

Characterising the effect of therapeutic intervention to discover pathogenic immune interactions in multiple sclerosis



Redwan Farooq
St Peter's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Hilary 2026

Abstract

Multiple sclerosis (MS) is a chronic inflammatory disease of the central nervous system (CNS), in which current therapies incompletely address the mechanisms underlying sustained disability progression. A major unresolved question is how peripheral immune dysregulation relates to compartmentalised inflammatory processes within the CNS over the disease course. In this thesis, I used high-efficacy therapies as *in vivo* perturbational probes to study pathogenic immune interactions in MS.

To address the challenge of resolving subtle treatment-related shifts in continuous cellular phenotypes, I developed an analytical pipeline for multimodal single-cell data integrating manifold-based perturbation analysis, latent transcriptional program inference, and explainable machine learning. This enabled robust identification of treatment-perturbed immune cell states and their quantification in orthogonal datasets spanning tissue compartments, disease stages, and therapeutic contexts.

By applying this pipeline to longitudinal peripheral blood profiling during B cell depletion with ocrelizumab, I identified broad reprogramming of the immune landscape, including secondary depletion of a distinct proinflammatory, cytotoxic, tissue-homing memory T cell phenotype (T_c1 -TH). This cell state was enriched in the blood of donors with highly active relapsing MS and was not reduced in a treatment non-responder with persistent breakthrough disease.

Across independent public datasets, the same transcriptional signature was enriched in T cells from cerebrospinal fluid and at the leading edge of chronic MS brain lesions, supporting a link between peripheral immune activation and

compartmentalised CNS inflammation. Complementary immune profiling during natalizumab treatment demonstrated enrichment of circulating memory B cells and T_c1-TH cells during integrin α 4-mediated migration blockade, providing convergent evidence for a common pathogenic axis disrupted by two mechanistically distinct high-efficacy therapies.

Collectively, this work supports a model of a pathogenic B–T cell axis in MS across tissue compartments and disease stages, centred on interactions between memory B cells and a chronically activated, CNS-homing cytotoxic T cell state. Further work to delineate the precise molecular pathways mediating these cellular interactions within chronically inflamed tissue niches may provide a framework for the rational development of more efficacious, safer, and durable therapies for MS.

Declaration of generative AI use

The use of generative artificial intelligence (AI) in the preparation of this thesis complies with University of Oxford guidance, including divisional and departmental policies.

Generative AI was used solely as an assistive tool for the following:

- GitHub Copilot (Microsoft): used to assist with debugging, refining, formatting, and documentation of analysis code
- Gemini 3.1 Pro (Google): used to facilitate background research and general subject understanding
- GPT-5.4 Thinking (OpenAI): used to assist with grammatical proofreading and provide structural feedback on draft text

All hypotheses, experimental designs, analytical strategies, and scientific arguments presented within this thesis are entirely my own work, except where otherwise indicated. No original AI-generated code or text was used in this work. All AI-modified code was verified for accuracy and has been made available for review.

Acknowledgements

It has been a great privilege to spend the last three and a half years developing new knowledge and skills in the course of my doctoral research. I am extremely grateful to the Association of British Neurologists and the Guarantors of Brain for providing me with this opportunity through a Clinical Research Training Fellowship, and to the OAK Foundation, Aims2Cure, and the Medical Research Council for their generous financial support of this work.

I would like to express my sincere gratitude to my supervisors, Professor Lars Fugger and Dr Kate Attfield, for their invaluable mentorship and support. Their generosity with their time and expertise, insightful feedback, and careful guidance, combined with the intellectual freedom they gave me to explore and develop my own ideas, have been instrumental in shaping this work.

I am also grateful to collaborators within the University of Oxford and beyond, including Professor Fabian Theis, Professor Ralf Gold, Professor Gabriele DeLuca, Professor Jacqueline Palace, and Dr Marco Pisa, for their constructive feedback on the methodological and clinical aspects of the work presented in this thesis.

It has been a pleasure to work with many wonderful colleagues within the Oxford Centre for Neuroinflammation, and I greatly appreciate their support and friendship over the past few years. I would like to extend special thanks to Brian Cutler for his diligence and skill in performing the laboratory experiments for single-cell datasets generated in this work, and to Subita Balaram Kuttikkatte for her expert curation of samples in our tissue bank, without whom I would have had no data to analyse!

Many other people have facilitated this work indirectly. I would like to thank the

clinical and research staff within the MS Clinical Trials Unit at the John Radcliffe Hospital for their assistance with obtaining consent from research participants and collecting blood samples, as well as the laboratory managers and administrators of the high-performance computing cluster within the MRC Weatherall Institute for Molecular Medicine for their skilful maintenance of the core research infrastructure.

My family has been a constant source of support, and I would like to thank them for keeping me grounded and reminding me of the importance of balance in life. To my parents, thank you for always encouraging me to pursue my passions, and for instilling in me the value of hard work and perseverance. To my wife, Olivia, thank you for your unwavering love, support, patience, and understanding throughout this chapter, for enduring more than your fair share of monologues about statistics and immunology, and for inspiring a few good ideas along the way! And to my daughter, Maisie, thank you for all the cuddles and laughter that brought so much joy to the final stretch of this journey.

Finally, I dedicate this thesis to all the patients who have participated in this research. I sincerely hope that my work will contribute to a better understanding of MS, and ultimately lead to improved treatments and outcomes for those living with this disease.

Contents

List of figures	xiv
List of tables	xv
List of abbreviations	xvii
1 Background	1
1.1 Multiple sclerosis	2
1.1.1 Epidemiology	2
1.1.2 Natural history	2
1.1.3 Pathogenesis	5
1.1.3.1 Genetic risk	5
1.1.3.2 Environmental risk	7
1.1.3.3 Peripheral immune activation	10
1.1.3.4 CNS compartmentalisation and tissue-resident immunity	12
1.1.3.5 Emerging paradigms and unresolved mechanisms	14
1.1.4 Therapeutic strategies	14
1.1.4.1 Immunomodulation	16
1.1.4.2 Migration blockade	18
1.1.4.3 Cell depletion	19
1.1.4.4 Immune reconstitution	21
1.1.4.5 Limitations of current therapies	22
1.2 Single-cell technologies and analytical strategies	24
1.2.1 Multimodal single-cell profiling	24
1.2.2 Heterogeneity and nuisance variation in clinical cohorts	26
1.2.3 Modelling technical noise	30
1.2.4 Resolving cellular phenotypes	33
1.2.5 Quantifying therapeutic perturbations	36
1.2.6 Interpreting phenotype–perturbation relationships	39
1.3 Conclusions and thesis aims	41
2 Methods	45
2.1 Consent and ethical approval	46
2.2 Cohorts	46
2.2.1 Discovery cohort	46
2.2.2 Treatment non-responder	46
2.2.3 Public datasets	47
2.3 Clinical data extraction	47
2.4 Sample processing	48
2.4.1 Sample collection and storage	48
2.4.2 Single-cell library preparation and sequencing	49

2.5	Data analysis	51
2.5.1	Discovery cohort	51
2.5.1.1	Data preprocessing	51
2.5.1.2	Cell type annotation	55
2.5.1.3	Quantifying treatment effect	56
2.5.1.4	Gene program inference and annotation	60
2.5.1.5	Identifying gene programs linked to treatment effect	61
2.5.1.6	Denoising and differential expression analysis	63
2.5.2	Validation cohorts	64
2.5.2.1	Data preprocessing	64
2.5.2.2	Cell type annotation	66
2.5.2.3	Gene program projection	66
2.6	Statistics	66
2.7	Visualisation	68
2.8	Data availability	68
2.9	Code availability	69
3	Characterising peripheral immune perturbations following B cell depletion	71
3.1	Introduction	72
3.1.1	Anti-CD20 therapies in MS	72
3.1.2	Immunological effects of anti-CD20 therapies	73
3.1.3	Hypothesis	75
3.2	Results	76
3.2.1	Measuring dynamic longitudinal changes in peripheral immune cells following ocrelizumab	76
3.2.2	Early repopulating B cells are immature and antigen-inexperienced	82
3.2.3	Delayed depletion of a proinflammatory, cytotoxic and tissue-homing (T_c1 -TH) memory T cell phenotype	88
3.2.4	T_c1 -TH signature correlates with clinical disease activity	99
3.2.5	Heterogeneous enrichment of a proinflammatory (Classical-Inflammatory) monocyte phenotype	102
3.3	Discussion and conclusions	108
4	Tracking pathogenic immune cell states across tissue compartments	113
4.1	Introduction	114
4.1.1	Intercompartmental differences in immune landscape between blood and CSF	114
4.1.2	Immune landscape of MS brain tissue	117
4.1.2.1	Focal lesions	117
4.1.2.2	Perivascular lymphocytic cuffs	120
4.1.2.3	Meningeal tertiary lymphoid structures	121
4.1.3	CNS-peripheral immune trafficking	122
4.1.4	Hypothesis	123

4.2	Results	123
4.2.1	T _c 1-TH signature is enriched in CSF memory T cells in MS	123
4.2.2	ASCs are enriched in CSF B cell compartment MS	127
4.2.3	T _c 1-TH signature is enriched within memory T cells at the leading edge of chronic MS brain lesions	128
4.2.4	ASCs predominate lesional B cell infiltrate in chronic MS .	133
4.3	Discussion and conclusions	134
5	Characterising altered immune cell trafficking during natalizumab treatment	139
5.1	Introduction	140
5.1.1	Natalizumab in MS	140
5.1.2	Immunological effects of natalizumab	141
5.1.3	Hypothesis	144
5.2	Results	144
5.2.1	Measuring altered immune cell trafficking following natalizumab	144
5.2.2	Multiple overlapping CD4 ⁺ memory T cell states are enriched in circulation during natalizumab treatment	147
5.2.3	T _c 1-TH signature is enriched in circulating CD8 ⁺ memory T cell compartment during natalizumab treatment	149
5.2.4	Memory B cells are enriched in the circulating B cell compartment during natalizumab treatment	153
5.3	Discussion and conclusions	156
6	Discussion	161
6.1	Novel approach to resolving treatment-related cellular phenotypic perturbations	162
6.2	Model of a pathogenic B–T cell axis in MS	163
6.3	Implications for therapeutic advances	166
6.4	Limitations	170
6.5	Future directions	171
	References	175
	Appendices	203
	A Supplementary files	203
	B Software packages	205
	C Publications and presentations	207

List of figures

1.1	Natural history of MS	4
1.2	Single-cell data integration strategies	29
1.3	Probabilistic modelling of CITE-seq data using totalVI	32
1.4	Perturbation analysis using MELD	38
2.1	Sample processing workflow	50
2.2	CITE-seq data preprocessing pipeline	54
2.3	Overview of analysis strategy for characterising shifts in cellular phenotypes due to treatment	57
2.4	Empirical MELD hyperparameter optimisation	58
2.5	Validation of normality assumption for perturbation empirical null distributions	60
2.6	cNMF K selection plots	61
3.1	Discovery cohort of pwMS with longitudinal multimodal profiling of the peripheral immune repertoire during ocrelizumab treatment	77
3.2	Discovery dataset integration	78
3.3	Discovery dataset cell type annotation	78
3.4	Treatment-induced shifts in peripheral immune landscape following ocrelizumab	79
3.5	Treatment-induced shifts in B cells following ocrelizumab	82
3.6	B cell-specific gene programs	83
3.7	Distribution of gene program scores across B cell subtypes	84
3.8	B cell gene program co-expression network	84
3.9	B cell gene program SHAP plots	86
3.10	Reprogramming of the peripheral B cell compartment during ocrelizumab treatment	87
3.11	Treatment-induced shifts in T cells following ocrelizumab	88
3.12	T cell-specific gene programs	90
3.13	Distribution of gene program scores across T cell subtypes	91
3.14	T cell gene program co-expression network	92
3.15	T cell gene program SHAP plots	93
3.16	Circulating proinflammatory, cytotoxic, tissue-homing (T_c1 -TH) memory T cells are less abundant following B cell depletion	94
3.17	Phenotypic characterisation of cells with high T_c1 -TH program score	95
3.18	CD20 expression in cells with high T_c1 -TH program score	98
3.19	Temporal dynamics of T_c1 -TH cell depletion following anti-CD20 therapy	99
3.20	Pre-treatment T_c1 -TH cell abundance across disease groups	100
3.21	Comparative longitudinal immune profiling of treatment non-responder	101
3.22	Treatment-induced shifts in monocytes following ocrelizumab	102
3.23	Monocyte-specific gene programs	103

3.24	Distribution of gene program scores across monocyte subtypes . . .	104
3.25	Monocyte gene program co-expression network	104
3.26	Monocyte gene program SHAP plots	106
3.27	Circulating proinflammatory (Classical-Inflammatory) monocytes are heterogeneously enriched following B cell depletion	107
4.1	CNS immunological niches in pwMS	118
4.2	Validation cohort of pwMS and healthy control donors with paired PBMC / CSF scRNA-seq samples	124
4.3	CSF/blood mononuclear cell validation dataset integration and cell type annotation	125
4.4	T _c 1-TH signature is enriched among memory T cells in CSF of pwMS	126
4.5	ASC signature is enriched among B cells in CSF of pwMS.	129
4.6	Validation cohort of pwMS and healthy control donors with post- mortem brain snRNA-seq samples from lesion subregions/healthy white matter	130
4.7	Brain tissue lymphocytes validation dataset integration and cell type annotation	131
4.8	T _c 1-TH signature is enriched among memory T cells in chronically inflamed brain tissue of pwMS	132
4.9	ASC signature is enriched among B cells in chronically inflamed brain tissue of pwMS	134
5.1	Validation cohort of pwMS with pre- and post-natalizumab PBMC scRNA-seq samples	145
5.2	Natalizumab PBMC validation dataset integration and cell type annotation	145
5.3	Treatment-induced shifts in peripheral immune landscape following natalizumab	146
5.4	Treatment-induced shifts in T cells following natalizumab	148
5.5	T cell gene program SHAP plots	150
5.6	CD4 ⁺ memory T cell effector states are enriched in circulation during natalizumab treatment	151
5.7	T _c 1-TH signature is enriched within circulating CD8 ⁺ memory T cells during natalizumab treatment	152
5.8	Treatment-induced shifts in B cells following natalizumab	153
5.9	B cell gene program SHAP plots	154
5.10	Memory B cells are enriched in circulation during natalizumab treatment	155

List of tables

1.1	Overview of currently available DMTs for MS.	15
3.1	Clinical and demographic details of treatment-naive pwMS with longitudinal PBMC samples during ocrelizumab treatment.	81
B.1	Key software packages.	205

List of abbreviations

ADT	antibody-derived tag
AHSCT	autologous hematopoietic stem cell transplantation
APC	antigen-presenting cell
ASC	antibody-secreting cell
BBB	blood–brain barrier
BCR	B cell receptor
BMI	body mass index
BTKi	Bruton’s tyrosine kinase inhibitor
CAR	chimeric antigen receptor
cDNA	complementary DNA
CIS	clinically isolated syndrome
CITE-seq	cellular indexing of transcriptomes and epitopes by sequencing
CLR	centered log ratio
cNMF	consensus NMF
CNS	central nervous system
CSF	cerebrospinal fluid
DMT	disease modifying therapy
EAE	experimental autoimmune encephalomyelitis
EBV	Epstein–Barr virus
EDMUS	European database for MS
EDSS	expanded disability status scale
EGA	European genome–phenome archive
FACS	fluorescence-activated cell sorting
FDR	false discovery rate
GEO	gene expression omnibus
GFAP	glial fibrillary acidic protein

GO	gene ontology
GWAS	genome-wide association study
HERV	human endogenous retrovirus
HHV-6	human herpesvirus 6
HTO	hashtag oligonucleotide
ILC	innate lymphoid cell
IRT	immune reconstitution therapy
JCV	John Cunningham virus
KDE	kernel density estimate
kNN	k -nearest neighbours
MAIT	mucosal-associated invariant T
MELD	manifold enhancement of latent dimensions
MHC	major histocompatibility complex
MNN	mutual nearest neighbour
MOG	myelin oligodendrocyte glycoprotein
MRI	magnetic resonance imaging
mRNA	messenger RNA
MS	multiple sclerosis
MSE	mean squared error
NAWM	normal-appearing white matter
NfL	neurofilament light chain
NK	natural killer
NMF	non-negative matrix factorization
OCT	optical coherence tomography
OUH	Oxford University Hospitals
PBMC	peripheral blood mononuclear cell
PCA	principal component analysis
PDF	probability density function

PIRA progression independent of relapse activity
PML progressive multifocal leukoencephalopathy
PPMS primary progressive MS
pwMS people with MS
QC quality control
qPCR quantitative polymerase chain reaction
RAW relapse-associated worsening
RES rapidly-evolving severe
RIS radiologically isolated syndrome
RNA-seq RNA sequencing
RRMS relapsing remitting MS
S1PR sphingosine-1-phosphate receptor
scRNA-seq single-cell RNA sequencing
scVI single-cell variational inference
SHAP Shapley additive explanations
siRNA small interfering RNA
SNP single nucleotide polymorphism
snRNA-seq single-nucleus RNA sequencing
SPMS secondary progressive MS
starCAT star cell annotator
T_c1-TH type 1 cytotoxic tissue-homing T
T_{fh} follicular helper T
T_h1 type 1 helper T
T_h17 type 17 helper T
T_{reg} regulatory T
T_{rm} tissue-resident memory T
TCR T cell receptor
TNF tumor necrosis factor

List of tables

totalVI total variational inference

UMAP uniform manifold approximation and projection

UMI unique molecular identifier

VAE variational autoencoder

Background

1.1	Multiple sclerosis	2
1.1.1	Epidemiology	2
1.1.2	Natural history	2
1.1.3	Pathogenesis	5
1.1.3.1	Genetic risk	5
1.1.3.2	Environmental risk	7
1.1.3.3	Peripheral immune activation	10
1.1.3.4	CNS compartmentalisation and tissue-resident immunity	12
1.1.3.5	Emerging paradigms and unresolved mechanisms	14
1.1.4	Therapeutic strategies	14
1.1.4.1	Immunomodulation	16
1.1.4.2	Migration blockade	18
1.1.4.3	Cell depletion	19
1.1.4.4	Immune reconstitution	21
1.1.4.5	Limitations of current therapies	22
1.2	Single-cell technologies and analytical strategies	24
1.2.1	Multimodal single-cell profiling	24
1.2.2	Heterogeneity and nuisance variation in clinical cohorts	26
1.2.3	Modelling technical noise	30
1.2.4	Resolving cellular phenotypes	33
1.2.5	Quantifying therapeutic perturbations	36
1.2.6	Interpreting phenotype–perturbation relationships . .	39
1.3	Conclusions and thesis aims	41

This chapter situates the research presented in this thesis within the current literature, reviewing recent advances, identifying outstanding questions, and outlining the aims and structure of the work.

1.1 Multiple sclerosis

1.1.1 Epidemiology

Multiple sclerosis (MS) is the most common inflammatory disease of the central nervous system (CNS), affecting more than 2.8 million people worldwide, with a rising prevalence¹. It is a leading cause of non-traumatic neurological disability in young adults². Its prevalence is approximately three times higher in females, and globally, the highest rates are reported in Europe and the Americas². The mean age at diagnosis is 32 years, although the disease can begin in childhood or older age groups².

1.1.2 Natural history

MS is a clinically heterogeneous disease due to high inter-individual variability in the topographical distribution of pathology throughout the CNS, involving the brain, spinal cord or optic nerves. In practice, the disease course is most usefully described by the temporal trajectory of symptoms and disability accumulation.

Classically, MS has been categorised based on its natural history into distinct subtypes^{3,4}. In the more common subtype, relapsing remitting MS (RRMS) (85–90% of cases), patients experience intermittent and partially reversible clinical manifestations (relapses). Conversely, progressive MS is characterised by insidious accumulation of irreversible neurological deficits; a progressive pattern from the onset of disease is termed primary progressive MS (PPMS) (10–15% of cases), whereas progressive disease after a relapsing onset is termed secondary progressive MS (SPMS).

Multiple recent lines of evidence, most notably an analysis of a large longitudinal clinical dataset using probabilistic machine learning⁵, have challenged this dichotomy and reframed the disease as a continuous spectrum in which both patterns coexist throughout its course. This is supported by pathological evidence of a strong interdependence of inflammation and neurodegeneration in MS at all disease stages^{6,7}. In this modern conception of the disease, smouldering inflammation-driven neuronal loss occurs from the earliest stages through to advanced end-stage disease, manifesting as disability progression independent of relapse activity (PIRA)⁸. Superimposed on this background, recurrent bouts of acute CNS inflammation caused by infiltrating peripheral immune cells result in the accumulation of lesions exhibiting focal loss of myelin and axonal damage, which present clinically as relapses if they occur at an eloquent neuroanatomical site at sufficient severity to cause symptoms. Over time, there is a gradual trend for decreased relapse activity and increased PIRA⁵, potentially reflecting increasing compartmentalisation of the pathogenic inflammatory response within the CNS⁹ (Figure 1.1).

In keeping with this improved understanding of the disease continuum, the 2024 revision of the international McDonald diagnostic criteria¹⁰ presents a unified framework for the diagnosis of MS across the disease spectrum, encompassing both relapsing and progressive patterns, as well as presymptomatic and prodromal stages of the disease historically known as radiologically isolated syndrome (RIS) and clinically isolated syndrome (CIS), respectively.

1. Background

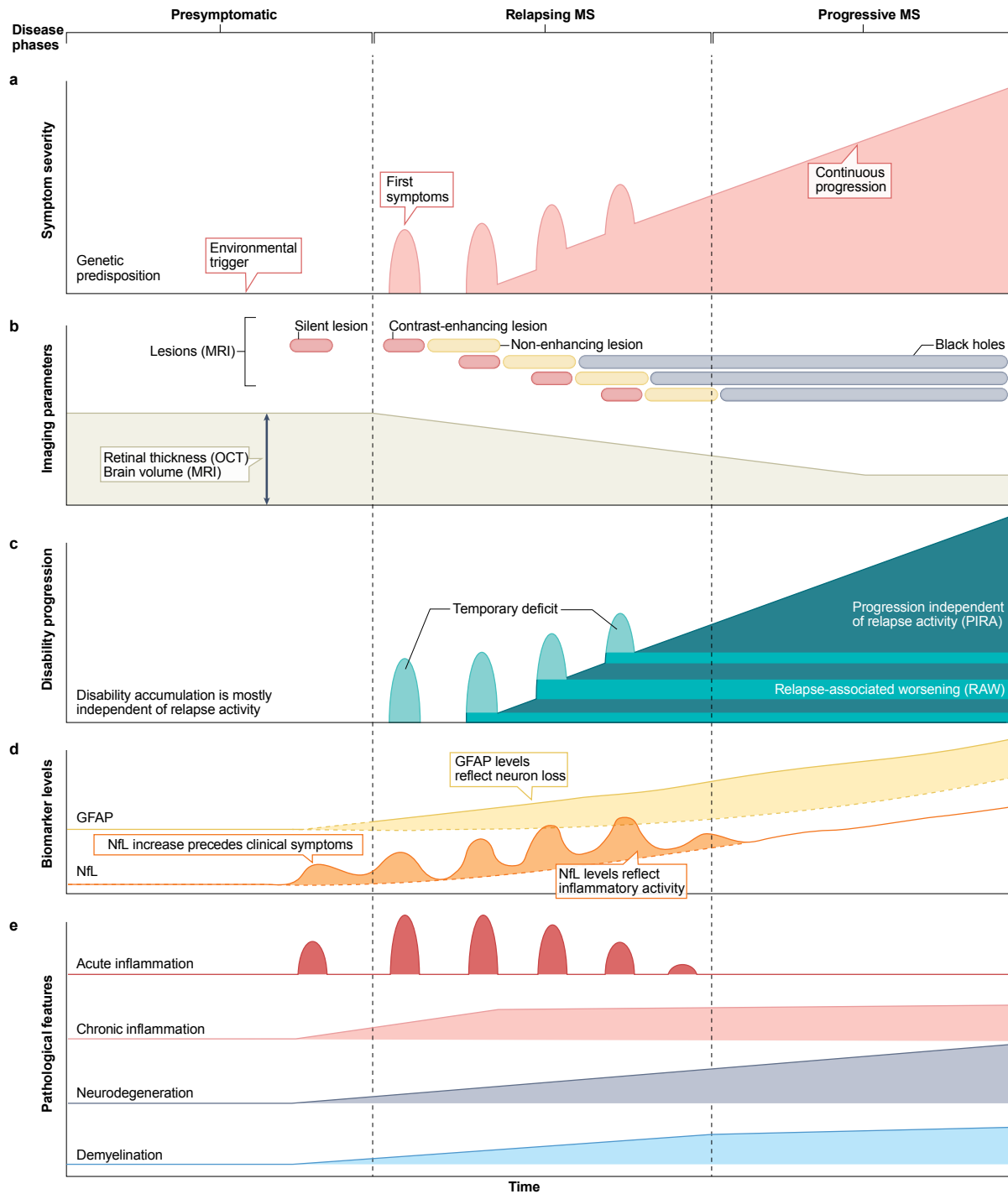


Figure 1.1. Natural history of MS. Schematic diagram illustrating the natural history of MS, highlighting a clinicopathological disease continuum reflected across symptom progression (a), imaging findings on magnetic resonance imaging (MRI) and optical coherence tomography (OCT) (b), disability accrual (c), emerging blood biomarkers neurofilament light chain (NfL) and glial fibrillary acidic protein (GFAP) (d), and pathological features (e) over time. Adapted with permission from Woo MS *et al.*⁷.

1.1.3 Pathogenesis

The pathogenesis of MS remains incompletely resolved. Based primarily on studies in experimental autoimmune encephalomyelitis (EAE), an animal model of CNS inflammatory disease, MS has classically been conceptualised as T cell-driven autoimmunity against undetermined CNS antigens, possibly precipitated by an infectious trigger. However, contemporary genetic, epidemiological, immunological and histopathological evidence from people with MS (pwMS) increasingly supports a more nuanced model of a continuum of disease stages unified by immune dysregulation, but characterised by evolving cellular interactions and anatomical compartmentalisation, rather than a single aberrant causal cell type with a response directed against a specific self-antigen. Understanding how genetic susceptibility and environmental exposures converge on this dysregulated immune state remains a major challenge. This complexity has been further highlighted by modern high-dimensional profiling techniques, which expose immune heterogeneity, plasticity, and context-dependent behaviour and thereby challenge traditional models of MS pathogenesis¹¹.

1.1.3.1 Genetic risk

Studies of genetic susceptibility to MS strongly support the view that it is an immune-mediated disease. The most substantial contribution arises from variation within the major histocompatibility complex (MHC); in general, class II alleles (e.g. HLA-DRB1*15:01, HLA-DRB1*13:03, HLA-DRB1*08:01, and HLA-DQB1*03:02) confer increased risk whereas class I alleles (e.g. HLA-B*38:01, HLA-A*02:01, HLA-B*44:02, and HLA-B*55:01) confer decreased risk¹². Notably, MHC class II is

1. Background

canonically involved in antigen presentation to CD4⁺ T cells, and this association has historically led to a focus on these cells as central pathogenic drivers¹³. However, it has emerged that MHC class II expression is not restricted to professional antigen-presenting cells (APCs) and can be induced on additional immune populations, including activated T cells, under conditions of chronic inflammation^{14,15}; therefore, the strong genetic association with class II alleles does not necessarily imply a cell type-specific driver mechanism in MS.

The largest MS genome-wide association study (GWAS) to date identified more than 200 non-MHC risk alleles, most of which map to genetic loci located within and proximal to genes expressed in a broad range of both peripherally-derived immune cells and CNS-intrinsic microglia¹². Rather than altering coding sequences, most of these variants appear to influence gene expression and chromatin accessibility in activated immune cells¹¹, consistent with a model in which MS risk reflects modulation of immune cell states rather than production of functionally aberrant proteins. Because most risk alleles have modest effect sizes, there is ongoing uncertainty over whether these signals pinpoint discrete processes in specific pathogenic cell types or reflect system-wide shifts that predispose to dysregulated responses to environmental triggers.

Notably, a more recent GWAS studying genetic determinants of MS severity identified two novel loci not previously identified in earlier studies that were focussed on disease susceptibility¹⁶. Moreover, heritability tissue enrichment analysis of severity-linked variants revealed significant enrichment exclusively in CNS tissues, in contrast to the strong immune tissue enrichment of susceptibility-linked variants. This discrepancy may suggest that although the immune system plays a critical role

in orchestrating the underlying causal pathogenic processes in MS, the functional consequences in terms of disability severity are predominantly determined by factors modulating CNS resilience.

1.1.3.2 Environmental risk

Among environmental factors, Epstein–Barr virus (EBV) infection has emerged as the most consistently implicated in MS pathogenesis¹⁷. Seropositivity for EBV is virtually universal among pwMS, compared to 90–95% among all adults. A landmark longitudinal study in a large cohort of EBV-seronegative adults demonstrated that EBV infection preceded MS onset and thereafter significantly increased the risk of developing the disease¹⁸, suggesting a prerequisite priming or triggering role. The timing and severity of primary infection appears to modulate risk, with infection in late adolescence or early adulthood, especially when manifesting as infectious mononucleosis, significantly increasing the risk of developing the disease^{19,20}. The biological basis of this association remains an area of intense investigation, but the specificity and magnitude of the effect distinguishes EBV from other proposed environmental risks.

EBV is a ubiquitous γ -herpesvirus that establishes lifelong latent infection in humans, predominantly within memory B cells. During primary infection in epithelial cells and naive B cells, viral DNA enters the host cell nucleus as an episome (unintegrated circular DNA fragment), and viral proteins (e.g. EBNA1–EBNA6, LMP1, LMP2A and LMP2B) manipulate host gene expression and signalling pathways to promote cell activation, survival, proliferation, and differentiation, thereby enabling viral entry into the long-lived memory B cell compartment. This

1. Background

latent reservoir is maintained through homeostatic proliferation of memory B cells, during which viral genomes replicate alongside host DNA, and can periodically reactivate to enter lytic cycles, resulting in viral shedding and infection of new B cells^{21–23}. Importantly, the virus has evolved mechanisms to shield infected cells from immune clearance during the latent phase; for example, the glycine–alanine repeat domain of EBNA1 impairs proteasomal processing and reduces viral antigen presentation to cytotoxic CD8⁺ T cells²⁴, impairing effective antiviral responses. Moreover, expression of latent-phase viral proteins in B cells can alter differentiation; for example, EBNA3 has been shown to repress regulators of plasma cell differentiation, favouring a sustained latent state in memory B cells over terminal differentiation²⁵.

Epidemiological and immunological studies in pwMS reveal that immune responses to EBV are both heightened and dysregulated compared with healthy individuals. Individuals with MS exhibit broader adaptive immune responses to EBV than EBV-seropositive healthy controls^{26,27}, suggesting that the normal equilibrium between virus and host immunity is perturbed. Importantly, this finding was demonstrated in MS-discordant monozygotic twin pairs, discounting a confounding effect due to genetic factors.

Within this context, EBV has multiple recognised effects on the immune system that could plausibly drive MS. One hypothesis that has been strongly pursued is molecular mimicry, wherein immune responses directed against EBV antigens cross-react with structurally similar CNS targets. For example, EBNA1 has been shown to have structural homology with various CNS antigens (e.g. GlialCAM, CRYAB, MBP, and ANO2) and cross-reactive antibodies and T cell receptors (TCRs) have

been identified in pwMS^{28–30}. Crucially, although pwMS have a broader range of cross-reactive responses, similar responses directed towards each individual CNS antigen are detectable in healthy control donors; therefore, the presence of these cross-reactive adaptive immune responses alone appears insufficient to cause disease. Additionally, MS risk loci appear to modulate EBV-specific immunity; for example, CD4⁺ T cells from individuals carrying the HLA-DRB1*15:01 allele demonstrated increased autoproliiferation in the presence of EBV-infected memory B cells in an experimental model³¹. In a different disease context, EBNA2 has been recently shown to upregulate antigen presentation function in EBV-infected autoreactive B cells³². However, the precise mechanisms by which EBV contributes to MS pathogenesis remain unclear.

Human herpesvirus 6 (HHV-6), another γ -herpesvirus, has also been implicated in MS risk, although the strength and consistency of this association is weaker than for EBV³³. Moreover, numerous studies have found an association between MS and higher expression in blood and cerebrospinal fluid (CSF) of genes derived from human endogenous retroviruses (HERVs), a broad group of ancient retroviruses which have become integrated into human germline DNA (accounting for up to 8% of the human genome)³⁴. Additional environmental factors, such as smoking, vitamin D deficiency, and obesity, are reproducibly associated with MS susceptibility and outcome, yet their mechanistic links to immune dysregulation in the disease context are poorly defined¹¹.

Taken together, the convergence of genetic risk in immune pathways and strong evidence of an infectious trigger suggests that MS arises when a genetically primed immune system encounters a specific environmental challenge, with the strongest

1. Background

evidence implicating EBV, and establishes a persistent dysregulated immunological circuit rather than a homeostatic inflammatory response.

1.1.3.3 Peripheral immune activation

The origin of immune dysregulation in MS is widely thought to lie in the periphery rather than the CNS. This view is supported by several converging observations, including the strong genetic enrichment of risk variants at loci implicated in peripheral immune function¹² and the efficacy of therapies that primarily act on peripheral immune cells.

Historically, MS pathogenesis models were heavily influenced by EAE, in which myelin-specific CD4⁺ T cells play a central role and can transfer disease between animals³⁵. Indeed, human studies demonstrate enrichment of CD4⁺ T cells with type 1 helper T (T_h1) and type 17 helper T (T_h17) transcriptional and cytokine profiles in the blood, CSF, and active CNS lesions of pwMS, supporting a contributory role for these lineages in disease pathogenesis^{36,37}. In addition, CD4⁺ regulatory T (T_{reg}) cells have been shown to have impaired suppressive function in pwMS³⁸, which may facilitate the escape of self-reactive adaptive immune cells from peripheral tolerance mechanisms.

However, several lines of evidence from human studies challenge this canonical view of pathogenesis. Direct evidence for disease-defining, clonally expanded myelin-specific TCRs in MS remains elusive, and T cells capable of recognising myelin antigens can be detected at comparable frequencies in healthy individuals, indicating that self-antigen recognition alone is insufficient to define pathogenicity in human disease^{39–41}. Moreover, histopathological analyses consistently demonstrate that

CD8⁺ T cells dominate the lymphocytic infiltrates in MS lesions, often outnumbering CD4⁺ T cells by several-fold, and exhibit clonal expansion and markers of chronic activation and proinflammatory cytotoxic effector function^{42–45}. These observations raise the possibility that CD8⁺ T cells may act as either direct mediators of tissue injury or as critical amplifiers of local inflammation. Notably, CD8⁺ T cells are specialised for recognition of intracellular antigens presented on MHC class I molecules, a functional profile that aligns with responses to virus-infected or stressed cells and contrasts with the classical role of CD4⁺ T cells in orchestrating immune responses to processed extracellular antigens.

B cells, although less abundant within parenchymal lesions than T cells, have also emerged as key orchestrators of the pathogenic immune response in MS. Despite prominent intrathecal immunoglobulin production, detectable clinically as the diagnostic hallmark of CSF-restricted oligoclonal bands on gel electrophoresis, there is limited evidence for a pathogenic autoantibody as a central driver of MS pathology. In fact, the identification of specific autoantibodies against the CNS antigens aquaporin-4 and myelin oligodendrocyte glycoprotein (MOG) over the last two decades have instead led to the recognition and reclassification of distinct antibody-mediated CNS inflammatory disorders⁴⁶ with overlapping clinical features but distinct biology, genetic and environmental risk, and treatment response⁴⁷. The lack of broadly prevalent disease-specific autoantibodies argues strongly against a primary humoral role of B cells in MS pathogenesis^{48–50}. Instead, B cells are increasingly recognised for their non-humoral functions, including antigen presentation, costimulatory signalling, and cytokine production, all of which can promote or modulate T cell responses and shape the adaptive immune repertoire.

1. Background

The role of memory B cells as a latent reservoir for EBV adds a layer of complexity, raising the possibility that chronic antiviral immune responses directed against EBV-infected memory B cells could provide a sustained source of immune activation, with collateral damage to the CNS.

Finally, innate lymphoid cells can modulate these adaptive immune responses. For example, immature CD56^{bright} natural killer (NK) cells have been shown to suppress aberrant T cell responses through direct cell-mediated cytotoxicity, a function which is impaired in pwMS compared with healthy controls⁵¹. Similarly, a recent study showed that NKG2C⁺ and NKG2D⁺ NK cell-mediated control of GlialCAM-reactive EBV-infected B cells is severely impaired in pwMS⁵². In both contexts, evasion of NK cell-dependent regulation of the adaptive immune response was found to be dependent on upregulation of HLA-E on the surface of T and B cells, which dampens NK cell activation through binding to the inhibitory receptor NKG2A. Therefore, genetic variants and environmental factors that alter these innate–adaptive interactions may act, at least in part, by disrupting homeostatic regulatory mechanisms that normally restrain maladaptive pathogenic immune responses.

1.1.3.4 CNS compartmentalisation and tissue-resident immunity

In later stages of MS, pathological processes become increasingly compartmentalised within the CNS, and dependence on continuous peripheral immune recruitment appears to diminish. Histopathological studies have demonstrated two key anatomical substrates of this compartmentalised response: (i) expanded perivascular lymphocytic cuffs^{53–55}, and (ii) meningeal tertiary lymphoid structures, which

commonly overlies areas of focal grey matter pathology^{56–60}. These structures are associated with more aggressive disease and poorer clinical outcomes^{57–59}. In parallel, although the formation of new focal lesions declines over time, many established lesions exhibit persistent inflammatory activity at their borders, known as chronic active or mixed active/inactive lesions^{6,61,62}. These lesions are increasingly recognised as the pathological correlates of PIRA^{61,63}.

The immune cell populations implicated in CNS-compartmentalised inflammation differ in several respects from those dominating in relapsing disease. The majority of lymphocytes within chronic lesions are tissue-resident memory T (T_{rm}) cells, expressing markers that promote long-term retention within the CNS and limit recirculation in the periphery^{54,55}. These cells exhibit clonal expansion and markers of chronic activation, proinflammatory signalling and cytotoxicity, suggesting local reactivation and ongoing antigenic stimulation. B cells, including memory B cells and plasma cells, are also found within meningeal lymphoid structures and perivascular spaces^{60,64}, where they may support local antigen presentation and cytokine production rather than causing antibody-mediated injury. Finally, microglia, the resident myeloid cells of the CNS, play a central role in sustaining this compartmentalised inflammatory milieu. In chronic active lesion rims and the surrounding normal-appearing tissue, these cells adopt distinct activation states characterised by loss of homeostatic signatures and upregulation of proinflammatory, phagocytic, iron-handling and oxidative stress-associated programs, consistent with histopathological evidence of disease-associated microglia that accumulate myelin debris, iron and reactive oxygen species at the lesion edge^{65–68}.

1.1.3.5 Emerging paradigms and unresolved mechanisms

A central emergent view consolidated by recent work is that the immunopathogenesis of MS involves heterogeneous and plastic cell states, not static, lineage-defined populations¹¹. Taken together, current evidence supports a model in which MS arises from a dysregulated immune network, initiated in the periphery but progressively reshaped by compartmentalised immune interactions within the CNS.

While genetic variants and environmental exposures conferring disease risk clearly converge on immune pathways, the precise cellular states and interactions that act as causal drivers of disease remain incompletely defined. Moreover, although improvements in our understanding of the natural history increasingly support a shared biology between relapsing and progressive disease, we have not yet identified a unifying immunophenotype that bridges these two aspects of MS. This mechanistic gap underscores both the limitations of inferring pathogenic mechanisms of human disease from animal models and the challenges of discovering causal biology directly from noisy observational human data.

1.1.4 Therapeutic strategies

Despite the complex and incompletely resolved immunobiology of MS, significant advances have been made in its treatment over the past several decades. The modern landscape of licensed disease modifying therapies (DMTs) in MS comprises a wide range of pharmacological agents with often broad and diverse effects on the immune system⁶⁹ (Table 1.1). Despite the heterogeneous mechanisms of action of these drugs, their clinical successes and failures can provide key insights into the underlying pathogenic mechanisms.

Table 1.1. Overview of currently available DMTs for MS.

Drug/Class	Approved Indication	Route ¹	Frequency ¹	Type	Primary Mechanism(s) of Action
Low- or moderate-efficacy					
Interferon β	CIS, RRMS, SPMS	Intramuscular, subcutaneous	Variable (3 times a week to every 2 weeks)	Small molecule	Shifts balance between pro- and anti-inflammatory cytokines, inhibits proliferation and blocks trafficking of inflammatory cells into the CNS
Glatiramer acetate	CIS, RRMS	Subcutaneous	Variable (3 times a week to once weekly)	Synthetic	Shifts balance between proinflammatory and regulatory T cells, inhibits APCs by competition with myelin antigens
Fumaric acid esters (dimethyl fumarate, diroximel fumarate, monomethyl fumarate)	RRMS	Oral	Twice daily	Small molecule	Promotes anti-oxidant responses via activation of NRF2, downregulation of aerobic glycolysis via inhibition of GAPDH
Teriflunomide	RRMS	Oral	Once daily	Small molecule	Reduces proliferation of activated lymphocytes, inhibits dihydro-orotate dehydrogenase
High-efficacy					
S1PR modulators (fingolimod, siponimod, ozanimod)	RRMS, SPMS	Oral	Once daily	Small molecule	Inhibits egress of lymphocytes from secondary lymphoid organs by impairing S1PR binding
Cladribine	RRMS, SPMS	Oral	Once yearly (2 courses)	Synthetic	Impairs DNA synthesis and repair
Mitoxantrone	RRMS, SPMS	Intravenous	Once every 3 months	Synthetic	Promotes DNA cross-linking and strand breaks
Natalizumab	RRMS, SPMS ²	Once monthly	Once every 4 weeks	Monoclonal antibody (humanised)	Inhibits migration of lymphocytes across BBB (anti-integrin $\alpha 4$)
Alemtuzumab	RRMS	Intravenous	Once yearly (2-4 courses)	Monoclonal antibody (humanised)	Paralymphocyte depletion (anti-CD20)
Rituximab	RRMS (off label)	Intravenous	Once every 6 months	Monoclonal antibody (chimeric)	B cell depletion (anti-CD20)
Ubituximab	RRMS, PPMS	Intravenous	Once every 6 months	Monoclonal antibody (chimeric)	B cell depletion (anti-CD20)
Oclatumumab	RRMS	Subcutaneous	Once every 4 weeks	Monoclonal antibody (humanised)	B cell depletion (anti-CD20)

Adapted with permission from Attfield KE *et al.*¹¹ and Olejnik P *et al.*⁶⁹.

¹Standard administration protocol
²Approved for SPMS in US only

1. Background

1.1.4.1 Immunomodulation

The first generation of MS DMTs that emerged in the early 1990s can be broadly conceptualised as immunomodulators. Although their precise mechanisms are pleiotropic and incompletely resolved, these agents are thought to act by shifting immune responses away from proinflammatory effector programs, as well as modulating cytokine signalling and antigen presentation¹¹. These therapies include the pioneering injectable therapies interferon- β ⁷⁰ and glatiramer acetate⁷¹, as well as the subsequently licensed, orally administered fumaric acid esters^{72,73} (e.g. dimethyl fumarate, diroximel fumarate, and monomethyl fumarate). The efficacy of these drugs demonstrates that altering peripheral immune tone in MS can attenuate inflammatory activity in the CNS, although their relatively modest benefit compared to modern high-efficacy DMTs suggests that a non-specific global dampening of inflammation may be insufficient to fully extinguish the pathogenic dysregulated immune networks that drive disease.

