate gene analysis

Candidate gene analysis was performed with a focus on cell type-specific DEGs. There are many benefits to a candidate gene approach. For example, researchers can propose and test specific hypotheses about gene functions based on prior knowledge. Testing a limited number of genes also requires fewer resources than more comprehensive analyses. In this analysis, priority was given to cell type-specific DEGs because I wanted to gain insight into unique responses between cells. The analyses focussed on oligodendrocytes and excitatory neurons as these cell types had the highest number of nuclei and thus I was more confident in the results of these groups. These cell types also had a large number of cell type-specific DEGs. Candidate genes were selected on the basis of known functions in the cell type and all candidate genes were within the top 10 upregulated or downregulated cell type-specific DEGs by $\log_2$FC.

4.6.1 Candidate genes in the general TB response

4.6.1.1 Excitatory neurons (APOL6, CCL2, VIPR2)

APOL6, which was upregulated in the excitatory neurons, is part of the human *apolipoprotein L (APOL)* gene family, which codes for *APOL1-6* (126)(Figure 3.10 A). This gene is involved in the innate immune response to viruses, bacteria and protozoan parasites (126), and is likely related to humoral chemokine signalling (127). Initially thought to be solely involved in lipid transport, *APOL6* has gained attention because of its emerging role in diseases (128). One such study showed that an

overexpression of *APOL6,* along with *APOL1* and *APOL3* can cause cytotoxicity in cells (126). Another 2005 study involving human cancer cell lines found that *APOL6* caused mitochondrial-mediated apoptosis when overexpressed (129). *APOL6* has also been recently studied in endocrine diseases such as diabetes, where it was found to mediate β cell damage through regulation of islet inflammation (130). These studies collectively demonstrate how dysregulation of *APOL6* can lead to cell damage.

APOL6 has also been studied in pulmonary TB, and has been identified as part of a collection of blood signatures to distinguish active from latent TB (131). Although there are no documented studies of *APOL6* expression in CNS-TB or in excitatory neurons, it is tempting to speculate that *APOL6* may specifically be involved in neuronal cytotoxicity and apoptosis in CNS-TB. This observation is supported by a 2022 study, using samples from blood and CSF, which found that patients with CNS-TB had higher expression of markers related to apoptosis and suppression of survival (132).

CCL2, which was also found to be upregulated in excitatory neurons, is a leucocyte activating cytokine, along with *CCL3* and *CXCL8* (133)(Figure 3.10 A). *CCL2* is well described in TB and plays a prominent role in the restriction of TB infection in the lung (134). *CCL2* is known to be activated by TNF and works by recruiting monocytes and T-cells to the site of mycobacterial infection (134) with one study describing *CCL2* as the most potent chemotactic factor in monocyte recruitment (135).

CCL2 is constitutively expressed in neurons and glia in different regions of the healthy brain (136). In terms of neuronal function, one study proposed that *CCL2* can directly regulate synaptic transmission (137). In addition, *CCL2* has been implicated in other neurological diseases. For example, the gene is involved with the recruitment of microglia and macrophages in chronic traumatic encephalopathy (138). In another study, mice were intracerebrally inoculated with Theiler's murine encephalomyelitis virus (TMEV) and this was found to induce *CCL2* overexpression in neurons. This response was necessary to recruit immune cells to the site of infection (139). However, *CCL2* is upregulated in patients with AD and an overexpression of *CCL2* results in proliferation of pathogenic tau species alongside deleterious glial neuroinflammatory changes (140). Collectively, these studies suggest a prominent role for *CCL2* in various neurological diseases.

CCL2 has also been studied in the context of CNS-TB. One such study showed that *CCL2,* along with other chemokines were released by infected microglia in patients with CNS-TB (26). In addition, increased *CCL2* has been labelled as a marker for CNS-TB severity (141) and other forms of disseminated TB (142). A neuroprotective role has also been described for *CCL2*. Earlier studies have shown how CCL2 effectively reduces neuronal death following N-methyl-D-aspartate (NMDA) challenge, in addition to decreased glutamate release and neuronal protection following oxygen-glucose deprivation (136, 143).

Since *CCL2* was upregulated in the excitatory neurons of TB-infected mice, one could conclude that in addition to the recruitment of immune cells, *CCL2* may be regulating synaptic transmission in CNS-TB. The link between *CCL2* overexpression and neurodegenerative diseases invites comparisons with CNS-TB. Prior studies have established a relationship between CNS-TB and neurodegenerative diseases, with excitotoxicity suggested as a likely contributor (47, 48, 49). In addition, one South African study showed that poor cognitive performance in CNS-TB is common, noting that global cognitive performance was significantly lower in CNS-TB patients compared to the control group (144),

suggesting a neurodegenerative component in the pathogenesis of CNS-TB. However, given that *CCL2* has also been shown to have a neuroprotective function when overexpressed (143), further studies should attempt to characterise the function of *CCL2* in the brain.

The third candidate gene, *Vasoactive Intestinal Peptide Receptor 2 (VIPR2),* was downregulated in excitatory neurons (Figure 3.10 A). Despite not being previously studied in TB, *VIPR2* is known to encode for VPAC2, a G-protein coupled receptor which is involved in neuronal signalling cascades (145). Like *CCL2*, *VIPR2* has also been implicated in the pathogenesis of neurodegenerative disorders making it a plausible candidate gene. One 2019 study developed an agonist to *VIPR2*, called LBT-3627. The study showed that this *VIPR2*-agonist was neuroprotective for PD-associated dopaminergic neurodegeneration (146), and suggested that *VIPR2*-agonists may be useful therapeutic tools to restore T-cell functioning, decrease neuroinflammation and attenuate neurodegeneration. In addition to neurodegeneration, *VIPR2* has also been studied in other brain pathologies. Evidence from human genetics studies have shown that microduplications at 7q.36.3, which contains the *VIPR2* gene, are linked to schizophrenia along with autism spectrum disorder (147). Taken together, these studies highlight the important role of *VIPR2* in neurological disease. In the context of this study, it is notable that *VIPR2* was downregulated. Therefore, it is plausible that *VIPR2* may be an additional contributor to the neurodegenerative phenotype seen in CNS-TB.

4.6.1.2 Oligodendrocytes (BATF2, OASL1)

Amongst the oligodendrocyte population, two candidate genes were identified. *BATF2 or basic leucine zipper transcription factor 2*, upregulated in the data, is a component of the activator protein 1 family of transcription factors, which play an important role in the differentiation of dendritic cells and T-cells (148)(Figure 3.10 G). Specifically, *BATF2* has been shown to induce the expression of C-C Chemokine Receptor Type 5 (CCR5), a chemokine that promotes myeloid differentiation (149).

BATF2 has been previously studied in TB and has been proposed as a single biomarker for active TB (150). Interestingly, one study argued that *BATF2* is detrimental in type 1-controlled infectious diseases, of which TB is an example. The study showed how *BATF2*-deficient mice ($BATF2^{-/-}$), infected with TB or Listeria Monocytogenes, showed reduced tissue pathology compared to control mice. Therefore, increased levels of *BATF2* are detrimental in pulmonary TB infection (151). Prior studies have also suggested a role for *BATF2* in neurological diseases. Blood RNA sequencing found that overexpression of *BATF2* was associated with autism (152). Moreover, another study suggested that depletion of *BATF2* in humans is associated with severe epilepsy and mental retardation (153). This underscores the importance of appropriate regulation of *BATF2* in the brain.

Despite no reported studies on *BATF2* in the context of CNS-TB or in oligodendrocytes, one may posit a relationship. Considering that an overexpression of *BATF2* leads to worse outcomes in mice with pulmonary TB, it is tempting to propose that the upregulation of *BATF2* seen in oligodendrocytes may be contributing in part to CNS-TB pathogenesis in mice.

OASL1, the second candidate gene in oligodendrocytes was upregulated in mice with TB (Figure 3.10 G). *OASL1*, or 2'–5' oligoadenylate synthetase (OAS)-like 1 is known to be a negative regulator of IFN-1 production, which is necessary for innate immune responses (154). For example, one study showed

how *OASL1* knockout mice infected with Lymphocytic choriomeningitis virus Clone 13 (LCMV CL-13) had sustained levels of IFN-1, allowing for faster control of viremia and improved functional T-cell responses, in comparison to WT mice (155). The authors concluded from this that *OASL1* in the early stages of viral infections is essential to enable viral persistence.

OASL1 has also been studied in pulmonary TB and was found to facilitate TB survival in macrophages (156). Using a viable intracellular bacteria assay, the study showed that the number of viable TB bacilli decreased in *OASL* knocked-down THP-1 cells, in comparison to the control group at 3 days post infection. This suggests that *OASL* expression was promoting TB survival in THP cells. *OASL1* has also been studied in brain diseases. *OASL1* deficiency was shown to limit the development of multiple-sclerosis (MS) associated autoimmunity, and was suggested as a potential target for future MS treatment (157). *OASL1* attenuation was also shown to promote an exaggerated inflammatory response. One study used human induced pluripotent stem-derived microglia with decreased *OASL1* expression, and found increased activity of TNF and IFN-γ. These studies highlight the critical role of *OASL1* in modulating the degree of immune responses (158).

To my knowledge, *OASL1* has not been studied in oligodendrocytes or in the context of CNS-TB. However, since *OASL1* was selectively upregulated in oligodendrocytes in the data, this may suggest that oligodendrocytes are involved in attenuating the immune response in mice with CNS-TB.

4.6.2 Candidate genes in the immunocompromised TB response

4.6.2.1 Excitatory neurons (GBP3, SERPING1, LEF1)

Within the excitatory neuron population, three candidate genes were chosen for the immunocompromised TB response. The first gene, *GBP3*, upregulated in our analysis, is known to disrupt the intracellular environment in which pathogens grow (Figure 3.11 A)(159, 160). Prior studies using the gram negative bacterium *Francisella novicida* have shown the importance of *GBP3* (and *GBP1*) in driving inflammasome activation as a response to infection (161). It is also known that macrophages which do not express *GBPs 1, 6, 7, 10* are not able to contain the effects of Mycobacterium Bovis BCG vaccine, highlighting the general role of *GBPs* in the immune response (162).

GBP's are also well-studied in pulmonary TB, with *GBP1* considered to be a diagnostic marker (163). In addition to pulmonary TB, a role for *GBP3* has also been suggested in neurological diseases. *GBP3* is highly elevated and is reported to promote tumorigenicity of glioblastoma, the most common malignant brain tumour (164, 165), ultimately leading to poorer outcomes. This has generated interest in using *GBP3* as a potential therapeutic tool for glioblastoma (164).