Notably, the failure of some immunomodulatory agents in pwMS has been mechanistically informative. A striking example of this is the paradoxical worsening of disease activity induced by anti-tumor necrosis factor (TNF) therapies⁷⁴. Anti-TNF therapies have been shown to be protective in EAE⁷⁵ and elevated levels of TNF- α are detectable in MS serum and CSF⁷⁶. However, a key functional genomics study identified that a known MS-associated single nucleotide polymorphism (SNP) leads to increased expression of a soluble form of TNF receptor 1 which blocks TNF signalling, mirroring the effect of anti-TNF therapies⁷⁷. These findings therefore highlight a key limitation of inferring mechanisms of complex human disease from

animal models and the challenges inherent in translating preclinical findings from these model systems to effective therapies for pwMS.

Another targeted immunomodulatory drug briefly licensed for treating MS was daclizumab, a monoclonal antibody that binds to the α subunit of the high-affinity receptor for IL-2, a key trophic cytokine expressed by both activated CD4⁺ and CD8⁺ T cells following cognate antigen recognition by the TCR. By doing so, it impairs IL-2 signalling and reduces peripheral T cell activation. Moreover, the resultant increased availability of IL-2 is thought to amplify its signalling through the intermediate-affinity IL-2 receptor found on NK cells, which does not contain an α subunit⁷⁸. Importantly, IL-2 signalling in NK cells has been shown to drive expansion of anti-inflammatory CD56^{bright} subsets⁷⁹, which play an immunoregulatory role through direct cytotoxic killing of highly activated T cells. Despite promising efficacy in clinical trials^{80,81}, daclizumab was withdrawn due to safety concerns after reports emerged of severe autoimmune disorders, including fatal cases of encephalitis, following treatment. Although speculative, it has been postulated that the occurrence of autoimmune disorders may have been precipitated by daclizumab-mediated suppression of T_{reg} cells, which constitutively highly express IL-2R α and are dependent on IL-2 signalling for their development and maintenance⁷⁸. Therefore, the failure of daclizumab illustrates the importance of understanding and preserving immune interactions required to maintain homeostasis, alongside targeting those that are potentially pathogenic, in the context of identifying novel therapeutic targets.

1. Background

1.1.4.2 Migration blockade

A major conceptual and clinical advance in the 2000s was the development of therapies that primarily act by altering immune cell trafficking. The earliest of these drugs to be licensed in MS was natalizumab, a monoclonal antibody against integrin $\alpha 4$, which forms part of a heterodimeric complex with either integrin $\beta 1$ or integrin $\beta 7$ to form the cellular adhesion molecules VLA-4 or LPAM-1, respectively. VLA-4 plays a critical role in lymphocyte transmigration across the blood–brain barrier (BBB), and natalizumab inhibits the binding of this complex to VCAM-1 on vascular endothelial cells to prevent migration into the CNS. In a pivotal trial, natalizumab treatment was shown to produce large reductions in both clinical relapse rates and radiologically-detectable disease activity⁸². It is widely used to treat highly active RRMS due to its high efficacy and rapid onset of therapeutic benefit. Crucially, the success of natalizumab provides strong evidence that continuous recruitment of peripherally-originating immune cells into the CNS is a major driver of relapses and new lesion formation.

Subsequently, an alternative class of lymphocyte migration blocking DMTs emerged with the licensing of sphingosine-1-phosphate receptor (S1PR) modulators^{83,84}, which interfere with a key receptor required for egress of lymphocytes from secondary lymphoid organs, thereby reducing the pool of circulating cells available for recruitment to the CNS. Drugs in this class (e.g. fingolimod, siponimod, ozanimod, and ponesimod) are broadly licensed for use in active forms of MS, as defined by the presence of clinical relapses or radiologically detectable inflammation.

An important trade-off highlighted by migration blocking DMTs is the increased risk of opportunistic CNS infection resulting from impairment of the normal immune surveillance of this compartment. Natalizumab, in particular, is associated with potentially fatal cases of progressive multifocal leukoencephalopathy (PML) due to reactivation of John Cunningham virus (JCV)⁸⁵, the risk of which increases with longer treatment duration⁸⁶ and therefore often imposes a *de facto* limit on treatment duration due to an unfavourable safety profile in a subset of pwMS with JCV seropositivity or prior immunosuppressant use. This highlights the lack of specificity of this class of DMT, impairing protective immune function in addition to disrupting pathogenic circuits, and emphasises the need for more targeted approaches to achieve durable disease control without exposing pwMS to unacceptable risks.

1.1.4.3 Cell depletion

A second major shift in the therapeutic landscape transpired in the 2010s, demarcated by the treatments designed to deplete specific immune cell subsets using either cytotoxic/antiproliferative drugs or monoclonal antibodies against lineage-specific cell surface proteins.

Cytotoxic/antiproliferative DMTs that have shown therapeutic efficacy in MS include mitoxantrone⁸⁷, teriflunomide^{88,89}, and cladribine⁹⁰. These drugs all exhibit broad effects across both B and T lymphocyte compartments.

In contrast to broadly acting immunoablative drugs, targeted monoclonal antibodies have provided clearer mechanistic insights into MS immunobiology by their selective effect on cell lineages defined by expression of specific cell surface

1. Background

proteins. Perhaps the most consequential class of drugs in this category has been anti-CD20 therapies, which target B cells. Preliminary evidence for the efficacy of these drugs in MS was first noted through the off-license use of the human/mouse chimeric anti-CD20 monoclonal antibody rituximab⁹¹. Subsequently, the landmark trials of anti-CD20 therapies in MS used the humanised monoclonal antibody ocrelizumab^{92,93}. Importantly, in addition to showing high efficacy in relapsing disease, ocrelizumab was the first DMT to show modest benefit at slowing PIRA in the context of PPMS. Since then, other anti-CD20 therapies have been approved for use in MS (e.g. ublituximab and ofatumumab)^{94,95}, with differences primarily related to pharmacodynamics or route of administration. Importantly, the remarkable clinical effectiveness of these therapies in MS despite the lack of disease-specific autoantibodies supports the view that B cell pathogenicity in MS is through non-humoral functions⁹⁶. This view is further supported by the paradoxical worsening of disease activity observed in a trial of atacicept⁹⁷, a recombinant fusion protein that neutralises the B cell survival factors BAFF and APRIL, which preferentially depletes antibody-secreting cells (ASCs) and naive B cells with relative sparing of memory B cells⁹⁸.

It is notable that in contrast to targeted B cell depletion, an early trial of a monoclonal antibody specifically targeting CD4⁺ T cells was not found to show clinical benefit in pwMS⁹⁹, offering another striking example of the immunological discrepancy between MS and EAE, which is largely driven by a pathological myelin antigen-directed CD4⁺ T cell response and can be ameliorated by anti-CD4 therapies¹⁰⁰.

1.1.4.4 Immune reconstitution

The final major conceptual category of MS therapeutics is immune reconstitution therapies (IRTs), which aim to achieve a transient, profound adaptive immune cell depletion, followed by a reset of the immune repertoire. The goal with these therapies is to achieve a more durable reprogramming of the immune system, under the assumption that the dysfunctional immunological niches responsible for disease that arose through a combination of genetics, environment and stochastic effects are unlikely to reappear in a newly reconstituted adaptive immune repertoire. The primary appeal of this strategy is the promise of a durable therapeutic effect without the need for continuous, long-term immunosuppression, albeit with more profound early immunoablation and consequent risk of opportunistic infection.

Alemtuzumab, an anti-CD52 monoclonal antibody which depletes mature circulating lymphocytes (especially B and T cells, but sparing plasma cells and haematopoietic stem cells), showed high efficacy in clinical trials^{101,102} with many pwMS achieving durable control of relapses and radiological disease activity following 2–4 annual courses of treatment^{103,104}. However, the immune reconstitution phase following treatment has been associated with a high (up to 48%) incidence of secondary autoimmune diseases¹⁰⁵, which can occasionally be serious and therefore necessitates prolonged regular monitoring following treatment. This, combined with reports of a higher incidence of serious cardiovascular events temporally linked to treatment, has effectively limited the use of alemtuzumab in practice to highly active RRMS cases refractory to multiple first-line DMTs.

Autologous hematopoietic stem cell transplantation (AH SCT) represents the

1. Background

most extreme case of IRT for MS. In brief, the procedure involves collection of an individual's own haematopoietic stem cells mobilised from the bone marrow, followed by administration of a cytotoxic conditioning drug to ablate the immune system, and subsequent re-infusion of the harvested stem cells which engraft and reconstitute the immune repertoire. Observational studies^{106,107}, as well as an early clinical trial¹⁰⁸, have shown this treatment results in profound and durable reduction in risk of disability progression (including PIRA), especially in individuals with relatively early stage disease¹⁰⁹, and multiple follow-up trials are currently underway comparing AHSCT with the highest efficacy DMTs currently in use for MS. Although it is increasingly used for highly aggressive, treatment-resistant MS, and the adoption of less intense, non-myeloablative conditioning regimens have improved its safety and deliverability, AHSCT remains unlikely to become a scalable solution for the broader MS population. However, the clinical experience with AHSCT suggests that a sufficiently strong disruption of the dysregulated causal immunophenotype in early disease may enable a durable reduction in relapse-independent disability progression, highlighting a window during which the eventually compartmentalised immunological drivers of PIRA remain peripherally modifiable.

1.1.4.5 Limitations of current therapies

Despite the number and mechanistic breadth of current MS DMTs, most licensed therapies are primarily effective against the acute focal inflammatory activity that underpins clinical relapses and the accrual of new lesions detectable by MRI; their impact on PIRA, if any, remains modest^{84,93}. This mismatch has focussed

attention on CNS-compartmentalised pathogenic processes, including the biology of smouldering inflammation within chronic MS lesion rims and meningeal tertiary lymphoid structures. There is a critical unmet need for DMTs that can disrupt pathogenic immunological axes on both sides of the BBB.

In this context, CNS-penetrant small molecule immunomodulators are an emerging class of drugs under active investigation. Bruton's tyrosine kinase inhibitors (BTKis) are a key example that have been recently evaluated in both relapsing and progressive MS. These drugs target a key intracellular signalling pathway downstream of B cell receptor (BCR) stimulation, as well as mediating activation and proinflammatory functions in myeloid cells (including microglia); importantly, the B cell effects of BTKis do not result in cellular depletion as with anti-CD20 therapies¹¹⁰. The mixed results of clinical trials of BTKis across relapsing and progressive disease contexts to date^{111–113} suggest that B cell pathogenicity in MS may involve mechanisms independent of antigen-specific BCR signalling and highlights that a more nuanced understanding of the precise cellular states driving compartmentalised inflammation in MS and how they arise will be key to discovering the most appropriate targets for immunomodulation across disease stages.

Following on from their success in cancer, there has been a surge in interest regarding the use of chimeric antigen receptor (CAR)-T therapies for various immune-mediated diseases, including MS. Specifically, CAR-T therapies against CD19, which act by targeted cytotoxic depletion of CD19⁺ stages of the B cell lineage, are now being explored in early clinical settings, with some initial case reports demonstrating feasibility¹¹⁴. Importantly, the ability of CAR-T cells to cross the BBB is a key theoretical advantage over monoclonal antibody-based

1. Background

cytoablative therapies for targeting the compartmentalised processes driving PIRA.

Taken together, the heterogeneous modern therapeutic repertoire strongly supports the idea that MS arises from a dysregulated network of immune interactions rather than a specific causal cell type. Importantly, this concept provides a firm rationale for using therapies as probes in a systems immunology framework. By understanding how mechanistically divergent interventions reshape the landscape of immune cell states *in vivo*, it may be possible to triangulate the cellular niches most plausibly linked to pathogenesis across compartments and disease stages. Crucially, a deeper characterisation of these cell states may reveal targetable pathways that exploit natural plasticity in the immune system to effectively and durably restore homeostatic balance without the need for indiscriminate blockade of cross-compartment immune trafficking or depletion of broad cell types.

1.2 Single-cell technologies and analytical strategies

1.2.1 Multimodal single-cell profiling

Recent advances in multimodal single-cell technologies offer an unprecedented opportunity to study the immune system directly in the context of human disease, enabling characterisation of immune cell populations at a resolution appropriate to the underlying biology of continuous and dynamic cellular states, in contrast to bulk profiling approaches that average signals across heterogeneous populations, potentially obscuring rare or context-dependent phenotypes¹¹⁵.

Single-cell RNA sequencing (scRNA-seq) provides a high-dimensional readout of pan-transcriptome gene expression at the level of individual cells, allowing

unbiased (i.e. not limited to a preselected panel of genes) examination of the transcriptional programs that define cellular identity and functional state. Modern high-throughput protocols, most commonly using microfluidic droplet-based technologies to encapsulate individual cells with uniquely barcoded gel beads¹¹⁶ or plate-based split-pool combinatorial barcoding¹¹⁷, make it feasible to profile tens of thousands to millions of cells per experiment, which is essential for capturing rare or transient immune cell states. Importantly, scRNA-seq does not assume discrete cell types *a priori*, but instead permits cellular phenotypes to be inferred from patterns of coordinated gene expression, aligning naturally with emerging views of immune dysregulation in MS as a state-based rather than lineage-restricted process.

However, transcriptional measurements alone provide an incomplete description of cellular phenotype. Gene expression is inherently dynamic and responsive to short-term environmental cues, and transcript abundance does not always correlate directly with protein expression, particularly for surface markers and signalling receptors that undergo extensive post-transcriptional regulation or exhibit long protein half-lives^{118,119}. This limitation is especially relevant in immunology, where the definition of cell identity and function has historically relied on protein-level measurements.

Cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) addresses this limitation by enabling the simultaneous quantification of transcriptomic and surface protein information from the same cells¹²⁰. In this approach, antibodies conjugated to oligonucleotide barcodes provide a quantitative sequencing-based readout of selected proteins. Protein-level measurements serve two important roles in this context. First, they provide a more stable and often

1. Background

less noisy proxy for core aspects of immune identity, enabling validation of transcriptionally-defined states¹²¹. Second, they allow direct cross-referencing with established immunophenotyping derived from flow or mass cytometry or immunohistochemistry, facilitating contextualisation of high-dimensional biological discoveries within the existing literature.

Taken together, multimodal single-cell profiling with CITE-seq integrates the sensitivity and breadth of transcriptomic measurements with the specificity and interpretability of protein-level data. This combination is particularly well suited to the study of immune dysregulation in MS, where pathogenic processes are likely to be reflected in subtle shifts in cellular state rather than coarse changes in cell counts or frequencies. At the same time, it facilitates reconciliation of high-dimensional data-driven phenotypes with decades of prior knowledge from classical immunology techniques, which is essential for robust biological inference.

1.2.2 Heterogeneity and nuisance variation in clinical cohorts

A key challenge in applying high-dimensional single-cell technologies to observational studies of human disease arises from substantial inter-individual biological heterogeneity combined with practical constraints on sample size. In typical single-cell studies using clinical samples, donor-specific effects comprise a significant proportion of the variance structure, producing large shifts in cell state measurements that are reproducible within individuals but highly variable across donors. Although this variation is biologically real, it is often orthogonal to the biological question of interest, such as the effect of a therapeutic intervention on shared disease-associated cellular phenotypes, and can therefore obscure subtle

signals or generate spurious associations. In addition, technical factors (e.g. sample processing, library preparation, sequencing depth, and reagent batches) contribute further heterogeneity. Collectively, donor-specific biological differences and technical batch effects are referred to hereafter as *nuisance variation*, motivating the need for data integration strategies that mitigate these effects while preserving subtle, conditional shifts in cell state¹²².

Numerous integration methods have been developed for single-cell data, which can be broadly grouped into conceptually related classes based on how nuisance variation is modelled and corrected. One popular class of approaches uses explicit matching of shared cell states to estimate and remove offsets related to sources of nuisance variation (Figure 1.2a). This category is exemplified by `mnncorrect`¹²³ and its scalable implementation `fastMNN`, as well as the related anchor-based methods implemented in Seurat, which assume that at least some cellular states are conserved between samples or experimental batches and can be identified through mutual nearest neighbour (MNN) pairs. Integration is achieved by aligning these matched states to produce a corrected expression matrix or reduced dimensional embedding, without imposing a global latent representation.

A second major class of methods directly quantifies and corrects nuisance variation with linear models applied to the reduced dimensional data embeddings. The exemplar method in this category is Harmony¹²⁴, which uses soft clustering assignment and estimation of cluster-specific vectors of nuisance variation in principal component analysis (PCA) space to adjust cell embeddings and shrink these effects iteratively until convergence (Figure 1.2b). The key assumptions are that nuisance variation is locally approximately linear and separable from the

1. Background

biological signal of interest, allowing removal of this variation without substantially distorting local relationships between similar cell states in the embedding space.

More recently, there has been an explosion of deep learning-based integration approaches¹²⁵, typically using a conditional variational autoencoder (VAE), a neural network-based architecture which models the observed cells as existing in a global latent space intended to capture the underlying biological cell state, from which the measured expression data are probabilistically generated according to a specified distribution whose parameters are functions of the latent embedding. Within this class, tools such as single-cell variational inference (scVI)¹²⁶ can learn complex, non-linear sources of nuisance variation based on user-specified covariates that condition the generative model parameters during VAE training (Figure 1.2c). These approaches implicitly assume that the biological relationships between similar cell states can be accurately captured by the common latent space, and prioritise alignment between covariates within that representation.

Deep learning-based approaches have proven extremely powerful for atlas-level integration across tissues, species, and technologies, where the primary objective is to identify related cell types and states across numerous diverse biological contexts and datasets¹²⁷. However, the strengths of these methods in aggressively harmonising cells from disparate samples into a shared latent space may become a weakness in the setting of a clinical cohort study focused on detecting subtle phenotype shifts between conditions, based on samples from a limited number of donors¹²¹.

Therefore, the safest strategy to address nuisance variation without compromising biological discovery power in clinical cohort studies is to prioritise optimised

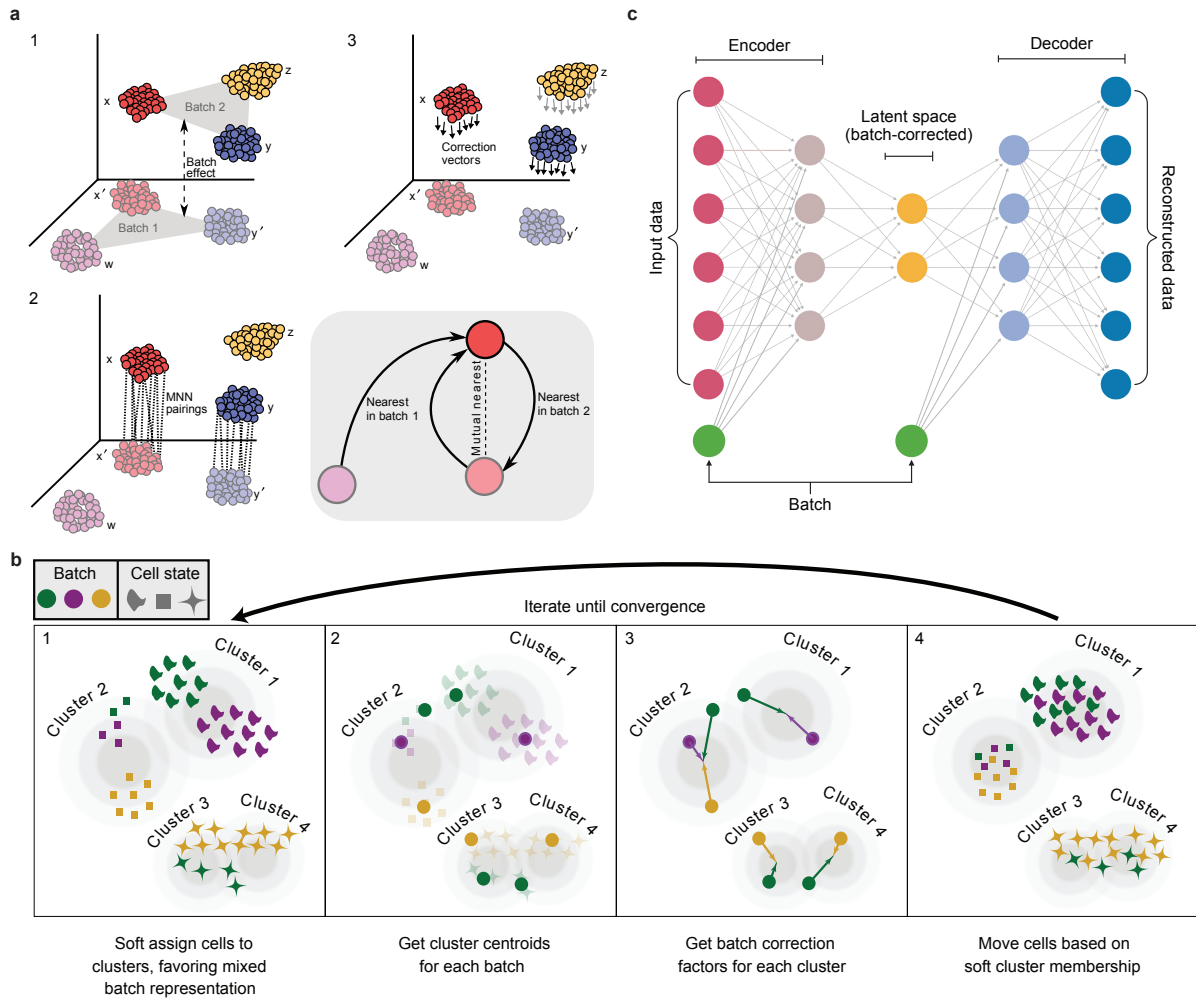


Figure 1.2. Single-cell data integration strategies. Schematic diagrams illustrating the key principles underlying major classes of single-cell data integration methods; *batch* is used here as a general term to refer to sources of nuisance variation. **a**, Explicit matching of shared cell states across batches using MNN pairs, which are used to estimate local batch correction vectors in either raw expression or PCA space. Adapted with permission from Haghverdi L *et al.*¹²³. **b**, Iterative soft cluster assignment and fitting linear models to estimate local batch correction vectors in PCA space. Adapted with permission from Korsunsky I *et al.*¹²⁴. **c**, Conditional VAE architecture with batch covariates incorporated into the model to learn a latent representation of cell state while accounting for batch effects. Created with BioRender.com.

1. Background

experimental design. Longitudinal studies are a powerful paradigm for addressing inter-donor heterogeneity by using statistical models that explicitly account for donor-specific baselines. Moreover, cell hashing¹²⁸ offers the ability to pool multiple samples into the same processing batch, ensuring the experimental variables of interest (e.g. longitudinal pre- and post-treatment samples from the same donor) are independent of nuisance factors. The nuisance variation can thereafter be corrected using a conservative integration strategy such as the Harmony algorithm, which has been consistently shown in multiple benchmarking studies to perform well relative to other methods at conserving subtle biological variation in simple data integration tasks^{127,129,130}.

1.2.3 Modelling technical noise

In addition to between-sample heterogeneity, single-cell measurements are affected by multiple sources of technical noise that arise from the experimental assay. Widely recognised issues in droplet-based scRNA-seq, such as ambient RNA contamination, doublets, and low quality or dying cells, can introduce artefactual signal if not appropriately handled. Several computational strategies exist for mitigating these noise sources during quality control (QC) steps, and their use is considered best practice in modern single-cell analysis workflows¹²².

There are additional modality-specific sources of technical noise in CITE-seq data. Single-cell RNA measurements are inherently sparse, reflecting capture of only a small fraction (typically 10–15%) of the transcriptome in each cell. This stochastic under-sampling gives rise to dropout, whereby transcripts that are genuinely expressed may not be detected, resulting in excess zero counts that

are indistinguishable from true biological absence. In contrast, antibody-derived tag (ADT) measurements are typically less affected by dropout due to the limited size of antibody panels, but exhibit a distinct and equally important source of technical noise arising from non-specific antibody binding, particularly via Fc receptors¹³¹.

Among currently available frameworks for analysing CITE-seq data, total variational inference (totalVI)¹³² is specifically designed to model RNA and ADT measurements within a unified probabilistic framework that accounts for the distinct noise characteristics of each modality. This method utilises a VAE to jointly model the RNA and ADT measurements using a generative process (Figure 1.3). Crucially, ADT measurements are modelled as arising from a mixture of background and foreground components, enabling non-specific antibody binding to be distinguished from true protein expression. In addition, jointly modelling RNA and ADT data from the same cells improves sensitivity for detecting low-level but biologically meaningful protein expression in specific cellular states, while simultaneously enabling imputation of missing proteins due to differences in antibody panels across experiments. While several alternative approaches exist for normalising ADT measurements and performing integrative analysis with RNA data^{120,131,133–135}, none of these methods provide a statistical model suitable for joint RNA and protein differential expression testing. Crucially, applying standard differential expression methods developed for RNA data to either raw or transformed ADT measurements can lead to spurious results driven by variation in the non-specific background component^{121,132}.

1. Background

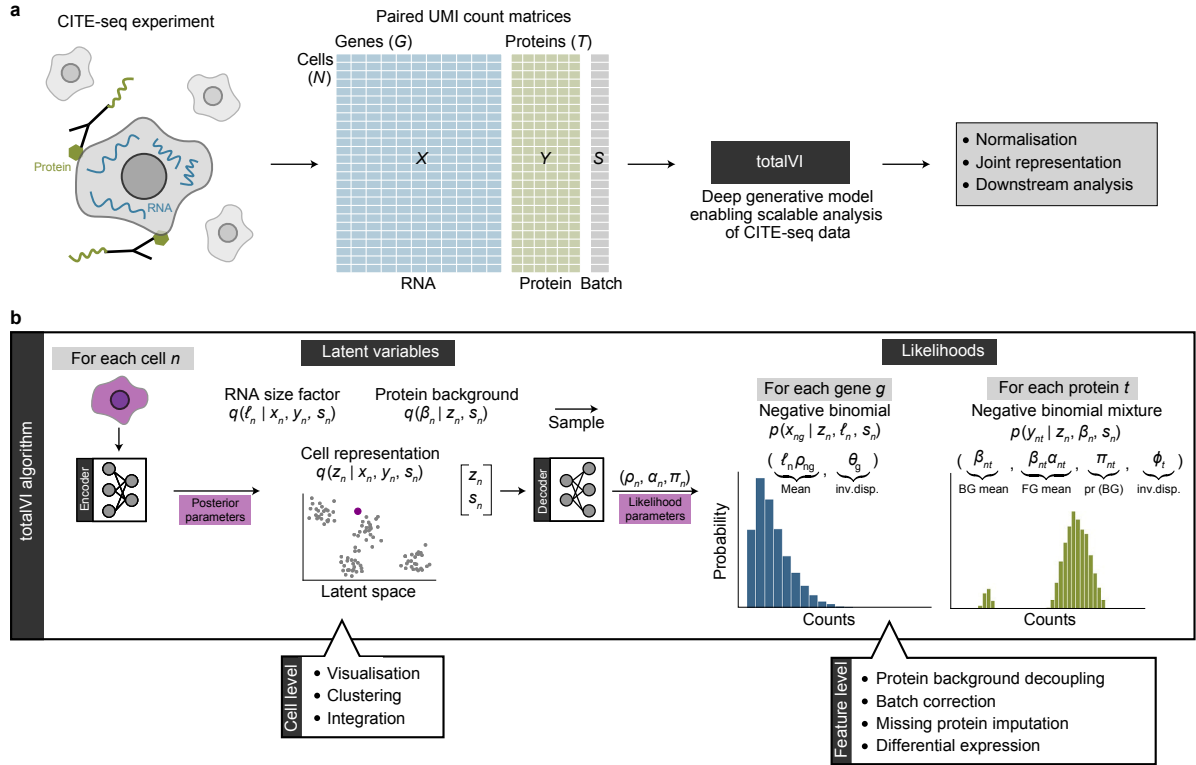


Figure 1.3. Probabilistic modelling of CITE-seq data using totalVI. Schematic diagram illustrating the totalVI model. **a**, Model inputs comprise paired RNA and ADT count matrices, along with user-specified batch annotations and other technical covariates. **b**, Conditional VAE framework used to learn a joint low-dimensional representation of cell state from both RNA and ADT modalities while accounting for technical noise. During model training, for each cell n , the encoder infers latent variables for (i) cell state, z_n , (ii) RNA library size, ℓ_n , and (iii) protein-specific background, β_n . The decoder reconstructs the observed counts from these latent variables and the specified covariates. RNA counts are modelled using a negative binomial distribution, whereas ADT counts are modelled using a mixture of negative binomial distributions with foreground (FG) and background (BG) components to distinguish true protein expression from technical noise. Adapted with permission from Gayoso *et al.*¹³².

1.2.4 Resolving cellular phenotypes

A central element in analysing single-cell data is the representation of cellular phenotype. Typical analysis workflows use unsupervised clustering to group cells into discrete partitions based on similarity of high-dimensional expression profiles¹²². This approach has thus far enabled data-driven discovery of specific tissue- and disease-associated cell populations that were not adequately resolved using a gating-based phenotyping paradigm traditionally used in flow or mass cytometry.

However, the limitations of this approach become apparent when the biological processes of interest exist at the level of continuous, plastic, and overlapping cellular states (e.g. activation, exhaustion, cytotoxicity, or migration) rather than cell lineages (e.g. CD4⁺ or CD8⁺ T cells). Discrete clustering imposes several constraints that are misaligned with this paradigm. First, clusters may conflate multiple coexisting biological processes into a single label, obscuring the modular structure of cellular phenotype. Second, biological processes that vary smoothly across cells, or span cluster boundaries, are poorly captured by hard partitions, reducing sensitivity to subtle but biologically meaningful state shifts that may underlie dysregulated immune networks and their modulation by therapies. Finally, cluster boundaries are inherently arbitrary and specific to the dataset and choice of parameters, making the resulting biological discoveries difficult to reproduce and validate.

An alternative and increasingly adopted paradigm is to represent cellular phenotype in terms of gene programs, which are weighted combinations of coexpressed genes that reflect underlying biological processes. In this model,

1. Background

individual cells are represented by the simultaneous activity of multiple programs, each contributing additively to the observed transcriptional profile. This representation naturally accommodates cellular plasticity and heterogeneity while aligning with a modern view of systems immunology in which immune cell lineage and function arise from the combined activity of multiple, partially independent transcriptional programs.

Methods to infer gene programs from single-cell data rely on factorisation of the original $N \times G$ expression matrix, $\mathbf{X}$, of counts for $g \in \{1, \dots, G\}$ genes in $i \in \{1, \dots, N\}$ cells into (i) an $N \times K$ matrix, $\mathbf{U}$, of the weights of $k \in \{1, \dots, K\}$ components in each cell, and (ii) a $K \times G$ matrix, $\mathbf{W}$, of the weights of the genes in each component. The simplest factorisation method is PCA, which performs a linear algebraic decomposition of $\mathbf{X}$ into orthogonal factors ordered by the proportion of total variance explained. Numerous other factor analysis methods have been developed for single-cell data^{133,136–141}. Several of these methods are designed for supervised factor analysis based on known sample-level covariates or decomposition based on prior knowledge of biological processes/pathways, rather than unsupervised data-driven discovery of novel latent factors. Among unsupervised methods, the gene weights for the factors derived using many approaches can be positive or negative, which poses challenges for biological interpretability, implying biological processes characterised by negative expression of a gene.

Non-negative matrix factorization (NMF) facilitates biological interpretability of unsupervised factors by finding the best approximate factorisation of $\mathbf{X}$ using only non-negative weights (i.e. $\mathbf{U}, \mathbf{W} \geq 0$):

$$\underset{\mathbf{U}, \mathbf{W}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{U}\mathbf{W}\|_F \quad (1.1)$$

In Equation 1.1, $\|\cdot\|_F$ denotes the Frobenius norm, representing the reconstruction error of the approximate matrix factorisation. This represents an optimisation problem, which yields slightly different solutions across different random initialisations. To account for this, consensus NMF (cNMF)¹³⁷, a widely used implementation of NMF-based factor analysis for single-cell data, repeats the factorisation across multiple initialisations and aggregates the solutions to derive a robust set of consensus factors. A key advantage of this approach is the ability to represent cellular phenotype as a continuously varying measure along multiple partially overlapping axes of biological function. Additionally, the method naturally lends itself to portability across datasets. Once gene programs have been learned in a discovery dataset using cNMF, they can be projected onto independent datasets by fixing the program definitions, $\mathbf{W}^*$, and estimating only the program activities, $\mathbf{U}'$, for each cell in a query expression matrix, $\mathbf{X}'$:

$$\underset{\mathbf{U}'}{\operatorname{argmin}} \|\mathbf{X}' - \mathbf{U}'\mathbf{W}^*\|_F \quad (1.2)$$

Star cell annotator (starCAT)¹⁴² implements this strategy, enabling consistent quantification of program activity across different cohorts and clinical contexts and facilitating direct biological validation of potentially disease- or treatment-linked cell states in independent and orthogonal datasets.

1.2.5 Quantifying therapeutic perturbations

Single-cell omics technologies are increasingly used as high-resolution readouts of experimental perturbations, including *in vitro* stimulation, genetic perturbation, and high-throughput drug screening^{143,144}. This paradigm can be extended to the study of *in vivo* therapeutic effects in human disease, where treatments can be viewed as perturbations to a disease-associated pathological immune network. In this setting, the core analytical challenge is to detect reproducible changes in disease-associated cellular phenotypes induced by therapy while accounting for inter-individual variability and the lack of experimental control over potential confounding factors. At the same time, optimal analysis of treatment-related perturbations should provide a readout at the same resolution as factor analysis-based quantification of phenotype (i.e. per cell) to preserve sensitivity to subtle state shifts.

A range of strategies have been developed to assess condition-associated changes in cell frequencies, known as differential abundance analysis¹²². Several approaches have focused on comparing cluster frequencies between conditions^{145,146}, but these methods inherit the limitations of discrete clustering and are insensitive to within-cluster heterogeneity.

More refined methods attempt to model shifts along the cell state manifold, which can be conceptualised as the continuous high-dimensional landscape of cell states from which the observed data are sampled. A range of methods exist which apply different models to this problem, including diffusion maps¹⁴⁷, VAE latent spaces¹⁴⁸, or k -nearest neighbours (kNN) graphs^{149–151}, with varying strengths and

weaknesses based on assumptions about the nature of the underlying biological processes (e.g. developmental trajectories or plastic functional states).

Among these, graph-based methods represent an intuitive and robust approach for capturing the cell state manifold. Milo¹⁵¹ and manifold enhancement of latent dimensions (MELD)¹⁵⁰ are prominent examples of methods in this class, with both demonstrating strong performance in a recent benchmarking study¹⁵². Milo operates on a graph of overlapping neighbourhoods of cells, effectively partitioning the cellular manifold into small local regions and testing for differences in proportions of cells between conditions within each region. Importantly, this method provides a flexible generalised linear model-based statistical framework that can accommodate complex experimental designs and explicitly accounts for biological replicates. However, its resolution is limited by the size of the neighbourhoods (typically tens to hundreds of cells) which is misaligned with the more granular per-cell readout of phenotype derived from factor analysis.

Conversely, MELD models perturbation effects directly at the level of individual cells. This is achieved by treating the observed cells as a sample from an underlying probability distribution of cellular states across the manifold approximated by a kNN graph. Condition labels (e.g. pre- or post-treatment) are encoded as signals on this graph and smoothed using graph signal processing to produce condition-specific kernel density estimates (KDEs) across the manifold. These KDEs are used to derive a continuous score for each cell that quantifies the relative likelihood of its state under each condition (Figure 1.4). This readout aligns naturally with per-cell gene program activities, enabling perturbations to be quantified as shifts along interpretable biological axes. However, the original implementation of

1. Background

MELD used heuristic score thresholds to identify perturbed cell states, lacking a statistical framework for quantifying uncertainty in density estimates or accounting for biological replicates. Therefore, realising the full potential of MELD in the context of clinical cohort studies requires a more statistically principled approach to distinguish conserved, treatment-associated effects from stochastic or idiosyncratic variation.

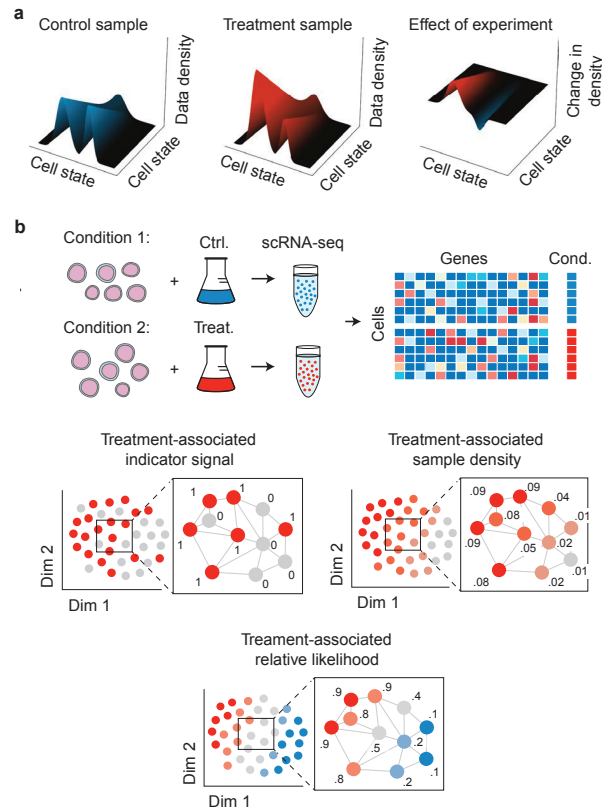


Figure 1.4. Perturbation analysis using MELD. Schematic diagram illustrating the key principles underlying MELD. **a**, Observed cells are treated as samples from an underlying probability distribution over the cell state manifold; the perturbation effect is quantified as the change in density between conditions. **b**, The cell state manifold is approximated using a kNN graph, and each experimental condition is encoded as a signal on this graph; the signal is smoothed and L1-normalised to derive condition-specific density estimates, which are then compared across conditions to obtain the relative likelihood of observing each cell state in each condition. Adapted with permission from Burkhardt DB *et al.*¹⁵⁰.

1.2.6 Interpreting phenotype–perturbation relationships

The ultimate analytical objective in a single-cell study of *in vivo* therapeutic immune perturbations in human disease is to link the effect of treatment to specific cellular phenotypes in a manner that captures potentially complex biological dependencies while remaining interpretable. In this regard, a core assumption underlying the work in this thesis is that the disease-relevant immunological effects of mechanistically broad and heterogeneous MS DMTs are unlikely to operate linearly along a single phenotypic axis. Rather, as is often the case in complex biological systems, these effects likely comprise non-linear and interdependent relationships between multiple aspects of cellular lineage and state. Linear models (e.g. multiple linear regression), which assume additive and independent effects, are therefore poorly suited to this problem and may miss subtle but conserved biological effects. In contrast, tree-based ensemble methods, such as random forest regression¹⁵³, provide a flexible alternative, enabling modelling of non-linear relationships without parametric assumptions, while maintaining robustness to noise and correlated predictors.

In the context of an exploratory (i.e. hypothesis-generating) study design, the goal is not to maximise the predictive power of the phenotype–perturbation model, but to use it as a tool to prioritise the cell states that are most consistently perturbed by treatment. Shapley additive explanations (SHAP)¹⁵⁴ are a recently described feature importance measure that can be applied to tree-based machine learning models to enhance their interpretability. Based on principles of cooperative game theory, this method decomposes model predictions into additive feature-level contributions, thereby providing a quantitative measure of how much each active

1. Background

gene program contributes to the effect of treatment on each cell state, independent of other programs.

Given a trained model, f , the SHAP value, $\phi_{i,k}$, quantifying the additive effect of a given input feature (i.e. gene program), k , to the output prediction (i.e. therapeutic perturbation effect) in a given cell state, i , is defined as follows:

$$\phi_{i,k}(f, \mathbf{x}) = \sum_{S \subseteq \{1, \dots, K\} \setminus \{k\}} \frac{|S|!(K - |S| - 1)!}{K!} [f(\mathbf{x}_{S \cup \{k\}}) - f(\mathbf{x}_S)] \quad (1.3)$$

In Equation 1.3, $\mathbf{x}$ is the vector of input features for cell i , S denotes a subset of the K total features excluding k , and $f(\mathbf{x}_S)$ represents the expected model prediction using the observed values of the features in S , with the remaining features marginalised over the background distribution. Consequently, the right-hand component of the summand represents the marginal contribution of feature k to the model prediction, and the left-hand component represents a weighting factor corresponding to the proportion of feature permutations in which the subset S precedes feature k . This weighted computation of marginal feature contribution over all possible subsets of input features ensures a fair and consistent attribution of importance across correlated features.

Therefore, the combination of random forest regression models and SHAP applied to single-cell resolution quantification of cell state and therapeutic perturbation effect may provide a principled framework for robust modelling of treatment-related changes to cellular phenotype, facilitating the generation of biologically interpretable hypotheses regarding dysregulated immune programs and their modulation by

current treatments, and serving as a conceptual bridge to designing more precise and durable therapeutic strategies.

1.3 Conclusions and thesis aims

Despite significant advances over recent decades, the precise immunobiology of MS, particularly the origin and phenotype of pathogenic cells responsible for driving chronic, smouldering CNS inflammation, remains poorly understood. Consequently, the current landscape of immunomodulatory therapies is non-specific and broadly immunosuppressive. While these DMTs have shown remarkable efficacy in preventing clinical relapses and the accumulation of new radiologically-detectable CNS lesions, they have limited effect on PIRA and therefore ultimately fail to halt the accumulation of neurological disability in the long term. Developing novel therapies to address this crucial unmet need in MS will likely require a deeper and more nuanced understanding of the causal immunophenotype, to enable more precise and durable targeting of the pathogenic axis leading to CNS-compartmentalised inflammation, while sparing bystander immune cells to mitigate side effects of chronic immunosuppression.

Although animal models have heavily influenced the historical view of MS immunopathogenesis and the development of therapeutic approaches over the last three decades, there is increasing recognition of the limitations of these models in precisely recapitulating the complex pathogenic mechanisms of human disease. Consequently, there is an urgent need for studies using samples from pwMS to decipher the disease biology. In this regard, using the paradigm of reverse translation to study the *in vivo* immunological effects of high-efficacy but low-

1. Background

specificity DMTs as perturbational probes offers a key opportunity to disentangle complex dysregulated cellular interactions in the disease context.

High-throughput multimodal single-cell technologies, together with emerging nuanced analysis strategies, represent a powerful tool for reverse translational research in MS. However, significant challenges remain in applying these methods to enable biological discovery in clinical patient cohorts. Specifically, there is an unmet need for an analysis pipeline that can: (i) resolve subtle immune network effects of treatment by moving beyond binning cells into arbitrary discretely partitioned groups and instead modelling shifts on a continuous manifold of cell states, (ii) account for complex non-linear biological interactions and noise inherent in high-dimensional datasets from observational clinical cohorts, (iii) disentangle biologically interpretable functional modules delineating treatment-perturbed cell states to accelerate generation of mechanistic hypotheses and prioritise potential targetable pathways, and (iv) facilitate validation of findings in independent and orthogonal datasets to confirm their robustness and biological relevance.