In the context of CNS-TB, a 2024 study using whole blood RNA-seq suggested a role for *GBP3* in delineating CNS-TB from other neurological diseases (73). Although *GBP3* has not been specifically studied in excitatory neurons, its upregulation in this analysis could represent a compensatory attempt from neurons to contain the TB infection in response to the functional loss of TNF. However,

as seen with the aforementioned neurological diseases, an upregulated *GBP3* can also have negative consequences that require deeper exploration.

Serpin1 peptidase inhibitor, clade G, member 1 (*SERPING1*), the second candidate gene which is upregulated in this analysis, is known to encode for the C1 inhibitor protein, a protease that performs an inhibitory role in the complement pathway (Figure 3.11 A)(166). The complement pathway refers to a group of inactive proteins that circulate in the blood. After stimulation by antibody complexes (classical pathway) or by pathogens (alternative pathway), complement proteases activate other complement precursors, which ultimately enhance the functions of antibodies and phagocytes (167).

SERPING1 is well-known as a biomarker that is overexpressed in patients with pulmonary TB (168). Interestingly, *SERPING1* is also part of a six gene blood signature (along with *GBP2, FCGR1B, TUBGCP6, TRMT2A, and SDR39U1*) that have been shown to accurately predict TB in TB-infected patients up to 12 months prior to TB diagnosis (169).

In the brain, *SERPING1* is known to be secreted by neurons and inhibits the complement pathway in a similar mechanism to the lungs (170). *SERPING1* has also been implicated in neurological diseases such as PD. One study found that *SERPING1* was upregulated in mice chronically intoxicated with 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP). MPTP is a neurotoxin that can produce irreversible brain damage in a similar manner to PD (171). Therefore, it has been postulated that *SERPING1* plays a significant role in dopaminergic cell death in the substantia nigra (SN) of mice chronically infected with MPTP (170).

Even though *SERPING1* has been studied in TB and has a functional role in neurons, no studies to my knowledge have specifically explored *SERPING1* in the context of CNS-TB, making it a suitable candidate gene. In our study, *SERPING1* was upregulated in the $TNF^{-/-}$ TB-infected mice with respect to WT controls, suggesting that in the immunocompromised environment there may be increased inhibition of the complement pathway, which may impact disease pathogenesis. This is supported by a 2022 study that explored the relationship between complement levels and CNS-TB mortality (172). The study showed that CNS-TB patients with a low serum level of complement C3 had a higher rate of mortality compared to those with a higher C3 level (172). Therefore, one could posit that the increased *SERPING1* expression in TB-infected, immunocompromised mice, which inhibits the complement pathway, may contribute to worse disease outcomes.

Lymphoid enhancer-binding factor 1 (*LEF1*), the third candidate gene, which is upregulated in excitatory neurons is a transcription factor that plays an important role in several biological processes, including embryonic development, stem cell maintenance, immune cell differentiation, and cancer progression (Figure 3.11 A)(173). Despite not being previously reported in the pathogenesis of TB or CNS-TB, *LEF1* was chosen as a candidate gene because of its reported role in neurogenesis. In the developing mouse brain, *LEF1* has been shown to drive the generation of hippocampal neurons (174). In addition, earlier studies in 2004 suggested that *LEF1* over expression in adult mice lead to proliferation and differentiation of neuronal precursors, but the authors argued that this was confined to thalamic neurons (175). This is in contrast to our data which showed *LEF1* upregulation to be present in cortical excitatory neurons. In support of *LEF1* activity beyond the thalamus, a 2017 study showed that *LEF1* is required for the neurogenesis of hypothalamic neurons in zebrafish and mice

(176). The upregulation of *LEF1* in this study raises interesting possibilities about the role of neurogenesis in immunocompromised mice with CNS-TB. Further studies would be needed to elucidate the exact mechanisms of this relationship.

4.6.2.2 Oligodendrocytes (NLRC5, CSPG4)

Two candidate genes were identified in oligodendrocytes, both of which were upregulated (Figure 3.11 F). *Nucleotide-binding oligomerization domain-like receptors caspase recruitment domain 5* (*NLRC5*) was the first candidate gene explored. *NLRC5* is a pattern recognition receptor involved in the innate immune response and also known to negatively regulate IFN-1 and NF-κβ pathways (177, 178).

Interestingly, upregulation of *NLRC5* in the brain has been associated with both neuroprotection and neurodegeneration. *NLRC5* overexpression was found to alleviate inflammation by decreasing the levels of TNF, IL-6, IL-1β, reactive oxygen species (ROS), malondialdehyde (MDA), Bax and caspase-3 (179). Higher levels of *NLRC5* was also shown to mitigate apoptosis in PC12 cells via suppression of the NF-κβ pathway (179). PC12 cells are commonly used to model an ischaemic stroke in rats under the conditions of oxygen-glucose deprivation and reoxygenation (180). In contrast to the above, a 2023 study found that *NLRC5* was overexpressed in the nigrostriatal axis of mice with MPTP-induced PD (181). Consequently, *NLRC5* deficiency was shown to improve motor deficits and decrease striatal inflammation, reinforcing the hypothesis that *NLRC5* overexpression is associated with neurodegeneration (181). Other studies have implicated *NLRC5* with diseases such as multiple sclerosis (MS), epilepsy, schizophrenia and bipolar disorder. (182, 183, 184, 185). Although not yet studied in TB or CNS-TB, *NLRC5* has been implicated in oligodendrocyte precursor cells (OPCs), making them interesting candidates. One study showed that *NLRC5* is expressed in all oligodendrocyte lineage cells and can be used as biomarkers of (MS) (182). Another study used immunostaining analysis to suggest that *NLRC5* is involved in neurogenesis of hippocampal neurons (186).

The upregulation of *NLRC5* in this data analysis suggests that the immune response is attenuated in TB-infected immunocompromised mice through inhibition of the NF-κβ pathway. In addition, the potential association with neurogenesis is in keeping with the findings of *LEF1* in excitatory neurons. Thirdly, the *NLRC5* overexpression in our data is another indicator of a neurodegenerative component in the pathogenesis of CNS-TB,.

Chondroitin sulfate proteoglycan 4 (*CSPG4*), the second candidate gene codes for a transmembrane proteoglycan that is known to be highly expressed in OPCs and is implicated in regulating OPC proliferation, migration, and maturation (Figure 3.11 F)(187). Despite not being studied in TB or CNS-TB, *CSPG4* was chosen as a candidate gene because of its role in neuronal repair. OPCs have been shown to repair the growth cones of damaged axons, which is thought to assist with restoration of nerve connections and function (188). OPCs have also been shown to form synapse-like structures with damaged neurons (189). Interestingly, knockout of *CSPG4* does not inhibit myelin function, but decreases the ability to form new mature oligodendrocytes that can then be involved in myelin repair (190). Considering that *CSPG4* was upregulated in immunocompromised mice, this suggests that, in the immunocompromised state, OPCs may be performing a reparative role in supporting damaged neurons.

4.6.3 Candidate genes in TB-infected mice over time

4.6.3.1 Excitatory neurons (IL-7, GADD45α)

In the comparison between TB-infected mice over time, two candidate genes were identified in excitatory neurons, namely *IL-7* and *GADD45A (Figure 3.12 A)*. *IL-7*, upregulated in this analysis, is a hematopoietic cytokine and is required for T-cell development, survival and homeostasis (191). *IL-7* promotes T-cell survival by upregulating the expression of anti-apoptotic genes such as *B-cell lymphoma 2 (BCL-2*) and *myeloid-cell-leukemia sequence-1 (MCL-1)* – both of which are targeted by naïve T-cells. The protein products produced in response to these genes inhibit Bax and Bak, which are both pro-apoptotic proteins (191).

An anti-apoptotic role for *IL-7* has also been demonstrated in mouse brains. Embryonic brains that were treated with *IL-7* showed increased neuronal survival along with greater neurite outgrowth, highlighting the neuro-protective role of *IL-7* (192). The upregulation of *IL-7* in this analysis could suggest that as the TB infection progresses, anti-apoptotic pathways become more active.

In this study, *Growth Arrest and DNA Damage 45α* (*GADD45α*), the second candidate gene was downregulated in excitatory neurons *(Figure 3.12 A)*. The *GADD45* family consists of three members, *GADD45α, GADD45β, and GADD45γ*, all of which are stress-responsive nuclear proteins (193). In the brain, *GADD45α* has been showed to be induced by neuronal activity (194). Despite not being explicitly studied in TB or CNS-TB, *GADD45α* dysregulation has been implicated in a number of neurological diseases, making it a gene of interest. Some of the brain diseases associated with *GADD45α* include PD. One 2014 study created an *in vitro* model of PD by overexpressing 1-methyl-4-phenylpyridinium (MPP), a derivative of MPTP which produces neurotoxicity in dopaminergic neurons. The study noted profound exacerbation of cell damage upon downregulation of *GADD45α* (195). Another study found that *GADD45β,* a close relative of *GADD45α*, was depleted in the prefrontal cortex of individuals with schizophrenia (196). Furthermore, one publication pointed towards an inverse relationship between *GADD45α* and neuroinflammation. The study found that *GADD45α* knockout induces neuroinflammation, with strategies to increase *GADD45α* being suggested as potential therapeutic targets in AD (197). Since *GADD45α* is downregulated in this analysis, this may indicate that neuroinflammation increases over the course of CNS-TB infection.

In summary, the candidate gene analysis highlights that both neurons and oligodendrocytes appear to be involved in the immune response to CNS-TB. In the general TB response, neurons were associated with cytotoxicity and apoptosis. A link between CNS-TB and neurodegenerative diseases was also noted. In oligodendrocytes, candidate genes were associated with attenuated and pathological immune outcomes. Interestingly, in the immunocompromised TB-infected mice, neurogenesis was found to be associated with both excitatory neurons and oligodendrocytes. OPCs are also likely to be involved in repair of damaged neurons. Lastly, in the neuronal response over time, anti-apoptotic pathways were prevalent as well as neuroinflammation. It should also be noted that many of the candidate genes identified have already been described in the context of pulmonary TB, which increases the confidence in these results. Candidate genes that have not been previously identified hold promise for future research in both pulmonary TB and CNS-TB. The candidate gene analysis builds on previous work by suggesting cell type-specific roles for particular DEGs in CNS-TB.

One of the limitations of candidate gene analysis is that they are hypothesis-driven based on known gene functions in the literature, which may have resulted in a biased selection of genes. Secondly, only a limited number of genes were chosen for each condition due to time constraints an therefore, other functionally significant DEGs may have been overlooked. To address some of these limitations, pathway analysis was also performed.

4.7 Differentially expressed genes: Pathway analysis

Pathway analysis allows for a more broader overview of the molecular functions and biological processes in a particular condition by identifying known associations from a list of prescribed genes. In contrast to candidate gene analysis, pathway analysis incorporates all the genes in a given list, making it less biased. Pathway analysis was conducted using the outputs from g:Profiler. When identifying pathways for further discussion, priority was given to pathways that were enriched in excitatory neurons, oligodendrocytes or both.

4.7.1 Pathway analysis in the general TB response

4.7.1.1 GTPase activity (GO:0003924)

In the general TB response condition, *GTPase* activity was positively enriched in all cell types. In addition, most of the top 10 upregulated genes in each cell type in the general TB response were linked to *GTPase* activity, making this an important pathway for further investigation. *GTPase*s (guanosine triphosphatases) are a superfamily of enzymes that hydrolyze GTP (guanosine triphosphate) to GDP (guanosine diphosphate). Cycling between an 'active' GTP and an 'inactive' GDP state, *GTPase*s can control key aspects of cellular processes, including transport, cytoskeletal dynamics and cell growth (198). *GTPase*s can be further divided into different subcategories, such as 1) Ras – which are involved in regulating cell growth, differentiation and survival, 2) Dynamins – which control membrane fission and fusion processes such as endocytosis, 3) Heterotrimeric G Proteins – which transmit signals from G protein coupled receptors to effectors, 4) Translation *GTPase*s – which regulate protein synthesis and 5) Nuclear *GTPase*s – which facilitate nuclear transport and other nuclear processes (199). The vast array of cellular processes mentioned above reflects the diverse roles that *GTPase* activity may play in CNS-TB.

Indeed, *GTPase*s have also attracted interest in the pathogenesis of pulmonary TB. Studies have shown how *GTPase*s are involved in phagosome maturation, which is a key component of the innate immune response to TB in the lung. In particular, inhibition of phagosome maturation has been demonstrated to be due to disruption of the maturation phases controlled by the early endocytic small *GTPase Rab5* and the late endosomal small *GTPase Rab7* (200).

To my knowledge, no published works have directly explored the link between *GTPase* activity and CNS-TB. However, one can speculate potential roles based on existing studies on *GTPase* activity. For example, strong evidence suggests that *GTPase*s are involved in immune cell recruitment. A 2014 study showed how *Rho-GTPase*s (part of the *Ras* family), are key modulators of cellular cytoskeleton

dynamics and were suggested as potential pharmacological targets for limiting leukocyte motility in inflammatory disorders (201). Since CNS-TB is characterised by a mass influx of inflammatory mediators in the brain (22), it is possible that *GTPase*s are involved in this process. Secondly, *GTPase*s have been implicated in the regulation of the blood brain barrier (BBB). This was supported by a 2017 study, which posited that certain *GTPase*s, such as *RhoA, Rac*, and *Cdc42*, play critical roles in both formation and disruption of BBB (202). As described previously, TB likely enters the brain via a disruption of the BBB through different proposed mechanisms (22, 23, 24, 25). Since *GTPase*s may play a role in BBB integrity, it is tempting to speculate that *GTPase*s are involved in either maintaining or disrupting the BBB in CNS-TB.

In the context of neurons, the *GTPase*s *Rac1* and *Cdc42* have been shown to promote neuronal development, differentiation and neuroregeneration (203). In view of the current understanding that CNS-TB is driven by excessive inflammation, leading to neuronal damage, the potential role of *GTPase*s in neuroregeneration serve as interesting areas of future study in CNS-TB.

*GTPase*s have also been implicated in myelination, the principal function of oligodendrocytes. A 2022 study found that *Ral GTPase*s were essential regulators of spinal cord myelination and homeostasis (204). Another study has also implicated Ras *GTPase*s with myelin deficiency in neurodegenerative diseases (205). These studies underscore the diverse role of *GTPase*s in oligodendrocytes. Considering that myelination was a common pathway affected by CNS-TB infection, it is likely that this pathway is mediated at least in part by *GTPase*s in CNS-TB.

4.7.1.2 Myelin sheath (GO:0019911)

The myelin sheath, the protective layer around neurons produced by oligodendrocytes in the brain allows for faster conduction of neural signals. In this analysis, DEGs that contributed to the structural integrity of the myelin sheath were downregulated in the general TB response. The downregulation was particularly noted in oligodendrocytes, L2/3 IT CTX excitatory neurons and inhibitory neurons.

The relationship between CNS-TB and myelin has not been examined directly in the literature. Interestingly, however, the concept of pathogen-associated demyelination has been described, where one study explored the mechanisms by which TMEV lead to demyelination in mice (206). Based on current models of CNS-TB, one can speculate reasons for degradation of the myelin sheath. Firstly, as described earlier, demyelination in CNS-TB may be related to dysregulation of oligodendrocyte function. However, a second mechanism could be related to excitotoxicity, a feature of CNS-TB (48), which can damage neurons and consequently the surrounding myelin sheath. The predilection of demyelination in L2/3 IT CTX neurons and not the broader excitatory neuronal group is notable, suggesting that not all neuronal cells undergo demyelination in CNS-TB. The consequence of demyelination, and subsequent limitations of neuronal function in CNS-TB could also explain why many CNS-TB patients experience cognitive decline, as described previously (144). In addition, demyelination could account for the decrease in the proportion of oligodendrocytes in Day 17 TB-infected mice in comparison to the Day 17 saline group, which was observed in this study.

4.7.1.3 Neurodegenerative diseases (mmu05022)

Neurodegenerative pathways were a prominent feature in this analysis and the associated DEGs were downregulated in most cell types, with the exception of L4/5 IT CTX neurons. It should be stressed that these DEGs, for example *NDUFB9*, are known to be downregulated in neurodegenerative diseases like MS and AD (207). Thus, the finding that those DEGs are downregulated in this analysis suggests a positive enrichment for neurodegenerative disease. The observation that the excitatory neuron subtypes were not enriched for neurodegenerative pathways (whereas the broader excitatory population was enriched) could imply that other neuronal subtypes not studied in this analysis are more strongly associated with neurodegeneration in CNS-TB.

Apart from the processes of demyelination and excitotoxicity described above, there are other reasons why patients with CNS-TB might show neurodegenerative changes. Firstly, a hallmark of CNS-TB is microglial activation (26). While microglia play a protective role by phagocytosing pathological proteins, excessive uptake of such proteins can lead to detrimental effects. This overactivation can trigger neuroinflammation, which, if unresolved, may contribute to neurodegeneration (208). The high activity of microglia in CNS-TB was supported by this study, which showed that despite few captured nuclei per sample, microglia-PVM had the second largest number of DEGs in the general TB response after oligodendrocytes. Finally, there is a strong association between neurodegenerative diseases and disruption of the BBB (209). Since BBB dysregulation is a feature of CNS-TB (22, 23, 24), this could also account for the enrichment of neurodegenerative diseases in this analysis.

4.7.1.4 Epstein-Barr virus (mmu05169)

An unexpected pathway associated with the DEGs identified in the general TB response was EBV. EBV is a member of the herpesvirus family and is one of the most common viruses in humans (210). EBV, unlike TB, primarily infects B-cells and is transmitted through saliva (210). Despite these differences, there are interesting similarities between the two pathologies that warrant further discussion.

Firstly, EBV, like TB has been shown to modulate the BBB. A 2021 study used an *in vitro* model of EBV infection in human brain microvascular endothelial cells (HBECs) and found that EBV lead to endothelial activation and dysregulation of the BBB (211). Furthermore, EBV has been suggested as a potential driver of neurodegeneration in AD through adaptive immune changes and mitogenic signalling (212). Another interesting similarity between EBV and CNS-TB was noted on neuroimaging. A 2022 case study showed EBV meningitis mimicked TBM (a form of CNS-TB) on magnetic resonance imaging (MRI) (213). Therefore, there are many similarities between EBV and CNS-TB on molecular and phenotypic levels. This is also reinforced by EBV being associated with every cell type in the general TB response.

4.7.2 Pathway analysis in the immunocompromised TB response

4.7.2.1 Neurogenesis (*GO:0022008, 0030182)*

Perhaps the most remarkable finding in the pathway analyses was the enrichment of neurogenesis pathways in certain cell types of immunocompromised mice with CNS-TB. Neurogenesis, the process by which new neurons are generated in the brain, was traditionally believed to take place in adults in the hippocampal dentate gyrus (HDG) and subventricular zone (SVZ)(214). Neural stem cells (NSCs) and neural progenitor cells (NPCs) in these regions have been reported to differentiate and produce neurons. Other evidence, however, suggest that newly-generated neurons can be found in the cerebral cortex (215). The presence of new neurons in the cerebral cortex can occur via different mechanisms. Firstly, neurons generated in the SVZ and HDG can physically migrate to the cortex (216). Secondly, neuronal progenitors giving rise to new neurons have been discovered in the cerebral cortex and have been shown to be upregulated in response to brain ischaemia (217). This is termed brain damage-dependent neurogenesis (BDDN).

In the immunocompromised TB response, it was revealed that neurogenesis/neuron differentiation pathways were enriched in L4/5 excitatory neurons and oligodendrocytes[1]. The upregulation of neurogenesis in this study could be attributed to BDDN. Remarkably, however, it appears that this process was largely restricted to L4/5 IT CTX neurons. Beyond this, neuronal development was notably downregulated in the broader excitatory group and L2/3 IT CTX neurons, demonstrating polar responses between the neuronal subtypes. This could suggest that neurogenesis in the immunocompromised TB-infected brain is localized to L4/5 IT CTX neurons. However, this observation should be treated with caution as only two neuronal subgroups were analysed.

The upregulation of neurogenesis pathways in oligodendrocytes raises interesting questions about the roles that these cells may play in neurogenesis. One 2014 study showed that whilst the majority of OPCs differentiate into oligodendrocytes, a much smaller percentage can differentiate into astrocytes and neurons, depending on factors such as the location in the brain and the stage of CNS development (218). Therefore, it is tempting to speculate that in the immunocompromised state, CNS-TB infection in mice could promote OPC differentiation into neurons.