Accordingly, for the work presented in this thesis, I aimed to:

1. Develop and optimise an analysis strategy to measure subtle treatment-induced immunological shifts at cellular resolution in patient samples
2. Generate a multimodal longitudinal single-cell dataset of peripheral blood mononuclear cells (PBMCs) from pwMS during treatment
3. Use this reverse translational paradigm for unbiased discovery and phenotyping of candidate pathogenic cell types/states
4. Validate hypotheses generated from findings in my discovery cohort in

independent and orthogonal datasets

In Chapter 2, I detail the primary patient cohorts, public datasets and experimental/analytical methods used in this thesis. In particular, I provide details of a novel machine learning pipeline to measure subtle *in vivo* treatment-induced shifts in continuous cellular states, while robustly accounting for inter-donor heterogeneity, stochastic variation and technical noise. In Chapter 3, I apply this analysis framework to a newly-generated longitudinal PBMC CITE-seq dataset during ocrelizumab treatment, using the therapy as a targeted probe to characterise pathogenic B cell interactions in the disease context, and identifying a likely disease-associated and B cell-dependent T cell phenotype. In Chapter 4, I validate the findings in the blood from Chapter 3 across different tissue compartments and disease stages using independent and orthogonal publicly available datasets. In Chapter 5, I apply my novel analysis approach to a publicly available PBMC dataset with samples before and after natalizumab treatment, demonstrating a convergence in the immunological effects of ocrelizumab and natalizumab, which highlights a potential common mechanism of therapeutic benefit across the two drugs. Finally, in Chapter 6, I integrate the findings across Chapter 3, Chapter 4 and Chapter 5 with the existing literature to propose a model of a pathogenic B–T cell axis in MS and discuss the limitations of this model, as well as its implications for future therapeutic advances.

Methods

2.1	Consent and ethical approval	46
2.2	Cohorts	46
2.2.1	Discovery cohort	46
2.2.2	Treatment non-responder	46
2.2.3	Public datasets	47
2.3	Clinical data extraction	47
2.4	Sample processing	48
2.4.1	Sample collection and storage	48
2.4.2	Single-cell library preparation and sequencing	49
2.5	Data analysis	51
2.5.1	Discovery cohort	51
2.5.1.1	Data preprocessing	51
2.5.1.2	Cell type annotation	55
2.5.1.3	Quantifying treatment effect	56
2.5.1.4	Gene program inference and annotation	60
2.5.1.5	Identifying gene programs linked to treatment effect	61
2.5.1.6	Denoising and differential expression analysis	63
2.5.2	Validation cohorts	64
2.5.2.1	Data preprocessing	64
2.5.2.2	Cell type annotation	66
2.5.2.3	Gene program projection	66
2.6	Statistics	66
2.7	Visualisation	68
2.8	Data availability	68
2.9	Code availability	69

This chapter outlines the clinical samples, publicly available datasets, and experimental and analytical methods underpinning the results presented in Chapter 3, Chapter 4, and Chapter 5.

2.1 Consent and ethical approval

Participants were recruited from an internal tissue bank (Oxford Centre for Neuroinflammation, University of Oxford) of pwMS being initiated on ocrelizumab treatment within Oxford University Hospitals (OUH) NHS Foundation Trust since 2019. Informed consent was obtained from all participants to collect blood samples for research purposes. The study was approved by the local clinical research ethics committee (NHS REC 21/SC/0353).

2.2 Cohorts

2.2.1 Discovery cohort

Participants included in the discovery cohort were pwMS being initiated on ocrelizumab treatment with no prior history of DMT. All MS diagnoses were confirmed in accordance with the 2017 revisions of the McDonald criteria by MS neurologists within OUH, with supporting MRI scans reviewed by neuroradiologists. Patients with RRMS were classified as having high relapse frequency (RR_{high}) if they met clinical criteria for rapidly-evolving severe (RES) MS: two disabling relapses within a 12-month period with radiological evidence of disease activity on serial MRI scans (new/gadolinium-enhancing lesions or a significant increase in T2 lesion burden).

2.2.2 Treatment non-responder

An additional participant with RRMS and no prior history of DMT with suboptimal response to ocrelizumab (judged by a multidisciplinary team of MS neurologists within OUH as treatment failure) was identified for inclusion.

2.2.3 Public datasets

Publicly available single-cell sequencing datasets were identified for validation of findings in the peripheral blood from the discovery cohort across different tissue compartments and therapeutic modalities: (i) a published scRNA-seq compendium of CSF and blood mononuclear cells from pwMS, other neurological diseases and healthy controls compiled from multiple studies^{155–159} (available from Synapse under accession syn51730532), (ii) single-nucleus RNA sequencing (snRNA-seq) data from two studies of post-mortem brain tissue of pwMS and healthy control white matter^{66,67} (available from gene expression omnibus (GEO) under accession GSE180759 and from European genome–phenome archive (EGA) under accession EGAS50000000354), and (iii) CITE-seq data from a previously published study of PBMCs from pwMS and healthy controls, including a subset of participants with longitudinal samples before and after treatment with natalizumab¹⁶⁰ (available from GEO under accession GSE144744).

2.3 Clinical data extraction

Demographic and clinical data were extracted for all participants manually from electronic health records, including:

1. Demographic data: sex, date of birth, and body mass index
2. MS history: date of symptom onset, date of diagnosis, disability score (i.e. expanded disability status scale (EDSS) or European database for MS (EDMUS) grading scale), date of last relapse, number of relapses in the 12- and 24-month intervals before ocrelizumab, date of baseline MRI scan,

and baseline MRI activity

3. Medical history: comorbidities and medications

4. Treatment record: dates of ocrelizumab infusions, number and date of any relapses/pseudorelapses after commencing ocrelizumab, follow-up MRI activity, and disability progression after commencing ocrelizumab

2.4 Sample processing

2.4.1 Sample collection and storage

Samples were collected by research nurses in the MS Clinical Trials Unit, University of Oxford and processed for storage by Subita Balaram Kuttikkatte (Oxford Centre for Neuroinflammation, University of Oxford). Blood samples were collected from all participants prior to commencing treatment (baseline), two weeks after the first ocrelizumab infusion (early), and immediately prior to each subsequent infusion (late). No pre-treatment sample was available for the treatment non-responder participant. The dosing schedule for ocrelizumab comprised an initiation phase consisting of two 300 mg infusions given two weeks apart, followed by extended interval 600 mg maintenance infusions, with the timing of the next infusion determined on an individual basis by detectable CD19⁺ B cell repopulation (≥ 10 cells/ μ L) on monthly blood lymphocyte subset monitoring by flow cytometry. PBMCs were isolated from fresh whole blood samples using density gradient centrifugation and cryopreserved in liquid nitrogen.

2.4.2 Single-cell library preparation and sequencing

Single-cell library preparation was performed by Brian Cutler (Oxford Centre for Neuroinflammation, University of Oxford) and Kate Attfield (Oxford Centre for Neuroinflammation, University of Oxford) (Figure 2.1). In brief, PBMCs were thawed at 37°C, counted, assessed for viability and processed using Chromium Next GEM Single Cell 3' Reagent Kit v3.1 (discovery cohort) or Chromium GEM-X Single Cell 3' Reagent Kit v4 (treatment non-responder) (10x Genomics). Cells from each sample were first labelled with CITE-seq antibodies from TotalSeq-A Human Universal Cocktail v1.0 (BioLegend) or TotalSeq-B Human Universal Cocktail v1.0 (BioLegend), and corresponding TotalSeq-A or TotalSeq-B anti-human hashtag antibodies (BioLegend). Longitudinal samples from each donor were then pooled and 16,000–32,000 cells were loaded per well of a Chromium microfluidic chip, using 2–3 wells per donor; this experimental design was chosen to maximise cell recovery while mitigating against within-donor technical batch effects. Library preparation was performed according to the manufacturer instructions, with a modification in Step 2.2 for samples processed using the TotalSeq-A antibodies, as described previously¹⁶⁰. Complementary DNA (cDNA) library concentrations and sizes were measured by Qubit (Thermo Fisher Scientific) and TapeStation (Agilent Technologies), respectively.

Processed cDNA libraries for RNA, ADT and hashtag oligonucleotide (HTO) were pooled and sequenced across multiple flow cells using either the Illumina NextSeq 500 platform (Weatherall Institute of Molecular Medicine, University of Oxford) or NovaSeq X Plus platform (Novogene UK), using read configurations for

2. Methods

each library type as recommended by 10x Genomics and a minimum sequencing depth of 20,000 reads per cell for RNA libraries, 10,000 reads per cell for ADT libraries, and 2,000 reads per cell for HTO libraries.

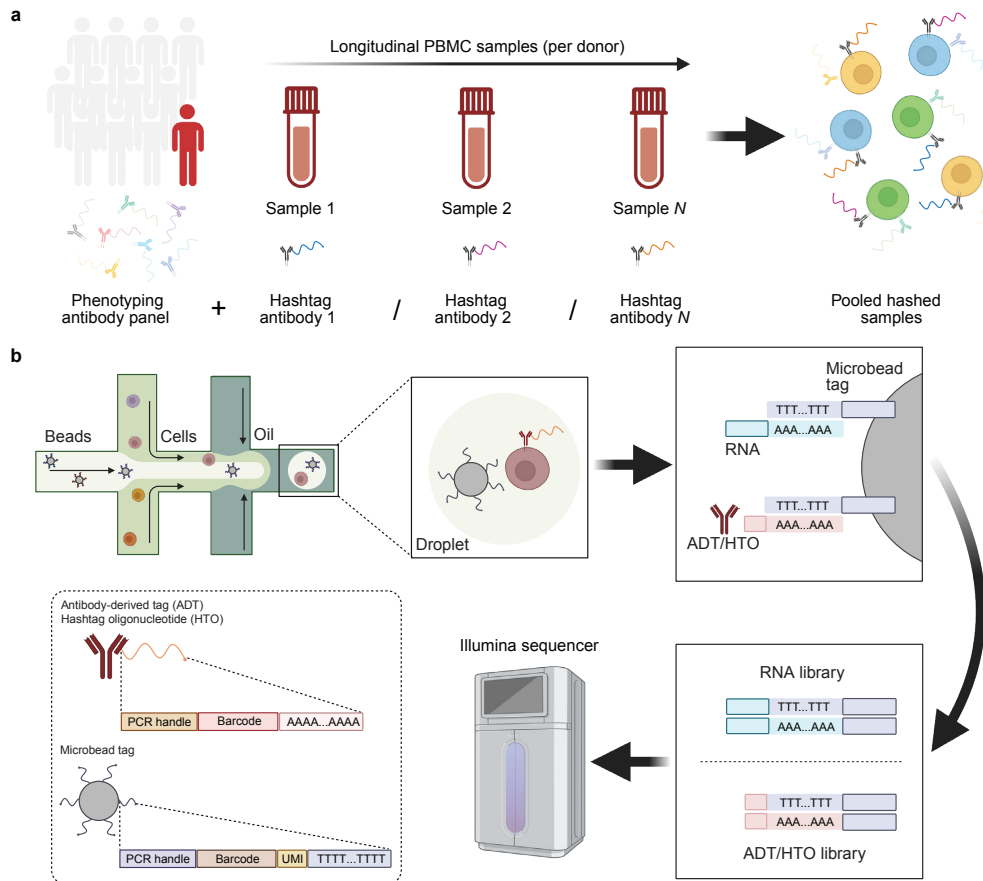


Figure 2.1. Sample processing workflow. Schematic diagram illustrating the sample processing workflow for the discovery cohort dataset. **a**, Longitudinal PBMC samples for each donor were labelled with a cell surface protein phenotyping antibody panel, along with a unique hashtag antibody per sample. Cells from all samples from the same donor were then pooled together and processed across 2–3 wells of a single microfluidic chip to mitigate technical sources of nuisance variation. **b**, Microfluidic droplet-based cell isolation and library preparation was performed using commercially available kits. RNA, ADT and HTO libraries were pooled and sequenced using Illumina sequencing platforms. The workflow illustrated here demonstrates ADT / HTO library preparation using TotalSeq-A antibodies, which feature antibody-conjugated synthetic oligonucleotides with poly(A) tails and are captured alongside messenger RNA (mRNA) molecules by the poly(dT) sequences on gel bed primers; in contrast, TotalSeq-B antibodies feature a specific capture sequence that hybridises to a complementary sequence on gel bead primers. Created with BioRender.com.

2.5 Data analysis

2.5.1 Discovery cohort

2.5.1.1 Data preprocessing

Data preprocessing was performed using a series of custom Snakemake¹⁶¹ pipelines for multisample multimodal single-cell experiments to execute the steps detailed in this section (Figure 2.2). Raw data for each sequencing run was demultiplexed to FASTQ files for each library using `bcl2fastq` (Illumina). For each sample, FASTQ files for each library type across different sequencing runs were aggregated. To generate raw count matrices, STARsolo¹⁶² was used to align RNA reads to a human (GRCh38) reference transcriptome (using option `--soloFeatures GeneFull_Ex50pAS`), and BarCounter¹⁶³ was used to align ADT and HTO reads to the oligonucleotide sequences for the antibody-conjugated tags used in the experiments.

For each sample, raw count matrices for all feature types were concatenated. The EmptyDrops¹⁶⁴ algorithm was used to call cell-containing barcodes based on the RNA modality only (using the following parameters: minimum UMI count 500, maximum number of cell-containing droplets 20,000, library size-ranked droplet indices for ambient estimation 45,000–90,000). CellBender¹⁶⁵ was applied to model ambient contamination, for both RNA and ADT modalities (excluding isotype control antibodies), from empty droplet barcodes (using options `--learning-rate 0.00005 --fpr 0`) and this was subtracted from the raw count matrix. Hashed samples were demultiplexed with demuxmix¹⁶⁶ (using default parameters); this step also allowed identification of inter-sample doublets as barcodes assigned to more than one HTO. Simulation-based prediction of additional heterotypic intra-

2. Methods

sample doublets was performed with `scDblFinder`¹⁶⁷ (using default parameters and providing the list of the known intra-sample doublets from the previous step using the `knownDoublets` parameter). The union of cell barcodes identified as doublets using both approaches were excluded.

Cell-level QC was performed separately for each sample using automated outlier-based (≥ 3 median absolute deviations) thresholds as implemented in `scater`¹⁶⁸. For RNA data, cells were excluded for low library size (total unique molecular identifier (UMI) count), low library complexity (number of genes with non-zero UMI count) or high percentage mitochondrial gene counts. For ADT data, cells were excluded for high isotype control antibody count, likely reflecting droplets containing antibody aggregates.

Transcript Ensembl IDs from the RNA count matrices were used to extract canonical gene symbols and gene type annotations with `ensemblDb`¹⁶⁹ (using Ensembl annotation database v86 for *Homo sapiens*). Transcripts annotated as types other than ‘protein_coding’, ‘lincRNA’, ‘IG_C_gene’, or ‘TR_C_gene’ were removed. Isotype control antibody tags were removed from ADT count matrices.

Further processing was performed using `Seurat`¹⁷⁰. First, post-QC filtered count matrices from all samples in the experiment were concatenated to construct a combined `Seurat` object with a separate assay for each modality. RNA counts were library size normalised per cell, scaled to counts per 10,000 and log-transformed with `NormalizeData` (using `method = 'LogNormalize'` and `scale.factor = 1e4`). Cell cycle scores were computed using `CellCycleScoring` and built-in lists of cell cycle-related genes. The top 5,000 highly variable features across samples were selected with `FindVariableFeatures` (using `selection.method`

= 'vst'). To limit the influence of uninteresting variation in highly expressed housekeeping genes on downstream clustering and integration steps, ribosomal and mitochondrial genes were removed from the list of highly variable features. ADT counts were separately normalised across cells from each sample using a centered log ratio (CLR) transformation with `NormalizeData` (using `method = 'CLR'` and `margin = 2`). To avoid problems with missing data from antibodies unique to each of the two panels used in the experiment, the intersection of antibody tags present in both panels were selected as the highly variable features. Normalised expression values for the highly variable features in both modalities were mean-centred and scaled to unit variance using `ScaleData`; for the RNA modality, this step also included regressing out percentage mitochondrial gene counts and cell cycle scores per cell (using the `vars.to.regress` parameter). The top 50 principal components for each modality were computed from the scaled expression values using `RunPCA`. Integration was performed with `IntegrateLayers` (using a custom wrapper function for the Harmony¹²⁴ algorithm as the `method` parameter and specifying sample, i.e. the pool of hashed PBMC samples loaded in a single Chromium chip well, donor and assay version as covariates). The resulting integrated embeddings for each modality were used in all downstream analysis steps requiring a low dimensional data representation as input, including computing uniform manifold approximation and projection (UMAP) embeddings and constructing kNN graphs.

2. Methods

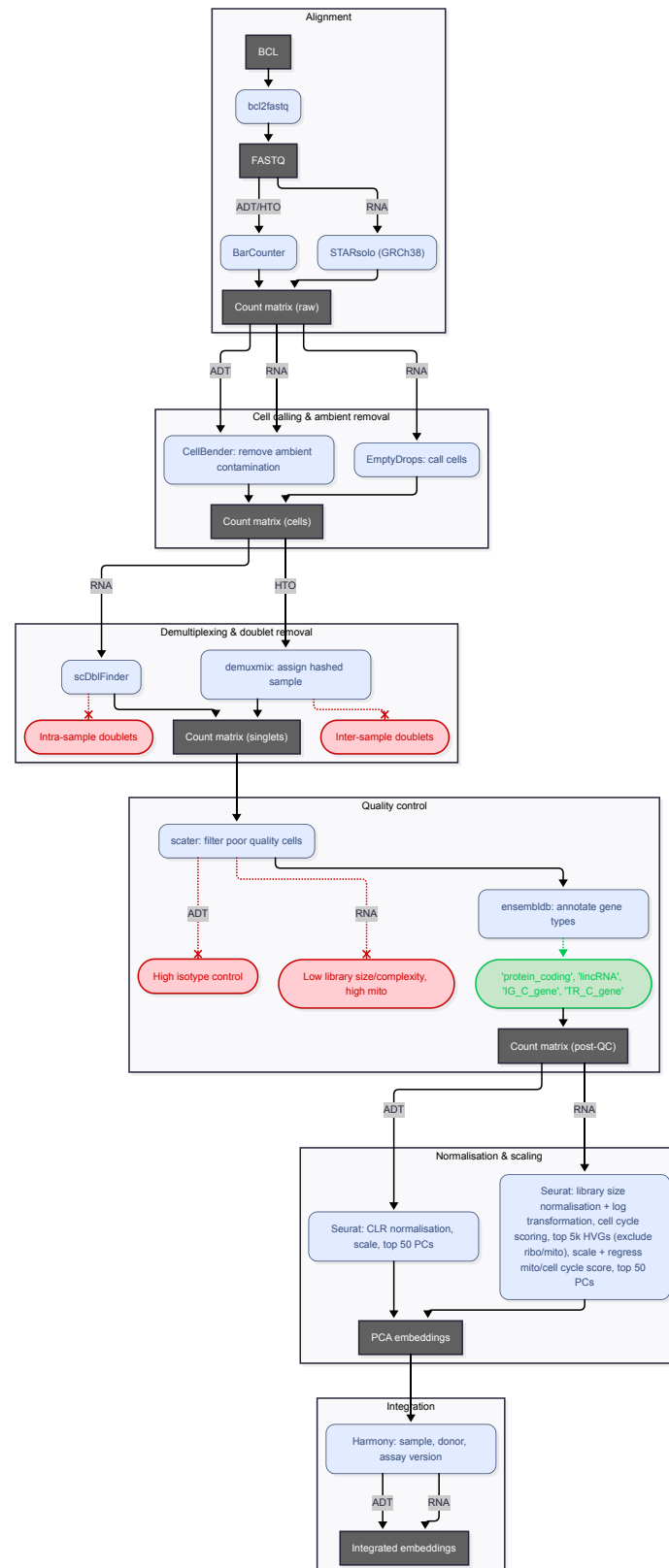


Figure 2.2. CITE-seq data preprocessing pipeline. Flowchart summarising preprocessing steps used for CITE-seq data, including sequencing alignment, cell calling, ambient removal, demultiplexing, doublet detection, quality control, normalisation, feature selection, scaling, dimensionality reduction, and integration.

2.5.1.2 Cell type annotation

A semi-supervised consensus approach was used to derive robust cell type annotations. First, using the RNA modality, reference-based automated cell type annotations were obtained using three different methods and reference datasets: (i) Seurat mutual nearest neighbours-based reference mapping to the human PBMC reference atlas from Azimuth¹³⁴, (ii) CellTypist¹⁷¹ logistic regression-based reference mapping to the Allen Institute for Immunology Human Immune Health Atlas¹⁷², and (iii) SingleR¹⁷³ Spearman correlation-based reference mapping to a bulk RNA sequencing (RNA-seq) reference derived from fluorescence-activated cell sorting (FACS)-sorted immune cell populations¹⁷⁴. Second, unsupervised clusters were obtained separately for each modality (RNA and ADT) using Scanpy¹⁷⁵, by first computing a kNN graph with `scanpy.pp.neighbors` (using cosine distance and $k = 30$), then performing graph-based clustering with `scanpy.tl.leiden` (using `resolution` parameter values in the integer range $[1, 9]$). Next, `scTriangulate`¹⁷⁶ was used to derive automated consensus clusters (using default parameters and providing unsupervised cluster labels with resolutions 2, 5, and 8 from each modality, and the highest granularity annotations from each reference mapping using the `query` parameter). Subsequently, the consensus cluster labels were manually harmonised into consistent cell type labels at 3 levels of granularity. Lastly, subclustering was performed for each cell type label to identify any remaining misclassified cells and artefactual clusters by manual inspection of marker genes and proteins.

The final cell type labels were added to the post-QC filtered combined multimodal

2. Methods

Seurat object and marker genes were computed for coarse (level 2) cell types and fine (level 3) cell types (within each parent coarse cell type) with `FindAllMarkers` (using `method = 'wilcox'` and `only.pos = TRUE`). Subsequently, the main (level 1) cell type labels were used to exclude cells not annotated as B cells, dendritic cells, innate lymphoid cells (ILCs), monocytes, NK cells or T cells from all downstream analyses.

2.5.1.3 Quantifying treatment effect

MELD¹⁵⁰ was used to quantify treatment effect at cellular resolution (Figure 2.3a, see Section 1.2.5). This algorithm requires two user-specified hyperparameters: (i) the number of neighbours k used to construct the kNN graph representing the cell state space manifold, and (ii) a smoothing parameter β that controls the scale of the KDE for the graph signal encoding treatment condition. These hyperparameters were selected through empirical optimisation using simulated ground truth probability density functions (PDFs), testing values of $k \in \{5, 10, \dots, 50\}$ and $\beta \in \{10, 20, \dots, 200\}$. The selected optimal values minimised the mean squared error (MSE) between MELD-derived likelihood estimates and the known ground truth (Figure 2.4).

Cells from different time points were assigned condition labels: (i) baseline (pre-treatment), (ii) early (two weeks after first dose), or (iii) late (all subsequent time points). To account for heterogeneity between biological replicates, MELD was run separately for each donor n . The RNA modality integrated embedding was used to construct the kNN graph, and the treatment condition-related graph signals from each donor were smoothed across the entire graph, yielding an interpolated

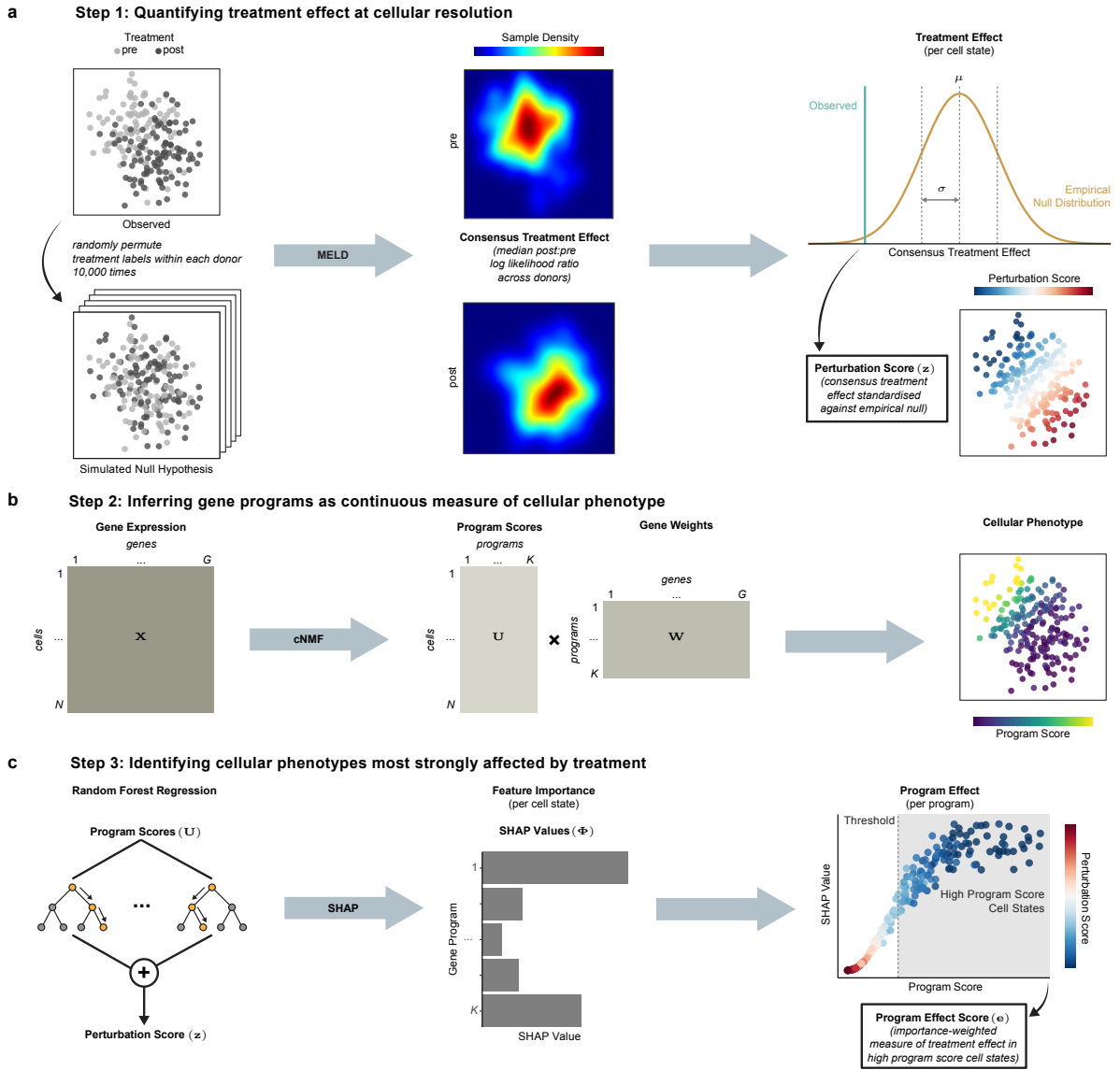


Figure 2.3. Overview of analysis strategy for characterising shifts in cellular phenotypes due to treatment. **a**, MELD is used to estimate the empirical probability density of cells from each pre- and post-treatment time point over the joint low dimensional manifold of cell state space. Estimated likelihoods for each cell state at each time point are computed separately based on the sampling distribution of cells from each donor and used to compute relative likelihoods compared to the pre-treatment baseline; the median across donors is used as the consensus treatment-induced perturbation signal. This signal is standardised against a donor-stratified permutation-based empirical null distribution to compute statistically robust, denoised *perturbation scores*. **b**, cNMF is used to infer gene programs (representing cell identity or biological function) within major immune cell populations, with *program scores* assigned to each cell state for each program as a continuous measure of cellular phenotype. **c**, Non-linear relationships between continuous program scores and perturbation scores across cell states are modelled using random forest regression, and SHAP-based feature importances obtained from the trained models are used to calculate *program effect scores*, representing an importance-weighted measure of treatment effect on the cellular phenotype represented by each gene program.

2. Methods

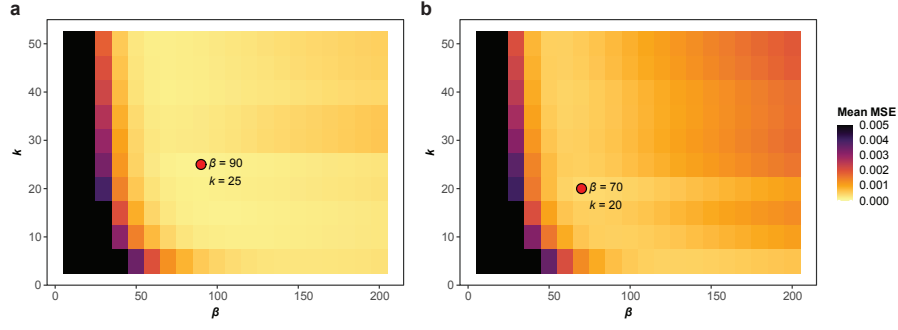


Figure 2.4. Empirical MELD hyperparameter optimisation. Tile plots showing empirical MELD hyperparameter optimisation for the discovery (a) and natalizumab validation (b) datasets, using a grid search over a range of values of β (x -axis) and k (y -axis). Optimal hyperparameters (red circles) were selected based on minimisation of the MSE between the algorithm-estimated KDE and simulated ground truth PDFs over the kNN graph, using 20 random simulations for each hyperparameter combination.

donor-specific estimated conditional likelihood, $l_{i,n}$, for each cell state i . Within each donor, the likelihoods for the early/late (post-treatment) conditions were normalised by the likelihoods for the baseline (pre-treatment) condition and log-transformed to obtain log likelihood ratios relative to pre-treatment; subsequently, for each cell state, the median log likelihood ratio across donors, $\tilde{\lambda}_i$, was calculated:

$$\tilde{\lambda}_i = \text{median}_n \left(\log \left(\frac{l_{i,n}^{\text{post}}}{l_{i,n}^{\text{pre}}} \right) \right) \quad (2.1)$$

This metric represents a consensus treatment-related perturbation signal, emphasising cell states that consistently shift in a coordinated manner in response to treatment across the majority of the cohort, while penalising noisy biological fluctuations or true outlier effects observed in only a few individuals.

A permutation test was used to quantify uncertainty in the estimates of treatment effect. Specifically, 10,000 donor-stratified random permutations of the treatment condition labels were performed to simulate the null hypothesis that there was no true difference in the distribution of cell states between conditions. MELD was

re-run on each permuted dataset as described above, and median log likelihood ratios across donors were calculated as per Equation 2.1 to produce an empirical null vector per cell state, $\tilde{\lambda}_i^*$. For each cell state, the observed median log likelihood ratio for each post-treatment time point was standardised against its null distribution and defined as the *perturbation score*, z_i :

$$z_i = \frac{\tilde{\lambda}_i - \text{mean}(\tilde{\lambda}_i^*)}{\text{std}(\tilde{\lambda}_i^*)} \quad (2.2)$$

This metric represents a statistically robust, denoised measure of consensus treatment effect per cell state, accounting for both biological (stochastic or idiosyncratic inter-donor heterogeneity) and technical (sampling bias and manifold topology-related) sources of noise.

Using the perturbation scores as an empirical Z -statistic, two-tailed p -values for each cell state were calculated from the standard normal distribution; the validity of the normality assumption was confirmed by analysis of skewness and kurtosis for the permutation-based null distributions (Figure 2.5). However, as the conditional likelihoods used to calculate the perturbation scores were obtained from graph signal smoothing, the underlying hypothesis tests are not independent, leading to positive correlation for p -values between cell states within local neighbourhoods. To mitigate this spatial autocorrelation, a weighted Benjamini–Hochberg procedure¹⁷⁷ was used to control the false discovery rate (FDR). Specifically, the Euclidean distance to the k^{th} nearest neighbour for each cell state was used as a proxy for local neighbourhood sparsity (inverse of density). These distances were normalised to have a mean of 1 and used to weight the p -values, assigning lower weight to

2. Methods

dense regions of the cell state manifold, before computing adjusted p -values.

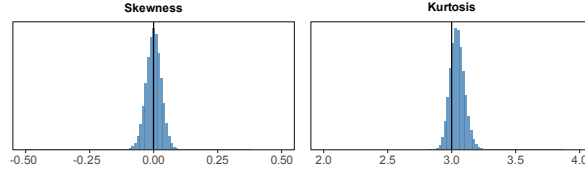


Figure 2.5. Validation of normality assumption for perturbation empirical null distributions. Histograms showing global distribution of skewness and kurtosis for donor-stratified permutation-based empirical null distributions per cell state per condition; theoretical values for a normal distribution are indicated with black vertical lines.

2.5.1.4 Gene program inference and annotation

cNMF¹³⁷ was used to infer gene programs in the discovery cohort dataset which were used as a continuous measure of cellular phenotype (Figure 2.3b, see Section 1.2.4). The full dataset was subset to each main (level 1) cell type label and the following steps were run separately for each cell type. First, the **Preprocess** pipeline was used to remove mitochondrial genes from the raw count matrix, perform library size normalisation and counts per million scaling, select the top 2,000 highly variable features and batch-correct the raw counts for these genes with the built-in modified implementation of the Harmony¹²⁴ algorithm (using the same covariates used for integration). The outputs of this module were then used as inputs to the cNMF pipeline. This was run using total number of components K in the integer range $[5, 40]$ for 20 iterations each. The final number of components for each cell type was chosen based on inspection of plots of stability and reconstruction error (Figure 2.6). Finally, the consensus gene program spectra (Z -scored weights of genes in each program) and program scores (L1-normalised weights of each program in each cell) were then computed for the chosen K , after filtering out outlier spectra based on inspection of the histogram of mean Euclidean distance to their nearest neighbours

among the other spectra.

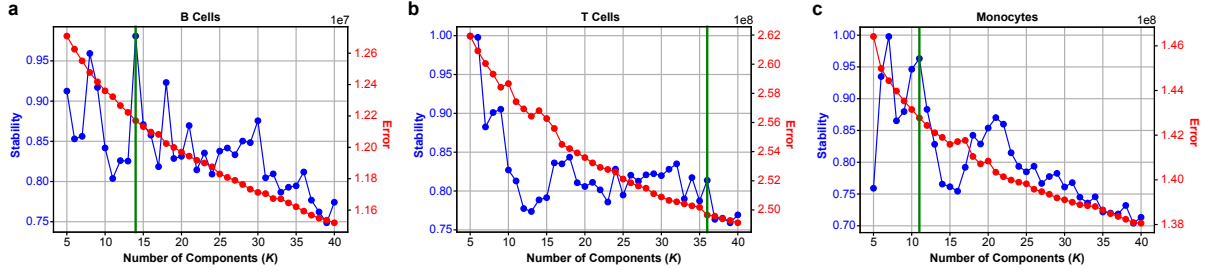


Figure 2.6. cNMF K selection plots. Scatter/line plots showing solution stability (silhouette distance; blue) and reconstruction error (Frobenius distance; red) across separate runs of cNMF using a varying number of components (K ; x -axis) for B cells (a), T cells (b), and monocytes (c). The final number of components selected for each cell type is indicated with green vertical lines.

Pathway enrichment analysis was performed for the top 100 highest weighted genes for each program using gene ontology (GO) gene sets (collection 5, subcollection GO) obtained from the Molecular Signatures Database (MSigDB v.2025.1.Hs). All genes in the full count matrix used for running the cNMF pipeline were used as the set of background genes, and query gene sets were filtered to exclude those containing fewer than 20 or more than 200 genes. Overrepresentation analysis was performed in GSEAPy¹⁷⁸ with `gseapy.enrichr`.

Gene programs were manually annotated based on inspection of the following: (i) top 100 highest weighted genes, (ii) distribution of program scores across fine (level 3) cell types, and (iii) enriched gene sets.

2.5.1.5 Identifying gene programs linked to treatment effect

To identify which gene programs were most strongly linked to treatment-related changes in cell states, random forest regression models were trained to predict per-cell perturbation scores, $\mathbf{z}$, from gene program scores, $\mathbf{U}$ (Figure 2.3c). Random forests were chosen because they are able to capture complex non-linear relationships

2. Methods

and are robust to correlations among predictors (multicollinearity), which are important properties given that gene program scores exhibit heavily skewed and bounded distributions with threshold effects, and often co-vary across related cell states.

As the goal was to estimate relative feature importance rather than out-of-sample prediction accuracy, models were trained using all available cells within each main (level 1) cell type, without data splitting. Models were implemented in scikit-learn¹⁷⁹ with `sklearn.ensemble.RandomForestRegressor` (using `n_estimators = 500`, `max_depth = 10`, `min_samples_leaf = 100`, `max_features = 'sqrt'`, and `bootstrap = True`).

Feature importance measures were derived using SHAP¹⁵⁴, which decompose a model's prediction into additive contributions from each input feature. For each set of cell type-specific gene programs, a matrix of SHAP values, Φ , was computed with `shap.TreeExplainer` (using `feature_perturbation = 'interventional'`). For each cell state i , the SHAP value for program k , $\phi_{i,k}$, quantifies how much the program score contributes to its predicted perturbation score, independent of other programs (see Section 1.2.6). Scatter plots of program scores against SHAP value, coloured by perturbation score, were used to visualise these relationships and programs showing high SHAP values in cells with high program scores and strong perturbation scores were interpreted as being strongly associated with treatment effects. For each program, the set of cells with program score > 0.2 , I_k , was considered to exhibit high program activity:

$$I_k = \{i \mid u_{i,k} > 0.2\} \quad (2.3)$$

Using these cells, two summary measures were computed per program: (i) the mean absolute SHAP value, reflecting program importance, and (ii) the mean perturbation score, reflecting treatment effect. The product of these measures was defined as the *program effect score*, e_k , an importance-weighted measure of treatment-related perturbation in cells with high program activity:

$$e_k = \text{mean} \left(\left| \phi_{I_k, k} \right| \right) \cdot \text{mean} \left(\mathbf{z}_{I_k} \right) \quad (2.4)$$

2.5.1.6 Denoising and differential expression analysis

totalVI¹³² was used to denoise both RNA and ADT data and perform differential expression analyses. For computational efficiency, only the highly variable features for the RNA modality were used for model training. The model was trained with donor ID as the batch variable and sample, assay version, percentage mitochondrial gene counts and cell cycle scores as covariates (using `n_latent = 50` and `max_epochs = 200`). Denoised, batch-corrected and library size-normalised protein expression values were obtained from the trained model using `scvi.model.TOTALVI.get_normalized_expression` (using `library_size = 1e4`, `include_protein_background = False`, `n_samples = 50`, `return_mean = True`, and providing a reference donor to the `transform_batch` parameter to correct for uninteresting donor-specific variation) then log-transformed. These values were used for gating the CD20^{dim} T cell population and for all plots visualising gene and protein expression levels. Differential expression analysis between groups of cells was performed with `scvi.model.TOTALVI.differential_expression` (using `mode = 'change'`, `delta = 0.1`, `use_protein_background = False`,

`batch_correction = True`, and `fdr_target = 0.05`).

2.5.2 Validation cohorts

2.5.2.1 Data preprocessing

For samples from the treatment non-responder, data preprocessing steps were performed starting from raw sequencing data as described for the discovery cohort samples (see Section 2.5.1.1). To account for higher throughput in the newer Chromium GEM-X Single Cell assay used for these samples, the EmptyDrops algorithm for calling cell-containing barcodes was run using the following modified parameters: maximum number of cell-containing droplets 35,000 and library size-ranked droplet indices for ambient estimation 80,000–160,000.

For external datasets, post-QC filtered count matrices and cell-level metadata tables were downloaded from public data repositories and used to construct AnnData objects. Gene names were harmonised across datasets with `ensembladb` (using Ensembl annotation database version 86 for *Homo sapiens*) as follows: (i) for transcript Ensembl IDs, canonical gene symbols and gene type annotations were extracted, (ii) for gene symbols, Ensembl IDs and gene type annotations were extracted, (iii) for any non-canonical gene symbols not found in the Ensembl database, `AnnotationDbi` (using the *org.Hs.eg.db* meta-database) was used to try to identify corresponding Ensembl IDs by querying known gene symbol aliases and these were then used to extract canonical gene symbols and gene type annotations from the Ensembl database, (iv) genes annotated as types other than ‘protein_coding’, ‘lincRNA’, ‘IG_C_gene’, or ‘TR_C_gene’ were removed, and (v) all gene names in the original count matrices were converted to canonical gene symbols and counts

for each unique gene were aggregated.

Further processing was performed using Scanpy. Counts were library size normalised per cell and scaled to counts per 10,000 with `scanpy.pp.normalize_total` (using `target_sum = 1e4`), then log-transformed with `scanpy.pp.log1p`. Cell cycle scores were computed using `scanpy.tl.score_genes_cell_cycle` and Seurat’s built-in lists of cell cycle-related genes. The top 5,000 highly variable features across samples were selected with `scanpy.pp.highly_variable_genes` (using default parameters and the appropriate `batch_key` for each dataset). Ribosomal and mitochondrial genes were removed from the list of highly variable features. Normalised expression values for the highly variable features were mean-centred and scaled to unit variance using `scanpy.pp.scale` (using `max_value = 10`), after regressing out percentage mitochondrial gene counts (not for snRNA-seq datasets) and cell cycle scores per cell with `scanpy.pp.regress_out`. The top 50 principal components for each modality were computed from the scaled expression values using `scanpy.pp.pca`. Integration was performed with `scanpy.external.pp.harmony_integrate` (specifying sample/donor and, where relevant, original study or assay version as covariates). The resulting integrated embeddings were used in all downstream analysis steps requiring a low dimensional data representation as input, including computing UMAP embeddings and constructing kNN graphs.

For snRNA-seq datasets from post-mortem brain tissue, only cells annotated as lymphocytes in the original studies (labels ‘lymphocytes’, ‘BC’, or ‘TC’) were included in subsequent analysis.

2.5.2.2 Cell type annotation

Preprocessed AnnData objects were converted to Seurat objects and cell type annotations were transferred from the discovery cohort dataset to all validation datasets using Seurat mutual nearest neighbour-based reference mapping. Further subclustering was performed within automatically transferred cell type labels to identify any misclassified cells and artefactual clusters by manual inspection of marker genes. For the datasets containing CSF mononuclear cells, an additional cell type label was added for macrophages/microglia-like cells which were not present in the PBMC reference dataset.

The final cell type labels were added to the Seurat objects and main (level 1) cell type labels were used to exclude cells not annotated as B cells, dendritic cells, ILCs, monocytes, NK cells, or T cells from all downstream analyses.

2.5.2.3 Gene program projection

Precomputed gene programs inferred from the discovery dataset were projected into the validation datasets with starCAT¹⁴². starCAT was run separately for each main (level 1) cell type using the counts per million-scaled variance-normalised reference spectra for gene programs obtained from running cNMF in the discovery dataset to compute program scores in validation dataset cells.

2.6 Statistics

No statistical methods were used to predetermine the number of participants in the discovery cohort. Participants were selected based on meeting the clinical inclusion criteria and availability of samples. The final number of participants in the cohort

was comparable to previously published studies using single-cell methods to study treatment effects in patient samples^{160,180,181}.

Statistical tests for cell type marker gene identification, perturbation scores, gene set enrichment analysis for gene programs and gene/protein differential expression analysis were performed as described in Section 2.5.1.2, Section 2.5.1.3, Section 2.5.1.4, and Section 2.5.1.6, respectively.