4.7.2.2 Glutamate activity (GO:0004970, GO:0035254, GO:0007215)

The role of excitotoxicity driven by excessive glutamate in CNS-TB has been described earlier in this study (48). However, it is notable that in the immunocompromised state, DEGs in TB-infected mice are associated with both increased and decreased glutamate activity. One possible explanation for these findings could relate to the dynamics between TNF and glutamate. A prior study showed that TNF inhibits the re-uptake of glutamate by astrocytes, likely resulting in an increase of ambient glutamate in the extracellular space, potentially worsening excitotoxicity (219). Since the knockout of TNF was used as a proxy for immunocompromised mice, this could explain why glutamate receptor

[1] Inhibitory neurons were also enriched for neurogenesis but due to the low number of nuclei and DEGs, inhibitory neurons were excluded from this comparison.

activity was downregulated in some neurons amidst the general background of excitotoxicity. The above findings also raise questions about the utility of using TNF knockout to model an immunocompromised brain, which will be discussed later.

4.7.2.3 Negative regulation of the innate immune response (GO:0045824):

The negative regulation of the innate immune response is seen amongst excitatory neurons in immunocompromised mice with TB compared to controls. This finding is not unexpected since TNF is crucial to the innate immune response (38). Therefore, when the function of TNF is removed via knockout, a downregulation of the innate immune response is expected. However, it should be noted that this downregulation does not imply that the entire immune response is attenuated, as excitotoxicity and neurodegeneration are still present.

4.7.3 Pathway analysis in TB-infected mice over time

In the response of TB-infected mice over time, recurring pathways were featured, such as neurodegeneration and myelin sheath integrity. The presence of these pathways indicate that as TB infection progresses, this can result in worsened neurodegeneration and functionality of the myelin sheath. These findings are not unexpected since both neurodegeneration and demyelination have been linked with excitotoxicity and neuroinflammation (48, 209).

4.7.3.1 Negative regulation of apoptotic process (GO:0043066)

In addition to the above pathways, DEGs related to the negative regulation of apoptosis were downregulated in this comparison, indicating that apoptosis may be upregulated. Apoptosis is a type of programmed cell death process that plays an essential role in maintaining tissue homeostasis and development in multicellular organisms. It is characterized by a series of biochemical events leading to specific cellular changes (220). Apoptosis differs from necrosis, which is a form of unordered, traumatic cell death in response to severe and irresistible trauma, resulting in inflammation. Apoptosis under normal circumstances allows for the elimination of damaged or unnecessary cells without harming surrounding tissues (220). It is essential for processes such as embryonic development, regulation of the immune system and the removal of cancerous cells (220).

Two major types of apoptosis have been previously been described. Intrinsic apoptosis refers to when internal signals, such as stress or DNA damage result in mitochondrial release of pro-apoptotic proteins, such as cytochrome c. In contrast, extrinsic apoptosis is driven by external signals resulting in programmed cell death (220). Interestingly, TNF is one of the major drivers of the extrinsic apoptotic pathway and therefore its functional inhibition in the immunocompromised mice could account for the lack of apoptotic pathways seen in that condition (221).

The potential upregulation of apoptosis over time in this study could be a reflection of increased inflammation and an imbalance between death and survival signals in CNS-TB. These findings were supported by a 2022 study, which showed that patients with a longer duration of CNS-TB, had higher

expression of caspase-3, a pro-apoptotic marker. The study also found that patients who died had higher concentrations of caspase-3, suggesting that an upregulation of apoptosis in CNS-TB could be detrimental in disease outcome (132).

It should be noted that pathway analysis does have limitations. For example, comparisons and inferences are generated from existing associations in the literature, which may be biased. Therefore, pathway findings should be correlated with direct gene analysis to improve the quality of interpretations.

4.8 Synthesizing candidate gene and pathway analyses

There was a general concordance between the findings of the candidate gene and pathway analyses. For example, both analyses identified neuroinflammation, excitotoxicity and neurodegeneration as key processes in CNS-TB. These findings are in keeping with another scRNA-seq study on CNS-TB, which also noted pro-inflammatory and neurodegenerative pathways in CNS-TB (93). In addition, the potential for neurogenesis in immunocompromised TB-infected mice was also identified in both analyses, which is a novel finding that at the time of writing has not been mentioned in prior studies. The combined analyses also suggested a cell-type specific role for oligodendrocytes in the immune response to CNS-TB, which have not been previously reported.

One interesting area of discordance between the candidate and pathway analyses arose in the third comparison (TB-infected mice over time), where candidate gene analysis implicated a role for neuron survival and pathway analyses was suggestive of neuronal apoptosis. To reconcile these findings, one could argue that pro-apoptotic and anti-apoptotic pathways are both active in the disease course, but pro-apoptotic pathways were more prevalent which is why they appear in the pathway analysis. A more extensive candidate gene analysis may have also identified genes in favour of apoptosis.

4.9 Limitations and future work

While this research has added insight into the transcriptomic changes in CNS-TB on a single nucleus level, there was many limitations to this analysis.

4.9.1 Limitations related to experimental design

In relation to the experimental design, a relatively small sample size was used in the analysis. This meant that individual cell types had to be pooled together to increase statistical power. Secondly, the number of nuclei detected in each sample may have been influenced by variations in the amount and site of tissue harvest. This could explain why very few astrocytes and microglia-PVM were detected in this analysis.

Another consideration in the experimental design relates to using mice as a proxy of the human brain in CNS-TB. Even though animal models of neurological disease have been instrumental in

understanding brain diseases, there are important neurological differences between mice and humans. These include the increased degree of cerebral complexity, the protracted time development and unique genetics of human brains (222). A consequence of these differences is that not all aspects of human brain diseases can be represented in animal models, which limit the generalisability of the findings in this study. Future studies should build on this work by eventually using snRNA-seq with human CNS-TB samples.

CNS-TB in mice was modelled by performing direct brain inoculation with TB. However, TB primarily enters the brain through haematogenous dissemination from the lung, which may have impacted the accuracy of the analysis. To more accurately model TB pathogenesis, future studies could consider using optogenetics and intravital microscopy (223). Optogenetics refers to the use of light to artificially manipulate the electrochemical properties of cells (224), whereas intravital microscopy is a novel *in vivo* technique used to visualise cells in real time (225). Together, these techniques could be used to manipulate cell membrane physiology and visualize *Mtb* entry into the brain, along with its interaction with the immune system (223).

$TNF^{-/-}$ was used to model the immunocompromised brain based on prior studies which showed that these mice were more susceptible to TB infection (37, 38). However, the role of TNF in CNS-TB is multifaceted and has been shown to play both a neuropathogenic and neuroprotective role in the disease (37, 226). Therefore the extent to which $TNF^{-/-}$ accurately models HIV infection in the brain is unclear, especially in view of the fact that HIV leads to overexpression of TNF (227). Future studies should consider comparing the findings of this analysis with other murine models of HIV, such as the Tg26 mouse model, which has been shown to be more representative of the long-term effects of HIV in addition to modelling the influence of anti-retrovirals (ARVs) (228).

4.9.2 Limitations in sample preparation and data processing

In terms of sample preparation, nuclei were used and not cells for logistical reasons as tissue had to be frozen. Even though nuclei provide a more restricted view of the total transcript milieu, scRNA-seq and snRNA-seq have been shown to demonstrate similar results in the brain (229).

The 10x genomics method was used for library generation but only measures the 3' end of transcripts, resulting in a loss of isoform information. Therefore, single cell tools such as Smart-seq3 could be used to capture the full transcript length, allowing for a broader understanding of transcript differences (230). The droplet microfluidics system used by 10X genomics has a disadvantage at high cell numbers due to limits on drop generation speed. Therefore, the sequencing of large numbers of cells requires much longer run times. Future studies should consider using Particle-templated instant partition sequencing (Pip-Seq), which allows for single-cell capturing and cDNA barcording using only a vortexer, which would save time, expertise and cost (231).

ScSplit was used for demultiplexing pooled samples and identification of doublets (96). As described earlier, future studies should also consider comparing different bioinformatics tools as there is no standard system that is universally agreed upon.

4.9.3 Limitations in data analysis

Prior to differential expression analysis with DESeq2, nuclei from cell subtypes were pooled together to increase overall confidence in the data. In this analysis, I have demonstrated the transcriptional variation between two neuronal subtypes. However, future studies should interrogate other neuronal subtypes in greater detail to gain a more comprehensive understanding of neuronal transcriptional activity. In additional, the pooling of nuclei did not allow for the identification of infiltrating peripheral immune cells that may have been present in the data. Future studies should consider using methods to identify these cell types in order for the transcriptional activity in these cells to be measured. Furthermore, the analysis did not separate between mature and progenitor cells. For example, the enrichment of neurogenesis amongst oligodendrocytes in immunocompromised TB- infected mice is likely because OPCs were classified as oligodendrocytes in the label transfer annotation.

Differential expression analysis was only performed on three comparative conditions due to limitations on time. However, many other analyses could have been performed, such as the TB response strictly in immunocompetent mice (TB vs saline in WT) and the late TB response (TB vs saline in Day 17 mice). The analysis also prioritised oligodendrocytes and excitatory neurons in the above conditions because of the large number of nuclei annotating to these cell types. However, future studies should consider evaluating the transcriptional activity of other cell types in more detail.

The pseudobulk approach used by DESeq2 aggregates gene counts across clusters or samples, but this method overlooks within-sample variability, which may lead to inaccurate estimates of variation when the variability between cells exceeds that between samples (108). This can result in underpowered analyses and false negative observations, especially when the number of cells vary between samples (108). Consequently, future studies could consider increasing the number of biological replicates to improve statistical power (232). Another option is to explore the use of single cell differential expression methods, which treat each cell/nucleus as a separate sample in the analysis. The challenge with these methods is that cells are not independent entities and function in one biological system. Therefore, it would not be appropriate to refer to these cells as separate samples, which may lead to inaccurate interpretations of results. In light of the above, one study has suggested that pseudobulk approaches such as DESeq2 can outperform single cell DE methods (233).

4.9.4 Limitations of using transcription to understand brain function

Lastly, there are other factors that can influence brain function beyond transcription. For example epigenetic factors can activate or silence genes based on different conditions(234). Methods such as Assay for Transposase-Accessible Chromatin with high-throughput sequencing (snATAC-seq) could be used to understand the role of transcription factors and other proteins in influencing chromatic structure and thus gene expression (235). In addition, not all transcribed mRNA are translated into protein, and therefore validation techniques such as immunohistochemistry could be used in future studies to strengthen the conclusions from the analyses (236).

4.10 Conclusion

The aim of the analyses was to use snRNA-seq to determine the transcriptional activity of CNS cells in CNS-TB in general and immunocompromised settings. In particular, the study expands on the literature by suggesting potential mechanisms by which excitatory neurons and oligodendrocytes contribute to the pathogenesis of CNS-TB. A link was suggested between CNS-TB, neurodegenerative diseases and demyelination and supportive evidence of neurogenesis amongst TB-infected mice in the immunocompromised state was provided. Moreover, candidate genes, some of which have not been previously studied in CNS-TB were identified and may possess diagnostic and therapeutic potential that should be explored in future work.