All other statistical tests were performed using R. Comparisons of program scores across samples with repeated measures from the same biological replicates (i.e. donors) were performed using linear mixed effects models, as implemented in lme4¹⁸², with Satterthwaite estimates of degrees of freedom, as implemented in lmerTest¹⁸³. For comparisons across disease type (RR_{high} , RR_{low} , or PP) and condition (baseline, early, or late) in the discovery dataset, the model formula used was `program_score ~ disease_type * condition + age + sex + (1 | donor)`. For comparisons across disease group (MS / CIS or healthy control) and tissue (PBMC or CSF) in the CSF validation dataset, the model formula used was `program_score ~ disease_group * tissue + age + sex + (1 | donor)`. For comparisons across tissue type (perilesional normal-appearing white matter (NAWM), chronic active lesion rim, chronic inactive lesion rim, lesion core, or healthy control white matter) in the post-mortem brain tissue validation dataset, the model formula used was `program_score ~ tissue_type + age + sex + (1 | donor)`. For comparisons across condition (pre- or post-treatment) in the natalizumab validation dataset, the model formula used was `program_score ~ condition + age + sex + (1 | donor)`. Marginal means, confidence intervals, and p -values for contrasts of interest from the linear models were computed using emmeans¹⁸⁴.

Multiple testing correction was performed for FDR control using the Benjamini–Hochberg method and adjusted p -values are reported in the text and figures. For statistical significance, an FDR threshold of 0.05 was selected for all hypothesis tests, except in tests of significantly perturbed cell states following treatment, where a more conservative threshold of 10^{-5} was selected due to the large number of hypothesis tests involved.

2.7 Visualisation

All data figures were created using R or Python. Scatter plots (including UMAP projections and dot plots), bar plots and box plots were generated with ggplot2 and extensions ggpubr, patchwork, and cowplot. Heatmaps were generated with ComplexHeatmap. Plots showing stability and reconstruction error for different values of K in cNMF runs were generated with Matplotlib. Diagrams illustrating study designs were created using BioRender.com.

2.8 Data availability

Raw sequencing data (FASTQ files) generated for this thesis cannot be made publicly available to protect the privacy of human participants. These data have been deposited in EGA (accession EGAS50000002112), and are available under controlled access for research use.

Processed data (count matrices and associated metadata) necessary to reproduce the analysis have been deposited in GEO (accession GSE316688) and are publicly available.

2.9 Code availability

All code and software environment specifications required to reproduce the analysis are publicly available on GitHub (<https://github.com/redwanfarooq/thesis>). Key software packages and versions used are listed in Table B.1.

Characterising peripheral immune perturbations following B cell depletion

3.1	Introduction	72
3.1.1	Anti-CD20 therapies in MS	72
3.1.2	Immunological effects of anti-CD20 therapies	73
3.1.3	Hypothesis	75
3.2	Results	76
3.2.1	Measuring dynamic longitudinal changes in peripheral immune cells following ocrelizumab	76
3.2.2	Early repopulating B cells are immature and antigen-inexperienced	82
3.2.3	Delayed depletion of a proinflammatory, cytotoxic and tissue-homing (T _c 1-TH) memory T cell phenotype	88
3.2.4	T _c 1-TH signature correlates with clinical disease activity	99
3.2.5	Heterogeneous enrichment of a proinflammatory (Classical-Inflammatory) monocyte phenotype	102
3.3	Discussion and conclusions	108

This chapter presents the results of a multimodal longitudinal observational study characterising *in vivo* shifts in the peripheral immune landscape of pwMS following B cell depletion at single-cell resolution. Using this reverse translational paradigm, it provides new insights into dysregulated immune interactions in MS and identifies a disease-associated T cell state that is selectively depleted following anti-CD20 therapy.

3.1 Introduction

3.1.1 Anti-CD20 therapies in MS

B cell depletion with anti-CD20 therapies has revolutionised the management of MS in recent years. In addition to high efficacy in preventing relapses and new MRI lesions⁹², one of the pivotal trials demonstrated that they modestly slow disability progression in patients with PPMS⁹³, establishing them as the first class of DMTs with proven clinical benefit in modulating PIRA. However, this benefit is incomplete, and as a consequence of robust suppression of relapse-associated worsening (RAW), PIRA has emerged as the major driver of disability progression in anti-CD20-treated pwMS at all stages of disease^{8,185}.

Despite the practice-changing impact of these therapies, a coherent mechanism of B cell pathogenicity in MS remains elusive. The original trials of anti-CD20 therapies in MS were motivated by the observation that intrathecal immunoglobulin synthesis, detectable as CSF-restricted oligoclonal bands, is ubiquitous in MS⁹¹. However, the therapeutic benefit of these drugs does not seem to be dependent on reducing oligoclonal bands and studies of the antigen-specificity of the CSF immunoglobulin repertoire in MS have revealed heterogeneous donor-specific targets comprising predominantly intracellular proteins^{186,187}, suggesting these antibodies may arise as a secondary response to tissue damage within the CNS rather than as a primary pathogenic driver.

Most current theories regarding the role of B cells in MS focus on their non-humoral functions, including antigen presentation and cytokine production, which may modulate the state of interacting immune cells to give rise to the causal

immunophenotype in MS⁴⁹. The increasing recognition that the efficacy of B cell depletion is dependent on persistent suppression of memory B cells⁹⁶ may additionally implicate a mechanism involving EBV, which establishes chronic latent infection in this cell population and represents the strongest environmental risk factor for MS susceptibility¹⁷. Consequently, studying the broader immunological consequences of B cell depletion has garnered much interest as a promising avenue to better understand immune dysregulation and the peripherally-modifiable drivers of disability progression in MS.

3.1.2 Immunological effects of anti-CD20 therapies

Numerous studies have extensively investigated the effects of anti-CD20 therapies on the B cell compartment in MS. Following administration, near-complete depletion of the circulating B cell repertoire, determined by quantification of CD19⁺ cells in blood using flow cytometry, is typically observed within 1–2 weeks⁴⁹. Longitudinal studies following treatment suggest that the timing of detectable CD19⁺ B cell repopulation is highly variable between individuals, with most showing initial re-emergence between 6–18 months¹⁸⁸.

Phenotypic characterisation of early repopulating B cells using flow cytometry¹⁸⁹ and mass cytometry^{190,191} has identified a predominance of transitional B cells, which represent an immature population of naive cells migrating into the circulation from the bone marrow⁴⁹. This finding suggests that B cell repopulation following treatment occurs predominantly through hematopoiesis rather than re-expansion of incompletely depleted cells. This is further supported by a recent study showing an inverse correlation between the number of detectable circulating B cells six

3. Characterising peripheral immune perturbations following B cell depletion

months following ocrelizumab treatment and the level of somatic hypermutation within the peripheral BCR repertoire, indicating that the repopulating B cells have near-germline BCR sequences¹⁹². Notably, circulating plasma cells have been shown to be preserved following anti-CD20 therapy, consistent with their loss of CD20 surface expression following terminal differentiation^{189–191}. Together, these findings align with the hypothesis that persistent depletion of memory B cells is responsible, either directly or indirectly, for the clinical benefit of anti-CD20 therapies.

In contrast to the B cell effects, the reported changes in other immune compartments following anti-CD20 therapies in MS are less consistent across studies.

Several groups have investigated the T cell effects of anti-CD20 therapy. Studies using FACS to measure quantitative changes in total T cells and major subpopulations reveal inconsistent findings^{193–197}, suggestive of a high degree of inter-individual variability over a conserved disease-modifying effect at the level of coarse cell type frequencies.

With the advent of more recent techniques for high-dimensional immunophenotyping, several groups have identified a relative decrease in specific memory T cell subpopulations, more prominent within CD8⁺ T cells, manifesting as an increase in the post-treatment CD4⁺:CD8⁺ memory T cell ratio^{189,198}. The recent discovery of a subset of circulating T cells with low-level surface CD20 expression (CD20^{dim}) that is rapidly depleted following anti-CD20 therapy has led to the hypothesis that the observed T cell changes following treatment may be a result of a direct drug effect on this subpopulation^{199–202}. However, there is contradictory evidence suggesting broader effects of ocrelizumab on the T cell compartment independent

of depleting CD20^{dim} T cells¹⁹⁸.

T cell populations depleted post-treatment have been shown to have a surface protein phenotype consistent with activation and increased migratory capacity^{198,203,204}. Moreover, TCR clonality analyses have demonstrated depletion of highly expanded $\alpha\beta$ and $\gamma\delta$ clonotypes¹⁹², and functional assays have shown reduced EBV²⁰⁴ and myelin²⁰⁵ peptide-reactivity within the T cell compartment following B cell depletion. Therefore, it is plausible that at least some of the clinical benefits of anti-CD20 therapies in MS may arise through either direct or indirect effects on pathogenic T cell states, underscoring the need for robust and unbiased immunophenotyping of the T cell compartment changes following treatment.

Among innate immune cells, the most consistent changes following anti-CD20 therapies have been reported in the myeloid compartment. Similar to T cells, quantitative alterations in monocyte subpopulations are inconsistent across studies^{193–195}, emphasising the need for more nuanced approaches to detect conserved shifts in cellular state following treatment. Circulating monocytes have been found to show increased proinflammatory cytokine production following anti-CD20 therapy, although this has been observed variably in classical CD14⁺⁺CD16⁻ or non-classical CD14⁺CD16⁺⁺ monocytes across studies^{180,181,189,206}. However, the causes and functional consequences of enriched circulating proinflammatory monocyte states following B cell depletion in MS remain uncertain.

3.1.3 Hypothesis

The immune system is a dynamic and interactive network, in which targeted depletion of major cell lineages (e.g. B cells) is likely to have broad ranging

secondary effects on other immune cell types. Several lines of evidence from the existing literature highlight non-B cell effects of anti-CD20 therapies that may be disease-relevant. Moreover, there is a growing body of evidence supporting complex dysregulated interactions between B cells, T cells and myeloid cells as critical drivers of chronic, compartmentalised CNS inflammation (see Section 1.1.3.4), which is at least partially modulated by depletion of circulating B cells. Therefore, I hypothesised that an unbiased and high-resolution analysis of the effect of B cell depletion therapy on the peripheral immune cell landscape of pwMS may help disentangle the underlying immunobiology, uncovering the specific cell states that arise from pathogenic B cell-dependent interactions in the disease context.

3.2 Results

3.2.1 Measuring dynamic longitudinal changes in peripheral immune cells following ocrelizumab

To investigate the network effects of anti-CD20 therapy on the peripheral immune system in MS, in collaboration with Brian Cutler (Oxford Centre for Neuroinflammation, University of Oxford) and Subita Balaram Kuttikkatte (Oxford Centre for Neuroinflammation, University of Oxford), we performed CITE-seq on PBMC samples from a discovery cohort ($n = 14$) of treatment-naïve pwMS commencing ocrelizumab (Figure 3.1). We collected samples prior to commencing treatment (baseline), two weeks after the first ocrelizumab infusion (early) and immediately prior to each subsequent infusion (late), which were timed to coincide with the earliest detectable B cell repopulation. For each individual, we obtained between 3–5 longitudinal samples during treatment. Participants were grouped into

three disease type categories: (i) RRMS with high relapse frequency (RR_{high} , $n = 5$), (ii) RRMS with low relapse frequency (RR_{low} , $n = 4$), and (iii) PPMS ($n = 5$) (see Section 2.2.1). Additional clinical and demographic characteristics of all participants are shown in Table 3.1.

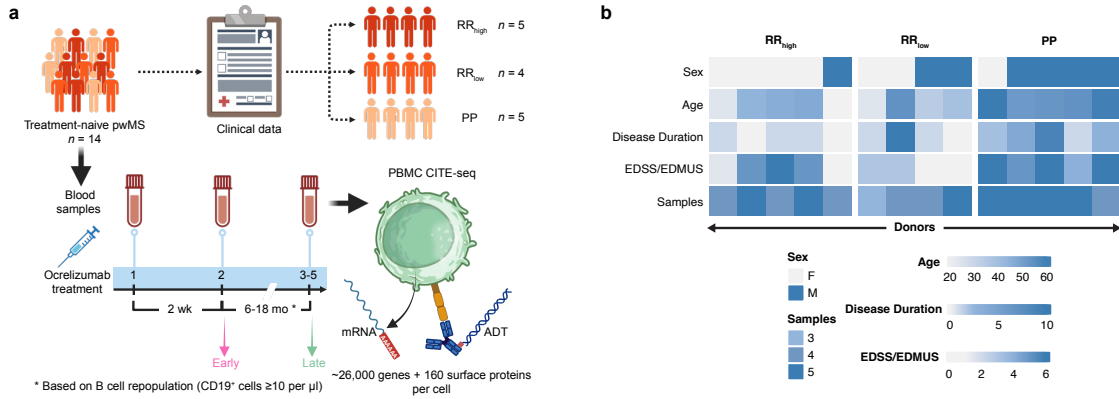


Figure 3.1. Discovery cohort of pwMS with longitudinal multimodal profiling of the peripheral immune repertoire during ocrelizumab treatment. **a**, Graphical summary of discovery cohort of treatment-naive pwMS ($n = 14$), grouped by clinical parameters into three categories: RRMS with high disease activity (RR_{high} , $n = 5$), RRMS with low disease activity (RR_{low} , $n = 4$) and PPMS (PP, $n = 5$). Longitudinal blood samples were collected before and during treatment with ocrelizumab, at both early (2 weeks after first dose) and late (6–18 months after previous dose based on individual B cell repopulation timing) time points. CITE-seq was performed on PBMCs isolated from whole blood, to measure the transcriptome and surface expression of 160 protein markers for each cell. **b**, Heatmap summarising key demographic and clinical characteristics (rows) for each individual (columns).

Following QC, I generated an integrated dataset of 264,512 cells comprising 5 main (level 1), 20 coarse (level 2), and 34 fine (level 3) cell types (Figure 3.2, Figure 3.3a–b). Cell types were determined using a semi-supervised consensus approach combining unsupervised clustering with automated label transfer from multiple publicly available PBMC reference datasets (see Section 2.5.1.2), and validated by manual inspection of marker genes (Figure 3.3c, Supplementary File 1, Supplementary File 2) and surface proteins (Figure 3.3d).

To quantify treatment effects, I used MELD¹⁵⁰ to compute perturbation scores for

3. Characterising peripheral immune perturbations following B cell depletion

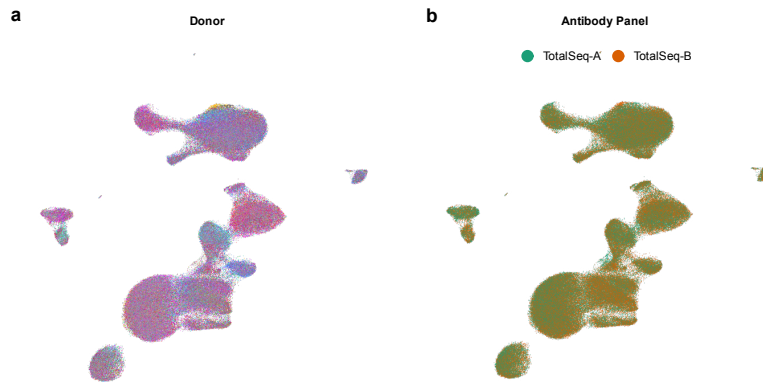


Figure 3.2. Discovery dataset integration. a, UMAP visualisation of 264,512 cells comprising the discovery dataset coloured by donor of origin (a) and CITE-seq antibody panel used (b).

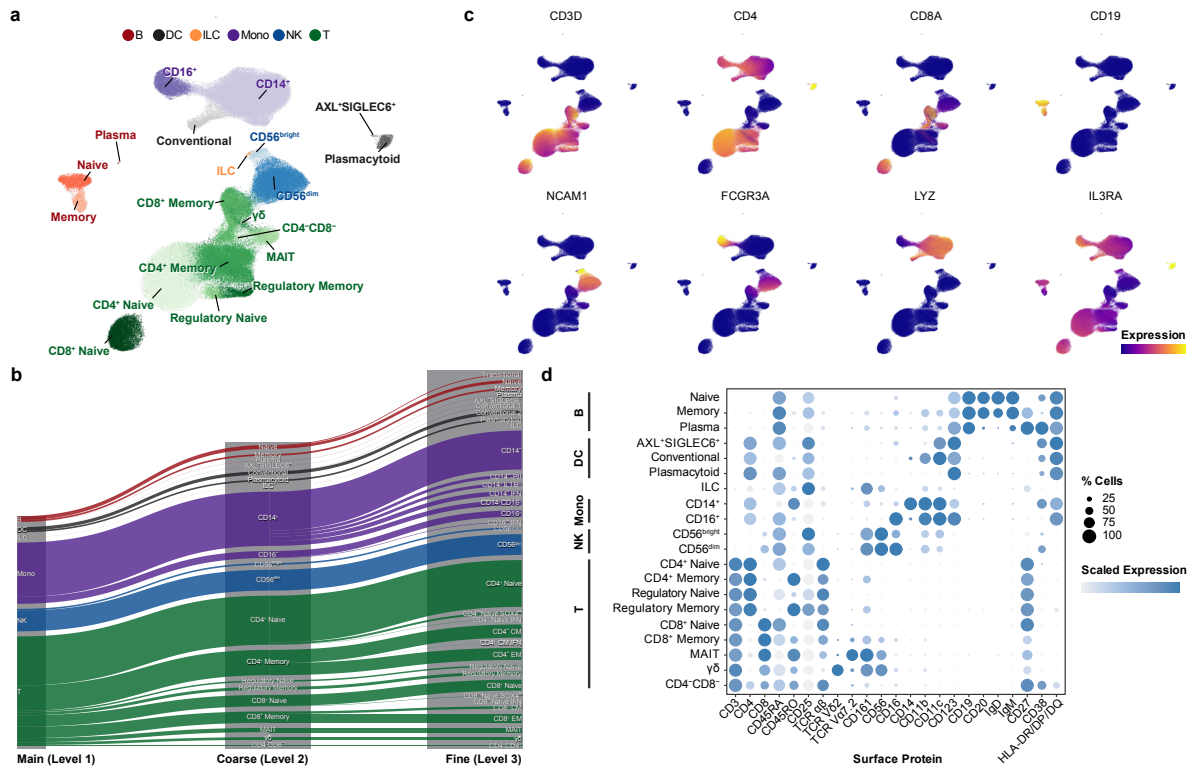


Figure 3.3. Discovery dataset cell type annotation. a, UMAP visualisation of 264,512 cells comprising the discovery dataset, coloured and labelled by cell type. b, Parallel sets diagram showing hierarchical cell type annotations in the discovery dataset; line width indicates number of cells in each category c, UMAP visualisation of discovery dataset cells coloured by normalised expression of canonical cell type marker genes; upper limits of colour scales are clipped at 99th percentiles. d, Dot plot showing expression of canonical immune surface protein markers (x-axis) in each coarse (level 2) cell type (y-axis); colour scale shows normalised expression scaled to unit variance and dot size shows percentage of cells in each group with expression above threshold levels manually selected for each protein based on inspection of the expression distribution across all cells.

each transcriptional cell state (Figure 2.3a; see Section 2.5.1.3), whereby negative scores represent depletion and positive scores represent enrichment relative to the donor-specific pre-treatment baseline (Figure 3.4). These scores represent a statistically robust, denoised measure of consensus treatment effect across donors per cell state, accounting for both biological (stochastic or idiosyncratic inter-donor heterogeneity) and technical (sampling bias or manifold topology-related) sources of noise.

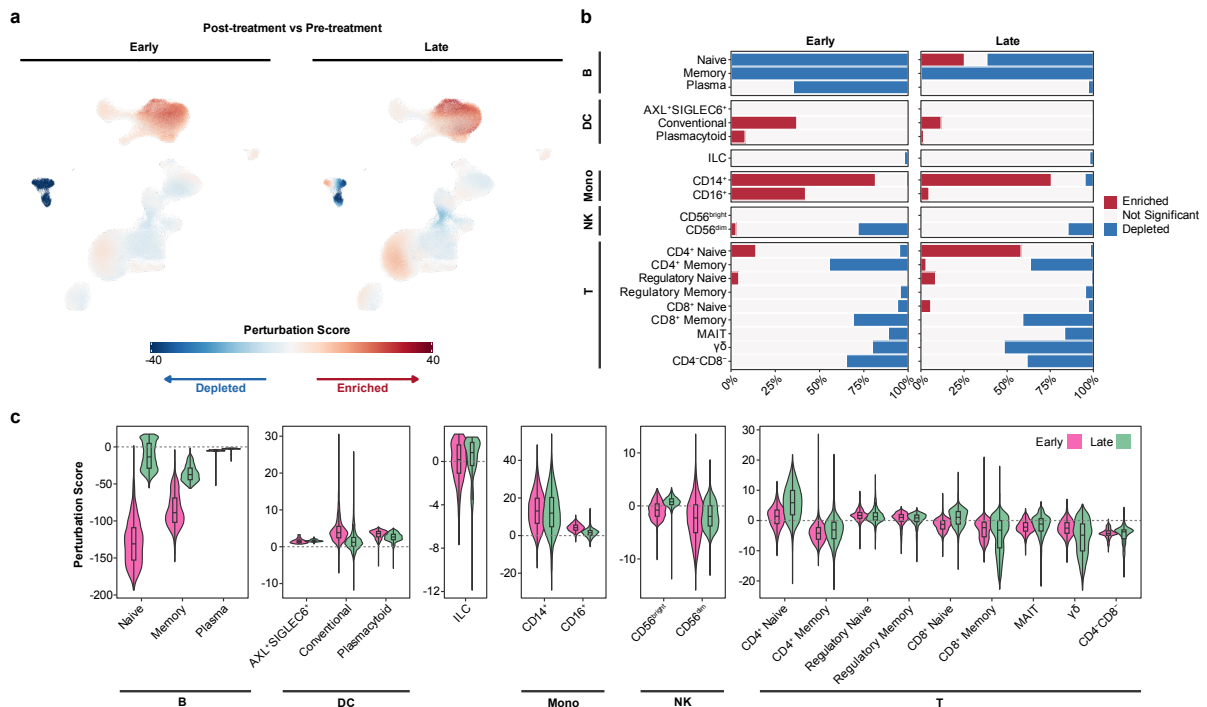


Figure 3.4. Treatment-induced shifts in peripheral immune landscape following ocrelizumab. **a**, UMAP visualisation of cell states coloured by treatment-related perturbation score at early and late post-treatment time points; positive values (red) represent enrichment and negative values (blue) represent depletion relative to pre-treatment baseline. **b**, Bar plots showing percentage (x -axis) of cell states within each coarse (level 2) cell type (y -axis) statistically significantly enriched (red) or depleted (blue) at the early and late post-treatment time points. **c**, Violin/box plots showing distribution of perturbation scores within each coarse (level 2) cell type, coloured by post-treatment time point. Statistical testing for perturbation scores was performed with Z -tests using an empirical permutation-based null distribution for each cell state.

As expected, I found the strongest treatment-related depletion in naive and

3. Characterising peripheral immune perturbations following B cell depletion

memory B cells at the early post-treatment time point, with a markedly weaker effect seen in plasma cells (including plasmablasts), which are known to downregulate surface CD20 expression. At the late post-treatment points, the majority of the B cell compartment remained strongly depleted, except for notable enrichment of a subset of naive B cells and unchanged plasma cell abundance relative to baseline.

Beyond these changes in the B cell compartment, I also identified phenotypic changes in several other immune cell populations. The strongest treatment effect outside the B cell compartment was an enrichment in monocyte populations, particularly CD14⁺ monocytes, at both the early and late post-treatment time points. I also observed clear shifts in the T cell compartment post-treatment, particularly at the late time points, with significant depletion of several subsets of memory and unconventional T cells, including CD4⁺, CD8⁺, $\gamma\delta$ and mucosal-associated invariant T (MAIT) cells, and enrichment of naive cell subsets, especially within CD4⁺ T cells.

Taken together, these findings add strong support to the growing body of evidence that, rather than having a B cell-restricted immunomodulatory effect, ocrelizumab treatment induces widespread, temporally dynamic, cell type-specific shifts in the peripheral immune system, with important implications for the possible mechanisms underlying its therapeutic benefit in pwMS.

Table 3.1. Clinical and demographic details of treatment-naïve pwMS with longitudinal PBMC samples during ocrelizumab treatment.

Donor ID	Age (years)	Sex	BMI (kg/m ²)	Disease Type	Disease Duration (years)	Pre-treatment Relapses ¹		Baseline MRI Activity ²	Post- treatment Relapses	Follow Up MRI Activity ²	EDSS/ EDMUS ³	Infusion Interval ⁴ (days)			
						24 months	12 months					2	3	4	5
OCRZ003	19	M	26.6	RR _{high}	0	4	4	new	0	none	1.0	14	189	303	—
OCRZ004	24	F	22.5	RR _{high}	1	2	2	new	0	none	1.5	14	226	259	—
OCRZ011	39	M	23.7	RR _{low}	0	1	1	enhancing	0	none	1.0	14	405	483	427
OCRZ012	46	F	32.7	RR _{high}	1	6	2	none	0	none	6.0	14	156	287	—
OCRZ013	44	F	26.8	RR _{high}	0	3	2	none	0	none	5.0	14	307	343	322
OCRZ014	47	F	25.5	RR _{high}	1	4	3	enhancing	0	none	5.0	14	328	290	270
OCRZ019	61	F	28.8	PP	2	0	0	enhancing	0	none	6.0	15	202	287	409
OCRZ020	24	F	23.8	RR _{low}	1	2	1	none	0	none	3.0	14	308	—	—
OCRZ022	60	M	24.8	PP _{low}	3	0	0	new	0	none	6.0	14	202	231	—
OCRZ023	52	M	29.8	PP	4	0	0	enhancing	0	none	5.0	15	331	319	245
OCRZ031	54	M	26.3	PP	1	0	0	new	0	none	4.0	14	287	266	371
OCRZ033	53	M	25.0	PP	9	0	0	enhancing	0	none	6.0	14	224	266	252
OCRZ043	34	M	27.7	RR _{low}	1	1	1	new	0	none	1.0	14	287	273	—
OCRZ045	54	F	22.9	RR _{low}	10	2	1	new	0	none	3.0	14	259	280	—
3530194 ⁵	34	M	47.4	RR _{low}	1	2	0	none	2	none	2.0	14	191	194	168

¹Number of relapses in 24- and 12-month intervals prior to starting ocrelizumab treatment²Stratified radiologic disease activity on MRI: none = no changes from previous scan, new = new lesions since previous scan, enhancing = new contrast-enhancing lesions since previous scan; baseline refers to the most recent available imaging prior to starting ocrelizumab treatment³Disability score at time of starting ocrelizumab treatment⁴Time since previous ocrelizumab infusion for each subsequent infusion after first⁵Treatment non-responder

3.2.2 Early repopulating B cells are immature and antigen-inexperienced

To gain a more granular understanding of the dynamics of B cell repopulation during extended interval treatment with ocrelizumab, I then performed an unbiased characterisation of the enriched subset of naive B cells observed at the late post-treatment time points (Figure 3.5).

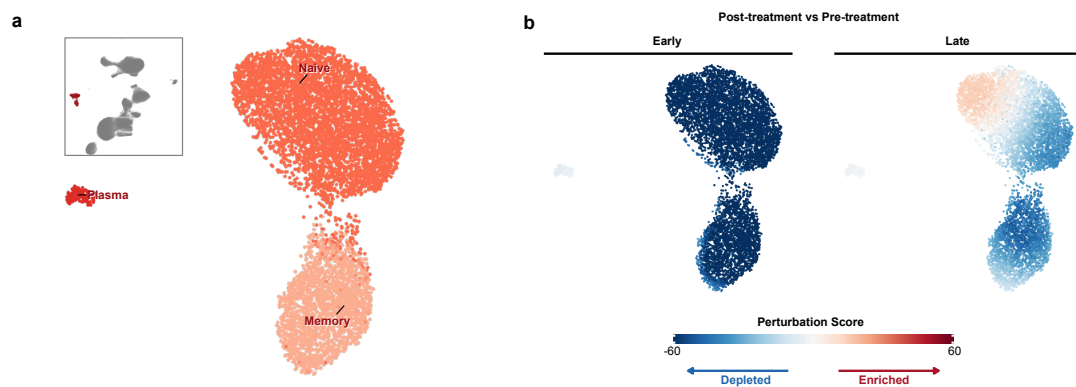


Figure 3.5. Treatment-induced shifts in B cells following ocrelizumab. a, UMAP visualisation of B cells, which are shown highlighted (dark red) in the UMAP visualisation of all cells (inset), coloured and labelled by coarse (level 2) cell type. b, UMAP visualisation of B cell states coloured by treatment-related perturbation score at early and late post-treatment time points; positive values (red) represent enrichment and negative values (blue) represent depletion relative to pre-treatment baseline.

In order to fully leverage the single-cell resolution dataset and maximise sensitivity to subtle phenotypic shifts due to treatment, I developed and optimised a novel analytical approach to quantify treatment-related effects on continuous cellular states without relying on partitioning cells into discrete subpopulations (Figure 2.3).

I used cNMF¹³⁷ to infer cell type-specific gene programs, with cells assigned continuous program scores reflecting how much each program contributes to its overall transcriptional profile (Figure 2.3b; see Section 2.5.1.4). I identified 14 gene programs within the B cell compartment, which I annotated based on a

combination of their top weighted genes, expression pattern across fine (level 3) cell types and pathway enrichment analysis (Figure 3.6, Figure 3.7, Figure 3.8, Supplementary File 3, Supplementary File 4). Two of these programs represented low-level contaminating transcripts from other cell types and an additional program was dominated by widely-expressed housekeeping genes; these were excluded from further analyses.

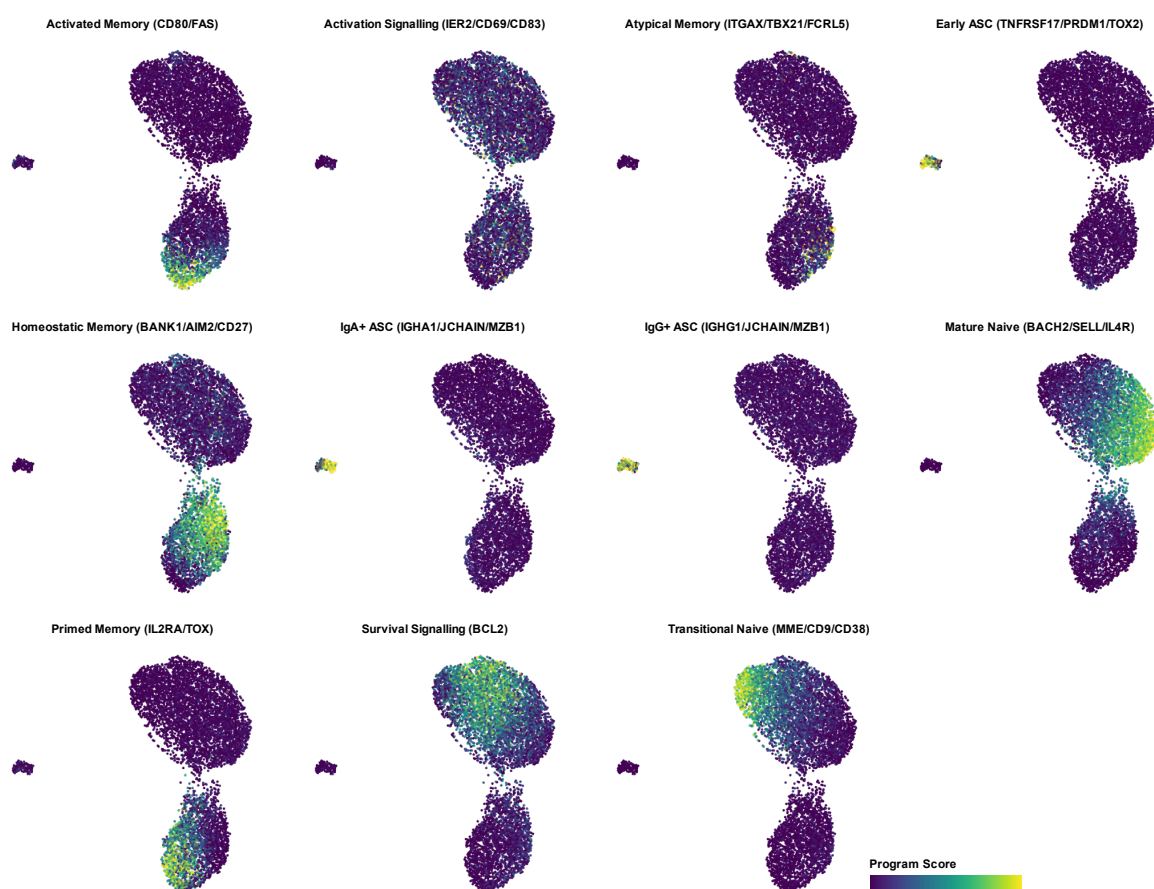


Figure 3.6. B cell-specific gene programs. UMAP visualisations of B cells in the discovery cohort dataset coloured by program score for each B cell-specific gene program; upper limits of colour scales are clipped at 99th percentiles.

I then correlated the two orthogonal continuous measures of treatment effect (perturbation scores) and cellular phenotype (program scores) by training a machine learning model (random forest regression) to predict perturbation scores based

3. Characterising peripheral immune perturbations following B cell depletion

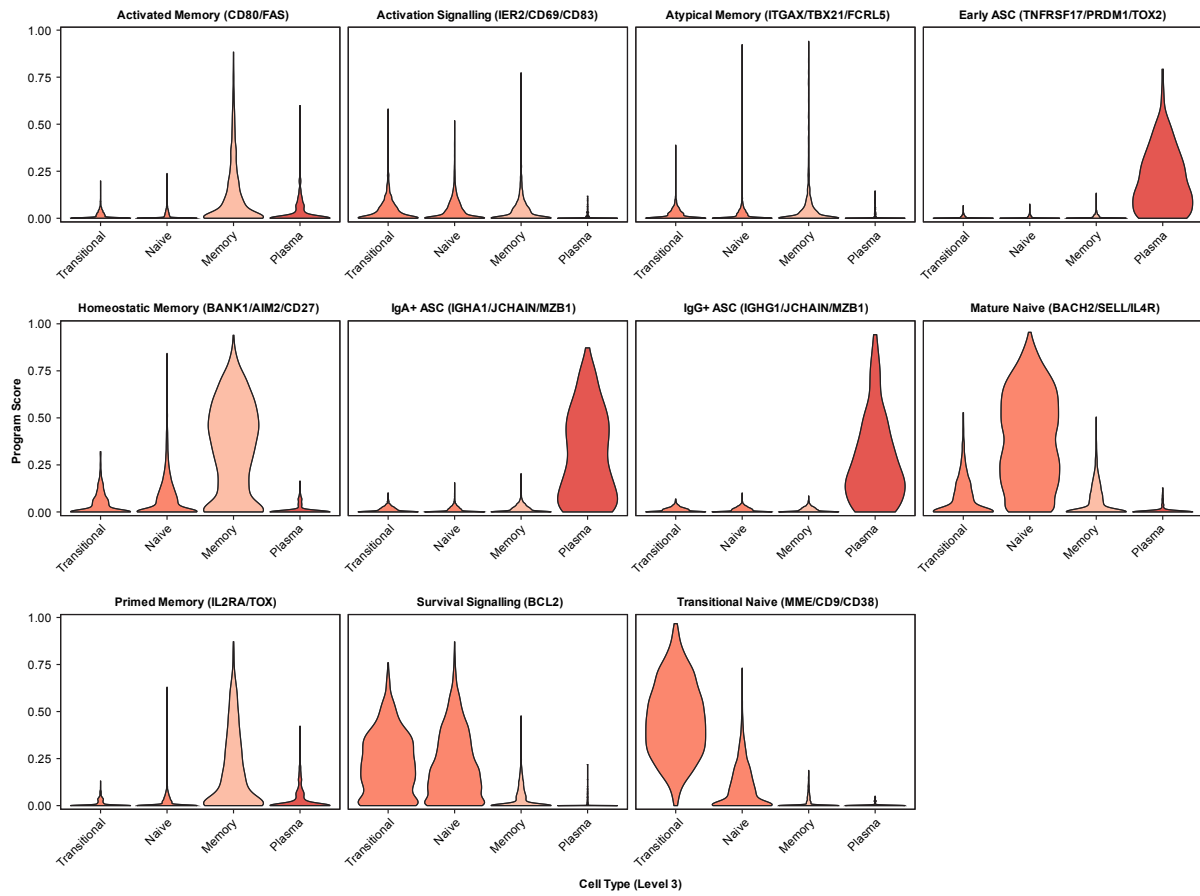


Figure 3.7. Distribution of gene program scores across B cell subtypes. Violin plots showing distribution of program scores (y -axis) for each B cell-specific gene program across fine (level 3) cell types (x -axis) in the discovery cohort dataset.

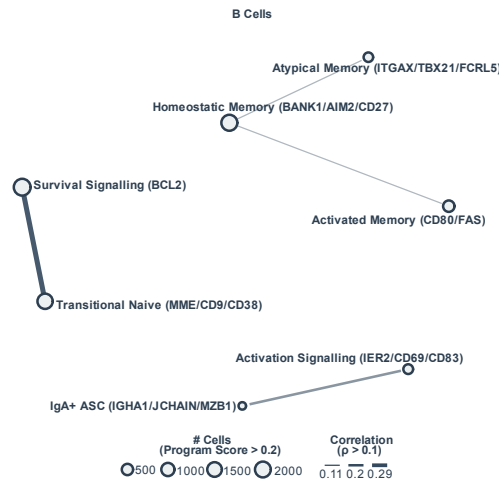


Figure 3.8. B cell gene program co-expression network. Graph showing co-expression patterns of B cell-specific gene programs in the discovery cohort dataset; nodes represent gene programs, node sizes indicate the number of cells with a high (>0.2) program score, and edge thicknesses indicate Spearman correlation coefficient (ρ) of the program scores and are only shown between pairs of programs if $\rho > 0.1$.

on program scores. The motivation for using random forest regression models in preference to simpler linear models is detailed in Section 2.5.1.5. From the trained model, I extracted feature importances using SHAP; SHAP values are a quantitative metric indicating the additive contribution of each gene program to the effect of treatment on each cell state, independent of other co-expressed programs. Finally, by combining the perturbation scores, program scores and SHAP values, I derived program effect scores to provide a summary metric for each program representing its importance-weighted association with treatment effect (Figure 2.3c; see Section 2.5.1.5).

Based on the program effect scores, I identified the *Transitional Naive* program as solely associated with enrichment during early B cell repopulation (Figure 3.9, Figure 3.10a). This transcriptional program was expressed in a subset of naive B cells (Figure 3.6, Figure 3.7) and the top weighted genes were enriched for known markers of B cell immaturity (*MME*, *CD9*, *CD38*, *SOX4*, *EBF1*, *TCL1A*, *VPREB1*, *VPREB3*), supporting their identity as recently differentiated, antigen-inexperienced B cells migrating out of the bone marrow into the circulation. To validate the phenotype associated with this transcriptional signature, I performed protein-level differential expression analysis to identify the surface markers of cells with high (> 0.2) program scores within naive B cells (Figure 3.10b, Supplementary File). This revealed high expression of canonical markers of transitional B cells (CD9, CD38, CD24, CD5, CD1d, IgM). I confirmed enrichment of this cell state during B cell repopulation at the individual level (Figure 3.10c) by using a linear mixed effects model (see Section 2.6) to compare the mean program usage in naive B cells from each donor between the late post-treatment and pre-treatment time points

3. Characterising peripheral immune perturbations following B cell depletion

(marginal mean difference 0.598, 95% CI [0.534, 0.661], $p = 3.77 \times 10^{-10}$).

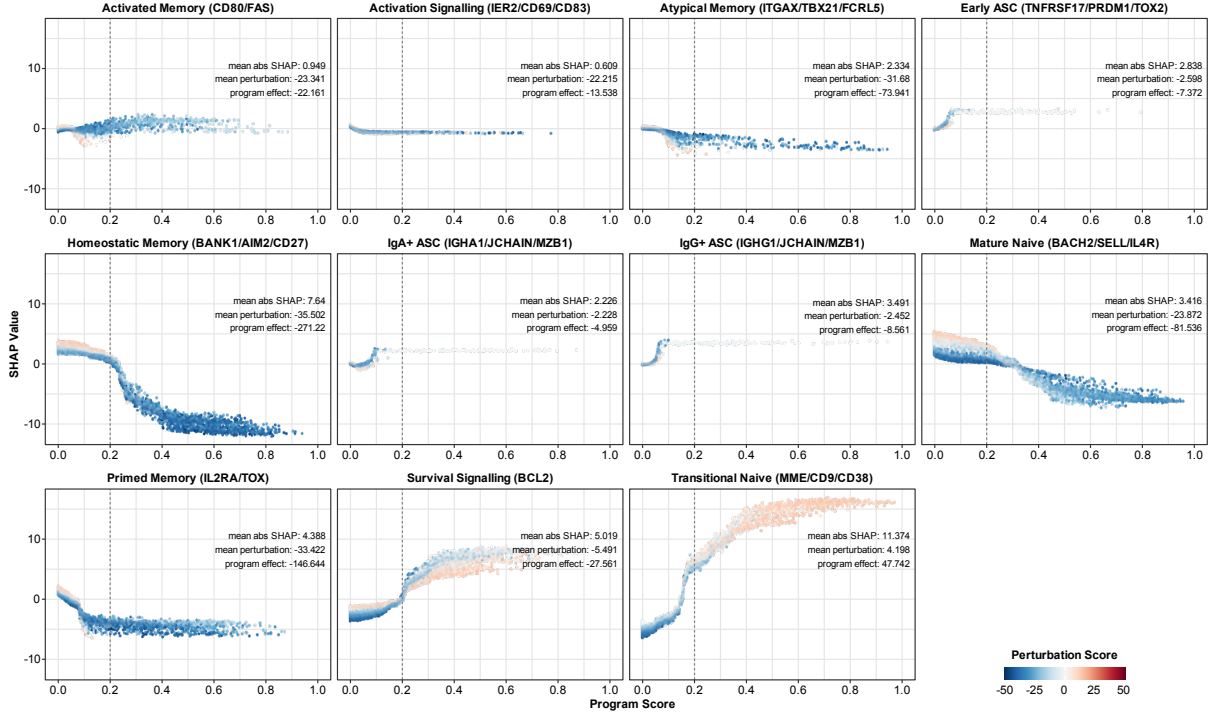


Figure 3.9. B cell gene program SHAP plots. Scatter plots of SHAP values (y -axis) for B cell-specific gene programs for all B cells in the discovery cohort, obtained from a random forest regression model trained to predict perturbation score based on program score, vs program score (x -axis), coloured by perturbation score; positive values (red) represent enrichment and negative values (blue) represent depletion relative to pre-treatment baseline. Dashed vertical lines indicate the threshold (0.2) used as a cutoff for cells with high program scores. Mean absolute SHAP values and mean perturbation scores for high program score cells are shown along with the computed program effect scores for each program.

Of the remaining programs, those most strongly associated with persistent depletion at the late post-treatment time points were programs linked to mature, quiescent memory B cell (*Homeostatic Memory*, *Primed Memory*, and *Atypical Memory*) and naive B cell (*Mature Naive*) states (Figure 3.6, Figure 3.7). By contrast, the *Early ASC*, *IgA+ ASC*, and *IgG+ ASC* programs, expressed exclusively in plasma cells (Figure 3.6, Figure 3.7), showed the weakest program effect scores, indicating that these cell populations are relatively spared from depletion by anti-CD20 therapies.