References:

1. World Health Organization. *Global tuberculosis report 2024* [Internet]. 2024 [cited 2024 Oct 31]. Available from: https://www.who.int/publications/i/item/9789240101531
2. Barberis, I. *et al*. The history of tuberculosis: from the first historical records to the isolation of Koch's bacillus. *J Prev Med Hyg* **58,** E9–E12 (2017).
3. World Health Organization. *Global Tuberculosis Report 2019*. Available at: https://www.who.int/tb/publications/en/ Accessed September 14, 2020.
4. Shinyapps.io. Available from: https://worldhealthorg.shinyapps.io/tb_profiles/ (2014).
5. Centers for Disease Control and Prevention (CDC). *TB Risk Factors*. Available at: https://www.cdc.gov/tb/topic/basics/risk.html (2019).
6. Lee, S. H. Tuberculosis infection and latent tuberculosis. *Tuberculosis and Respiratory Diseases* **79,** 201 (2016).
7. O'Garra, A. *et al*. The immune response in tuberculosis. *Annu. Rev. Immunol.* **31**, 475–527 (2013).
8. Leung, A.N. Pulmonary tuberculosis: the essentials. *Radiology* **210**, 307–322 (1999).
9. Luies, L. & du Preez, I. The echo of pulmonary tuberculosis: mechanisms of clinical symptoms and other disease-induced systemic complications. *Clin. Microbiol. Rev.* **33**, (2020).
10. Schluger, N.W. The pathogenesis of tuberculosis: the first one hundred (and twenty-three) years. *Am. J. Respir. Cell Mol. Biol.* **32**, 251–256 (2005).
11. Kumar, P. & Clark, M. *Kumar and Clark's Clinical Medicine*, 9th edn. Elsevier (2016).
12. Fontanilla, J. M*. et al*. Current diagnosis and management of peripheral tuberculous lymphadenitis. *Clin. Infect. Dis.* **53**, 555–562 (2011).
13. Mayosi, B.M., Burgess, L.J. & Doubell, A.F. Tuberculous pericarditis. *Circulation* **112**, 3608–3616 (2005).
14. Abbara, A. & Davidson, R.N. Etiology and management of genitourinary tuberculosis. *Nat. Rev. Urol.* **8**, 678–688 (2011).
15. Rohlwink, U.K*. et al*. Biomarkers of cerebral injury and inflammation in pediatric tuberculous meningitis. *Clin. Infect. Dis.* **65**, 1298–1307 (2017).
16. Jain, S.K*. et al*. Tuberculous meningitis: a roadmap for advancing basic and translational research. *Nat. Immunol.* **19**, 521–525 (2018).
17. Perez-Malagon, C. D*. et al*. Diagnostic and neurological overview of brain tuberculomas: a review of literature. *Cureus* (2021).
18. Wilkinson, R.J*. et al*. Tuberculous meningitis. *Nat. Rev. Neurol.* **13**, 581–598 (2017).
19. Donovan, J*. et al*. The neurocritical care of tuberculous meningitis. *Lancet Neurol.* **18**, 771–783 (2019).
20. Rieder, H.L*. et al*. Tuberculosis diagnosed at death in the United States. *Chest* **100**, 678–681 (1991).
21. Mason, S*. et al*. Tuberculous meningitis in infants and children: Insights from nuclear magnetic resonance metabolomics. *S. Afr. J. Sci.* **112**, 1–8 (2016).
22. Davis, A.G*. et al*. The pathogenesis of tuberculous meningitis. *J. Leukoc. Biol.* **105**, 267–280 (2019).
23. Be, N.A., Bishai, W.R. & Jain, S.K. Role of *Mycobacterium tuberculosis* pknD in the pathogenesis of central nervous system tuberculosis. *BMC Microbiol.* **12**, 7 (2012).

24. Yang, R. *et al*. Blood-brain barrier integrity damage in bacterial meningitis: the underlying link, mechanisms, and therapeutic targets. *Int. J. Mol. Sci.* **24**, 2852 (2023).
25. Sánchez-Garibay, C. *et al*. *Mycobacterium tuberculosis* infection induces BCSFB disruption but no BBB disruption in vivo: implications in the pathophysiology of tuberculous meningitis. *Int. J. Mol. Sci.* **23**, 6436 (2022).
26. Spanos, J.P., Hsu, N.J. & Jacobs, M. Microglia are crucial regulators of neuro-immunity during central nervous system tuberculosis. *Front. Cell. Neurosci.* **9**, 182 (2015).
27. Rock, R.B. *et al*. *Mycobacterium tuberculosis*-induced cytokine and chemokine expression by human microglia and astrocytes: effects of dexamethasone. *J. Infect. Dis.* **192**, 2054–2058 (2005).
28. Mo, S. *et al*. Single-cell transcriptome reveals highly complement-activated microglia cells in association with pediatric tuberculous meningitis. *Front. Immunol.* **15**, 1387808 (2024).
29. Randall, P.J. *et al*. Neurons are hosT-cells for *Mycobacterium tuberculosis*. *Infect. Immun.* **82**, 1880–1890 (2014).
30. Bowen, S. *et al*. The phagocytic capacity of neurons. *Eur. J. Neurosci.* **25**, 2947–2955 (2007).
31. Rohlwink, U.K. *et al*. Tuberculous meningitis in children is characterized by compartmentalized immune responses and neural excitotoxicity. *Nat. Commun.* **10**, 3767 (2019).
32. Swanson, K.V., Deng, M. & Ting, J.P.Y. The NLRP3 inflammasome: molecular activation and regulation to therapeutics. *Nat. Rev. Immunol.* **19**, 477–489 (2019).
33. Haug, S. A century of study: *Mycobacterium tuberculosis* in the brain. *Brain* **130**, 1–16 (2007).
34. Carver, P.L. & Millar, B.C. The neglected problem of tuberculosis of the central nervous system. *Lancet Infect. Dis.* **19**, 97–104 (2019).
35. Bhattacharya, S. & Mandal, R.K. Neurological manifestations of tuberculosis: an overview. *Neurol. Sci.* **39**, 295–304 (2018).
36. Pasparakis, M. *et al*. Immune and inflammatory responses in TNF alpha-deficient mice: a critical requirement for TNF alpha in the formation of primary B-cell follicles, follicular dendritic cell networks and germinal centers, and in the maturation of the humoral immune response. *J. Exp. Med.* **184**, 1397–1411 (1996).
37. Francisco, N. M. *et al*. TNF-dependent regulation and activation of innate immune cells are essential for host protection against cerebral tuberculosis. *J. Neuroinflammation* **12**, 125 (2015).
38. Hsu, N. J. *et al*. Myeloid and T-cells-derived TNF protects against central nervous system tuberculosis. *Front. Immunol.* **8**, 180 (2017).
39. Lin, P. L. *et al*. Tumor necrosis factor and tuberculosis. *J. Investig. Dermatol. Symp. Proc.* **12**, 22–22 (2007).
40. Török, M. E. Tuberculous meningitis: advances in diagnosis and treatment. *Br. Med. Bull.* **113**, 117–131 (2015).
41. Wong, H. Y. *et al*. CD137L, a driver of harmful inflammation in the nervous system. *Neural Regen. Res.* **18**, 2387–2388 (2023).
42. Blomgran, R. & Ernst, J. D. Lung neutrophils facilitate activation of naive antigen-specific CD4+ T-cells during *Mycobacterium tuberculosis* infection. *J. Immunol.* **186**, 7110–7119 (2011).

43. Shi, F. *et al*. Tuberculosis-specific antigen stimulated and unstimulated interferon-γ for tuberculous meningitis diagnosis: a systematic review and meta-analysis. *PLOS One* **17**, e0273834 (2022).
44. Russell, S. L. *et al*. Compromised metabolic reprogramming is an early indicator of CD8+ T-cells dysfunction during chronic *Mycobacterium tuberculosis* infection. *Cell Rep.* **29**, 3564–3579.e5 (2019).
45. Hemingway, C. *et al*. Childhood tuberculosis is associated with decreased abundance of T-cells gene transcripts and impaired T-cells function. *PLOS One* **12**, e0185973 (2017).
46. Choi, D. W. Excitotoxic cell death. *J. Neurobiol.* **23**, 1261–1276 (1992).
47. Pal, M. M. Glutamate: the master neurotransmitter and its implications in chronic stress and mood disorders. *Front. Hum. Neurosci.* **15**, 722323 (2021).
48. Nishizawa, Y. Glutamate release and neuronal damage in ischemia. *Life Sci.* **69**, 369–381 (2001).
49. Giza, C. C. & Hovda, D. A. The new neurometabolic cascade of concussion. *Neurosurgery* **75**, Suppl 4–33 (2014).
50. Szydlowska, K. & Tymianski, M. Calcium, ischemia and excitotoxicity. *Cell Calcium* **47**, 122–129 (2010).
51. Ardiansyah, E. *et al*. Tryptophan metabolism determines outcome in tuberculous meningitis: a targeted metabolomic analysis. *Elife* **12**, e85307 (2023).
52. Adeyemo, S., Sangotola, A. & Korosteleva, O. Modeling transmission dynamics of tuberculosis–HIV co-infection in South Africa. *Epidemiologia* **4**, 408–419 (2023).
53. Leeds, I. L. *et al*. Site of extrapulmonary tuberculosis is associated with HIV infection. *Clin. Infect. Dis.* **55**, 75–81 (2012).
54. Thuong, N. T. T. *et al*. Leukotriene A4 hydrolase genotype and HIV infection influence intracerebral inflammation and survival from tuberculous meningitis. *J. Infect. Dis.* **215**, 1020–1028 (2017).
55. Thwaites, G. E. *et al*. The influence of HIV infection on clinical presentation, response to treatment, and outcome in adults with tuberculous meningitis. *J. Infect. Dis.* **192**, 2134–2141 (2005).
56. Ellaurie, M., Calvelli, T. A. & Rubinstein, A. Human immunodeficiency virus (HIV) circulating immune complexes in infected children. *AIDS Res. Hum. Retroviruses* **6**, 1437–1441 (1990).
57. Kanmogne, G. D. *et al*. HIV-1 gp120 compromises blood-brain barrier integrity and enhances monocyte migration across blood-brain barrier: implication for viral neuropathogenesis. *J. Cereb. Blood Flow Metab.* **27**, 123–134 (2007).
58. Frankel, A. D. & Young, J. A. HIV-1: fifteen proteins and an RNA. *Annu. Rev. Biochem.* **67**, 1–25 (1998).
59. Price, T. O. *et al*. HIV-1 viral proteins gp120 and Tat induce oxidative stress in brain endothelial cells. *Brain Res.* **1045**, 57–63 (2005).
60. Pu, H. *et al*. HIV-1 Tat protein-induced alterations of ZO-1 expression are mediated by redox-regulated ERK1/2 activation. *J. Cereb. Blood Flow Metab.* **25**, 1325–1335 (2005).
61. Sporer, B. *et al*. Human immunodeficiency virus type-1 Nef protein induces blood-brain barrier disruption in the rat: role of matrix metalloproteinase-9. *J. Neuroimmunol.* **102**, 125–130 (2000).