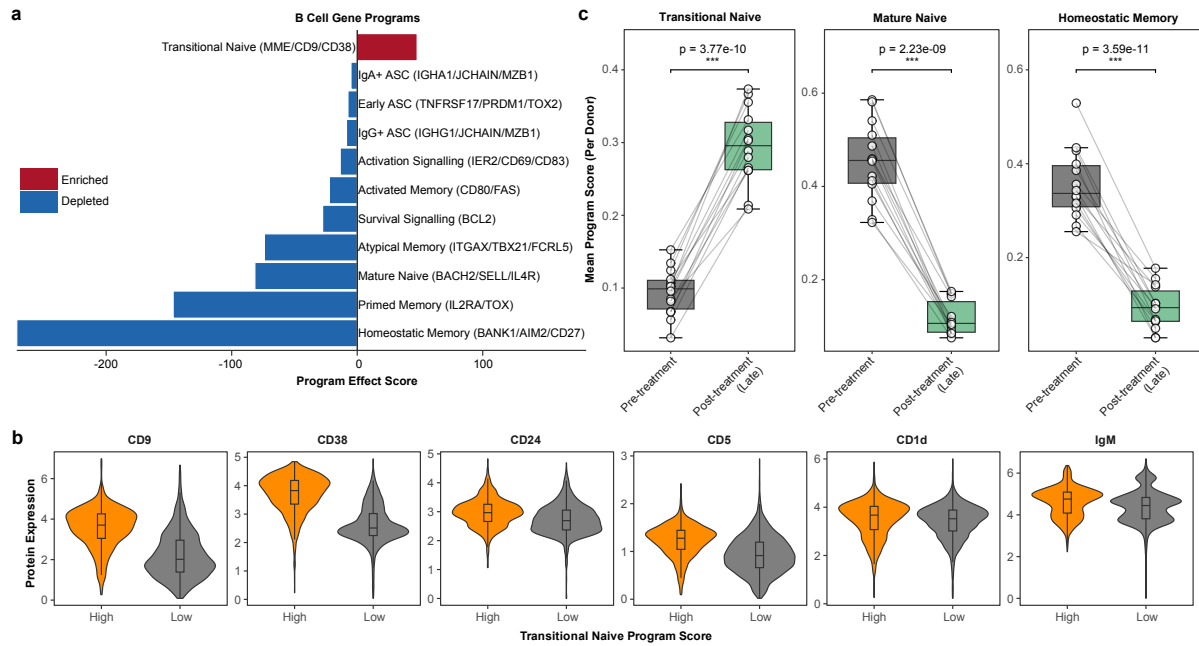


Figure 3.10. Reprogramming of the peripheral B cell compartment during ocrelizumab treatment. **a**, Bar plot showing program effect scores (x -axis) for B cell-specific gene programs (y -axis); positive scores (red) indicate that high program expression is associated with enrichment post-treatment and negative scores (blue) indicate that high program expression is associated with depletion post-treatment. **b**, Violin/box plots showing distribution of normalised expression of selected differentially expressed protein markers between naive B cells with high (> 0.2 , orange) vs low (≤ 0.2 , grey) Transitional Naive program scores. **c**, Box plots comparing mean program scores per donor (y -axis) between pre- and late post-treatment time points (x -axis) for selected gene programs; lines connect paired data from the same individual. Statistical testing was performed using a linear mixed effects model adjusted for age and sex; p -values were calculated with t -tests of marginal means using Satterthwaite degrees of freedom.

My findings demonstrate a reprogramming of the B cell compartment during extended interval ocrelizumab treatment, with the early repopulating B cells between infusions harbouring an immature, antigen-inexperienced phenotype, persistent depletion of memory B cells and preservation of ASCs.

3.2.3 Delayed depletion of a proinflammatory, cytotoxic and tissue-homing (T_c1 -TH) memory T cell phenotype

Next, I aimed to characterise the observed phenotypic shifts in the T cell compartment during ocrelizumab treatment (Figure 3.11), based on my hypothesis that B cell depletion would result in reduced activation of pathogenic T cells in the disease context.

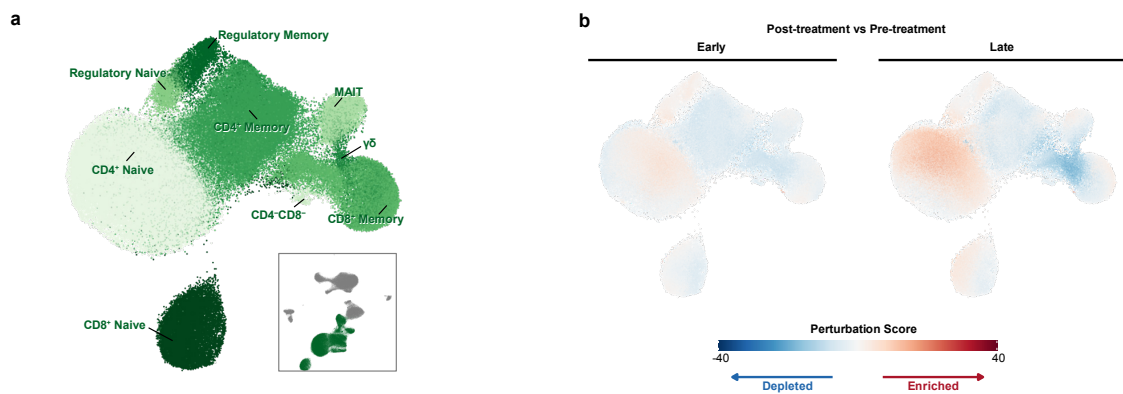


Figure 3.11. Treatment-induced shifts in T cells following ocrelizumab. a, UMAP visualisation of T cells, which are shown highlighted (dark green) in the UMAP visualisation of all cells (inset), coloured and labelled by coarse (level 2) cell type. b, UMAP visualisation of T cell states coloured by treatment-related perturbation score at early and late post-treatment time points; positive values (red) represent enrichment and negative values (blue) represent depletion relative to pre-treatment baseline.

Using cNMF¹³⁷, I identified and annotated 36 T cell-specific gene programs (Figure 3.12, Figure 3.13, Figure 3.14, Supplementary File 3, Supplementary File 4). Of these, five represented housekeeping programs and four comprised low-level contaminating transcripts from other cell types; these were excluded from further

analyses.

I then trained a random forest regression model to predict perturbation scores based on program scores, extracted feature importances using SHAP, and derived program effect scores (see Section 2.5.1.5).

I identified three gene programs, expressed predominantly in naive T cells (Figure 3.12, Figure 3.13), that were strongly associated with enrichment post-treatment (Figure 3.15, Figure 3.16a). The first, *Survival Signalling*, had top marker genes enriched for pro-survival and growth signalling pathways (*BCL2*, *PIM1*, *PDK1*, *IGF1R*). The second, *Naive CD4*, featured canonical marker genes of naive T cells (*IL7R*, *SELL*, *CCR7*, *TCF7*, *LEF1*, *BACH2*, *FOXO1*). The top genes for the third program, *Precursor Memory*, included genes encoding TCR co-stimulatory molecules (*CD28*, *ICOS*) and early activation markers (*CD38*, *IRF4*), suggestive of transcriptional priming for activation and naive-to-memory differentiation. Together, these programs suggest enrichment of antigen-inexperienced T cell states exhibiting tonic survival and growth signalling after ocrelizumab, consistent with reduced B cell-dependent T cell activation.

Conversely, I identified three gene programs, expressed predominantly in memory T cells (Figure 3.12, Figure 3.13), that were strongly associated with depletion post-treatment (Figure 3.16a–b). The largest association was observed for the *Cytotoxic Inflammatory* program, with top marker genes indicating cytotoxic effector polarisation (*CD8A*, *EOMES*, *GZMK*, *GZMA*, *GZMM*, *GZMH*, *CST7*, *NKG7*), inflammatory cytokine/chemokine signalling (*IFNG*, *IL32*, *CCL3*, *CCL4*, *CCL5*, *CXCR4*), and MHC class II-related genes (*HLA-DRB1*, *HLA-DRA*, *HLA-DRB5*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DMA*, *CD74*), which

3. Characterising peripheral immune perturbations following *B* cell depletion

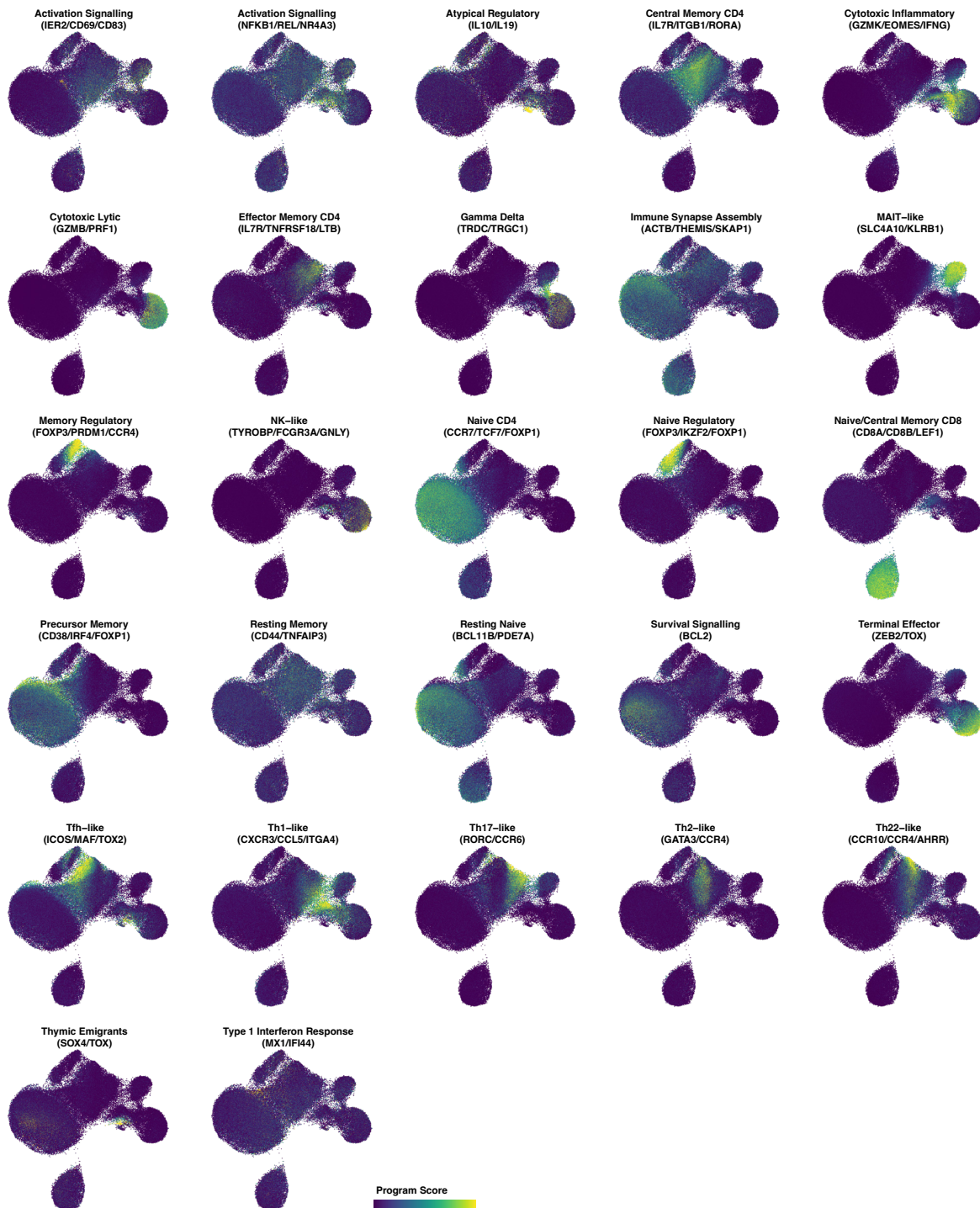


Figure 3.12. T cell-specific gene programs. UMAP visualisations of T cells in the discovery cohort dataset coloured by program score for each T cell-specific gene program; upper limits of colour scales are clipped at 99th percentiles.

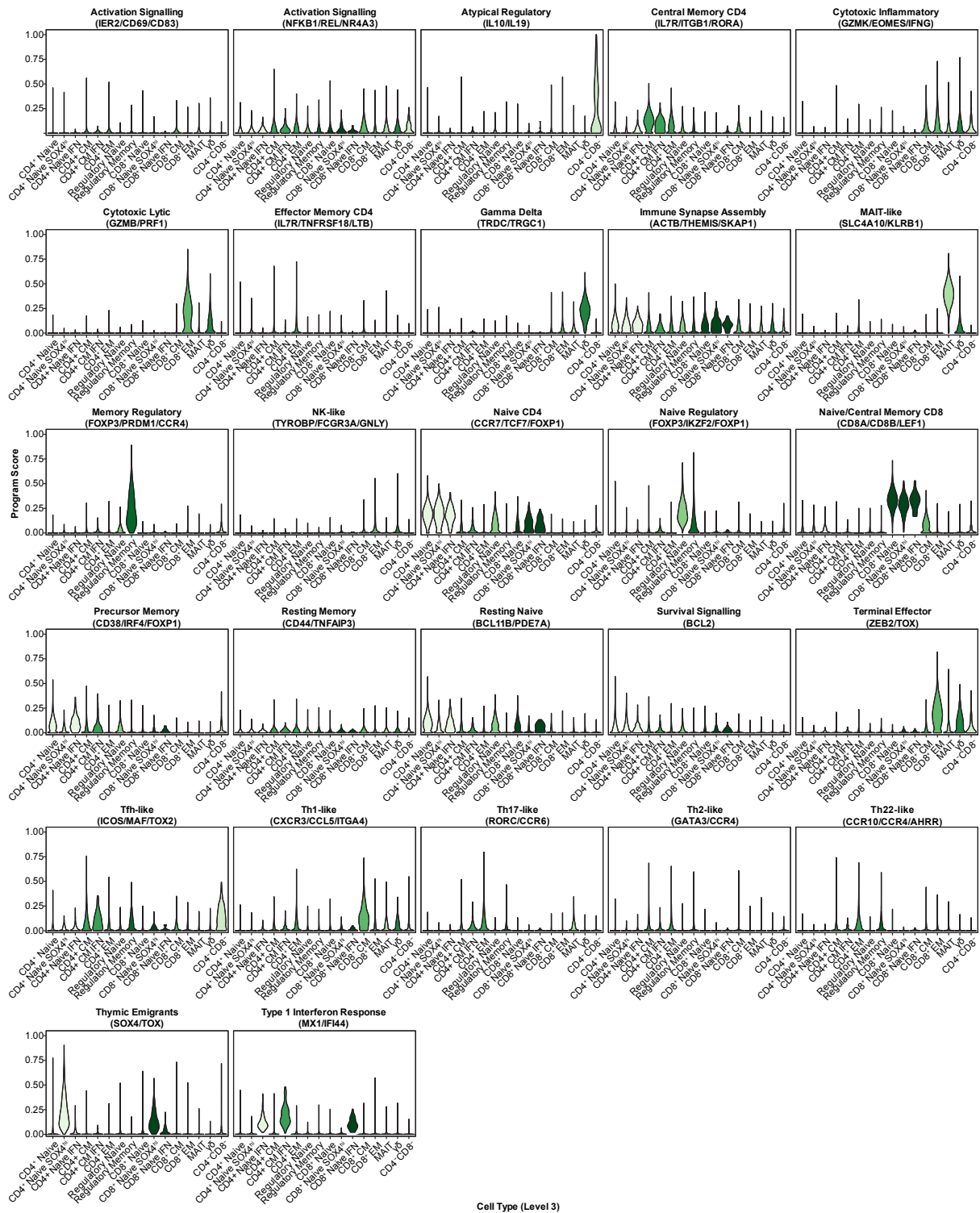


Figure 3.13. Distribution of gene program scores across T cell subtypes. Violin plots showing distribution of program scores (y -axis) for each T cell-specific gene program across fine (level 3) cell types (x -axis) in the discovery cohort dataset.

3. Characterising peripheral immune perturbations following B cell depletion

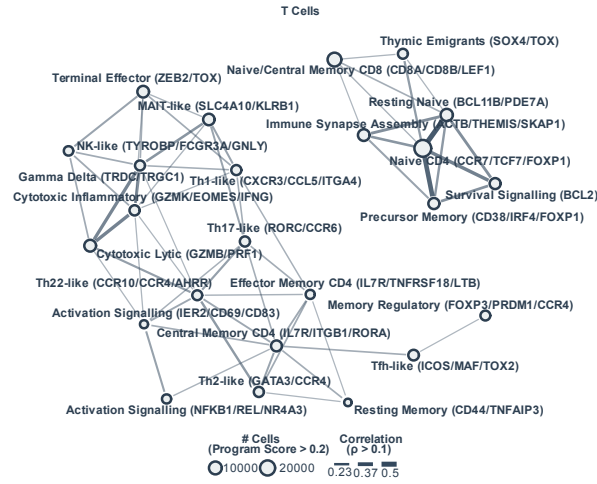


Figure 3.14. T cell gene program co-expression network. Graph showing co-expression patterns of T cell-specific gene programs in the discovery cohort dataset; nodes represent gene programs, node sizes indicate the number of cells with a high (>0.2) program score, and edge thicknesses indicate Spearman correlation coefficient (ρ) of the program scores and are only shown between pairs of programs if $\rho > 0.1$.

are upregulated in activated interferon- γ -responsive T cells. The second depletion-associated program was annotated *Th1-like*, based on several genes associated with T_H1 polarisation (*STAT4*, *JAK2*, *CXCR3*, *CD226*). Additionally, the top marker genes of this program were enriched for genes supporting adhesion/migration (*ITGA4*, *ITGA1*, *DOCK8*) and tissue residency potential (*PRDM1*, *ZNF683*, *NT5E*), suggesting a phenotype with increased capacity for endothelial transmigration into the CNS and differentiation into tissue-resident cells. The third depletion-associated program I identified was an *Activation Signalling* program, which was distinguished by numerous genes associated with signalling pathways downstream of TCR engagement, including AP-1 family genes (*FOS*, *FOSB*, *JUN*, *JUNB*), other immediate early genes (*IER2*, *EGR1*, *EGR3*, *NR4A1*, *NR4A3*), NF- κ B pathway genes (*NFKB1*, *NFKBIA*, *NFKBIZ*, *TNFAIP3*), and T cell activation markers (*CD69*, *CD83*, *TNF*).

To delineate the overall cellular phenotype within the T cell compartment most

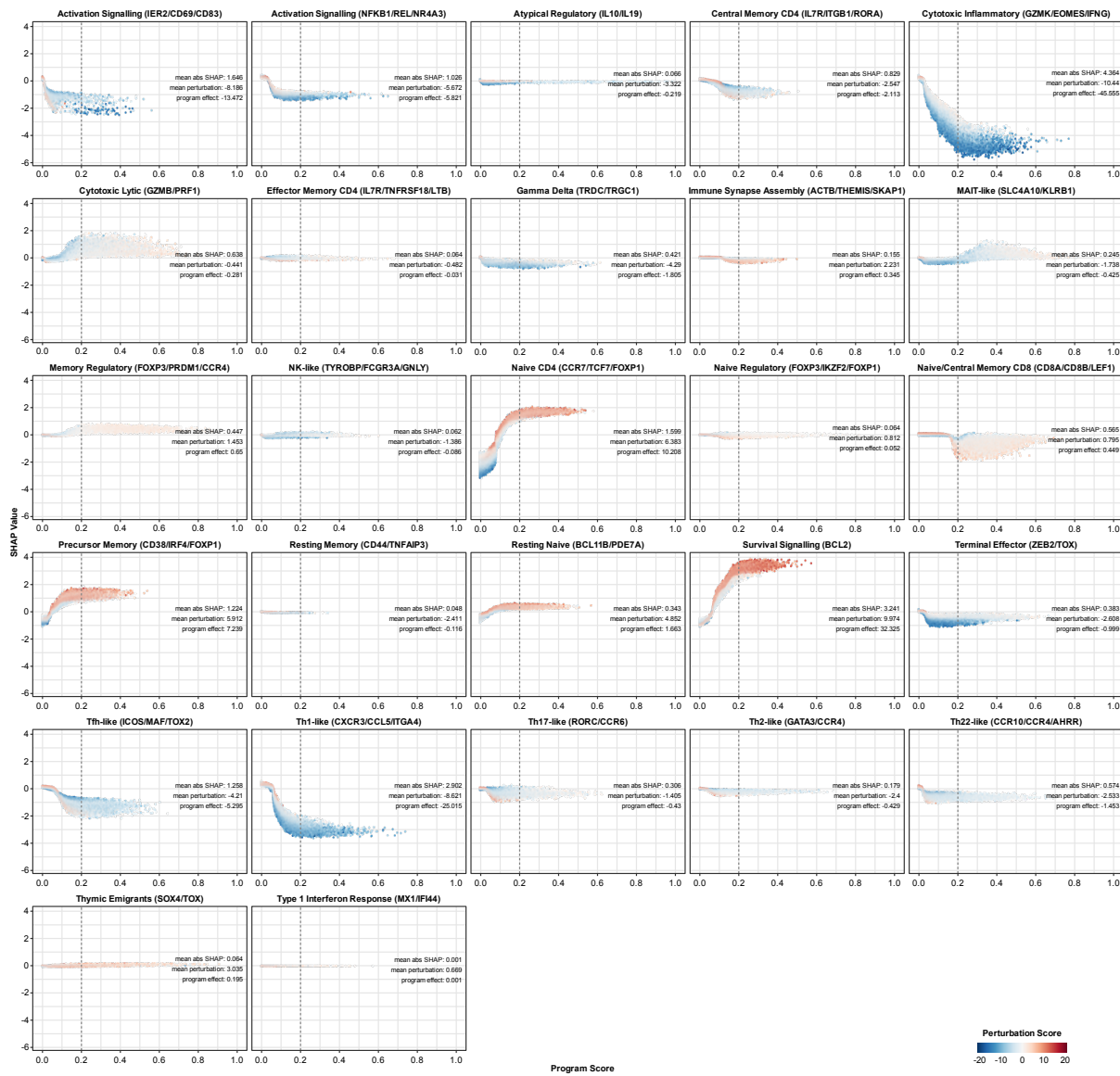


Figure 3.15. T cell gene program SHAP plots. Scatter plots of SHAP values (y -axis) for T cell-specific gene programs for all T cells in the discovery cohort, obtained from a random forest regression model trained to predict perturbation score based on program score, vs program score (x -axis), coloured by perturbation score; positive values (red) represent enrichment and negative values (blue) represent depletion relative to pre-treatment baseline. Dashed vertical lines indicate the threshold (0.2) used as a cutoff for cells with high program scores. Mean absolute SHAP values and mean perturbation scores for high program score cells are shown along with the computed program effect scores for each program.

3. Characterising peripheral immune perturbations following B cell depletion

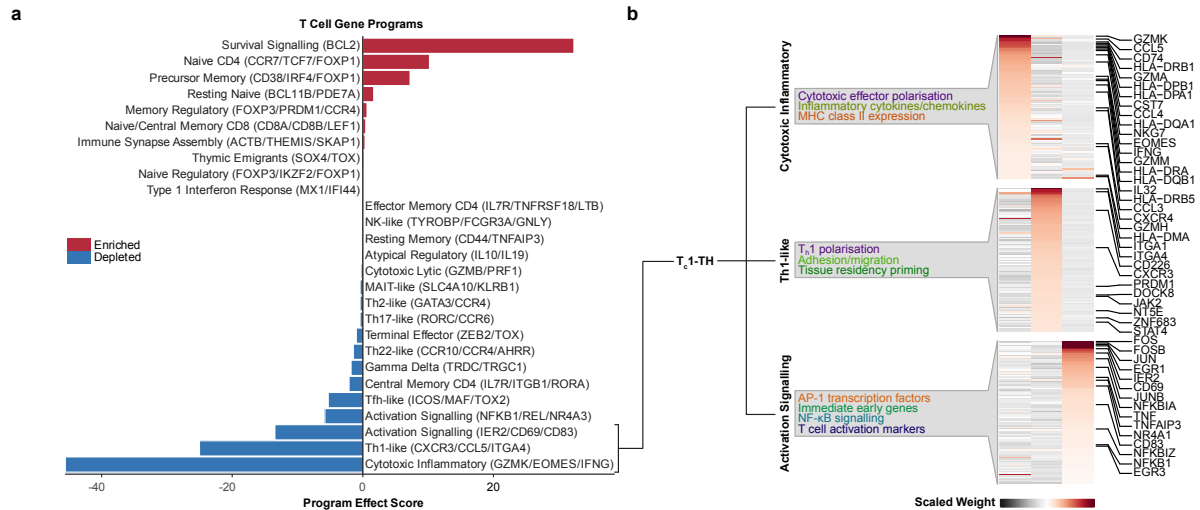


Figure 3.16. Circulating proinflammatory, cytotoxic, tissue-homing (T_{c1} -TH) memory T cells are less abundant following B cell depletion. a, Bar plot showing program effect scores (x -axis) for T cell-specific gene programs (y -axis); positive scores (red) indicate that high program expression is associated with enrichment post-treatment and negative scores (blue) indicate that high program expression is associated with depletion post-treatment. b, Heatmap of top 100 weighted genes per component program of the T_{c1} -TH signature with selected genes highlighted corresponding to key biological processes represented by each program; colour scale shows Z-scored gene weights.

strongly depleted following B cell depletion, I derived a composite score by summing the individual program scores for these three programs in each cell (Figure 3.17a), and confirmed that this score was selectively enriched in cell states more likely to be depleted post-treatment (Figure 3.17b). I hereafter describe the phenotype delineated by this composite score as type 1 cytotoxic tissue-homing T (T_{c1} -TH) cells, reflecting the functional modules captured by the component gene programs.

To further characterise the T_{c1} -TH phenotype, I analysed which cell types were enriched among cells with high scores (> 0.2) for this program (Figure 3.17c). Cells highly expressing the T_{c1} -TH signature were predominantly $CD8^+$ and $CD4^+$ memory cells (81%, $CD4:CD8$ ratio 0.45). I therefore performed protein-level differential expression analysis (see Section 2.5.1.6) to identify the cell surface markers associated with this transcriptional signature within memory T cells

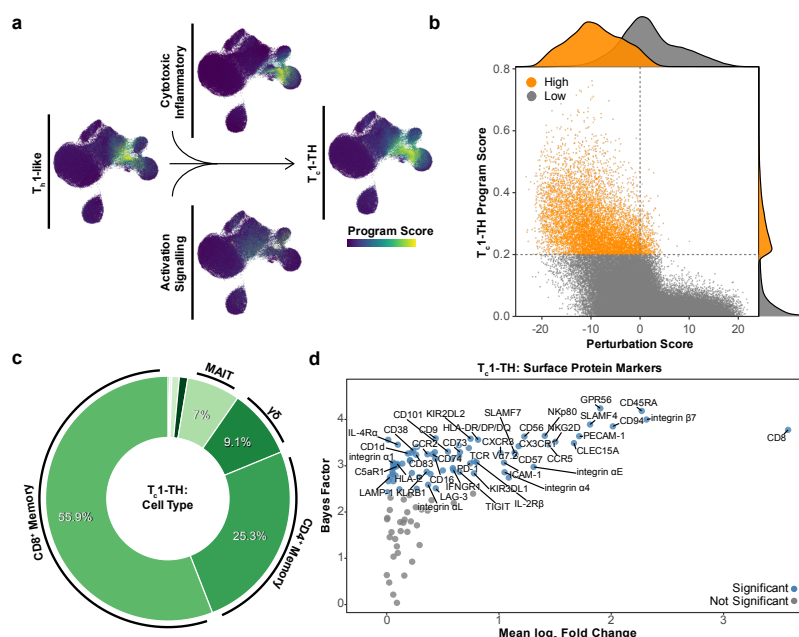


Figure 3.17. Phenotypic characterisation of cells with high T_c1-TH program score. **a**, Program scores for T_c1-TH and its three component gene programs projected on the UMAP visualisation of T cells in the discovery dataset; upper limits of colour scales are clipped at 99th percentiles. **b**, Scatter plot with marginal density plots of perturbation score (x -axis) and T_c1-TH program score (y -axis) per cell, with dashed lines indicating perturbation score of 0 (vertical) and the threshold (0.2) used for selecting cells with a high program score (horizontal); these cells are highlighted (orange). **c**, Distribution of coarse (level 2) cell type labels within T_c1-TH high cells. **d**, Volcano plot showing differentially expressed proteins which are more abundant in memory T cells with high T_c1-TH score, with mean log₂ fold change on the x -axis and Bayes factor (posterior evidence for differential expression) on the y -axis; statistically significant proteins (FDR threshold < 0.05) are highlighted in blue and selected key markers are labelled.

3. Characterising peripheral immune perturbations following B cell depletion

(Figure 3.17d, Supplementary File 5). This confirmed high expression of markers of innate-like cytotoxic effector function (CD8, CLEC15A, CD56, CD57, NKp80, CD94, NKG2D, KIR2DL3, KIR3DL1) and chronic activation (IL-2R β , IFNGR1, CD83, HLA-DR/DP/DQ, PD-1, TIGIT, LAG-3), as well as molecules linked to chemotaxis (CCR5, CXCR3, CX3CR1), adhesion (integrin α 4, integrin β 7, integrin α L, ICAM-1, PECAM-1), and tissue retention (integrin α E, integrin α 1, CD101, CD73, GPR56). Together, these findings support the interpretation that the T_c1-TH signature defines a subpopulation of activated circulating cytotoxic memory T cells that are primed to migrate into tissues and differentiate into T_{rm} cells upon re-encountering their cognate antigen.

As recent studies have identified a population of CD20^{dim} T cells that are also rapidly depleted along with B cells by anti-CD20 therapies^{199–202}, I sought to determine whether the post-treatment depletion of cells with high T_c1-TH signature was attributable to low-level CD20 expression on these cells. I used protein-level CD3 and CD20 expression combined with the cell type annotations to gate CD20^{dim} T cells (Figure 3.18a). I found that although memory T cells with a high T_c1-TH program score were more likely to be CD20^{dim} (odds ratio 1.3, Fisher exact test $p=0.0278$), the CD20^{dim} subpopulation only accounted for a minor proportion (1.23%) of these cells (Figure 3.18b).

The origin of CD20, canonically a B cell lineage-restricted protein, on the surface of T cells remains controversial. There is strong experimental evidence showing that T cells can directly acquire CD20 molecules from the surface of B cells through close cellular contact (trogocytosis)²⁰⁷. However, several studies have detected transcripts for the CD20 gene (*MS4A1*) in these cells^{27,199,208,209}, making the case

for an additional mechanism of true endogenous expression. I therefore investigated whether the higher CD20 expression was transcriptionally regulated by performing differential expression testing for the *MS4A1* gene in memory T cells with high vs low T_c1-TH signature. Intriguingly, I identified statistically significant higher expression of *MS4A1* in these cells (Supplementary File 5). However, on direct comparison of CD20 protein and gene expression to B cell subsets, I found the overall expression level in T_c1-TH high cells was substantially lower than that seen in plasma cells (Figure 3.18c–d), which significantly downregulate CD20 expression during differentiation and are notably spared from the direct effects of anti-CD20 therapies.

To further investigate the relevance of low-level CD20 expression in these cells on the observed treatment effect, I assessed the temporal dynamics of T_c1-TH cell depletion following ocrelizumab. I hypothesised that if T_c1-TH cells were directly targeted by ocrelizumab due to CD20 expression, they would be rapidly depleted following the first dose (i.e. at the early post-treatment time point), mirroring the time course of B cell depletion. I tested for depletion of this cell state longitudinally during the course of treatment using a linear mixed effects model (see Section 2.6) to compare mean program usage in memory T cells from each donor at each time point (Figure 3.19). I found T_c1-TH cells were not significantly depleted at the early post-treatment time point compared to baseline (marginal mean difference -0.0168 , 95% CI $[-0.0417, 0.0081]$, $p = 0.176$). However, they were significantly depleted at the late post-treatment time points compared to both pre-treatment (marginal mean difference -0.0616 , 95% CI $[-0.0865, -0.0367]$, $p = 1.16 \times 10^{-4}$) and early post-treatment (marginal mean difference -0.0448 , 95% CI $[-0.0697, -0.0199]$,

3. Characterising peripheral immune perturbations following B cell depletion

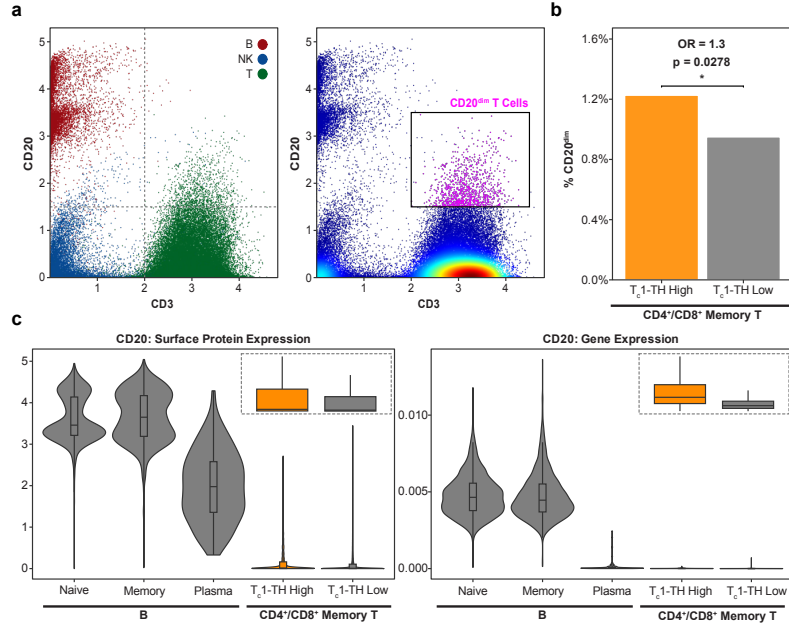


Figure 3.18. CD20 expression in cells with high T_c1-TH program score. **a**, Scatter plots demonstrating gating strategy for CD20^{dim} T cells using CITE-seq, showing normalised expression of CD3 (*x*-axis) vs CD20 (*y*-axis) in B cells, T cells and NK cells, coloured by main (level 1) cell type (left panel) or point density (right panel); lower limits of manually selected gates are indicated with dashed lines (left panel) and T cells gated as the CD20^{dim} population are highlighted (magenta, right panel). **b**, Bar plots comparing percentage of CD20^{dim} cells (*y*-axis) in memory T cells with high (> 0.2, orange) vs low (≤ 0.2, grey) T_c1-TH program scores (*x*-axis). **c**, Violin/box plots showing distribution of normalised expression of CD20 protein (left panel) and gene (right panel) across coarse (level 2) B cell subpopulations and T_c1-TH high/low memory T cells; zoomed in box plots comparing the memory T cell groups are shown inset and T_c1-TH high cells are highlighted (orange). Statistical testing was performed using a Fisher exact test.

$p = 1.75 \times 10^{-3}$) time points.

Collectively, these results suggest that B cell depletion with ocrelizumab is associated with secondary depletion of a proinflammatory, cytotoxic and tissue-homing memory T cell phenotype. Although a minor subset of these cells exhibit low-level CD20 expression and may be directly targeted by the drug, the delayed time course of the shift in the broader cell state following treatment is most consistent with an indirect mechanism, possibly reflecting loss of B cell-dependent antigenic stimulation or cytokine support.

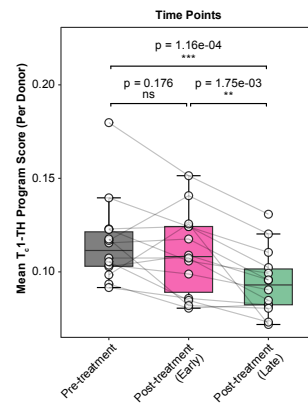


Figure 3.19. Temporal dynamics of T_c1-TH cell depletion following anti-CD20 therapy. Box plots comparing mean T_c1-TH program scores per donor (*y*-axis) between pre- and early/late post-treatment time points; lines connect paired data from the same individual. Statistical testing was performed using linear mixed effects models adjusted for age and sex; *p*-values were calculated with *t*-tests of marginal means (using Satterthwaite degrees of freedom) and adjusted for multiple testing using the Benjamini–Hochberg method.

3.2.4 T_c1-TH signature correlates with clinical disease activity

Given T_c1-TH cells are depleted following treatment with ocrelizumab, I hypothesised that they may have a pathogenic role in the untreated disease context. I therefore next explored whether their pre-treatment abundance was linked to clinical disease severity. I compared mean pre-treatment T_c1-TH program score within memory T cells between donors in different disease groups (Figure 3.20) and found higher T_c1-TH signature expression in the RR_{high} group compared to the RR_{low} group (marginal mean difference 0.0375, 95% CI [0.0103, 0.0646], *p* = 0.031), indicating that patients with more clinically aggressive MS have an enrichment of this cell phenotype within their circulating memory T cell repertoire.

Longitudinal samples from pwMS with breakthrough disease activity during ocrelizumab treatment are rare due to its high efficacy in RRMS, with a recent large real-world registry-based study finding a mean annualised relapse rate during

3. Characterising peripheral immune perturbations following B cell depletion

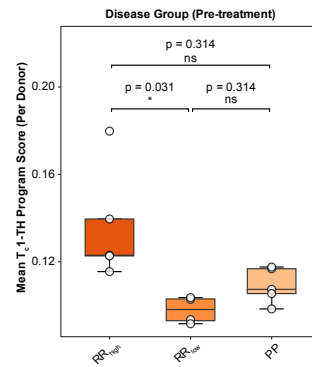


Figure 3.20. Pre-treatment T_c1-TH cell abundance across disease groups. Box plots comparing mean T_c1-TH program scores per donor (*y*-axis) between disease groups at the pre-treatment time point. Statistical testing was performed using linear models adjusted for age and sex; *p*-values were calculated with *t*-tests of marginal means and adjusted for multiple testing using the Benjamini–Hochberg method.

treatment of 0.05 and a 25.2% probability of 4-year disability progression²¹⁰. Despite this, in collaboration with Marco Pisa (Department of Neurology, John Radcliffe Hospital, OUH) and Gabriele DeLuca (Department of Neurology, John Radcliffe Hospital, OUH), we identified a treatment-naïve pwMS with poor therapeutic response despite depletion of CD19⁺ B cells below the lower limit of detection, experiencing breakthrough clinical disease activity and accumulation of disability (EDSS 2.0 to 6.0) over two years. As an additional exploratory study, in collaboration with Brian Cutler (Oxford Centre for Neuroinflammation, University of Oxford) and Subita Balaram Kuttikkatte (Oxford Centre for Neuroinflammation, University of Oxford), we performed CITE-seq on longitudinal PBMC samples from this patient to determine if there was a measurable difference in the pattern of treatment-related shift in T_c1-TH cells from the primary cohort of patients exhibiting good therapeutic response to ocrelizumab. After quality control and transferring cell type annotations from the discovery dataset, the final integrated dataset contained 24,195 cells (Figure 3.21a). I used starCAT¹⁴² to project the T cell-specific gene

programs from the discovery dataset onto the T cells from this additional donor, allowing me to compute T_c1 -TH program scores in these cells (see Section 2.5.2.3). Intriguingly, I found an absence of the longitudinal trend of decreasing T_c1 -TH signature enrichment in the T cell compartment at the late post-treatment time points relative to the early post-treatment time point in this donor (Figure 3.21b). This finding suggests that their poor treatment response was related to either inadequate suppression of this cell state or alternative immunological mechanisms playing a stronger role in driving their disease activity.

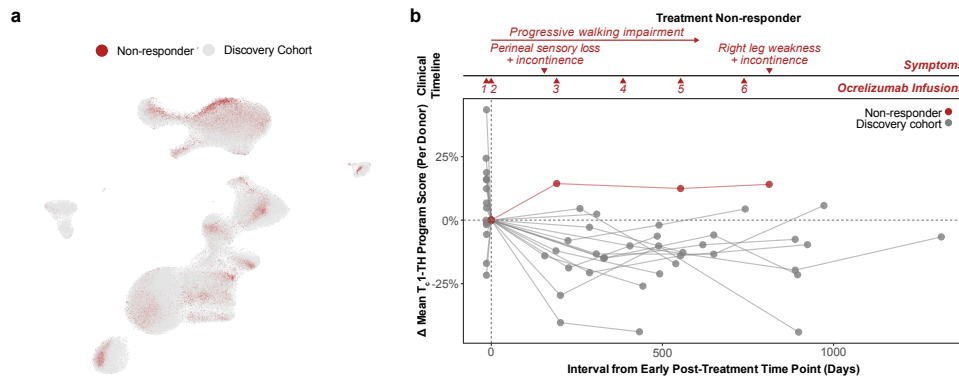


Figure 3.21. Comparative longitudinal immune profiling of treatment non-responder. **a**, UMAP visualisation of 24,195 cells from the treatment non-responder donor (red) projected onto the 264,512 cells from the discovery cohort donors (grey). **b**, Scatter/line plot showing longitudinal percentage change in mean T_c1 -TH program scores per donor (y -axis) over time (x -axis), relative to the early post-treatment time point; lines connect paired data from the same individual. Data from the donor with poor clinical response to treatment is highlighted (red), with annotations (top) indicating the timeline of symptom progression and ocrelizumab infusions for this donor.

Together, these findings support the clinical and pathogenic relevance of T_c1 -TH cells. Cells with this transcriptional signature are enriched in the circulation of pwMS with more aggressive relapsing disease, supporting a peripheral origin due to aberrant B cell interactions leading to an expanded circulating reservoir of proinflammatory T cells primed to migrate into the CNS. Concordantly, I

identified a failure in downregulation of this cell state in the circulation in an individual with poor therapeutic response to ocrelizumab; although this observation requires validation in other pwMS with anti-CD20 treatment failure, it supports the hypothesis that the treatment efficacy of ocrelizumab may be dependent on dampening this T cell phenotype.

3.2.5 Heterogeneous enrichment of a proinflammatory (Classical-Inflammatory) monocyte phenotype

Given the strongest perturbation scores outside the B cell compartment were observed within monocytes, I next sought to characterise the phenotypic shifts within this compartment during ocrelizumab treatment (Figure 3.22).

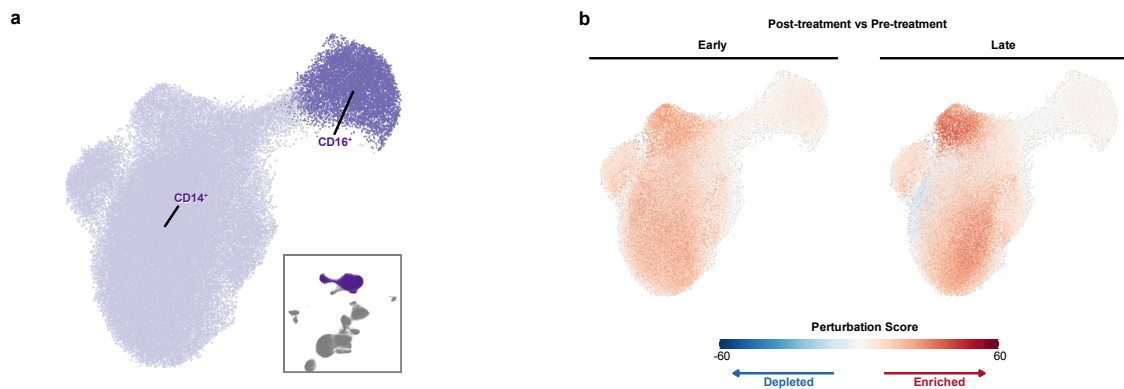


Figure 3.22. Treatment-induced shifts in monocytes following ocrelizumab. **a**, UMAP visualisation of monocytes, which are shown highlighted (dark green) in the UMAP visualisation of all cells (inset), coloured and labelled by coarse (level 2) cell type. **b**, UMAP visualisation of monocyte cell states coloured by treatment-related perturbation score at early and late post-treatment time points; positive values (red) represent enrichment and negative values (blue) represent depletion relative to pre-treatment baseline.

Using cNMF¹³⁷, I identified and annotated 11 monocyte-specific gene programs (Figure 3.23, Figure 3.24, Figure 3.25, Supplementary File 3, Supplementary File 4). Of these, one represented low-level contaminating transcripts from other cell types; this was excluded from further analyses.

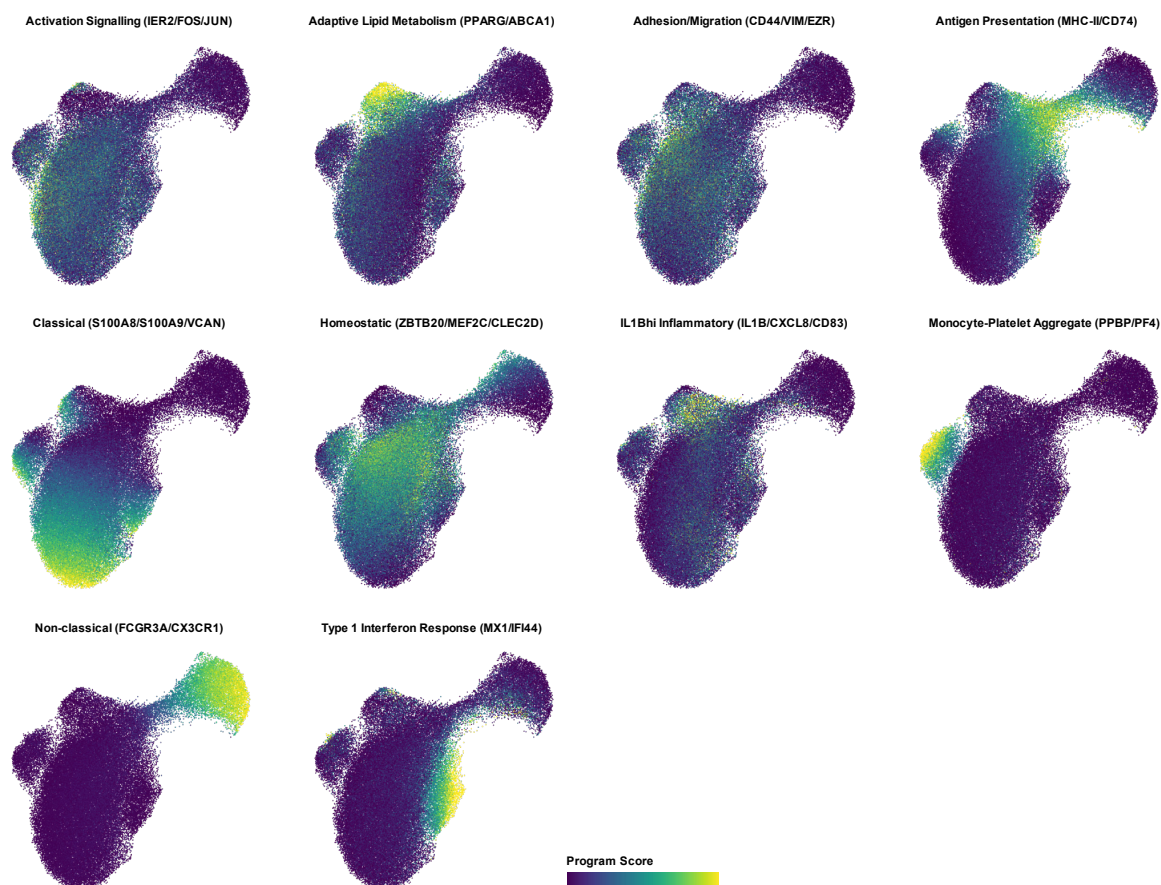


Figure 3.23. Monocyte-specific gene programs. UMAP visualisations of monocytes in the discovery cohort dataset coloured by program score for each monocyte-specific gene program; upper limits of colour scales are clipped at 99th percentiles.

3. Characterising peripheral immune perturbations following B cell depletion

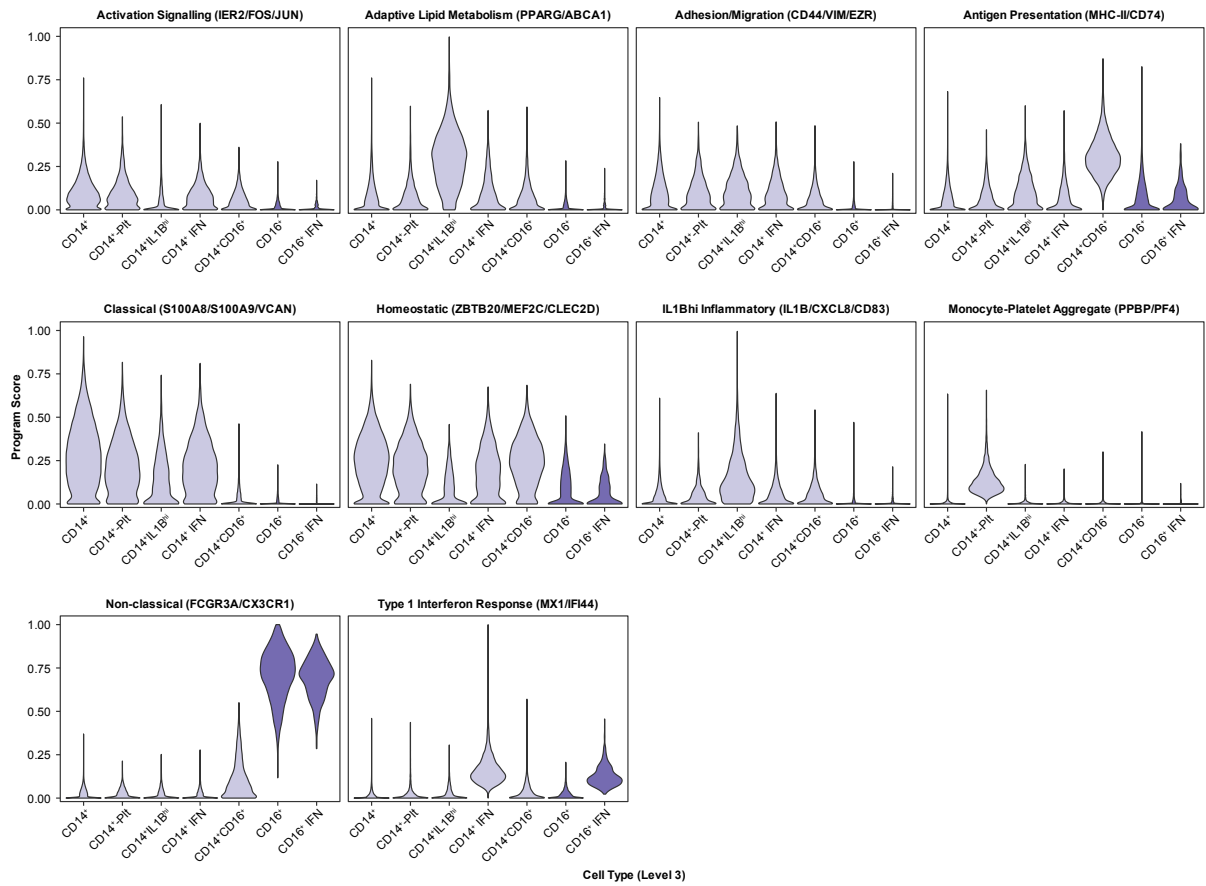


Figure 3.24. Distribution of gene program scores across monocyte subtypes. Violin plots showing distribution of program scores (y -axis) for each monocyte-specific gene program across fine (level 3) cell types (x -axis) in the discovery cohort dataset.

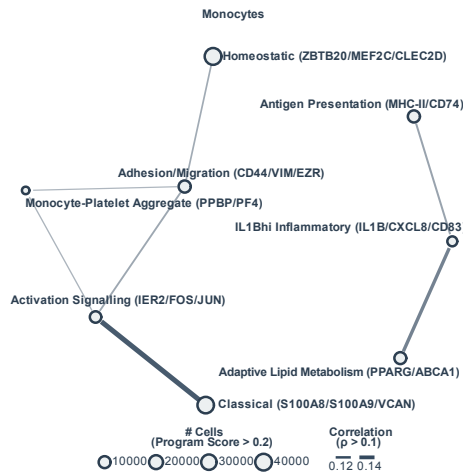


Figure 3.25. Monocyte gene program co-expression network. Graph showing co-expression patterns of monocyte-specific gene programs in the discovery cohort dataset; nodes represent gene programs, node sizes indicate the number of cells with high (>0.2) program score, and edge thicknesses indicate Spearman correlation coefficient (ρ) of the program scores and are only shown between pairs of programs if $\rho > 0.1$.

I then trained a random forest regression model to predict perturbation scores based on program scores, extracted feature importances using SHAP, and derived program effect scores (see Section 2.5.1.5).

Based on the program effect scores (Figure 3.26, Figure 3.27a), I identified that three gene programs, expressed predominantly in CD14⁺ monocytes (Figure 3.23, Figure 3.24), were most strongly associated with enrichment post-treatment. The top enrichment-associated program was *IL1Bhi Inflammatory*, with top marker genes dominated by prototypical myeloid inflammatory effector cytokine/chemokine (*IL1B*, *IL1A*, *TNF*, *CXCL8*, *CXCL2*, *CXCL3*, *CCL3*, *CCL4*) and inflammasome-associated (*NLRP3*) genes, as well as immediate early (*EGR1*, *EGR2*, *EGR3*, *IER2*, *IER3*) and NF- κ B pathway (*NFKBIA*, *NFKB1*, *REL*, *RELB*, *NFKB2*, *TNFAIP3*, *NFKBIZ*, *NFKBID*) genes. The second program was annotated *Classical* based on canonical classical monocyte lineage-associated marker genes (*S100A8*, *S100A9*, *S100A12*, *FCN1*, *VCAN*, *LYZ*). The third was an *Adaptive Lipid Metabolism* program, characterised by a core lipid-handling regulatory axis in monocytes (*PPARG*, *ABCA1*) and numerous genes involved in fatty-acid activation and membrane lipid metabolism (*ACSL3*, *GPAT3*, *PNPLA8*, *GPCPD1*). Additionally, the top markers of this program featured multiple intracellular signalling and regulatory genes (*MAP3K8*, *MAP2K1*, *BRAF*, *MALT1*, *FOXO3*, *NRIP1*), supporting the interpretation that it represents an activation-coupled adaptive metabolic program, consistent with recognised immunometabolic rewiring in inflammatory innate effector responses.

Therefore, the combination of the functional modules represented by these three programs delineates an activated, proinflammatory classical CD14⁺⁺CD16⁻

3. Characterising peripheral immune perturbations following B cell depletion

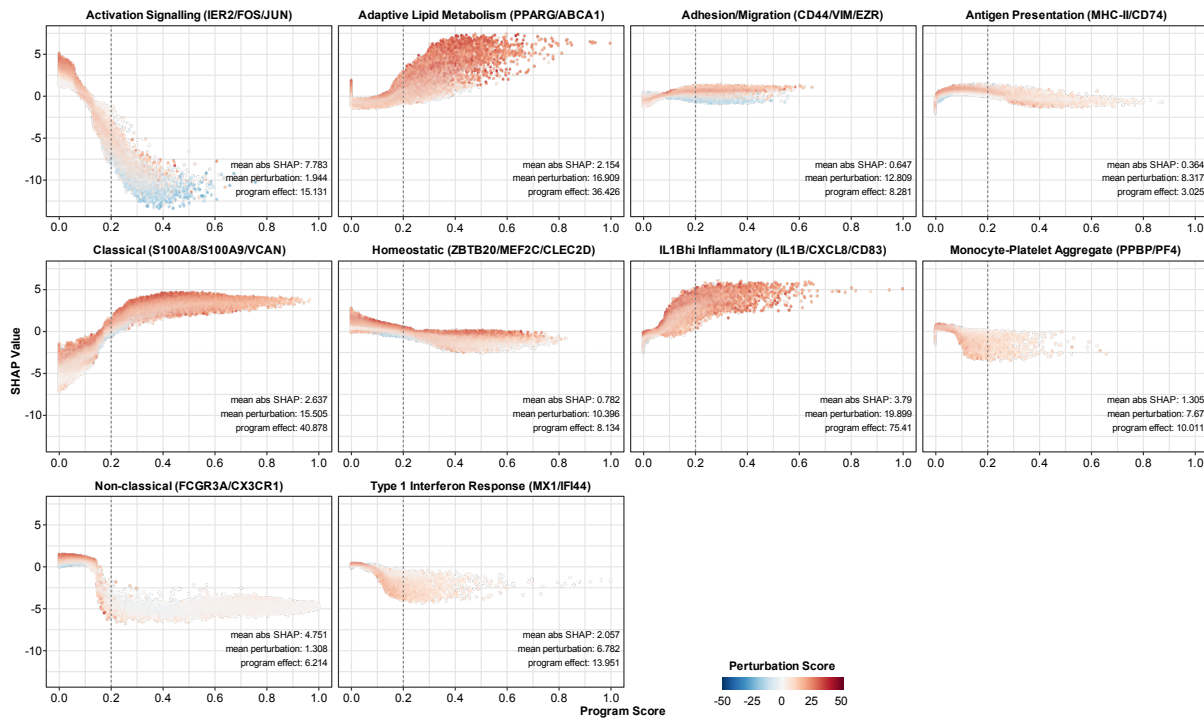


Figure 3.26. Monocyte gene program SHAP plots. Scatter plots of SHAP values (y -axis) for monocyte-specific gene programs for all monocytes in the discovery cohort, obtained from a random forest regression model trained to predict perturbation score based on program score, vs program score (x -axis), coloured by perturbation score; positive values (red) represent enrichment and negative values (blue) represent depletion relative to pre-treatment baseline. Dashed vertical lines indicate the threshold (0.2) used as a cutoff for cells with high program score. Mean absolute SHAP values and mean perturbation scores for high program score cells are shown along with the computed program effect scores for each program.

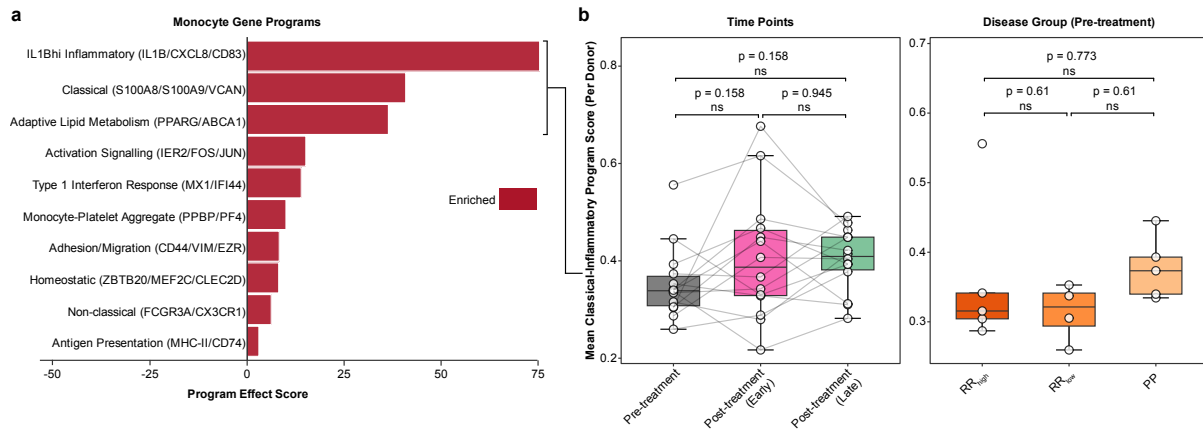


Figure 3.27. Circulating proinflammatory (Classical-Inflammatory) monocytes are heterogeneously enriched following B cell depletion. **a**, Bar plot showing program effect scores (x -axis) for monocyte-specific gene programs (y -axis); positive scores (red) indicate that high program expression is associated with enrichment post-treatment. **b**, Box plots comparing mean Classical-Inflammatory program scores per donor (y -axis) between pre- and early/late post-treatment time points (left panel) and between disease groups at the pre-treatment time point (right panel); lines connect paired data from the same individual. Statistical testing was performed using linear mixed effects models adjusted for age and sex; p -values were calculated with t -tests of marginal means (using Satterthwaite degrees of freedom) and adjusted for multiple testing using the Benjamini–Hochberg method.

monocyte state, hereafter described as Classical-Inflammatory. I derived a composite score to quantify this phenotype by summing the individual component program scores in each cell.

Subsequently, I tested for enrichment of this cell state following B cell depletion at the individual level using a linear mixed effects model (see Section 2.6) to compare the mean program scores in CD14⁺ monocytes from each donor at each time point (Figure 3.22b). Although this confirmed a trend of higher mean Classical-Inflammatory program score at the post-treatment time points relative to the pre-treatment baseline, there was substantial inter-donor heterogeneity in the longitudinal direction and magnitude of shift in this cell state, and no statistically significant overall effect of treatment was detected in any pairwise comparison: early post-treatment compared to baseline (marginal mean difference

3. Characterising peripheral immune perturbations following B cell depletion

0.151, 95% CI $[-0.027, 0.330]$, $p = 0.158$), late post-treatment compared to baseline (marginal mean difference 0.145, 95% CI $[-0.033, 0.323]$, $p = 0.158$), or late post-treatment compared to early post-treatment (marginal mean difference -0.006 , 95% CI $[-0.184, 0.172]$, $p = 0.945$). I additionally compared the mean pre-treatment Classical-Inflammatory program score between donors in different disease groups (Figure 3.22b), which also identified no statistically significant effect in any pairwise comparison.

Taken together, these findings suggest a heterogeneous enrichment of a circulating proinflammatory monocyte state following B cell depletion with ocrelizumab in a subset of pwMS, although there was no statistically significant association with treatment status or disease group within the discovery cohort in this study.

3.3 Discussion and conclusions

Depletion of B cells in pwMS is highly efficacious at reducing disease burden, indicating that these cells or the cells that they interact with are key pathogenic drivers. Using a novel analytical strategy to robustly correlate treatment effects with continuous transcriptional states across broad immune cell types, I delineated significant biological shifts in the B cell landscape and discovered a likely B cell-dependent memory T cell phenotype (T_c1 -TH) with proinflammatory, cytotoxic and tissue-homing characteristics. I found that these cells are enriched in the blood of patients with highly active RRMS at the pre-treatment baseline, in keeping with an expanded circulating reservoir of chronically activated cells with high pathogenic potential that are primed to migrate into the CNS. Moreover, I demonstrated a failure to downregulate this signature within the circulating memory T cell

compartment in longitudinally collected blood samples from a rare patient with poor therapeutic response to ocrelizumab, supporting my hypothesis that the therapeutic benefit of B cell depletion in MS may depend on dampening this T cell state.

I found a predominance of transitional naive cells within the repopulating B cell repertoire during extended interval ocrelizumab treatment, with persistent depletion of mature naive and memory B cells. Transitional naive B cells are immature and antigen-inexperienced, lacking the functional capacity to activate pathogenic T cells⁴⁹. Moreover, I observed no change in the abundance of ASCs in the longer term, with a weak and transient early depletion effect potentially accounted for by loss of recently differentiated plasmablasts which have not yet fully suppressed surface CD20 expression. The long-term preservation of ASCs supports the view that the therapeutic benefit of anti-CD20 therapies in MS is linked to mechanisms other than autoantibody production by B cells. My findings are supported by those of previous studies using flow cytometry¹⁸⁹ and mass cytometry^{190,191} to study repopulating B cells after anti-CD20 therapy. An important methodological contribution of this work is the use of an unsupervised machine learning approach to identify this phenotypic shift based on unbiased analysis of treatment-related changes in the continuous transcriptional landscape of cell states. Thus, the convergence of my findings with prior work serves as an important validation of the sensitivity and robustness of my analysis strategy.

I define T_c1-TH cells as a transcriptional phenotype of predominantly CD8⁺ memory T cells showing the strongest consistent depletion within the T cell compartment following treatment with ocrelizumab. I further demonstrate that

3. Characterising peripheral immune perturbations following B cell depletion

the temporal trajectory of this perturbation is consistent with a secondary effect due to loss of B cell support; importantly, I show that my findings cannot be solely accounted for by a direct drug effect on a minor subset of CD20^{dim} T cells which may have acquired low-level CD20 surface expression via trogocytosis during recent B cell contact or through weak transcriptional derepression of the CD20 gene (*MS4A1*) following activation. Previous studies investigating *in vivo* perturbations in non-B cell immune populations following anti-CD20 therapy in pwMS have also identified decreases in CD8⁺ memory T cell populations, although their phenotype has only been characterised using relatively limited panels of surface proteins or genes using multi-parametric flow cytometry, mass cytometry, or quantitative polymerase chain reaction (qPCR) assays^{198,203–205}, and analysis strategies constrained to discretely partitioned cell types. Crucially, the surface protein markers associated with the T_c1-TH transcriptional signature in my data show strong overlap with adhesion/migration (integrin α 4, integrin α L, CCR5, CXCR3, CXCR6) and activation (CD69, HLA-DR, PD-1) markers delineating depleted CD8⁺ T cell subpopulations in prior work, strengthening the evidence that these findings represent a convergent cell phenotype. Two previous studies^{180,181} have applied high-throughput single-cell sequencing-based approaches to profile immunologic changes after B cell depletion; however, neither study detected this cellular phenotype, possibly due to their reliance on cluster-based analyses which lack sensitivity to subtle but consistent shifts across the continuum of cell states.

Functional assays have shown that CD8⁺ T cells selectively exhibit reduced proliferation in response to EBV peptides following ocrelizumab treatment²⁰⁴, suggesting that B cell depletion leads to a preferential loss of EBV-specific cytotoxic

memory T cells. Bulk sequencing of TCR β chains from sorted CD8⁺ T cells has directly shown an enrichment of class I MHC-restricted EBV-specific sequences in the blood and CSF of pwMS compared with unaffected monozygotic twins or healthy control donors²⁶, as well as patients with other CNS neuroinflammatory diseases¹⁹². Moreover, these studies found selective reduction of these EBV-specific CD8⁺ T cells following treatment with ocrelizumab. Combined, these findings suggest that cells with a T_c1-TH phenotype in pwMS may arise in response to chronic latent EBV infection in memory B cells.

My finding of a heterogeneous enrichment of circulating proinflammatory monocytes partially aligns with previous work^{180,181,189,206}. In particular, the Classical-Inflammatory transcriptional signature delineated in this study is remarkably consistent with an IL-1B^{high} CD14⁺⁺CD16⁻ monocyte cluster shown to be enriched in the circulation following ocrelizumab treatment in a recently published longitudinal CITE-seq dataset¹⁸¹. Although the authors highlighted a six-fold increase in abundance of cells in this cluster post-treatment, this increase was largely driven by a small subset of individuals in their cohort; this observation aligns with my findings, and likely explains the absence of a statistically significant association with treatment status in this study. The mechanism of the increase in peripheral monocyte reactivity following anti-CD20 treatment remains unresolved. Plausible hypotheses include loss of anti-inflammatory cytokine signalling from a subset of regulatory B cells²⁰⁶, or reduced chemotactically-driven migration of these cells to the CNS following treatment¹⁸¹. Additionally, the inter-individual heterogeneity in the monocyte compartment shift calls into question whether this is part of a conserved pathogenic axis central to MS. Alternatively, it may represent

3. Characterising peripheral immune perturbations following B cell depletion

a bystander or homeostatic compensatory effect following B cell depletion or reflect variability in the involvement of peripheral innate immunity in CNS inflammation among individuals with the disease.

In conclusion, I confirmed that ocrelizumab induces rapid and persistent depletion of memory B cells with long-term sparing of ASCs. Additionally, by leveraging anti-CD20 therapy as an *in vivo* immune perturbation in the disease context, I identified secondary depletion of a chronically activated, proinflammatory, cytotoxic T cell state. Although the transcriptional and surface protein markers of this cell state, together with increased circulatory abundance in pwMS with higher baseline clinical disease severity, are highly consistent with a putative pathogenic role in disease, a stronger inference of causality is limited by the observational study design and detection of their depletion in the blood only, given the known CNS compartmentalisation of pathogenic processes in MS. Hence, independent cohort and cross-tissue validation studies are needed to confirm the robustness of this finding and delineate involvement of T_c1-TH cells across different disease stages.

Tracking pathogenic immune cell states across tissue compartments

4.1	Introduction	114
4.1.1	Intercompartmental differences in immune landscape between blood and CSF	114
4.1.2	Immune landscape of MS brain tissue	117
4.1.2.1	Focal lesions	117
4.1.2.2	Perivascular lymphocytic cuffs	120
4.1.2.3	Meningeal tertiary lymphoid structures	121
4.1.3	CNS-peripheral immune trafficking	122
4.1.4	Hypothesis	123
4.2	Results	123
4.2.1	T _c 1-TH signature is enriched in CSF memory T cells in MS	123
4.2.2	ASCs are enriched in CSF B cell compartment MS	127
4.2.3	T _c 1-TH signature is enriched within memory T cells at the leading edge of chronic MS brain lesions	128
4.2.4	ASCs predominate lesional B cell infiltrate in chronic MS	133
4.3	Discussion and conclusions	134

This chapter presents the results of two validation studies using independent and orthogonal publicly available single-cell datasets to examine enrichment of B cell phenotypes and the circulating disease-associated B cell-dependent T cell state identified in Chapter 3 across tissue compartments (including blood, CSF, and brain) and disease stages. The findings provide evidence for CNS migration and long-term residency of these cells, supporting a unified axis of immune dysregulation across the disease course.

4.1 Introduction

4.1.1 Intercompartmental differences in immune landscape between blood and CSF

The CNS was traditionally considered an immune-privileged site due to the presence of the BBB, which limits the free exchange of immune cells and soluble factors between the circulation and the parenchyma. However, it is now well established that the CNS is subject to continuous immune surveillance in both health and disease, mediated by controlled trafficking of immune cells across specialised vascular, lymphatic, and meningeal interfaces²¹¹. Under physiological conditions, this surveillance is tightly regulated, resulting in a distinct immune landscape within the CNS compared to the peripheral blood.

The CSF represents a key immunological interface between the peripheral immune system and the CNS. Clinically, it provides a uniquely accessible window into immune processes occurring behind the BBB, and is routinely sampled as part of the diagnostic workup for MS and a wide range of other neurological disorders.

The CSF compartment, in both health and disease, is characterised by markedly lower overall cellularity than peripheral blood, and is comprised predominantly of T cells, alongside smaller populations of B cells, myeloid cells with distinct border-associated macrophage and microglia-like phenotypes, and dendritic cells^{55,155,159,212}. Compared to peripheral blood, both CSF T cells and B cells are enriched in antigen-experienced memory subsets, with a bias towards activated phenotypes and increased expression of adhesion molecules and chemokine receptors that facilitate CNS trafficking.

CD4⁺ memory cells constitute the bulk of the CSF T cell compartment, with several studies demonstrating a bias towards T_h1 and T_h17 polarisation in both health and disease^{155,159,212}. In pwMS, the CD4⁺ compartment is enriched for follicular helper T (T_{fh}) cells compared to healthy control donors^{155,159,212,213}. Importantly, T_{fh} cells are usually found in secondary lymphoid organs, where they play a crucial role in germinal centre formation and maintenance, providing co-stimulation and cytokine support to promote B cell affinity maturation, somatic hypermutation, and class-switching to form memory B cells and ASCs; therefore, their increased abundance in MS CSF may facilitate compartmentalised B cell clonal expansion. Some studies have also reported relative increases in CD4⁺ cytotoxic effector T cells¹⁵⁹ and T_{reg}^{159,214} populations in MS CSF.

Despite the CD4⁺ predominance in the CSF T cell compartment, numerous studies have demonstrated that expanded T cell clonotypes within the CSF of pwMS are enriched for CD8⁺ T cells with an effector memory or T_{rm}-like phenotype, consistent with local antigen-driven proliferation or selective recruitment of expanded clones^{157,215,216}. Intriguingly, a key study combining scRNA-seq and TCR sequencing to directly compare the magnitude of clonal expansion across major T cell subpopulations in pwMS and healthy control donors did not identify any significant differences¹⁵⁷. Additionally, in this study, clonal analysis of TCRs did not reveal any shared clonotypes between different pwMS, but noted multiple overlapping clonotypes between pwMS and healthy control donors. However, genes related to T cell activation and cytotoxic effector polarisation were found to be more highly expressed in both expanded and unexpanded clonotypes in pwMS, across both CD4⁺ and CD8⁺ T cells. Additionally, relative expression of these gene

4. Tracking pathogenic immune cell states across tissue compartments

signatures in expanded versus unexpanded clonotypes was found to be higher in pwMS. Together, these findings suggest that the key immunological feature of MS within the CSF T cell compartment may be related to a cellular state shift within clonally expanded cells rather than disease-specific TCR clonotypes.

In pwMS, the B cell compartment within the CSF is expanded relative to health, and is skewed towards antigen-experienced and clonally expanded ASCs that contribute to intrathecal immunoglobulin production^{155,158,217,218}. Moreover, higher ASC proportion within CSF B cells have been found to be able to distinguish MS from other neuroinflammatory and infectious diseases²¹⁹, and correlate with the presence of contrast-enhancing lesions on MRI²¹⁸, despite the lack of evidence for disease-specific autoantibodies.

Importantly, while the CSF provides valuable insight into immune processes occurring within the CNS, it does not fully recapitulate the cellular and molecular composition of the parenchymal compartment, where the majority of tissue damage occurs in MS. Rather, it reflects a dynamic interface between the circulation and the CNS, capturing both infiltrating immune populations and those at the interface with CNS border structures. Additionally, a recent high-dimensional comparative study of the CSF immune landscape between MS, neurological infections, and other inflammatory and non-inflammatory neurological diseases demonstrated remarkable similarities in cellular composition and transcriptional features across these disease groups²²⁰, highlighting the possibility that global changes related to CNS inflammation may, in part, obscure disease-specific immunopathology within this compartment. Therefore, integration of immunological changes between the blood and CSF with complementary data from CNS tissue is essential for a

comprehensive understanding of disease pathogenesis.

4.1.2 Immune landscape of MS brain tissue

While the CSF provides a valuable window into immune activity at the CNS interface, the primary site of neuroinflammation and consequent tissue injury in MS is within the brain parenchyma. Immunopathological changes in MS brains comprise multiple distinct but interacting inflammatory niches, differing in their cellular composition and evolution across disease stages (Figure 4.1). In contrast to the relative accessibility of the CSF, analysis of brain tissue is largely restricted to post-mortem studies, but these have provided crucial insight into the spatial organisation of inflammatory processes and the cellular phenotypes involved in the chronic, smouldering inflammation driving PIRA, which predominates in advanced MS.

4.1.2.1 Focal lesions

Focal cortical and subcortical white matter lesions are the defining pathological hallmark of MS and represent discrete regions of demyelination, axonal injury and immune cell infiltration. Historically, these lesions were classified into four distinct immunopathological patterns on the basis of the specific composition of inflammatory cellular infiltrates, complement deposition, anatomical pattern of demyelination and features of oligodendrocyte dystrophy²²¹. Although this classification provided important early evidence for heterogeneity in lesion biology, more recent work has shifted the emphasis toward markers of lesion chronicity and activity. Chronic lesions identified in post-mortem MS brains are typically characterised by a demyelinated and hypocellular core, and a relatively hypercellular

4. Tracking pathogenic immune cell states across tissue compartments

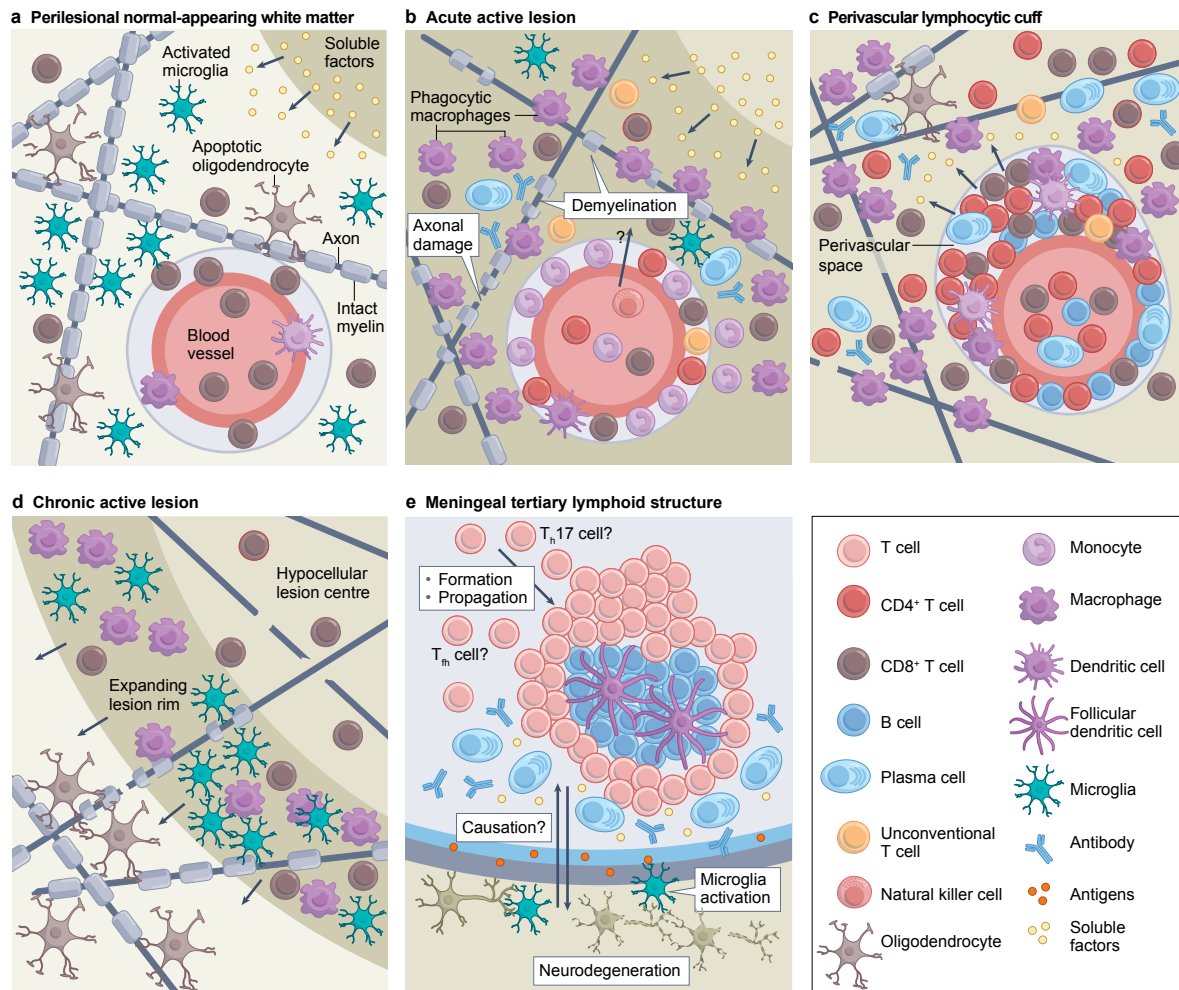


Figure 4.1. CNS immunological niches in pwMS. Schematic diagram illustrating distinct lesion stages and compartmentalised CNS immunological niches identified in studies of post-mortem brain tissue from pwMS. **a**, Perilesional NAWM contains infiltrates of predominantly $CD8^+$ T cells and activated microglia, along with dysfunctional neuronal and glial cells, without demyelination or axonal loss. **b**, Acute active lesions exhibit diffuse demyelination and axonal loss with phagocytic macrophages and $CD8^+$ T cell-predominant lymphocytic infiltrate. **c**, Peripherally-derived immune cells become compartmentalised within the CNS in perivascular cuffs from early stages of disease. **d**, Chronic active or mixed active/inactive lesions comprise a hypocellular demyelinated lesion core with an expanding rim of smouldering inflammation containing a prominent infiltrate of activated microglia and macrophages, alongside predominantly $CD8^+$ T cells. **e**, Meningeal tertiary lymphoid structures, which resemble germinal centres in secondary lymphoid tissues, are found in a subset of pwMS with progressive disease and are enriched in follicular dendritic cells and T cells which may support local B cell maturation, clonal expansion, and differentiation into ASCs. Adapted with permission from Attfield KE *et al.*¹¹.

rim representing the leading edge of demyelination and neurodegeneration. Chronic lesion rims vary in their cellular composition and inflammatory activity, and are broadly divided into: (i) chronic active (also known as mixed active/inactive) and (ii) chronic inactive lesions. In chronic active lesions, smouldering inflammatory activity persists at the lesion edge, consisting mainly of activated microglia and macrophages phagocytosing myelin debris, along with a T cell-predominant lymphocytic infiltrate. In contrast, chronic inactive lesions lack activated phagocytic myeloid cells and demonstrate less extensive lymphocytic infiltrates. Perilesional NAWM is also now widely recognised to be biologically altered compared to white matter from healthy control donors, containing inflammatory cellular infiltrates, along with neuronal and glial phenotypic alterations which extend beyond the histologically apparent lesion border. These distinct intra-lesion niches likely represent a spectrum of evolving pathology in chronic MS, characterised by an expanding rim of smouldering inflammation that may be sustained by inflammatory cell infiltrates from the surrounding NAWM, and which may eventually burn out. Despite the canonical temporal sequence of chronic lesion progression inferred from histopathological studies, *in vivo* radiological evidence of new contrast enhancement at the rims of previously quiescent lesions may suggest a more dynamic continuum of lesion stages, where inactive lesions may be re-infiltrated by activated macrophages and microglia and revert to the active stage⁶².

The modern lesion taxonomy is particularly relevant to understanding the mechanisms of PIRA, which has been linked to the presence of increased numbers of chronic active lesions with lipid- and iron-laden microglia and macrophages in post-mortem pathology studies^{61,63}; these are also detectable *in vivo* by the

4. Tracking pathogenic immune cell states across tissue compartments

presence of lesions with hypointense borders on iron-sensitive MRI sequences, known as paramagnetic rims^{63,222}. Recent modern high-resolution single-cell and spatial transcriptomics studies of the chronic lesion rim microenvironment have found evidence that T cells and disease-associated activated microglia engage in local cytokine-mediated interactions that may sustain low-grade inflammation^{66,68}. Among the adaptive immune cells present within these lesions, CD8⁺ T cells consistently predominate across studies^{42–45,63,68}. Moreover, a substantial fraction of these cells exhibit T_{rm} differentiation markers, suggesting that they are retained long-term within the tissue rather than continually recirculating through the peripheral immune compartment⁵⁵. Therefore, establishing the origin, antigen specificity, and cellular phenotype of these cells, as well as local intercellular communication networks, represents a key priority for understanding the cellular basis of chronic compartmentalised inflammation in MS.

4.1.2.2 Perivascular lymphocytic cuffs

Perivascular spaces, which are continuations of the CSF-containing subarachnoid space that surround small vessels penetrating the parenchyma, are another important immunological niche within the CNS. In MS, immune cells frequently accumulate around post-capillary venules within the brain and spinal cord, forming perivascular lymphocytic cuffs containing mixtures of T cells, B cells, and myeloid cells^{53–55}. Perivascular cuffs are increasingly understood as distinct microenvironments in which adaptive and innate immune cells can become sequestered and interact locally¹¹. Notably, this compartment is enriched for T_{rm} cells, with CD8⁺ T cells substantially outnumbering their CD4⁺ counterparts, closely mirroring the composition of lesional

lymphocytic infiltrates and contrasting with the CD4⁺ predominance of the broader CSF compartment⁵⁵. Their persistence across lesion stages suggests that they may contribute to both initiation and maintenance⁵⁴. Crucially, perivascular cuffs may provide a tissue niche in which circulating pathogenic memory T cells primed for tissue homing and residency may be reactivated and retained following antigen re-encounter within the CNS, thereby representing a plausible anatomical bridge between peripherally activated immune cell states and tissue-resident populations driving chronic neuroinflammation.

4.1.2.3 Meningeal tertiary lymphoid structures

Beyond the perivascular compartment, inflammatory processes in MS also become organised within the meninges, where tertiary lymphoid structures (also known as ectopic lymphoid follicles) have been described in a subset of pwMS with progressive disease^{56–60}. These structures resemble germinal centres in secondary lymphoid tissues in their organisation and cellular composition, and are thought to support local B cell activation, antigen presentation, and differentiation into ASCs.

The presence of meningeal tertiary lymphoid structures is also intimately linked to grey matter pathology, including subpial cortical demyelination and diffuse neuronal injury^{57–59}. Their existence offers an additional plausible mechanism by which immune activity becomes progressively insulated from peripheral immune modulation while remaining capable of driving ongoing tissue injury, which may explain why PIRA becomes more prominent in advanced disease and increasingly refractory to peripheral immunomodulation.

4.1.3 CNS–peripheral immune trafficking

The relationship between the peripheral immune system and the CNS is characterised by dynamic, bidirectional trafficking of immune cells across specialised anatomical interfaces²¹¹. Normally, this process underpins CNS immune surveillance, allowing antigen-experienced lymphocytes to access and respond to local infection or tissue injury. In MS, however, this process becomes dysregulated, resulting in enhanced recruitment, retention and reactivation of immune cells within the CNS.

Entry of immune cells into the CNS primarily occurs at the level of post-capillary venules, where interactions between adhesion molecules on lymphocytes and endothelial cells facilitate transmigration across the BBB. Key molecular pathways implicated in this process include interactions between adhesion molecules such as VLA-4 (integrin $\alpha4\beta1$) on the surface of lymphocytes and endothelial VCAM-1²¹¹, as well as chemokine gradients that direct the migration of cells expressing receptors such as CCR5, CXCR3, CXCR6, and CX3CR1²²³.

In addition to lymphocyte entry into the CNS, there is also evidence for egress of immune cells from the CNS via meningeal lymphatic vessels²²⁴ and drainage into deep cervical lymph nodes²²⁵, providing a pathway for antigen presentation and immune communication between compartments. This bidirectional trafficking supports the concept that the CNS is not an immunologically isolated organ, but rather part of an integrated immune network in which peripheral and tissue-resident immune responses are interlinked.

In this context, circulating immune cells expressing tissue-homing markers are more likely to access and persist within the CNS. This provides a conceptual

framework linking peripheral immune phenotypes to CNS-resident populations, and suggests that disease-relevant cellular states may be identifiable in the circulation prior to their accumulation within pathological tissue niches. Understanding how these circulating immune cell states relate to their tissue-resident counterparts is therefore essential for identifying the cellular drivers of disease and for understanding both the mechanisms and limitations of therapies that act primarily on the peripheral immune compartment.

4.1.4 Hypothesis

In Chapter 3, I identified a proinflammatory, cytotoxic and tissue-homing memory T cell signature (T_c1 -TH) in the blood that is enriched in clinically aggressive MS and depleted following treatment with ocrelizumab. Based on the transcriptional and surface protein phenotype of these cells, which was notably enriched for tissue-homing and retention markers, I hypothesised that these cells may represent a circulating reservoir of memory T cells with high pathogenic potential that may localise to the CNS, become reactivated, and contribute to chronic inflammatory pathology. Additionally, as this cellular state appears to be B cell-dependent in the periphery, co-localisation with disease-associated B cell phenotypes within the CNS would provide evidence to support an orchestrated pathogenic B–T cell axis.

4.2 Results

4.2.1 T_c1 -TH signature is enriched in CSF memory T cells in MS

To investigate the CNS-homing capability of circulating T_c1 -TH cells, I used a published compendium of scRNA-seq studies of mononuclear cells from both

4. Tracking pathogenic immune cell states across tissue compartments

peripheral blood and CSF across neurological diseases to obtain data from five public datasets^{155–159} with paired PBMC and CSF samples from the same donors (Figure 4.2).