62. Heinzinger, N. K. *et al*. The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing hosT-cells. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7311–7315 (1994).
63. Ferrucci, A., Nonnemacher, M. R. & Wigdahl, B. Extracellular HIV-1 viral protein R affects astrocytic glyceraldehyde 3-phosphate dehydrogenase activity and neuronal survival. *Journal of NeuroVirology* **19,** 239–253 (2013).
64. What is transcriptomics? PHG Foundation. Available at: https://www.phgfoundation.org/publications/explainers/what-is-transcriptomics/ (Accessed: 21 July 2024).
65. Hu, X. *et al*. Gene expression profiles in febrile children with defined viral and bacterial infection. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 12792–12797 (2013).
66. Ramilo, O. *et al.* Gene expression patterns in blood leukocytes discriminate patients with acute infections. *Blood* **109,** 2066–2077 (2006).
67. Zhang, Z. N. *et al*. Transcriptomic analysis of peripheral blood mononuclear cells in rapid progressors in early HIV infection identifies a signature closely correlated with disease progression. *Clin. Chem.* **59**, 1175–1186 (2013).
68. Obermoser, G. *et al.* Systems scale interactive exploration reveals quantitative and qualitative differences in response to influenza and pneumococcal vaccines. *Immunity* **38,** 831–844 (2013).
69. Ho, J. *et al.* A transcriptional blood signature distinguishes early tuberculosis disease from latent tuberculosis infection and uninfected individuals in a Vietnamese cohort. *Journal of Infection* **81,** 72–80 (2020).
70. Bloom, C. I. *et al.* Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. *PLoS ONE* **8,** (2013).
71. Zak, D. E. *et al.* A blood RNA signature for tuberculosis disease risk: A prospective cohort study. *The Lancet* **387,** 2312–2322 (2016).
72. Hai, H. T. *et al.* Whole blood transcriptional profiles and the pathogenesis of tuberculous meningitis. *eLife* **13,** (2024).
73. Huynh, J. *et al.* The ability of a 3-gene host signature in blood to distinguish tuberculous meningitis from other brain infections. *The Journal of Infectious Diseases* **230,** (2024).
74. Olsen, T. K. & Baryawno, N. Introduction to single-cell RNA sequencing. *Current Protocols in Molecular Biology* **122,** (2018).
75. Wen, F. *et al*. Comparison of single-nucleus and single-cell transcriptomes in hepatocellular carcinoma tissue. *Mol. Med. Rep.* **26**, 339 (2022).
76. 10x Genomics. Chromium Single Cell V(D)J Reagent Kits with Feature Barcoding Technology for Cell Surface Protein. Document Number CG000186 Rev A. (2019).
77. Zeisel, A. *et al.* Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347,** 1138–1142 (2015).
78. Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature Neuroscience* **19,** 335–346 (2016).
79. Keren-Shaul, H. *et al.* A unique microglia type associated with restricting development of alzheimer's disease. *Cell* **169,** (2017).
80. www.youtube.com. (n.d.). 01 Single Cell Introduction. Available at: https://www.youtube.com/watch?v=hWWkpe4Ewgo&t=839s&ab_channel=10xGenomics (Accessed: 10 Mar. 2021).

81. Sauder, D. N. Imiquimod: Modes of action. *British Journal of Dermatology* **149,** 5–8 (2003).
82. Abdelfattah, N. *et al.* Single-cell analysis of human glioma and immune cells identifies S100A4 as an immunotherapy target. *Nature Communications* **13,** (2022).
83. Zumla, A. *et al.* Host-directed therapies for infectious diseases: Current status, recent progress, and future prospects. *The Lancet Infectious Diseases* **16,** (2016).
84. Tobin, D. M. *et al*. Host genotype-specific therapies can optimize the inflammatory response to mycobacterial infections. *Cell* **148**, 434–446 (2012).
85. D'Souza, S. *et al*. Physicochemical and biological evaluation of curdlan-poly(lactic-co-glycolic acid) nanoparticles as a host-directed therapy against *Mycobacterium tuberculosis*. *J. Pharm. Sci.* **111**, 469–478 (2022).
86. Pisu, D. *et al.* Single cell analysis of *M. tuberculosis* phenotype and macrophage lineages in the infected lung. *Journal of Experimental Medicine* **218,** (2021).
87. Gierahn, T. M. *et al*. Seq-Well: Portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* **14**, 395–398 (2017).
88. Cronan M. R. *et al*. A non-canonical type 2 immune response coordinates tuberculous granuloma formation and epithelialization. *Cell* **184**, 1757–1774.e14 (2021).
89. Gideon, H. P. *et al*. Multimodal profiling of lung granulomas in macaques reveals cellular correlates of tuberculosis control. *Immunity* **55**, 827–846.e10 (2022).
90. Cai, Y. *et al*. Single-cell transcriptomics of blood reveals a natural killer cell subset depletion in tuberculosis. *EBioMedicine* **53**, 102686 (2020).
91. Wang, Y. *et al*. Systemic immune dysregulation in severe tuberculosis patients revealed by a single-cell transcriptome atlas. *J. Infect.* **86**, 421–438 (2023).
92. Bromley, J. D. *et al*. CD4+ T-cells are homeostatic regulators during Mtb reinfection. *bioRxiv* [Preprint] 2023.
93. Zhang X. *et al*. Single-cell analysis reveals changes in BCG vaccine-injected mice modeling tuberculous meningitis brain infection. *Cell Rep.* **42**, 112177 (2023).
94. Snell G.D. & Borges P.R. Determination of the histocompatibility locus involved in the resistance of mice of strains C57BL/10-x, C57BL/6-x, and C57BL/6Ks to C57BL tumors. *J. Natl Cancer Inst.* **14**, 481–484 (1953).
95. Botha T. & Ryffel B. Reactivation of latent tuberculosis infection in TNF-deficient mice. *J. Immunol.* **171**, 3110–3118 (2003).
96. Xu, J. *et al*. Genotype-free demultiplexing of pooled single-cell RNA-seq. *Genome Biol.* **20**, 271 (2019).
97. Steenken W., Oatway W.H. & Petroff S.A. Biological studies of the tubercle bacillus. *J. Exp. Med.* **60**, 515–540 (1934).
98. Habib, N. *et al*. Massively parallel single-nucleus RNA-seq with dronc-seq. *Nat. Methods* **14**, 955–958 (2017).
99. Single-cell gene expression - official 10X Genomics support [Internet]. [cited 2023 Nov 25]. Available from: https://www.10xgenomics.com/support/single-cell-gene-expression.
100. Fleming, S. J. *et al*. Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nat. Methods* **20**, 1323–1335 (2023).
101. Danecek, P. *et al*. Twelve years of SAMtools and BCFtools. *GigaScience* **10**, 2 (2021).
102. Garrison E. & Marth G. Haplotype-based variant detection from short-read sequencing [Internet]. 2012 [cited 2023]. Available from: https://arxiv.org/pdf/1207.3907.pdf.

103. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337.e4 (2019).
104. Hao, Y. *et al*. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
105. Yao Z. *et al*. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222–3241 (2021). doi: 10.1016/j.cell.2021.03.034.
106. Stuart T. *et al*. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
107. Love M. I., Huber W. & Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
108. Luecken, M.D. & Theis, F.J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* **15**, e8746 (2019).
109. DnA lab short read sequencing workshop. Multifactor Designs in DESeq2 [Internet]. [cited 2024 Oct 16]. Available from: https://youtu.be/X6p3E-QTcUc.
110. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: An R Package for The Visualization Of Intersecting Sets And Their Properties. *Bioinformatics* **33**, 2938–2940 (2017).
111. Kolberg L. *et al*. g:Profiler- interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res.* **51**(W1), W1–W10 (2023).
112. Yao, Z. *et al*. A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature* **598**, 103–110 (2021).
113. Stoeckius, M. *et al*. Cell hashing with barcoded antibodies enables multiplexing and doublet detection for single-cell genomics. *Genome Biol.* **19**, 224 (2018
114. Stoeckius M. *et al*. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
115. Zhang, Y. *et al*. Sample-multiplexing approaches for single-cell sequencing. *Cellular Mol. Life Sci.* **79**, 284 (2022).
116. Kang, H. M. *et al*. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89–94 (2018). doi: 10.1038/nbt.4042. Erratum in: *Nat. Biotechnol.* **38**, 1356 (2020).
117. Cardiello, J. F. *et al*. Evaluation of genetic demultiplexing of single-cell sequencing data from model species. *Life Sci. Alliance* **6**, e2023 (2023).
118. Neavin, D. *et al*. Demuxafy: Improvement in droplet assignment by integrating multiple single-cell demultiplexing and doublet detection methods. *Genome Biol.* **25**, 62 (2024).
119. Achim, K. *et al*. High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* **33**, 503–509 (2015).
120. Grabski, I. N. & Irizarry, R. A. A probabilistic gene expression barcode for annotation of cell types from single-cell RNA-seq data. *Biostatistics* **23**, 1150–1164 (2022).
121. Guo, H. & Li, J. scSorter: Assigning cells to known cell types according to marker genes. *Genome Biol.* **22**, 28 (2021).
122. de Lecea, L., del Río, J. A. & Soriano, E. Developmental expression of parvalbumin mRNA in the cerebral cortex and hippocampus of the rat. *Mol. Brain Res.* **32**, 1–13 (1995).