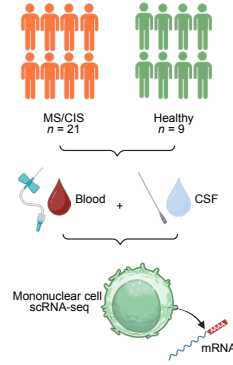


Figure 4.2. Validation cohort of pwMS and healthy control donors with paired PBMC / CSF scRNA-seq samples. Graphical summary of CSF/blood validation dataset of publicly available scRNA-seq studies with paired blood and CSF samples obtained from treatment-naive MS / CIS ($n = 21$) or healthy control donors ($n = 9$).

After preprocessing the raw count matrices and transferring cell type annotations from the discovery dataset (see Section 2.5.2.1 and Section 2.5.2.2), the final integrated dataset contained 178,027 cells from 30 treatment-naive individuals (Figure 4.3), with either established MS or its prodromal stage CIS ($n = 21$, 18 with paired PBMC and CSF samples) or healthy control donors ($n = 9$, all with paired PBMC and CSF samples).

Using starCAT¹⁴², I projected T cell-specific gene programs from the discovery dataset (see Section 2.5.2.3) and computed T_c1 -TH program scores in T cells from the validation dataset (Figure 4.4a–b). Using a linear mixed effects model (see Section 2.6), I identified significant enrichment of this signature within memory T cells in the CSF (Figure 4.4c) in MS / CIS (marginal mean difference 0.0438, 95% CI [0.0216,0.0661], $p = 7.65 \times 10^{-4}$) but not healthy control (marginal mean

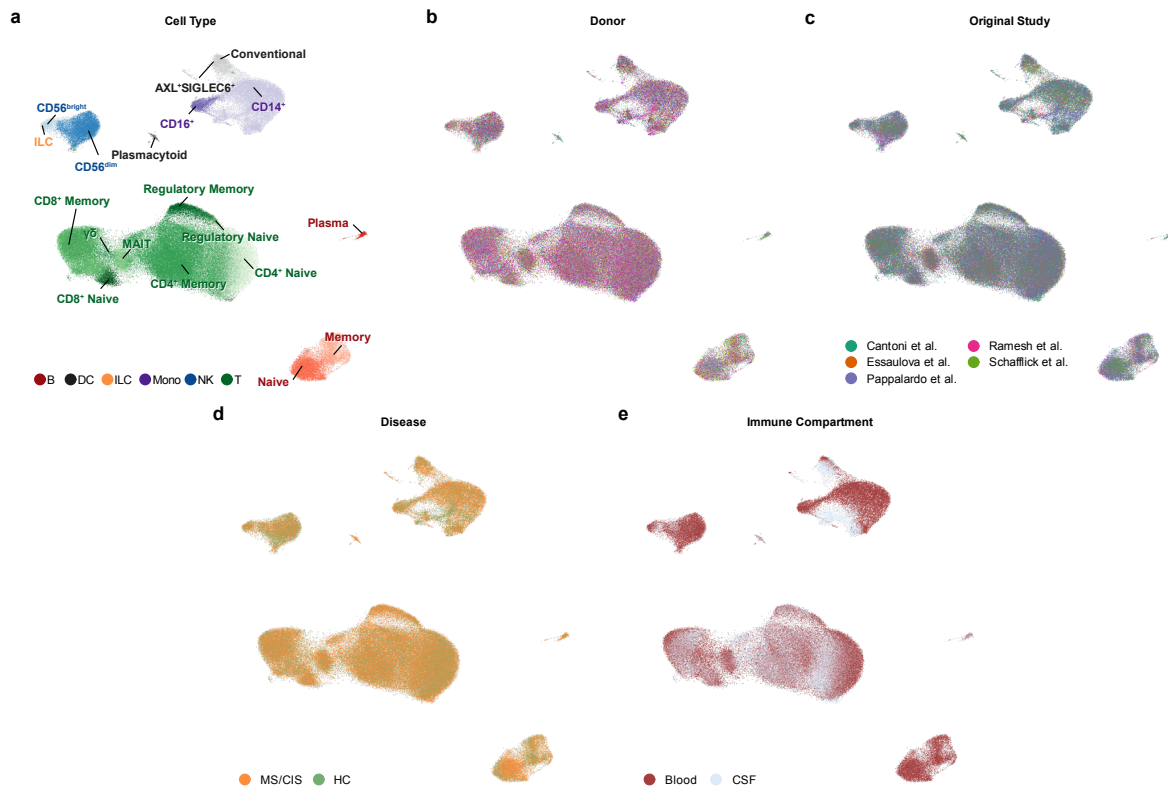


Figure 4.3. CSF/blood mononuclear cell validation dataset integration and cell type annotation. UMAP visualisation of 178,027 cells comprising the CSF/blood mononuclear cell validation dataset coloured by reference-based cell type annotation from the discovery dataset (a), donor (b), original study (c), disease type (d) and immune compartment (e).

4. Tracking pathogenic immune cell states across tissue compartments

difference 0.0012, 95% CI $[-0.0310, 0.0784]$, $p = 0.938$) donors. In addition, the magnitude of the intercompartment difference, calculated as the CSF : PBMC log ratio of mean T_c1-TH program score (Figure 4.4d), was greater in patients with MS / CIS compared to healthy control donors (marginal mean difference 0.578, 95% CI $[0.138, 1.02]$, $p = 0.012$).

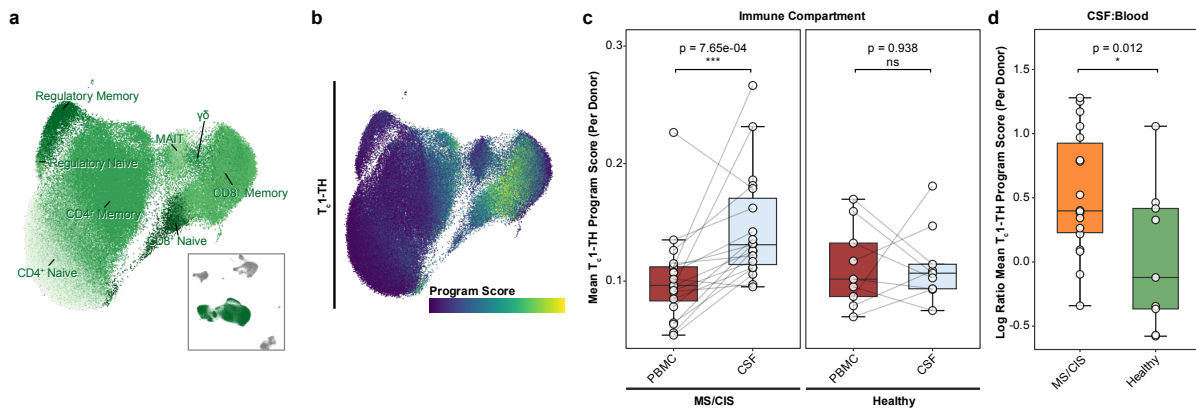


Figure 4.4. T_c1-TH signature is enriched among memory T cells in CSF of pwMS. **a**, UMAP visualisation of T cells, which are shown highlighted (dark green) in the UMAP visualisation of all cells (inset), coloured and labelled by coarse (level 2) cell type. **b**, Program score for the T_c1-TH signature projected on the UMAP visualisation of T cells in the CSF/blood mononuclear cell validation dataset; upper limits of colour scales are clipped at 99th percentiles. **c**, Box plots comparing mean T_c1-TH program scores per donor (y -axis) in each compartment (x -axis) between MS / CIS (left panel) and healthy control donors (right panel); lines connect paired data from the same individual. **d**, Box plots comparing CSF:blood log ratio of mean T_c1-TH program score per donor (y -axis) between MS / CIS and healthy control donors. Statistical testing was performed using linear mixed effects models (within-donor repeated measures) or standard linear models (no within-donor repeated measures) adjusted for age and sex; p -values were calculated with t -tests of marginal means (using Satterthwaite degrees of freedom for mixed effects models) and adjusted for multiple testing using the Benjamini–Hochberg method.

Taken together, these findings strongly support the hypothesis that cells with a T_c1-TH phenotype in the circulation cross the BBB and are enriched in the CSF memory T cell repertoire compared to the circulation in pwMS versus healthy control donors, possibly reflecting compartmentalised expansion of infiltrating cells or a greater chemotactic drive recruiting these cells to the CNS in the disease context.

4.2.2 ASCs are enriched in CSF B cell compartment MS

Next, I investigated the compositional differences in the B cell compartment between peripheral blood and CSF in MS / CIS compared to healthy control donors.

Using starCAT¹⁴², I projected B cell-specific gene programs from the discovery dataset (see Section 2.5.2.3) to B cells from the validation dataset. I derived composite gene program scores to quantify enrichment of major B cell subpopulations by summing the scores within each cell for specific programs expressed predominantly in naive B cells (*Transitional Naive* and *Mature Naive*), memory B cells (*Homeostatic Memory*, *Primed Memory*, *Atypical Memory*, and *Activated Memory*) or ASCs (*Early ASC*, *IgA+ ASC*, and *IgG+ ASC*). As expected, these composite gene program scores aligned closely with the automated cell type label transfer annotations from the discovery dataset (Figure 4.5a–d), but facilitated more precise quantitative modelling of B cell state shifts between compartments. Using a linear mixed effects model (see Section 2.6), I identified significant depletion of the naive B signature within B cells in the CSF (Figure 4.5e) in both MS / CIS (marginal mean difference -0.289 , 95% CI $[-0.340, -0.239]$, $p = 5.56 \times 10^{-12}$) and healthy control (marginal mean difference -0.194 , 95% CI $[-0.268, -0.120]$, $p = 1.27 \times 10^{-5}$) donors. Conversely, there was significant enrichment of the memory B signature within B cells in the CSF (Figure 4.5f) in both MS / CIS (marginal mean difference 0.207 , 95% CI $[0.157, 0.256]$, $p = 6.00 \times 10^{-9}$) and healthy control (marginal mean difference 0.183 , 95% CI $[0.112, 0.254]$, $p = 1.52 \times 10^{-5}$) donors. In contrast, the ASC signature was significantly enriched in the CSF (Figure 4.5g) in MS / CIS (marginal mean difference 0.101 , 95% CI $[0.074, 0.129]$, $p = 3.61 \times 10^{-9}$) but not healthy control

4. Tracking pathogenic immune cell states across tissue compartments

(marginal mean difference -0.0007 , 95% CI $[-0.0414, 0.0401]$, $p = 0.976$) donors. Comparing the magnitude of the intercompartment difference, calculated as the CSF : PBMC log ratio of mean program score (Figure 4.5h-j), only the ASC signature was found to be significantly different between patients with MS / CIS compared to healthy control donors (marginal mean difference 2.75 , 95% CI $[1.37, 4.12]$, $p = 4.12 \times 10^{-4}$).

In summary, these findings are concordant with previously reported inversion of the naive:memory B cell ratio within the CSF compared to peripheral blood, but suggest that this is a conserved intercompartmental immune compositional shift across both pwMS and healthy control donors. However, my analysis confirms a striking expansion of ASCs within the CSF B cell compartment in pwMS but not healthy control donors, in keeping with active B cell maturation and differentiation within the CNS in the disease context.

4.2.3 T_c1-TH signature is enriched within memory T cells at the leading edge of chronic MS brain lesions

To test the hypothesis that T_c1-TH cells may represent a circulating precursor state of pathogenic dysregulated T_{rm} cells in brain tissue in chronic MS, I identified two public datasets^{66,67} of snRNA-seq from post-mortem brain tissue of pwMS with annotated samples from different intra-lesion niches, including demyelinated core, rims of both chronic active and inactive lesions, and perilesional NAWM; both datasets also included white matter samples from healthy control donors with no known neurological disease (Figure 4.6).

After preprocessing the raw count matrices, extracting lymphocyte clusters and

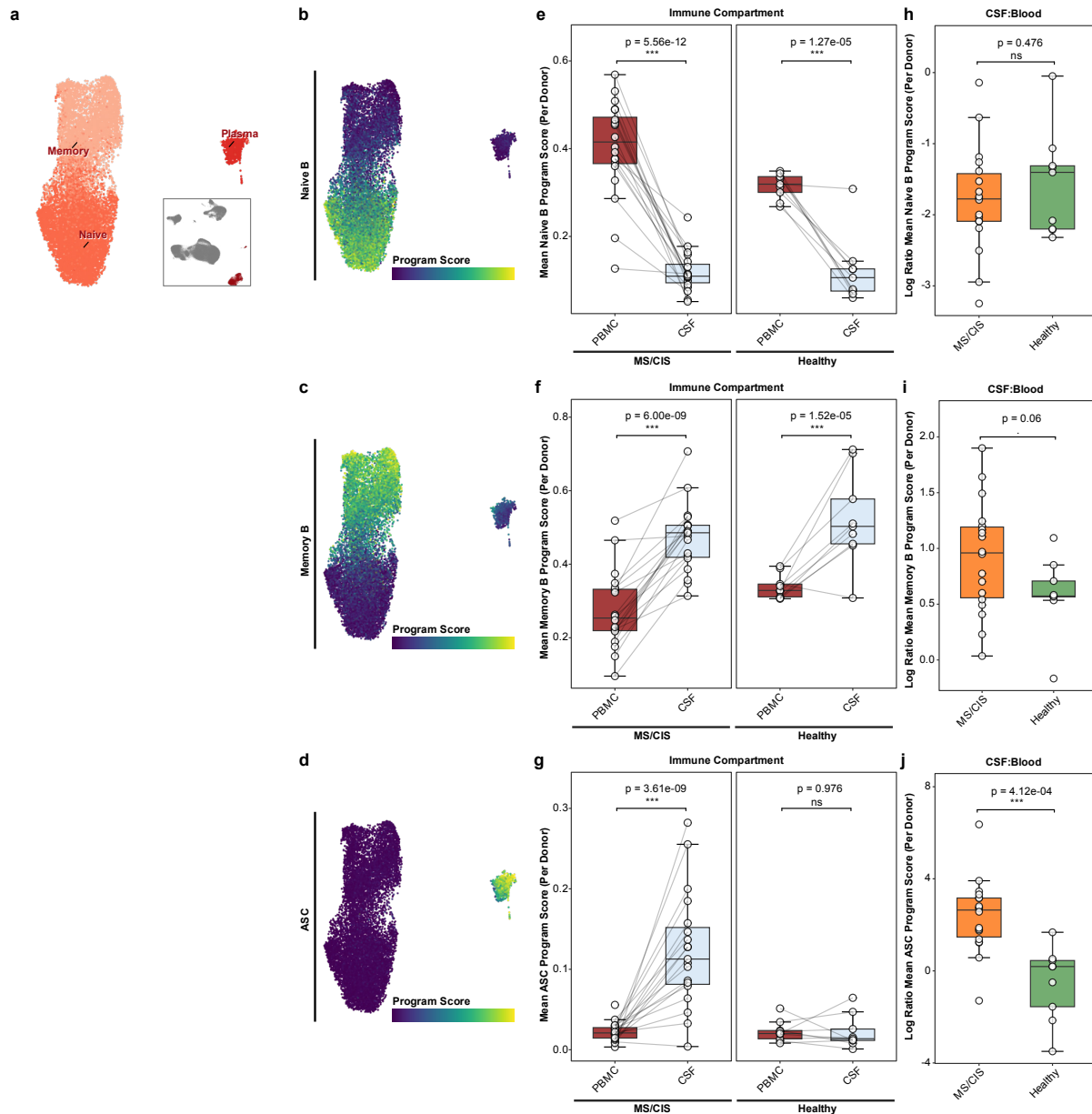


Figure 4.5. ASC signature is enriched among B cells in CSF of pwMS. **a**, UMAP visualisation of B cells, which are shown highlighted (dark red) in the UMAP visualisation of all cells (inset), coloured and labelled by coarse (level 2) cell type. **b–d**, Program scores for the naive B, memory B and ASC signatures, respectively, projected on the UMAP visualisation of B cells in the CSF/blood mononuclear cell validation dataset; upper limits of colour scales are clipped at 99th percentiles. **e–g**, Box plots comparing mean naive B, memory B, and ASC program scores, respectively, per donor (*y*-axis) in each compartment (*x*-axis) between MS / CIS (left panel) and healthy control donors (right panel); lines connect paired data from the same individual. **h–j**, Box plots comparing CSF: blood log ratio of mean naive B, memory B, and ASC program scores, respectively, per donor (*y*-axis) between MS / CIS and healthy control donors. Statistical testing was performed using linear mixed effects models (within-donor repeated measures) or standard linear models (no within-donor repeated measures) adjusted for age and sex; *p*-values were calculated with *t*-tests of marginal means (using Satterthwaite degrees of freedom for mixed effects models) and adjusted for multiple testing using the Benjamini–Hochberg method.

4. Tracking pathogenic immune cell states across tissue compartments

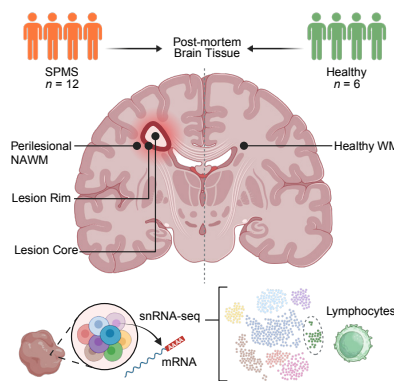


Figure 4.6. Validation cohort of pwMS and healthy control donors with post-mortem brain snRNA-seq samples from lesion subregions/healthy white matter. Graphical summary of brain tissue validation dataset of publicly available post-mortem brain snRNA-seq data obtained from SPMS ($n = 12$) and healthy control donors ($n = 6$). SPMS samples were taken from subregions of chronic active/inactive demyelinating white matter lesions (core, rim, perilesional NAWM) and samples of normal white matter were taken from healthy control donors. Analyses were restricted to lymphocyte clusters extracted from the full snRNA-seq datasets.

transferring cell type annotations from the discovery dataset (see Section 2.5.2.1 and Section 2.5.2.2), the final integrated dataset contained 1,417 cells from 18 individuals (Figure 4.7), either SPMS ($n = 12$) or healthy control donors ($n = 6$).

I first examined the overall composition of the combined T cell compartment across all brain tissue samples in the dataset based on automated cell type label transfer annotations from the discovery dataset (Figure 4.8a), which revealed a predominance $CD8^+$ and $CD4^+$ memory cells (68.5%, $CD4:CD8$ ratio 0.49). Additionally, 14.8% of T cells were annotated as $CD4^-CD8^-$ (double negative) T cells, which constitute a minor fraction of circulating T cells but have been shown to be expanded in brain tissue in the context of CNS inflammation²²⁶. Notably, 12.6% of T cells were annotated as $CD4^+$ naive T cells; this most likely represents an overestimate due to inaccurate automated cell type label transfer in the context of the known technical challenge of differentiating $CD4^+$ naive

and central memory subpopulations using transcriptomic data, which may be exacerbated by the additional sparsity and dropout inherent in snRNA-seq datasets compared to scRNA-seq.

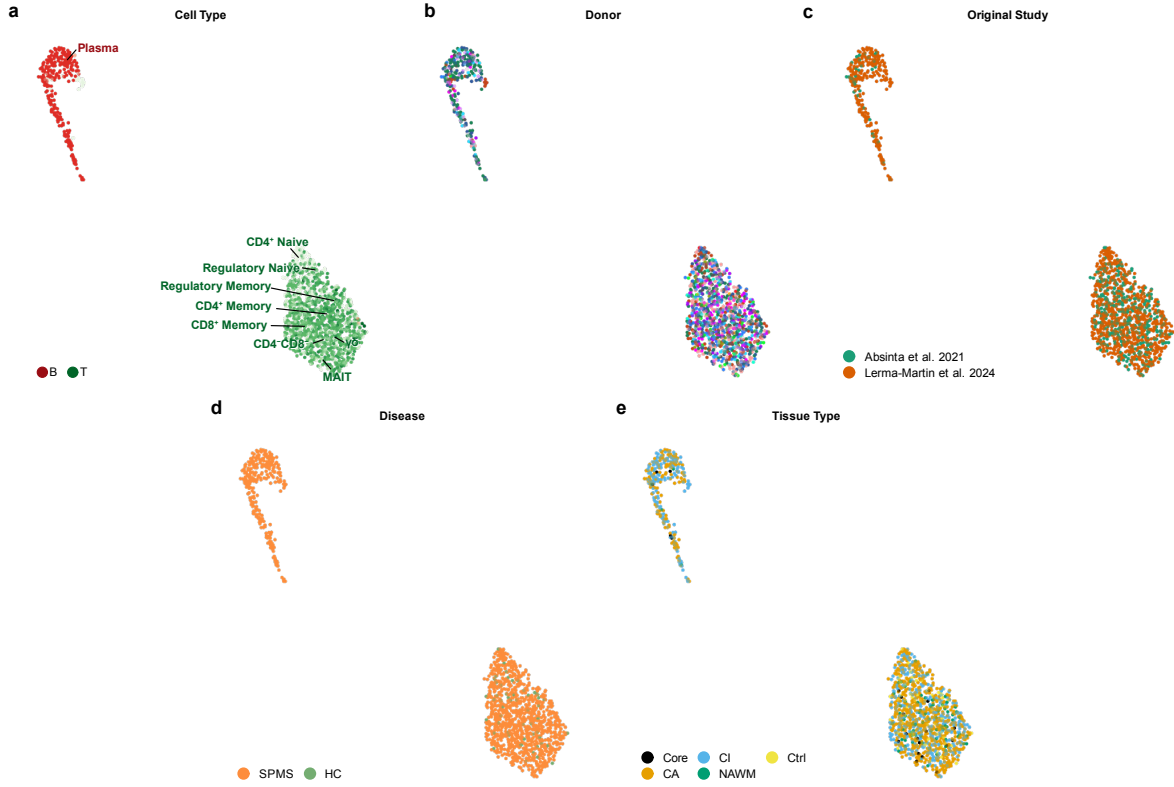


Figure 4.7. Brain tissue lymphocytes validation dataset integration and cell type annotation. UMAP visualisation of 1,417 cells comprising the brain tissue lymphocytes validation dataset coloured by reference-based cell type annotation from the discovery dataset (a), donor (b), original study (c), disease type (d) and tissue type (e).

I subsequently used starCAT¹⁴² to project T cell-specific gene programs from the discovery dataset (see Section 2.5.2.3) and computed T_c1 -TH program scores in T cells from the validation dataset. Due to the relatively low abundance of T cells in brain tissue samples, several donor–tissue type combinations had fewer than 10 cells detected, making donor-level mean aggregation of program scores unreliable. Therefore, I compared cell-level program scores across tissue types using a linear mixed effects model with donor as a random effect to mitigate against

4. Tracking pathogenic immune cell states across tissue compartments

pseudoreplication (see Section 2.6). Overall, I detected higher T_c1 -TH signature expression in memory T cells from MS lesion rims and perilesional NAWM compared to both MS lesion cores and white matter from healthy control donors (Figure 4.8b). Statistically significant differences were detected in pairwise comparisons between lesion cores and chronic active lesion rims (marginal mean difference 0.0739, 95% CI [0.0263, 0.122], $p = 0.0241$), chronic inactive lesion rims (marginal mean difference 0.0663, 95% CI [0.0153, 0.117], $p = 0.0341$), and perilesional NAWM (marginal mean difference 0.0704, 95% CI [0.0189, 0.122], $p = 0.0341$), as well as between healthy white matter and chronic active lesion rims (marginal mean difference 0.0616, 95% CI [0.0134, 0.110], $p = 0.0341$).

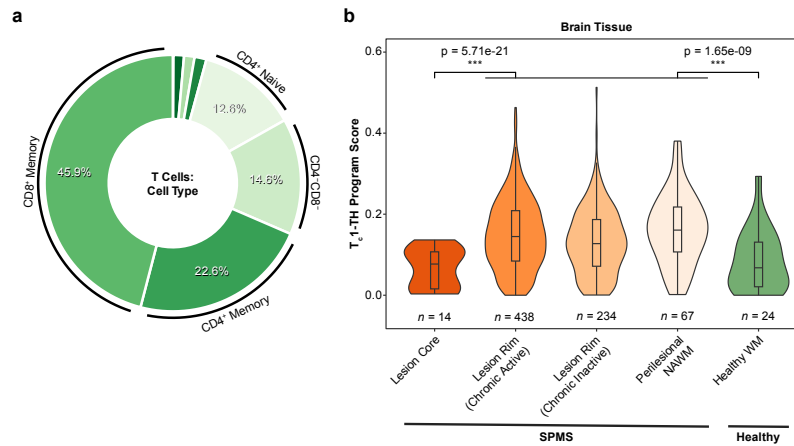


Figure 4.8. T_c1 -TH signature is enriched among memory T cells in chronically inflamed brain tissue of pwMS. **a**, Distribution of coarse (level 2) cell type labels within T cells from brain tissue lymphocytes validation dataset. **b**, Violin/box plots comparing T_c1 -TH program score (y -axis) in T cells from different subregions (x -axis); text labels above the x -axis indicate the number of T cells from each subregion in the dataset. Statistical testing was performed using linear mixed effects models adjusted for age and sex; p -values were calculated with t -tests of marginal means (using Satterthwaite degrees of freedom) and adjusted for multiple testing using the Benjamini–Hochberg method.

This finding, together with the surface protein marker profile of T_c1 -TH cells delineated in Chapter 3, supports the hypothesis that these cells represent a circulating T_{rm} precursor state.

4.2.4 ASCs predominate lesional B cell infiltrate in chronic MS

Given the expansion of ASCs within the B cell compartment of MS CSF, I sought to evaluate whether this compositional bias was also present within post-mortem brain tissue of pwMS with end-stage disease.

I first examined the overall composition of the combined B cell compartment across all brain tissue samples in the dataset based on automated cell type label transfer annotations from the discovery dataset (Figure 4.9a). There were markedly fewer B cells compared to T cells captured in the validation dataset, with no B cells detected in healthy control white matter samples. Strikingly, among the remaining samples, 98.9% of all B cells detected were annotated as plasma cells, suggesting either fully differentiated ASCs, or memory B cells actively differentiating and upregulating ASC programs, predominate the brain parenchymal B cell infiltrate in end-stage MS.

I subsequently used starCAT¹⁴² to project B cell-specific gene programs from the discovery dataset (see Section 2.5.2.3) and computed ASC program scores (see Section 4.2.2) in B cells from the validation dataset. As with T cells, the relatively low abundance of B cells in brain tissue samples precluded donor-level mean aggregation of program scores, necessitating cell-level comparison of program scores using a linear mixed effects model with donor as a random effect to mitigate against pseudoreplication (see Section 2.6). Overall, I found higher ASC signature expression in B cells from MS lesion rims compared to both lesion cores and perilesional NAWM (Figure 4.8b). However, no statistically significant differences were detected in pairwise comparisons between individual subregions following

multiple testing correction.

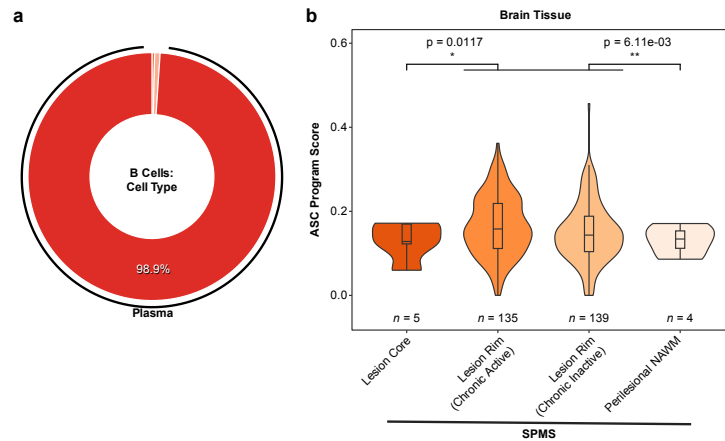


Figure 4.9. ASC signature is enriched among B cells in chronically inflamed brain tissue of pwMS. **a**, Distribution of coarse (level 2) cell type labels within B cells from brain tissue lymphocytes validation dataset. **b**, Violin/box plots comparing ASC program score (*y*-axis) in B cells from different subregions (*x*-axis); text labels above the *x*-axis indicate the number of B cells from each subregion in the dataset. Statistical testing was performed using linear mixed effects models adjusted for age and sex; *p*-values were calculated with *t*-tests of marginal means (using Satterthwaite degrees of freedom) and adjusted for multiple testing using the Benjamini–Hochberg method.

These findings support a predominance of ASCs in the B cell infiltrate of end-stage MS brains, which is likely enriched in chronically inflamed lesion rims. However, the low abundance of B cells in snRNA-seq from brain tissue limits the statistical power to detect subtle variation across different intra-lesion niches.

4.3 Discussion and conclusions

The validation analyses presented in this chapter support a potential pathogenic role for the T_c1-TH phenotype from multiple complementary perspectives. Overall, these findings suggest that circulating memory T cells with this phenotype may preferentially migrate to and persist within the CNS in pwMS. Importantly, the same transcriptional signature that delineates the circulating T cell state downregulated secondary to B cell depletion is enriched in the CSF of patients with

relapsing MS, as well as within chronically inflamed brain lesions from individuals with advanced progressive disease. Together, these observations suggest that the T_c1-TH phenotype may represent a conserved axis of immune dysregulation across tissue compartments and disease stages.

The enrichment of this phenotype within the CSF aligns with a growing body of evidence demonstrating that this compartment is enriched for clonally expanded cytotoxic T cells in MS compared to healthy controls^{157,215,216}. Notably, recent work integrating scRNA-seq with TCR sequencing in CSF T cells applied a range of both unbiased and targeted antigen discovery approaches to deeply interrogate clonally expanded CD8⁺ TCRs in MS²¹⁶. Despite this approach, antigen specificity remained unresolved for the majority of expanded clonotypes, likely reflecting the inherent technical challenges of computational sequence-based prediction or experimental determination of MHC-restricted TCR specificity. However, three distinct clonotypes exhibited donor genotype-concordant MHC class I-restricted reactivity to EBV peptides, with no evidence of cross-reactivity to other viral antigens or self-peptides. Importantly, the transcriptional profile of these clonally expanded CD8⁺ T cells, featuring markers of cytotoxic effector polarisation (*CD8A*, *CD8B*, *EOMES*, *GZMK*, *GZMA*, *GZMH*, *NKG7*, *PRF1*), chronic activation/antigen exposure (*PDCD1*, *TIGIT*, *LAG3*, *HOPX*, *DUSP2*), tissue homing (*CCL4*, *CCL5*, *CCR5*, *CXCR3*), and tissue residency (*ITGAE*, *CD69*), shows substantial overlap with the T_c1-TH signature, providing orthogonal support for the hypothesis that depletion of this signature within the circulating T cell repertoire following anti-CD20 treatment represents contraction of a B cell-dependent chronically activated, tissue-homing, and potentially EBV-directed cytotoxic T cell state that preferentially

4. Tracking pathogenic immune cell states across tissue compartments

migrates to the CNS and undergoes clonal expansion following antigen re-encounter.

The findings presented here also establish a link between peripherally originating T_c1-TH cells and tissue-resident T cell populations within the CNS. A large proportion of CD8⁺ T cells in MS brains^{43–45,55} and CSF^{157,215,216} have been found to exhibit a T_{rm} phenotype, functionally indicative of tissue retention and lack of re-circulation in the peripheral immune cell repertoire. These cells are characterised by high expression of markers including CD69, integrin α E, integrin β 7, integrin α 1, CD44, CCR5, CXCR6, CXCR3, CX3CR1, GPR56 and PD-1^{45,64}. Strikingly, the majority of these surface proteins are also markers I found to be differentially expressed in circulating T_c1-TH phenotype cells compared to other memory T cells in Chapter 3. Accordingly, analyses in this chapter demonstrate that memory T cells extracted from snRNA-seq of post-mortem MS brain tissue at the leading edge of chronic inflammatory lesions and adjacent NAWM are enriched for the same transcriptional signature, supporting the hypothesis that T_c1-TH cells represent a circulating precursor state of CNS-resident T_{rm} cells in MS. This interpretation is further supported by prior studies demonstrating the existence of circulating memory T cells primed for tissue residency through coexpression of transcription factors HOBIT (*ZNF683*) and BLIMP-1 (*PRDM1*)^{227,228}, both of which feature highly in the T_c1-TH transcriptional program.

Within the CNS, CD8⁺ T_{rm} cells have been shown to localise in proximity to EBV-infected B cells within perivascular cuffs and meningeal tertiary lymphoid structures^{60,64,229}, suggesting they may be differentiating from circulating EBV-specific CD8⁺ T cells which have migrated into the CNS and re-encountered their cognate antigen within these niches. Recently, high-resolution single-cell

spatial transcriptomics has delineated a pathogenic interferon- γ -mediated signalling axis between CD8⁺ T_{rm} cells and dysfunctional inflammatory microglia within chronic active lesion rims⁶⁸, highlighting a key role played by these cells in driving smouldering, compartmentalised neuroinflammation in MS.

Consistent with prior studies, the findings presented here demonstrate expansion of ASCs within the CSF in pwMS^{217–219}. The CSF compartment is known to be enriched for memory B cells relative to the blood. In later stages of MS, these cells become increasingly compartmentalised within CNS niches such as meningeal tertiary lymphoid structures^{56–60} and perivascular lymphoid cuffs^{53–55}. These structures may support local B cell maturation and differentiation into ASCs through interactions with dendritic cells, T cells, and the local microenvironment¹¹. Intrathecal ASCs are also clonally expanded^{158,218,230} and are the source of oligoclonal bands, which represent clonal immunoglobulin synthesis. However, the ubiquitous intracellular targets and heterogeneous donor-specific repertoire of these antibodies^{187,231} argue that they are more likely to be a consequence of tissue damage rather than a primary driver of pathology.

The predominance of terminally differentiated ASCs in chronic CNS lesions in end-stage SPMS demonstrated here is also consistent with existing literature⁴⁴. However, the lack of broadly prevalent disease-specific autoantibodies^{48–50} and paradoxical worsening of disease observed in a clinical trial of atacicept⁹⁷, a recombinant inhibitor of B cell survival factors BAFF and APRIL, which depletes circulating ASCs with relative sparing of memory B cells⁹⁸, argue against a primary pathogenic role for ASCs in MS. Notably, differentiation of memory B cells into ASCs has been associated with induction of the EBV lytic cycle²³², increasing

4. Tracking pathogenic immune cell states across tissue compartments

the availability of immunogenic viral antigens in their local environment. This observation may suggest a mechanism for local recruitment and activation of clonally expanded CNS-resident cytotoxic T_{rm} cells, which can then drive persistent chronic tissue injury through direct cytotoxic effects, or the secretion of proinflammatory mediators that recruit and activate additional immune cells, including microglia and macrophages^{66,68}.

Taken together, these findings build support for a model of MS pathogenesis in which a proinflammatory, cytotoxic, tissue residency-primed cytotoxic T cell phenotype arises in the periphery through B cell-dependent interactions, with subsequent co-migration into the CNS, clonal expansion, and differentiation, contributing to the establishment of a persistent pathogenic immune niche that drives compartmentalised inflammation. However, if $T_{\text{c}}1$ -TH cells represent a key conserved driver of disease in pwMS, they would be expected to be modulated not only by anti-CD20 therapies but also by other DMTs. Indeed, several MS DMTs have been shown to target memory B cells⁹⁶, modulate T cell activation and polarisation away from type 1 immunity (e.g. fumaric acid esters), or interfere with CNS migration of tissue-homing lymphocytes (e.g. natalizumab, S1PR modulators), raising the possibility that convergent therapeutic effects across mechanistically distinct treatments may reflect a shared impact on this cell state. Cross-therapy validation studies therefore represent an important additional avenue to further strengthen the evidence for the role of $T_{\text{c}}1$ -TH cells in MS pathogenesis.

Characterising altered immune cell trafficking during natalizumab treatment

5.1	Introduction	140
5.1.1	Natalizumab in MS	140
5.1.2	Immunological effects of natalizumab	141
5.1.3	Hypothesis	144
5.2	Results	144
5.2.1	Measuring altered immune cell trafficking following natalizumab	144
5.2.2	Multiple overlapping CD4 ⁺ memory T cell states are enriched in circulation during natalizumab treatment .	147
5.2.3	T _c 1-TH signature is enriched in circulating CD8 ⁺ memory T cell compartment during natalizumab treatment . .	149
5.2.4	Memory B cells are enriched in the circulating B cell compartment during natalizumab treatment	153
5.3	Discussion and conclusions	156

This chapter presents the results of a comparative analysis of *in vivo* treatment-induced shifts in the peripheral immune landscape of pwMS during natalizumab treatment, contextualised by the cell states identified to be perturbed following B cell depletion with ocrelizumab in Chapter 3. This contrastive approach provides convergent evidence for a dysregulated B–T cell axis disrupted by two highly effective DMTs with divergent mechanisms of action.

5.1 Introduction

5.1.1 Natalizumab in MS

Natalizumab was the first monoclonal antibody approved for use as a DMT in pwMS. Its target is integrin $\alpha 4$, which forms part of a heterodimeric complex with either integrin $\beta 1$ or integrin $\beta 7$ to form the cellular adhesion molecules VLA-4 or LPAM-1, respectively. VLA-4 plays a critical role in lymphocyte transmigration across the BBB, and natalizumab inhibits the binding of this complex to VCAM-1 on vascular endothelial cells to prevent migration into the CNS. This mechanism of action is distinct from the majority of MS DMTs, which work by depleting or functionally modulating circulating immune cells.

Natalizumab is currently licensed for use in highly active RRMS and is classified as a high-efficacy DMT, based on both trial data and real-world evidence demonstrating marked and sustained reduction in relapses, MRI activity, and disability worsening^{82,233}. Observational studies suggest its efficacy is comparable to ocrelizumab^{234,235}. It has a rapid onset of action, with significant decrease in disease activity within 3 months of treatment initiation²³⁶, supporting its use in RES MS. Additionally, in contrast to immune cell-depleting DMTs, which may have sustained effects over months to years as the immune system reconstitutes, the therapeutic effect of natalizumab is rapidly reversible within 8–16 weeks of treatment discontinuation²³⁷, allowing flexibility for switching between treatments. Conversely, this leads to a higher risk of early recurrent disease activity on treatment cessation, including the possibility of rebound worsening²³⁸.

A key safety concern with the use of natalizumab is an increased risk of the

potentially fatal CNS infection PML⁸⁵, caused by neurotropic reactivation of latent JCV infection, which is thought to be facilitated by impaired CNS immune surveillance by antiviral cytotoxic T cells. The primary risk factors for PML in natalizumab-treated patients are: (i) treatment duration, (ii) JCV seropositivity, and (iii) prior immunosuppressant use²³⁸. In practice, a validated framework⁸⁶ is used to quantify the annual risk of PML based on these factors at regular intervals during treatment, and individuals at higher risk are encouraged to switch to an alternative DMT. Consequently, this imposes a *de facto* limit on treatment duration due to an unfavourable safety profile in a subset of pwMS; in a large multinational observational study of natalizumab use in pwMS over 10 years, the median treatment duration was 3.3 years²¹⁰. Recent evidence suggests that PML risk may be substantially reduced by extended interval (6-weekly) dosing, without substantial loss of efficacy, offering an alternative strategy for risk mitigation^{239,240}.

Nonetheless, a more refined understanding of the peripherally-primed pathogenic immune cell states that are prevented from migrating across the BBB by natalizumab may facilitate the rational design of more precise therapies for MS that can disrupt this immunological axis while sparing normal CNS immune surveillance.

5.1.2 Immunological effects of natalizumab

During natalizumab treatment, there is an increase in total circulating lymphocyte counts, which reverts to the pre-treatment baseline following treatment cessation²³⁷. This reflects the accumulation of cells that transmigrate across the vascular endothelium into tissues in an integrin $\alpha 4$ -dependent manner. Although blockade of immune cell trafficking across the BBB, demonstrated by a concomitant marked

reduction in CSF lymphocyte counts under treatment^{212,241}, is a major driver of peripheral immune shifts, it is worth noting that natalizumab also impairs immune cell migration into secondary lymphoid tissues²⁴². Increased cell counts in the blood during treatment are observed across multiple lymphoid, but not myeloid, lineages, including T cells, B cells, and NK cells^{243–245}.

Within the T cell compartment, several studies have reported enrichment of memory T cells in the peripheral blood during treatment. Early studies using flow cytometry found increases in both CD4⁺ and CD8⁺ subpopulations, with no significant alteration in the circulating CD4:CD8 ratio^{244,245}. Moreover, circulating T cells from pwMS undergoing long-term treatment have been found to have higher expression of genes linked to T_h1 and T_h17 polarisation (*IFNG*, *CCR6*) and produce higher levels of corresponding proinflammatory cytokines (interferon- γ , TNF- α , IL-17) following *in vitro* stimulation assays^{246–248}; importantly, these findings are not reproduced following *in vitro* exposure to natalizumab²⁴⁶, in keeping with an enrichment of circulating proinflammatory T cells *in vivo* due to altered immune cell trafficking. Other studies have confirmed the quantitative expansion of CD4⁺ memory T cells during treatment, but failed to reproduce the functional phenotypic alterations²⁴⁹. Findings on the effect of natalizumab on the frequency of other circulating T cell populations, including T_{reg} cells, have been inconsistent across studies^{243,247,250}.

Studies applying high-dimensional single-cell multiomic approaches to characterise alterations in the immune landscape during natalizumab treatment have revealed more granular findings. A key study identified significant post-treatment enrichment of cells expressing the integrin $\alpha 4$ gene (*ITGA4*) in a cluster of CD4⁺ effector memory

T cells (T09) characterised by high coexpression of *KLRB1* and *LTB*, alongside *AQP3*, *TNFRSF25*, *ICOS*, *MAF*, and *RORA*¹⁶⁰. Moreover, the same transcriptional signature was found to be enriched in both focal lesions and parasulcal grey matter regions in a spatial transcriptomics dataset of post-mortem brain tissue from SPMS donors, consistent with the possibility that cells with this phenotype contribute to CNS-resident populations in end-stage disease but can be trapped in the periphery by natalizumab at earlier disease stages. A more recent study extended these findings using flow cytometry and scRNA-seq of unpaired CSF samples from pwMS with and without natalizumab treatment²¹². In natalizumab-treated pwMS, CSF samples demonstrated an inversion of the normal CD4⁺ T cell-predominance in this compartment. The most significantly depleted T cell cluster within the CSF compartment during treatment exhibited a T_{fh} phenotype. The majority of T cells persisting within the CSF compartment during treatment exhibited high expression of T_{rm}-associated genes (*CD69*, *CXCR6*, *ITGAE*, *ITGA1*, *ZFP36L2*, *ID2*, *DUSP1*); moreover, the residual CD4⁺ T cells exhibited an overlapping T_h1 / T_h17 transcriptional phenotype. Together, these findings support the interpretation that natalizumab preferentially restricts entry of specific T cell subsets into the CNS, leading to their relative accumulation in the peripheral circulation. However, it has limited effect on cells that have upregulated tissue residency programs and therefore persist within the CNS despite treatment.

Within the B cell compartment, natalizumab leads to enrichment of activated memory populations in the circulation^{242,251}, which is thought to reflect a combination of impaired migration into both secondary lymphoid tissues and the CNS. Moreover, the frequency of both memory B cells and plasmablasts within the

CSF of natalizumab-treated pwMS is substantially reduced²⁵², with a possible concomitant decrease in intrathecal clonal immunoglobulin synthesis²⁵³.

Collectively, the immunological effects of natalizumab reflect a redistribution of immune cell populations driven by blockade of transcompartmental trafficking, leading to peripheral accumulation of antigen-experienced, tissue-homing lymphocytes. As such, natalizumab may provide a complementary perturbational model to cell-depleting therapies, facilitating triangulation of the disease-relevant immunophenotype in MS.