123. Khatri, R. *et al*. Uncertainty estimation for single-cell label transfer. *Proceedings of Machine Learning Research* **179**, 1–14 (2022).
124. Michielsen, L. *et al*. Single-cell reference mapping to construct and extend cell-type hierarchies. *NAR Genomics Bioinform.* **5**, e3 (2023). doi: 10.1093/nargab/lqad070.
125. Gonzalez-Ferrer, J. *et al*. SIMS: A deep-learning label transfer tool for single-cell RNA sequencing analysis. *Cell Genomics* **4**, (2024).
126. Pant, J. *et al*. Apolipoproteins L1-6 share key cation channel-regulating residues but have different membrane insertion and ion conductance properties. *J. Biol. Chem.* **297**, 100951 (2021).
127. Wang, X. K. *et al*. Identification and validation of candidate clinical signatures of apolipoprotein L isoforms in hepatocellular carcinoma. *Sci Rep* **13**, 20969 (2023).
128. Elliott, D. A., Weickert, C. S. & Garner, B. Apolipoproteins in the brain: implications for neurological and psychiatric disorders. *Clin. Lipidol.* **51**, 555–573 (2010).
129. Liu, Z. *et al*. Apolipoprotein L6, a novel proapoptotic Bcl-2 homology 3-only protein, induces mitochondria-mediated apoptosis in cancer cells. *Mol. Cancer Res.* **3**, 21–31 (2005).
130. Paz-Barba, M. *et al*. Apolipoprotein L genes are novel mediators of inflammation in beta cells. *Diabetologia* **67**, 124–136 (2024).
131. Tabone, O. *et al*. Blood transcriptomics reveal the evolution and resolution of the immune response in tuberculosis. *J. Exp. Med.* **218**, e20210915 (2021).
132. Kalita, J. *et al*. Balancing between apoptosis and survival biomarkers in the patients with tuberculous meningitis. *Cytokine* **157**, 155960 (2022).
133. Hasan, Z. *et al*. CCL2 responses to *Mycobacterium tuberculosis* are associated with disease severity in tuberculosis. *PLoS One* **4**, e8459 (2009).
134. Hasan, Z. *et al*. Elevated ex vivo monocyte chemotactic protein-1 (CCL2) in pulmonary as compared with extra-pulmonary tuberculosis. *BMC Immunol.* **6**, 14 (2005).
135. Hasan, Z. *et al*. Relationship between circulating levels of IFN-γ, IL-10, CXCL9 and CCL2 in pulmonary and extrapulmonary tuberculosis is dependent on disease severity. *Scand. J. Immunol.* **69**, 259–267 (2009).
136. Sowa, J. E. & Tokarski, K. Cellular, synaptic, and network effects of chemokines in the central nervous system and their implications to behavior. *Pharmacol. Rep.* **73**, 1595–1625 (2021).
137. Ji, E. *et al*. The chemokine CCL2 promotes excitatory synaptic transmission in hippocampal neurons via GluA1 subunit trafficking. *Neurosci. Bull.* (2024).
138. Cherry, J. D. *et al*. CCL2 is associated with microglia and macrophage recruitment in chronic traumatic encephalopathy. *J. Neuroinflammation* 17, 370 (2020).
139. Howe, C. L. *et al*. Neuronal CCL2 expression drives inflammatory monocyte infiltration into the brain during acute virus infection. *J. Neuroinflammation* 14, 238 (2017).
140. Joly-Amado, A. *et al*. CCL2 overexpression in the brain promotes glial activation and accelerates tau pathology in a mouse model of tauopathy. *Front. Immunol.* 11, 997 (2020).
141. Hasan, Z. *et al*. Elevated serum CCL2 concomitant with a reduced mycobacterium-induced response leads to disease dissemination in leprosy. *Scand. J. Immunol.* 63, 241–247 (2006).
142. Mastroianni, C. M. *et al*. Chemokine profiles in the cerebrospinal fluid (CSF) during the course of pyogenic and tuberculous meningitis. *Clin. Exp. Immunol.* **114**, 210–214 (1998).

143. Bruno, V. *et al*. Neuroprotective activity of chemokines against N-methyl-D-aspartate or beta-amyloid-induced toxicity in culture. *Eur. J. Pharmacol.* **399**, 117–121 (2000).
144. Davis, A. G. *et al.* Cognitive impairment in tuberculous meningitis. *Clinical Infectious Diseases* **76,** 842–849 (2022).
145. Hughes, A. T., Guilding, C. & Piggins, H. D. Neuropeptide signaling differentially affects phase maintenance and rhythm generation in SCN and extra-SCN circadian oscillators. *PLoS One* **6**, e18926 (2011).
146. Olson, K. E. *et al.* Selective VIP receptor agonists facilitate immune transformation for dopaminergic neuroprotection in MPTP-intoxicated mice. *The Journal of Neuroscience* **35,** 16463–16478 (2015).
147. Ago, Y. *et al*. Probing the VIPR2 microduplication linkage to schizophrenia in animal and cellular models. *Front. Neurosci.* **15**, 717490 (2021).
148. Güler, R. *et al*. Targeting Batf2 for infectious diseases and cancer. *Oncotarget* **6**, 26575–26582 (2015).
149. Le, D. T. *et al*. BATF2 promotes HSC myeloid differentiation by amplifying IFN response mediators during chronic infection. *iScience* **26**, 106059 (2023).
150. Roe, J. K. *et al*. Blood transcriptomic diagnosis of pulmonary and extrapulmonary tuberculosis. *JCI Insight* **1**, e87238 (2016).
151. Güler, R. *et al*. Batf2 differentially regulates tissue immunopathology in Type 1 and Type 2 diseases. *Mucosal Immunol.* **12**, 390–402 (2019).
152. Voinsky, I. *et al*. Blood RNA sequencing indicates upregulated BATF2 and LY6E and downregulated ISG15 and MT2A expression in children with autism spectrum disorder. *Int. J. Mol. Sci.* **23**, 9843 (2022).
153. Zsurka, G. *et al*. Loss of the Immunomodulatory Transcription Factor BATF2 in Humans Is Associated with a Neurological Phenotype. *Cells* **12**, 227 (2023).
154. Lee, M. S. *et al*. OASL1 inhibits translation of the type I interferon-regulating transcription factor IRF7. *Nat. Immunol.* **14**, 346–355 (2013). Erratum in: *Nat. Immunol.* **14**, 355 (2013); *Nat. Immunol.* **14**, 877 (2013).
155. Oldstone, M. B. A. *et al.* Lymphocytic choriomeningitis virus Clone 13 infection causes either persistence or acute death dependent on IFN-1, cytotoxic T lymphocytes (CTLs), and host genetics. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E7814–E7823 (2018).
156. Reheman, A. *et al.* Involvement of 2′-5′ oligoadenylate synthetase-like protein in the survival of *Mycobacterium tuberculosis* avirulent strain in macrophages. *Anim. Dis.* **3**, 5 (2023).
157. Choi, B. Y. *et al.* 2′-5′ oligoadenylate synthetase-like 1 (OASL1) deficiency suppresses central nervous system damage in a murine MOG-induced multiple sclerosis model. *Neurosci. Lett.* **628**, 78–84 (2016).
158. Magusali, N. *et al.* A genetic link between risk for Alzheimer's disease and severe COVID-19 outcomes via the OAS1 gene. *Brain* **144**, 3727–3741 (2021).
159. Tretina, K. *et al*. Interferon-induced guanylate-binding proteins: Guardians of host defense in health and disease. *J. Exp. Med.* **216**, 482–500 (2019).
160. Pilla-Moffett, D. *et al*. Interferon-inducible GTPases in host resistance, inflammation and disease. *J. Mol. Biol.* **428**, 3495–3513 (2016).

161. Feng, S. *et al.* Pathogen-selective killing by guanylate-binding proteins as a molecular mechanism leading to inflammasome signaling. *Nat. Commun.* **13**, 4395 (2022).
162. Kim, B. H. *et al*. A family of IFN-γ-inducible 65-kD GTPases protects against bacterial infection. *Science* **332**, 717–721 (2011).
163. Shi, T. *et al.* Role of GBP1 in innate immunity and potential as a tuberculosis biomarker. *Sci. Rep.* **12**, 11097 (2022).
164. Xu, H. *et al.* GBP3 promotes glioblastoma resistance to temozolomide by enhancing DNA damage repair. *Oncogene* **41**, 3876–3885 (2022
165. Thakkar, J.P. *et al.* Epidemiologic and molecular prognostic review of glioblastoma. *Cancer Epidemiol. Biomarkers Prev.* **23**, 1985–1996 (2014).
166. Paréj, K. *et al*. The control of the complement lectin pathway activation revisited: both C1-inhibitor and antithrombin are likely physiological inhibitors, while α2-macroglobulin is not. *Mol. Immunol.* **54**, 415–422 (2013).
167. Janeway, C. A. *et al*. *Immunobiology: The Immune System in Health and Disease* (Garland Science, 2001).
168. Lubbers, R. *et al*. Expression and production of the SERPING1-encoded endogenous complement regulator C1-inhibitor in multiple cohorts of tuberculosis patients. *Mol. Immunol.* **120**, 187–195 (2020).
169. Penn-Nicholson, A. *et al*. RISK6, a 6-gene transcriptomic signature of TB disease risk, diagnosis and treatment response. *Sci. Rep.* **10**, 8629 (2020).
170. Gorelik, A. *et al*. Serping1/C1 Inhibitor affects cortical development in a cell autonomous and non-cell autonomous manner. *Front. Cell. Neurosci.* **11**, 169 (2017).
171. Seo, M. H. & Yeo, S. Association of increase in Serping1 level with dopaminergic cell reduction in an MPTP-induced Parkinson's disease mouse model. *Brain Res. Bull.* **162**, 67–72 (2020).
172. Im, H. *et al*. Low serum complement level is associated with higher mortality in tuberculous meningitis: a retrospective cohort study. *Encephalitis* **3**, 7–14 (2023).
173. Blazquez, R. *et al*. LEF1 supports metastatic brain colonization by regulating glutathione metabolism and increasing ROS resistance in breast cancer. *Int. J. Cancer* **146**, 3170–3183 (2020).
174. Galceran, J. *et al*. Hippocampus development and generation of dentate gyrus granule cells is regulated by LEF1. *Development* **127**, 469–482 (2000).
175. Wisniewska, M. B. *et al.* LEF1/beta-catenin complex regulates transcription of the Cav3.1 calcium channel gene (Cacna1g) in thalamic neurons of the adult brain. *J. Neurosci.* **30**, 4957–4969 (2010).
176. Xie, Y. *et al.* Lef1-dependent hypothalamic neurogenesis inhibits anxiety. *PLoS Biol.* **15**, e2002257 (2017).
177. Cui, J. *et al.* NLRC5 negatively regulates the NF-kappaB and type I interferon signaling pathways. *Cell* **141**, 483–496 (2010).
178. Neerincx, A. *et al.* A role for the human nucleotide-binding domain, leucine-rich repeat-containing family member NLRC5 in antiviral responses. *J. Biol. Chem.* **285**, 2622326232 (2010).
179. Zhang, Z. *et al*. NLRC5 alleviated OGD/R-induced PC12-cell injury by inhibiting activation of the TLR4/MyD88/NF-κB pathway. *J. Int. Med. Res.* **48**,300060520940455 (2020).

180. Li, N. *et al*. Chrysin, which targets PLAU, protects PC12 cells from OGD/R-stimulated damage through repressing the NF-κB signaling pathway. *Regen. Ther.* **19**, 69-76 (2022).
181. Liu, Z. *et al*. NOD-like receptor NLRC5 promotes neuroinflammation and inhibits neuronal survival in Parkinson's disease models. *J. Neuroinflammation* **20**, 96 (2023).
182. Falcão, A. M. *et al*. Disease-specific oligodendrocyte lineage cells arise in multiple sclerosis. *Nat. Med.* **24**, 1837–1844 (2018).
183. Bijalwan, M. *Pathogenesis of Theiler's Murine Encephalomyelitis Virus (TMEV) in an Experimental Model of Epilepsy*. (Texas A&M University, 2017).
184. Bergen, S. E. *et al.* Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. *Mol. Psychiatry* **17**, 880–886 (2012).
185. Pacifico, R. & Davis, R. L. Transcriptome sequencing implicates dorsal striatum-specific gene network, immune response and energy metabolism pathways in bipolar disorder. *Mol. Psychiatry* **22**, 441–449 (2017).
186. Li, P. *et al*. Neuronal NLRC5 regulates MHC class I expression in Neuro-2a cells and also during hippocampal development. *J. Neurochem.* **152**, 182–194 (2020).
187. Bromley-Coolidge, S., Iruegas, D. & Appel, B. Cspg4 sculpts oligodendrocyte precursor cell morphology. *bioRxiv* [Preprint] 2024.08.08.607226 (2024).
188. Li, N. & Leung, G. K. Oligodendrocyte precursor cells in spinal cord injury: A review and update. *Biomed. Res. Int.* **2015**, 235195 (2015). https://doi.org/10.1155/2015/235195.
189. Son, Y. J. Synapsing with NG2 cells (polydendrocytes), unappreciated barrier to axon regeneration? *Neural Regen. Res.* **10**, 346–348 (2015)
190. Kucharova, K. & Stallcup, W. B. Dissecting the multifactorial nature of demyelinating disease. *Neural Regen. Res.* **13**, 628-632 (2018).
191. ElKassar, N. & Gress, R. E. An overview of IL-7 biology and its use in immunotherapy. *J. Immunotoxicol.* **7**, 1–7 (2010).
192. Michaelson, M. D. *et al*. Interleukin-7 is trophic for embryonic neurons and is expressed in developing brain. *Dev. Biol.* **179**, 251–263 (1996).
193. Smaldone, G. *et al*. Members of the GADD45 protein family show distinct propensities to form toxic amyloid-like aggregates in physiological conditions. *Int. J. Mol. Sci.* **22**, 10700 (2021).
194. Brito, D.V.C. *et al*. Regulation of neuronal plasticity by the DNA repair associated Gadd45 proteins. *Curr. Res. Neurobiol.* **3**, 100031 (2022).
195. Wang, X.F. *et al*. Induction of GADD45α protects M17 neuroblastoma cells against MPP*. *IUBMB Life* **66**, 786–792 (2014).
196. Gavin, D.P. *et al*. Growth arrest and DNA-damage-inducible, beta (GADD45b)-mediated DNA demethylation in major psychosis. *Neuropsychopharmacology* **37**, 531–542 (2012).
197. Althobaiti, N.A. *et al*. Gadd45A-mediated autophagy regulation and its impact on Alzheimer's disease pathogenesis: Deciphering the molecular nexus. *Biochim. Biophys. Acta Mol. Basis Dis.* **1870**, 167353 (2024).
198. Etienne-Manneville, S., Hall, A. Rho GTPases in cell biology. *Nature* **420**, 629–635 (2002).
199. Satoh, T. Diverse physiological functions and regulatory mechanisms for signal-transducing small GTPases. *Int. J. Mol. Sci.* **21**, 7291 (2020).

200. Via, L. E. *et al*. Arrest of mycobacterial phagosome maturation is caused by a block in vesicle fusion between stages controlled by rab5 and rab7. *J. Biol. Chem.* **272**, 13326–13331 (1997).
201. Biro, M., Munoz, M. A. & Weninger, W. Targeting Rho-GTPases in immune cell migration and inflammation. *Br. J. Pharmacol.* **171**, 5491–5506 (2014).
202. Ramos, C. J. & Antonetti, D. A. The role of small GTPases and EPAC-Rap signaling in the regulation of the blood-brain and blood-retinal barriers. *Tissue Barriers* **5**, e1339768 (2017).
203. Kalpachidou, T. *et al*. Rho GTPases in the physiology and pathophysiology of peripheral sensory neurons. *Cells* **8**, 591 (2019).
204. DeGeer, J. *et al.* Ral GTPases are critical regulators of spinal cord myelination and homeostasis. *Cell Rep.* **40**, 111413 (2022).
205. Alcover-Sanchez, B. *et al*. R-Ras GTPases signaling role in myelin neurodegenerative diseases. *Int. J. Mol. Sci.* **21**, 5911 (2020).
206. Das Sarma, J. A mechanism of virus-induced demyelination. *Interdiscip. Perspect. Infect. Dis.* **2010**, 109239 (2010).
207. Lunnon, K. *et al.* Mitochondrial genes are altered in blood early in Alzheimer's disease. *Neurobiol. Aging* 2017, **53**, 36-47 (2017).
208. Gao, C. *et al.* Microglia in neurodegenerative diseases: mechanism and potential therapeutic targets. *Sig Transduct Target Ther* **8**, 359 (2023).
209. Knox, E.G. *et al.* The blood-brain barrier in aging and neurodegeneration. *Mol Psychiatry* **27**, 2659–2673 (2022).
210. Yu, H. & Robertson, E. S. Epstein-Barr Virus History and Pathogenesis. *Viruses* **15**, 714 (2023).
211. Indari, O., Chandramohanadas, R. & Jha, H. C. Epstein-Barr virus infection modulates blood-brain barrier cells and its co-infection with *Plasmodium falciparum* induces RBC adhesion. *Pathog. Dis.* **79**, ftaa080 (2021).
212. Schreiner, T. G. *et al*. New insights on the link between Epstein-Barr virus infection and cognitive decline in neurodegenerative diseases (Review). *Exp. Ther. Med.* **28**, 413 (2024).
213. Corrêa, D. G., Cañete, L. A. Q. & Cruz, L. C. H. D. Jr. Epstein-Barr virus meningitis mimicking tuberculous meningitis on neuroimaging. *Rev. Soc. Bras. Med. Trop.* **55**, e0254 (2022).
214. Gross, C. Neurogenesis in the adult brain: death of a dogma. *Nat Rev Neurosci* **1**, 67–73 (2000).
215. Bernier, P. J. *et al*. Newly generated neurons in the amygdala and adjoining cortex of adult primates. *Proc. Natl. Acad. Sci. U.S.A.***99**, 11464–11469 (2002).
216. Gould, E. *et al*. Neurogenesis in the neocortex of adult primates. *Science* **286**, 548–552 (1999).
217. Ohira, K. Injury-induced neurogenesis in the mammalian forebrain. *Cell Mol. Life Sci.***68**, 1645–1656 (2011).
218. Nishiyama, A., Suzuki, R. & Zhu, X. NG2 cells (polydendrocytes) in brain physiology and repair. *Front. Neurosci.* **8**, 133 (2014).

219. Fine, S. M. *et al.* Tumor necrosis factor alpha inhibits glutamate uptake by primary human astrocytes. Implications for pathogenesis of HIV-1 dementia. *J. Biol. Chem.* **271**, 15303–15306 (1996).
220. Elmore, S. Apoptosis: a review of programmed cell death. *Toxicol. Pathol.* **35**, 495–516 (2007).
221. Rath, P. C. & Aggarwal, B. B. TNF-induced signalling in apoptosis. *J. Clin. Immunol.* **19**,350–364 (1999).
222. Zhao, X. & Bhattacharyya, A. Human models are needed for studying human neurodevelopmental disorders. *Am. J. Hum. Genet.* **103**, 829–857 (2018).
223. Hokeness, S*. et al.* Tuberculous Meningitis: an Update on the Pathogenesis and Neuroimmunology. *Curr Trop Med Rep* **11**, 1–10 (2024).
224. Emiliani, V*. et al.* Optogenetics for light control of biological systems. *Nat Rev Methods Primers* **2**, 55 (2022).
225. Vaghela, R*. et al*. Actually seeing what is going on – intravital microscopy in tissue engineering. *Front. Bioeng. Biotechnol.* **9**, 627462 (2021).
226. Tsenova, L*. et al*. A combination of thalidomide plus antibiotics protects rabbits from mycobacterial meningitis-associated death. *J. Infect. Dis.* **177**, 1563–1572 (1998).
227. Pasquereau, S., Kumar, A. & Herbein, G. Targeting TNF and TNF receptor pathway in HIV-1 infection: from immune activation to viral reservoirs. *Viruses* **9**, 64 (2017).
228. Bryant, J. *et al*. 7,8-Dihydroxyflavone improves neuropathological changes in the brain of Tg26 mice, a model for HIV-associated neurocognitive disorder. *Sci. Rep.* **11**, 18519 (2021). Erratum in: *Sci. Rep.* **11**, 20212 (2021).
229. Bakken, T.E*. et al*. Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLoS One* **13**, e0209648 (2018).
230. Hagemann-Jensen, M. *et al*. Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat. Biotechnol.* **38**, 708-714 (2020).
231. Clark, I.C*. et al.* Microfluidics-free single-cell genomics with templated emulsification. *Nat Biotechnol* **41**, 1557–1566 (2023).
232. Zimmerman, K. D., Espeland, M. A. & Langefeld, C. D. A practical solution to pseudoreplication bias in single-cell studies. *Nat. Commun.* **12**, 738 (2021).
233. Van den Berge, K. *et al.,* Observation weights unlock bulk RNA-seq tools for zero inflation and single-cell applications. *Genome Biol.* **19**, 24 (2018).
234. Kiselev, I. S*. et al.* DNA methylation as an epigenetic mechanism in the development of multiple sclerosis. *Acta Naturae* **13**, 45–57 (2021).
235. Buenrostro, J. D*. et al.* ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–21.29.9 (2015).
236. Kim, S. W., Roh, J. & Park, C. S. Immunohistochemistry for pathologists: protocols, pitfalls, and tips. *J. Pathol. Transl. Med.* **50**, 411–418 (2016).