5.1.3 Hypothesis

During natalizumab treatment, cell states with a higher baseline tendency for integrin $\alpha 4$ -dependent egress from the circulation into tissues are preferentially retained in the blood. Given the demonstrated enrichment of T_c1-TH cells within the CSF and chronic brain lesions of pwMS (Chapter 4), along with their high expression of integrin $\alpha 4$ (Chapter 3), I hypothesised that disruption of blood–CNS trafficking of these cells may contribute to the therapeutic efficacy of natalizumab. Therefore, in contrast to the effect of ocrelizumab, the T_c1-TH signature would be expected to be enriched in the circulation during treatment with natalizumab, reflecting reduced ability of cells with this phenotype to enter the CNS.

5.2 Results

5.2.1 Measuring altered immune cell trafficking following natalizumab

To measure altered immune cell trafficking during natalizumab treatment, I used a large published CITE-seq dataset of PBMC from pwMS and matched healthy

control donors¹⁶⁰, which contained a subset of samples taken before and after treatment with natalizumab (Figure 5.1).

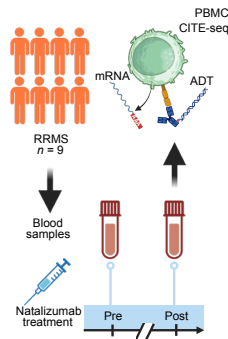


Figure 5.1. Validation cohort of pwMS with pre- and post-natalizumab PBMC scRNA-seq samples. Graphical summary of natalizumab validation dataset of publicly available PBMC CITE-seq study with pre- and post-natalizumab samples obtained from donors with RRMS ($n = 9$).

After processing the raw count matrices and transferring cell type annotations from the discovery dataset (Section 2.5.2.1 and Section 2.5.2.2), the final integrated dataset contained 56,252 cells from 9 pwMS with pre- and post-natalizumab samples (Figure 5.2).

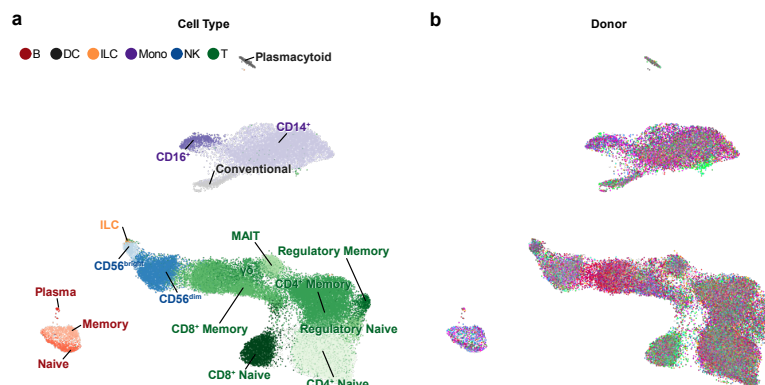


Figure 5.2. Natalizumab PBMC validation dataset integration and cell type annotation. UMAP visualisation of 56,252 cells comprising the natalizumab PBMC validation dataset coloured by reference-based cell type annotation from the discovery dataset (a) and donor (b).

I subsequently used MELD¹⁵⁰ to compute perturbation scores quantifying treatment effects on cell states (Figure 2.3a; see Section 2.5.1.3), whereby negative

5. Characterising altered immune cell trafficking during natalizumab treatment

scores represent depletion and positive scores represent enrichment relative to the donor-specific pre-treatment baseline (Figure 5.3).

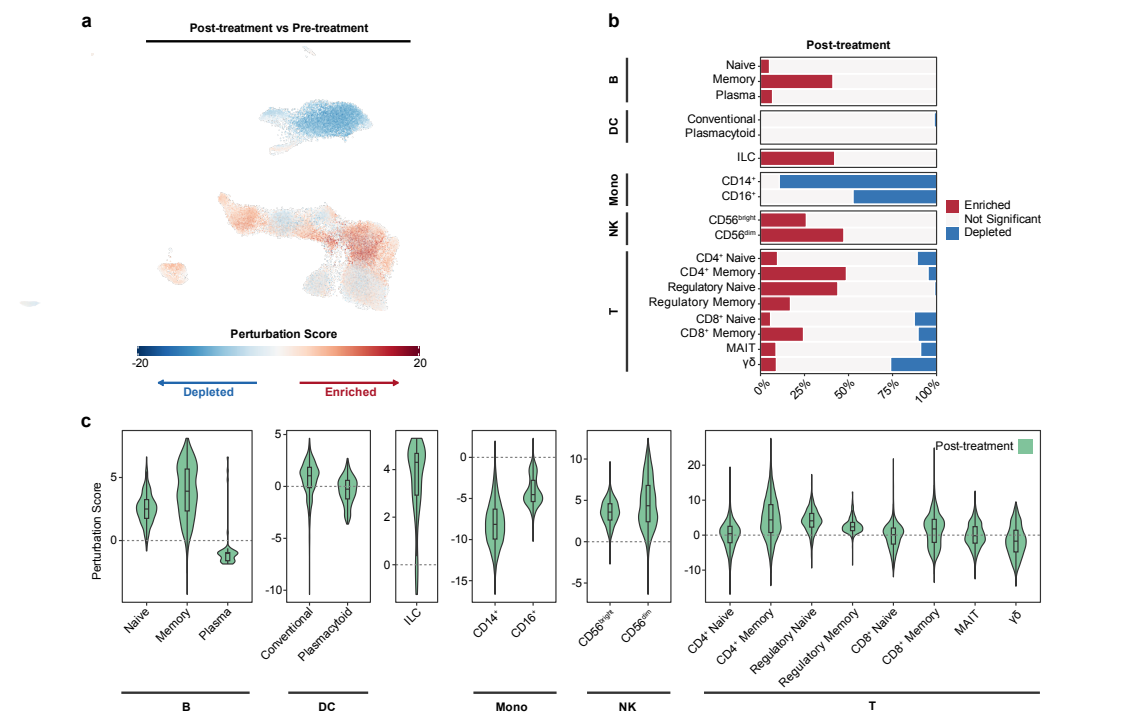


Figure 5.3. Treatment-induced shifts in peripheral immune landscape following natalizumab. **a**, UMAP visualisation of cell states coloured by treatment-related perturbation score; positive values (red) represent enrichment and negative values (blue) represent depletion relative to pre-treatment baseline. **b**, Bar plots showing percentage (x -axis) of cell states within each coarse (level 2) cell type (y -axis) statistically significantly enriched (red) or depleted (blue) after treatment. **c**, Violin/box plots showing distribution of perturbation scores within each coarse (level 2) cell type. Statistical testing for perturbation scores was performed with Z -tests using an empirical permutation-based null distribution for each cell state.

This revealed significant post-treatment enrichment of subpopulations of all lymphocyte populations in the circulation, including B cells, T cells, NK cells, and ILCs.

Within the B cell compartment, I found the strongest treatment-related enrichment in memory B cells, with a markedly weaker effect seen in naive B cells and plasma cells (including plasmablasts). I also observed clear shifts in the T cell compartment post-treatment, with the strongest enrichment seen in CD4⁺ memory

and regulatory naive subsets. Additionally, there was a compositional shift within CD8⁺ memory T cells, with distinct subpopulations exhibiting relative enrichment or depletion in the circulation during treatment.

Notably, both CD14⁺ and CD16⁺ monocytes exhibited significant depletion in the circulation post-treatment. This finding likely represents a compositional effect due to the significant increase in absolute lymphocyte counts in the peripheral blood during natalizumab treatment, which leads to a decrease in the proportion of monocytes within all PBMCs, as highlighted in previous studies using flow cytometry to measure both proportional and absolute shifts in circulating immune cell populations following natalizumab²⁵⁴.

Taken together, these post-treatment shifts in the peripheral immune cell repertoire are in keeping with the expected effect of impaired integrin α 4-dependent lymphocyte egress into tissues, including the CNS.

5.2.2 Multiple overlapping CD4⁺ memory T cell states are enriched in circulation during natalizumab treatment

To further delineate the cellular phenotypes with highest α 4 integrin-dependent migration within each major subpopulation of circulating lymphocytes, I next aimed to perform an unbiased phenotypic characterisation of enriched circulating T cell states during natalizumab treatment (Figure 5.4).

I used starCAT¹⁴² to project T cell-specific gene programs from the discovery dataset (see Section 2.5.2.3). I then trained a random forest regression model to predict perturbation scores based on program scores, extracted feature importances using SHAP, and derived program effect scores (Figure 2.3c; see Section 2.5.1.5).

5. Characterising altered immune cell trafficking during natalizumab treatment

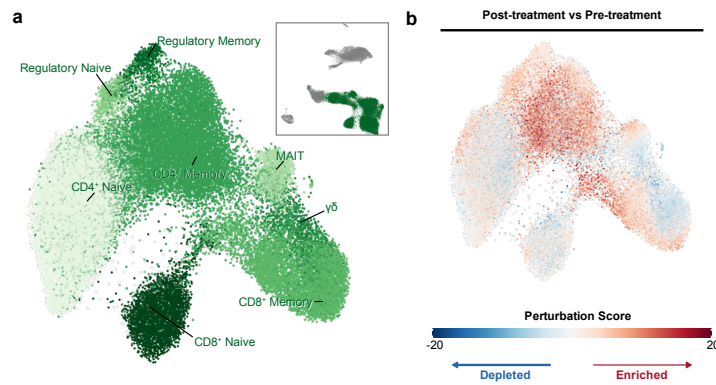


Figure 5.4. Treatment-induced shifts in T cells following natalizumab. a, UMAP visualisation of T cells, which are shown highlighted (dark green) in the UMAP visualisation of all cells (inset), coloured and labelled by coarse (level 2) cell type. b, UMAP visualisation of T cell states coloured by treatment-related perturbation score; positive values (red) represent enrichment and negative values (blue) represent depletion relative to pre-treatment baseline.

I identified three gene programs, expressed predominantly in memory T cells, that were strongly associated with enrichment post-treatment (Figure 5.5, Figure 5.6a–b). The first, *Th17-like*, had top weighted genes enriched for canonical markers of T_H17 cells (*RORC*, *CCR6*, *KLRB1*, *IL1R1*, *AHR*, *RORA*). The top weighted genes of the second program, *Tfh-like*, featured key transcriptional regulators of T_{fh} cell differentiation and function (*TOX2*, *TOX*, *ICOS*, *MAF*); this transcriptional phenotype, in the absence of alternative T cell lineage markers, is consistent with previously described circulating T_{fh} precursors that upregulate additional canonical T_{fh} marker genes (*CXCR5*, *BCL6*, *PDCD1*, *IL21*) following terminal differentiation within secondary lymphoid follicles^{255,256}. The third program was *Th1-like*, a component of the T_c1 -TH signature, with top weighted genes as described in Section 3.2.3. I confirmed enrichment of these cell states at the individual level (Figure 5.6c) by using linear mixed effects models (see Section 2.6) to compare the mean program usage in memory T cells from each donor between the post-

treatment and pre-treatment time points for each program: *Th17-like* (marginal mean difference 0.00444, 95% CI [0.00142, 0.00745], $p = 9.43 \times 10^{-3}$), *Tfh-like* (marginal mean difference 0.00528, 95% CI [0.00242, 0.00815], $p = 2.80 \times 10^{-3}$), and *Th1-like* (marginal mean difference 0.00602, 95% CI [0.00131, 0.01074], $p = 0.019$).

Together, these findings suggest that multiple cell states corresponding to canonical CD4⁺ memory T cell effector polarisation are strongly enriched in the circulating T cell repertoire during natalizumab treatment, suggesting that these cell phenotypes may have the highest baseline tendency for integrin α 4-dependent egress into tissues.

5.2.3 T_c1-TH signature is enriched in circulating CD8⁺ memory T cell compartment during natalizumab treatment

Although the unbiased analysis of circulating T cell phenotypic shifts during natalizumab treatment did not highlight all components of the T_c1-TH signature among the top perturbation-linked gene programs, these results may have been skewed by the overall CD4⁺ memory predominance of post-treatment enriched T cells in the blood. Notably, in Section 3.2.3, I found that cells with a high T_c1-TH program score were predominantly CD8⁺ memory T cells. Therefore, to test the specific hypothesis that T_c1-TH cells are also blocked from migrating to the CNS by natalizumab, I next looked for enrichment of this transcriptional signature within circulating CD8⁺ memory T cells during natalizumab treatment (Figure 5.7a).

I used a linear mixed effects model (Section 2.6) to compare the mean T_c1-TH program score per donor between the pre- and post-treatment time points (Figure 5.7b) and found a significant enrichment of this signature in circulating

5. Characterising altered immune cell trafficking during natalizumab treatment

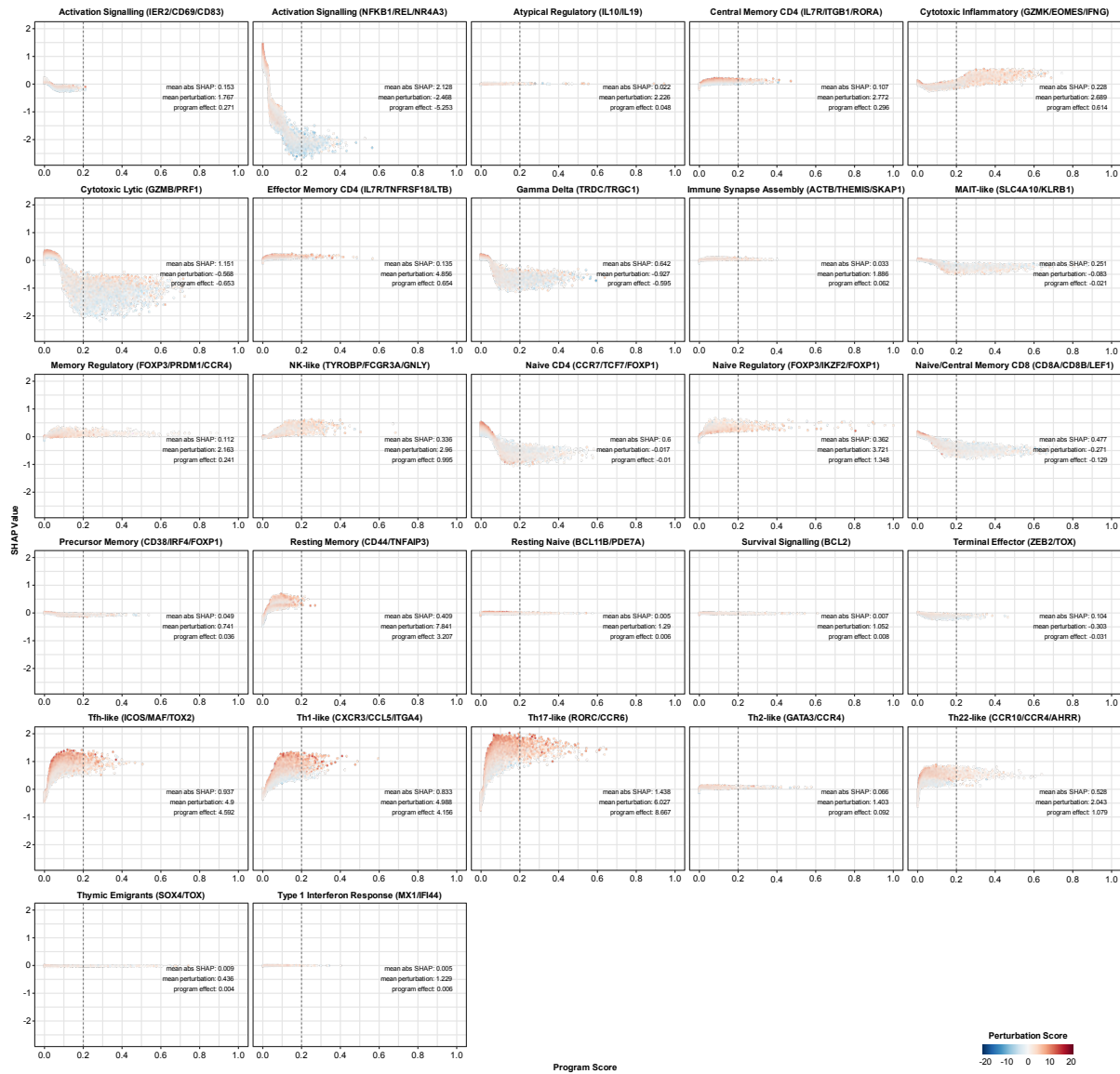


Figure 5.5. T cell gene program SHAP plots. Scatter plots of SHAP values (y -axis) for T cell-specific gene programs for all T cells in the natalizumab validation cohort, obtained from a random forest regression model trained to predict perturbation score based on program score, vs program score (x -axis), coloured by perturbation score; positive values (red) represent enrichment and negative values (blue) represent depletion relative to pre-treatment baseline. Dashed vertical lines indicate the threshold (0.2) used as a cutoff for cells with high program scores. Mean absolute SHAP values and mean perturbation scores for high program score cells are shown along with the computed program effect scores for each program.

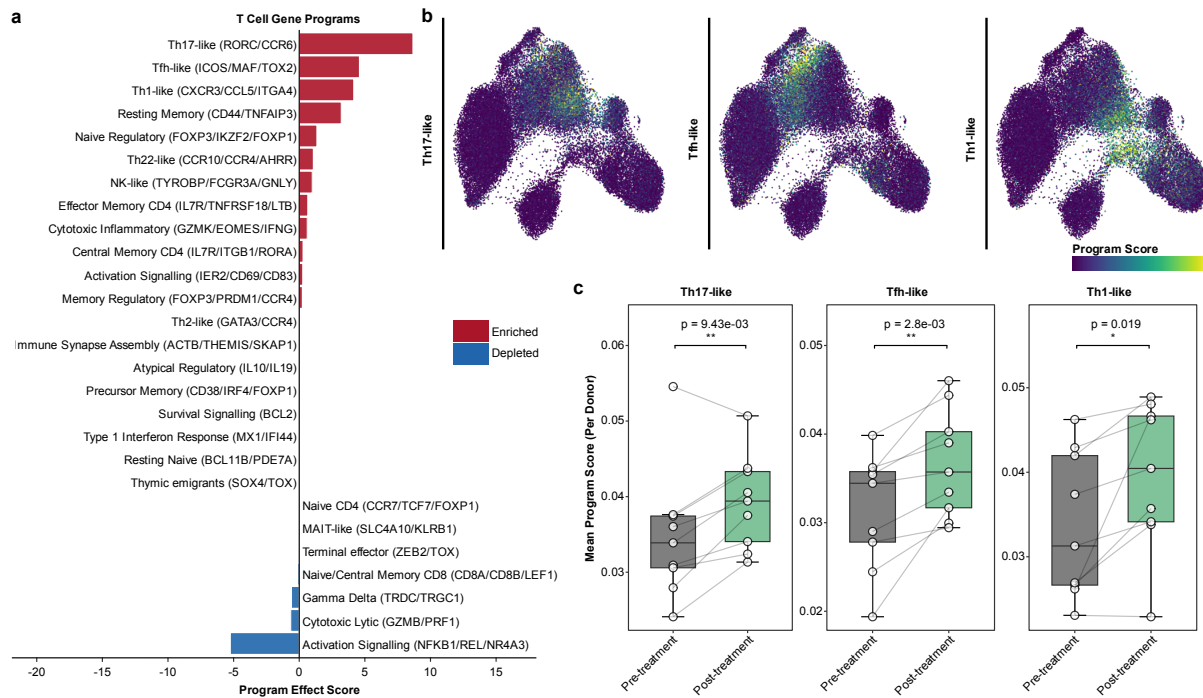


Figure 5.6. CD4⁺ memory T cell effector states are enriched in the circulation during natalizumab treatment. **a**, Bar plot showing program effect scores (x -axis) for T cell-specific gene programs (y -axis); positive scores (red) indicate that high program expression is associated with enrichment post-treatment and negative scores (blue) indicate that high program expression is associated with depletion post-treatment. **b**, Program scores for the *Th17-like*, *Tfh-like*, and *Th1-like* signatures projected on the UMAP visualisation of T cells; upper limits of colour scales are clipped at 99th percentiles. **c**, Box plots comparing mean program scores per donor (y -axis) between pre- and post-treatment time points (x -axis) for *Th17-like*, *Tfh-like*, and *Th1-like* signatures; lines connect paired data from the same individual. Statistical testing was performed using a linear mixed effects model adjusted for age and sex; p -values were calculated with t -tests of marginal means using Satterthwaite degrees of freedom.

5. Characterising altered immune cell trafficking during natalizumab treatment

CD8⁺ memory T cells at the post-treatment time point (marginal mean difference 0.0358, 95% CI [0.0159, 0.0556], $p = 3.19 \times 10^{-3}$), consistent with the hypothesis of reduced migration of cells with this phenotype into the CNS during natalizumab treatment.

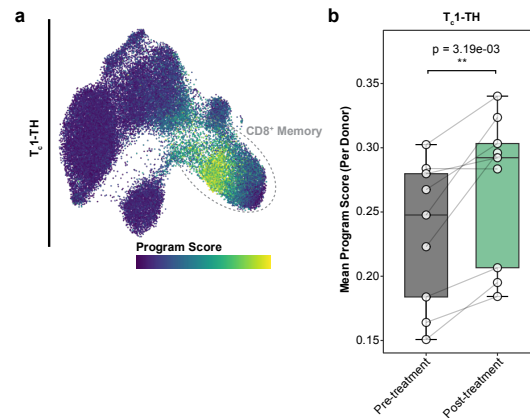


Figure 5.7. T_c1-TH signature is enriched within circulating CD8⁺ memory T cells during natalizumab treatment. a, Program score for the T_c1-TH signature projected on the UMAP visualisation of T cells; CD8⁺ memory T cells are indicated with a dashed gray circle and upper limits of colour scales are clipped at 99th percentiles. b, Box plots comparing mean T_c1-TH program score per donor (y -axis) between pre- and post-treatment time points (x -axis); lines connect paired data from the same individual. Statistical testing was performed using a linear mixed effects model adjusted for age and sex; p -values were calculated with t -tests of marginal means using Satterthwaite degrees of freedom.

The countereffect of ocrelizumab and natalizumab on the abundance of circulating T_c1-TH phenotype cells is a key observation which strongly supports the peripheral origin of the pathogenic adaptive immune response in MS. Furthermore, these data provide supportive evidence for a shared phenotype of proinflammatory, tissue-homing T cells that is depleted under treatment with ocrelizumab, and enriched in the circulation during blockade of lymphocyte trafficking across the BBB using natalizumab, providing novel insight into a potential common mode of therapeutic benefit across two highly effective DMTs with differing mechanisms of action.

5.2.4 Memory B cells are enriched in the circulating B cell compartment during natalizumab treatment

Next, I aimed to further characterise the observed phenotypic shifts in the B cell compartment during natalizumab treatment (Figure 5.8).

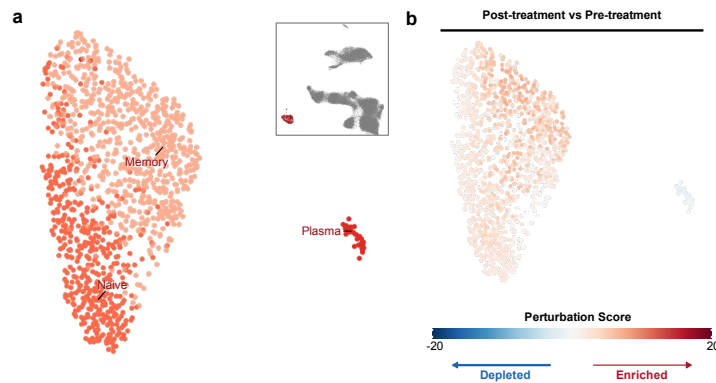


Figure 5.8. Treatment-induced shifts in B cells following natalizumab. a, UMAP visualisation of B cells, which are shown highlighted (dark red) in the UMAP visualisation of all cells (inset), coloured and labelled by coarse (level 2) cell type. b, UMAP visualisation of B cell states coloured by treatment-related perturbation score; positive values (red) represent enrichment and negative values (blue) represent depletion relative to pre-treatment baseline.

I used starCAT¹⁴² to project B cell-specific gene programs from the discovery dataset (see Section 2.5.2.3). I then trained a random forest regression model to predict perturbation scores based on program scores, extracted feature importances using SHAP, and derived program effect scores (see Section 2.5.1.5).

Based on the program effect scores (Figure 5.9, Figure 5.10a–b), the programs most strongly associated with enrichment at the post-treatment time point were those linked to memory B cell states (*Activated Memory*, *Homeostatic Memory*, and *Primed Memory*). In contrast, programs linked to naive B cell states (*Transitional Naive* and *Mature Naive*) and plasma cells (*Early ASC*, *IgA+ ASC*, and *IgG+ ASC*) showed much weaker program effect scores.

5. Characterising altered immune cell trafficking during natalizumab treatment

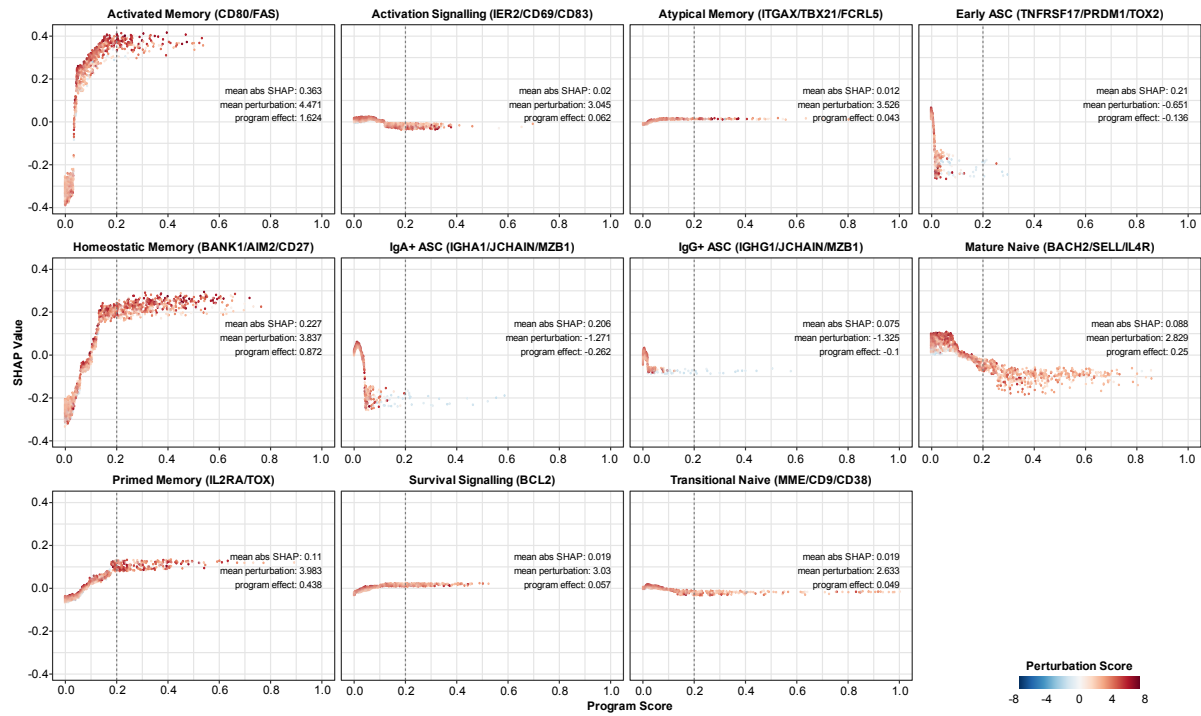


Figure 5.9. B cell gene program SHAP plots. Scatter plots of SHAP values (y -axis) for B cell-specific gene programs for all B cells in the natalizumab validation cohort, obtained from a random forest regression model trained to predict perturbation score based on program score, vs program score (x -axis), coloured by perturbation score; positive values (red) represent enrichment and negative values (blue) represent depletion relative to pre-treatment baseline. Dashed vertical lines indicate the threshold (0.2) used as a cutoff for cells with high program score. Mean absolute SHAP values and mean perturbation scores for high program score cells are shown along with the computed program effect scores for each program.

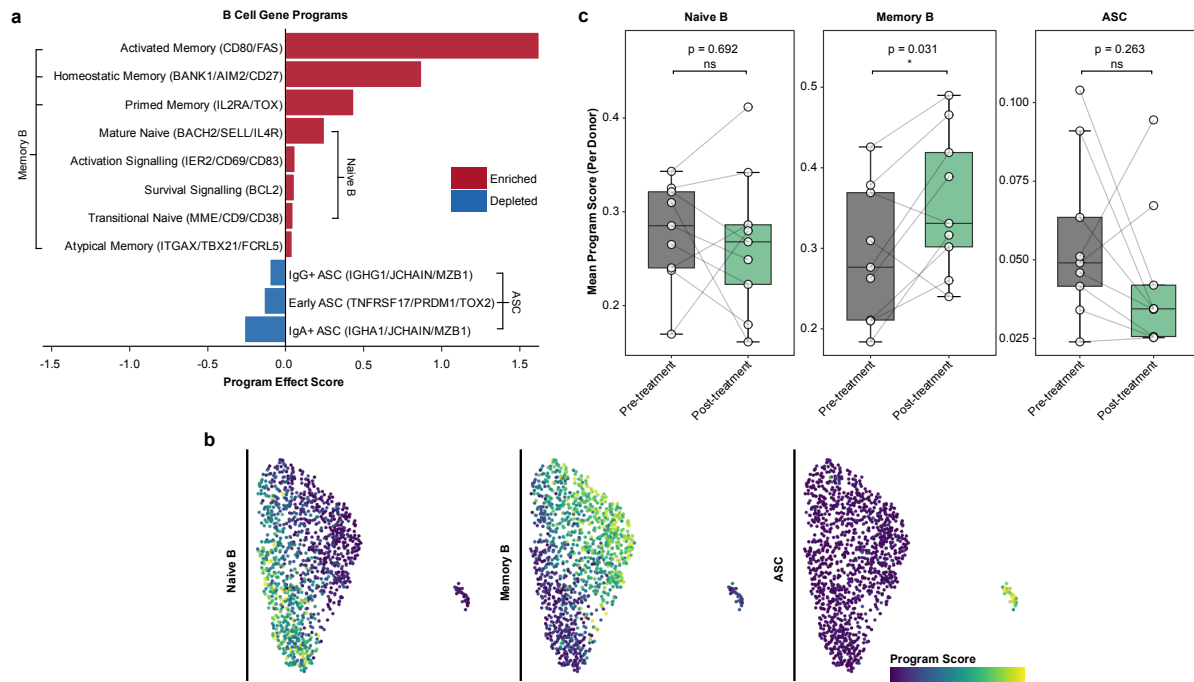


Figure 5.10. Memory B cells are enriched in the circulation during natalizumab treatment. **a**, Bar plot showing program effect scores (x -axis) for B cell-specific gene programs (y -axis); positive scores (red) indicate that high program expression is associated with enrichment post-treatment and negative scores (blue) indicate that high program expression is associated with depletion post-treatment. **b**, Program scores for the naive B, memory B, and ASC signatures projected on the UMAP visualisation of B cells; upper limits of colour scales are clipped at 99th percentiles. **c**, Box plots comparing mean program scores per donor (y -axis) between pre- and post-treatment time points (x -axis) for Naive B, Memory B, and ASC signatures; lines connect paired data from the same individual. Statistical testing was performed using a linear mixed effects model adjusted for age and sex; p -values were calculated with t -tests of marginal means using Satterthwaite degrees of freedom.

To verify these findings at the individual level, I computed composite gene program scores for naive B cells, memory B cells, and ASCs (see Section 4.2.2), then used a linear mixed effects model (see Section 2.6) to compare the mean program scores in B cells from each donor between the post-treatment and pre-treatment time points for each program (Figure 5.10c). I found that the memory B signature was significantly enriched at the post-treatment time point (marginal mean difference 0.0651, 95% CI [0.0078, 0.1224], $p = 0.031$); no significant differences were observed for the naive B (marginal mean difference -0.0108, 95% CI [-0.0713, 0.0497], $p = 0.692$), or ASC (marginal mean difference -0.0134, 95% CI [-0.0391, 0.0123], $p = 0.263$) signatures.

These findings suggest that natalizumab predominantly impairs egress of memory B cells from the circulation into the CNS in pwMS. Moreover, there was no significant enrichment of ASCs in the circulation during natalizumab treatment, despite the marked enrichment of these cells within the CSF and post-mortem lesional brain tissue (see Section 4.2.2 and Section 4.2.4), suggesting that their compartmentalised expansion in pwMS may not be dependent on integrin $\alpha 4$ -dependent migration of differentiated ASCs.

5.3 Discussion and conclusions

By characterising peripheral immune changes during treatment with natalizumab, which blocks immune cell trafficking across the BBB, I leverage a complementary *in vivo* perturbational model to test specific hypotheses generated based on the findings reported in previous chapters. Enrichment of specific cell states in the blood under natalizumab reflects the combined effect of impaired physiological

immune surveillance of the CNS and reduced infiltration of pathogenic immune cells, which has important implications for the interpretation of these findings.

Within the T cell compartment, the most prominent signals of enrichment during natalizumab treatment were observed in gene programs corresponding to *Th17-like*, *Tfh-like*, and *Th1-like* phenotypes. These represent canonical CD4⁺ T cell polarisation states associated with proinflammatory and B cell helper functions, which have been broadly implicated in the pathogenesis of MS¹¹. This is partially consistent with prior studies demonstrating enrichment of activated memory CD4⁺ T cells in the peripheral circulation under natalizumab, with increased expression of cytokines and transcriptional markers of T_h1 and T_h17 differentiation^{246–248}. Importantly, T_h1 and T_h17 cells have been shown to constitute a similar proportion of CSF CD4⁺ helper cells in both pwMS and healthy individuals¹⁵⁵, likely reflecting involvement in physiological CNS immune surveillance.

The enrichment of the *Tfh-like* program is notable in the context of MS pathogenesis, as T_{fh} cells have been shown to be selectively enriched within CSF CD4⁺ helper cells in pwMS^{155,159,212,213}. T_{fh} cells play a central role in germinal centres by supporting B cell maturation, and have been implicated in the formation and maintenance of meningeal tertiary lymphoid structures in MS, where they have been found to exhibit an overlapping phenotype with T_h1 / T_h17 cells^{257–259}. These structures represent key sites of compartmentalised B cell expansion and maturation within the CNS, particularly in progressive disease. Therefore, the preferential retention of cells expressing the *Tfh-like* transcriptional program in the blood during natalizumab treatment may reflect impaired access of circulating precursors of germinal centre T_{fh} cells to compartmentalised CNS niches.

The findings of my re-analysis of this publicly available dataset align with those of the original study; specifically, the key CNS-homing T cell cluster was characterised by an overlapping $T_{\text{h}}17$ / T_{fh} transcriptional phenotype¹⁶⁰. Importantly, the authors of the original study quantified post-treatment enrichment based on change in the proportion of cells expressing *ITGA4* transcripts per cluster, in contrast to my drug mechanism-agnostic approach of linking treatment-related shifts in density across the cell state manifold with continuous and overlapping gene programs. The convergence of findings across these distinct strategies provides further validation of the robustness of the analytical pipeline developed in this thesis, which is able to resolve coordinated shifts in transcriptional programs without relying on discrete clustering and has broad generalisability for single-cell therapeutic perturbation studies in clinical cohorts.

Despite the dominant $CD4^+$ memory T cell enrichment signals observed in the global analysis, my findings demonstrate enrichment of the $T_{\text{c}}1$ -TH transcriptional signature within the $CD8^+$ memory T cells during natalizumab treatment. This is consistent with the prior hypothesis that T cells with this phenotype have a high baseline tendency for integrin $\alpha 4$ -dependent CNS trafficking. The comparatively weaker global signal may reflect the predominance of $CD4^+$ T cells within the circulating and CSF compartments^{155,159,212}. However, this contrasts with the composition of lymphocytic infiltrates within brain tissue of pwMS, where $CD8^+$ T cells significantly outnumber their $CD4^+$ counterparts^{42–45}.

Within the B cell compartment, the most prominent effect of natalizumab was enrichment of memory B cell populations in the peripheral circulation, consistent with prior work^{242,251}. In contrast, no significant enrichment of ASCs was observed in

the circulation, despite enrichment of these cells within the CSF of pwMS^{155,158,217,218}. This discrepancy may suggest that the compartmentalised expansion of ASCs in MS arises predominantly through B cell maturation within the CNS, rather than through migration of differentiated plasmablasts from the periphery.

Taken together, the peripheral immune landscape shifts during natalizumab treatment build further support for a pathogenic B–T cell axis in MS. By preventing the migration of both memory B cells and proinflammatory, tissue-homing T cell populations into the CNS, natalizumab directly disrupts both arms of this axis, which likely contributes to its high clinical efficacy. However, its mechanism of action is inherently limited to targeting circulating immune cells, rather than cells that have already entered the CNS and become tissue-resident. Consistent with this, previous studies have shown that during treatment, the CSF compartment becomes proportionally enriched for CD8⁺ T cells expressing T_{rm} markers, suggesting persistence of compartmentalised immune populations²¹². In this context, the upregulation of genes associated with tissue residency priming in circulating T_c1-TH cells is consistent with the incomplete effect of natalizumab on PIRA, further emphasising the importance of targeting CNS-resident immune populations for future therapeutic strategies. More broadly, my findings highlight potential opportunities to target specific cellular interactions or molecular pathways within this B–T cell axis, rather than broadly suppressing CNS immune surveillance, as a strategy for achieving durable disease control while minimising treatment-associated risks for pwMS.

Discussion

6.1	Novel approach to resolving treatment-related cellular phenotypic perturbations	162
6.2	Model of a pathogenic B–T cell axis in MS	163
6.3	Implications for therapeutic advances	166
6.4	Limitations	170
6.5	Future directions	171

This chapter synthesises the work presented in this thesis, discusses the implications for understanding immune dysregulation and therapeutic strategies in MS, and outlines key limitations and future directions.

6.1 Novel approach to resolving treatment-related cellular phenotypic perturbations

A key element of the work in this thesis involves the development of a novel analytical pipeline to resolve treatment-related shifts in cellular phenotype using multimodal single-cell profiling of immune cells from patient samples (Chapter 2).

The previously described application of MELD¹⁵⁰, a modern single-cell perturbation analysis framework based on manifold learning and graph signal processing, is extended using a statistically principled approach, based on donor-stratified permutation-based empirical null distributions, to derive a perturbation score metric. This metric simultaneously accounts for biological and technical sources of nuisance variation inherent in single-cell data from observational cohorts of human participants, facilitating detection of subtle but conserved treatment-related shifts in the immune cell landscape at cellular resolution.

The denoised measure of perturbation is integrated with cNMF¹³⁷, a robust and interpretable algorithm for latent factor analysis of high-dimensional data, using explainable machine learning models. This approach enables the decomposition of treatment-related shifts in cellular phenotypes using continuous, overlapping, and biologically interpretable transcriptional programs. Importantly, these programs are highly portable across single-cell datasets, facilitating reproducible quantification of treatment-perturbed cell states across different tissue compartments and therapeutic contexts to test mechanistic hypotheses generated from the initial analysis.

6.2 Model of a pathogenic B–T cell axis in MS

By applying the analytical pipeline described in Section 6.1, the work presented in this thesis integrates findings from a longitudinal therapeutic perturbation study using ocrelizumab (Chapter 3) with multiple validation studies using independent and orthogonal public data, spanning tissue compartments and disease stages (Chapter 4), and a mechanistically divergent high-efficacy therapy, namely natalizumab (Chapter 5).

The findings converge on a compelling model of immune dysregulation in MS, centred on a pathogenic interaction between memory B cells and a distinct proinflammatory, cytotoxic, tissue-homing T cell phenotype (T_c1 -TH). Transcriptional profiling of these cells revealed a state characterised by markers of chronic activation/antigen stimulation, expression of cytotoxic effector molecules and proinflammatory cytokines, and upregulation of pathways associated with tissue homing and residency. Importantly, this signature was enriched in the circulation of patients with highly active relapsing disease and, notably, was not dampened following treatment in a patient with persistent breakthrough disease and disability progression on ocrelizumab, supporting potential pathogenicity over a bystander or regulatory role. Moreover, the same transcriptional signature is enriched within the CSF and at the leading edge of chronic lesions in end-stage progressive disease. These findings strengthen evidence for a connection between peripheral immune activation and chronic, compartmentalised CNS inflammation, suggesting that the T_c1 -TH phenotype may represent a circulating pool of potential pathogenic cells primed for CNS entry and long-term tissue residency. Within the

CNS, differentiation into T_{rm} cells, which are well-established to dominate the parenchymal and lesional lymphocytic infiltrates in advanced MS^{43–45,55}, may enable these populations to drive persistent smouldering inflammation through direct cytotoxicity, or cytokine-mediated recruitment and activation of CNS-resident microglia and other infiltrating immune cells^{68,260,261}.

Although the work in this thesis does not directly demonstrate the antigen specificity of cells exhibiting the $T_{\text{c}}1$ -TH phenotype, their transcriptional and surface protein profile aligns closely with emerging literature describing clonally expanded EBV-specific $CD8^+$ T cell repertoires in MS^{26,27,216,262}. Moreover, there is increasing evidence to support the presence of EBV-infected memory B cells, often co-localised with cytotoxic T cells, within chronic MS lesions and distinct CNS immunological niches^{60,64,263}. In this context, memory B cells may act as a latent EBV reservoir that provides persistent antigenic stimulation to dysfunctional antiviral T cells²⁶⁴. Therapeutic B cell depletion may therefore disrupt this axis by removing the antigenic source, leading to secondary contraction of the chronically stimulated $T_{\text{c}}1$ -TH pool. Aside from periodic latent–lytic cycling within memory B cells, clonal expansion and terminal differentiation of infected memory B cells into ASCs within the CNS compartment may further promote viral antigen shedding by triggering the lytic cycle through upregulation of the EBV gene *BZLF1* by ASC transcription factors BLIMP-1 and XBP-1²³². In addition to re-encounter with viral antigens, another plausible mechanism driving chronic activation and expansion of $T_{\text{c}}1$ -TH cells within the CNS may be molecular mimicry between EBV-derived peptides and CNS antigens^{28–30,264}.

This model also provides a coherent explanation for the discordant efficacy

of current DMTs in the relapsing and progressive components of the disease. In early relapsing disease, continued recruitment from the circulating precursor pool renders CNS inflammation susceptible to peripheral immune modulation, including B cell depletion and cross-BBB migration blockade. Over time, increasing compartmentalisation of these cells within the CNS, through upregulation of tissue residency programs, may reduce dependence on ongoing recruitment, accounting for the diminishing impact of therapies on PIRA. Crucially, the delineation of a circulating pool of T cells in MS demonstrating transcriptional and surface protein markers consistent with tissue residency priming represents a previously unrecognised link that anchors recent advances in our understanding of basic T_{rm} biology and fate determination^{227,228} in the context of a human disease, providing a plausible mechanistic rationale for the clinical consensus that early and aggressive immunomodulatory therapy is key to reducing long-term disability progression in MS^{265–268}.

Collectively, the findings of this thesis are consistent with an accumulating body of evidence that supports a conceptualisation of MS as principally resulting from a persistent dysregulated cytotoxic T cell response directed against chronic latent EBV infection in memory B cells^{269,270}. This view is supported by the conflation of multiple findings: (i) the existence of a broad, albeit potentially ineffective^{271,272}, cytotoxic T cell response against EBV in pwMS, which dominates lymphocytic infiltrates of chronically inflamed CNS tissue, (ii) the CNS infiltration and intrathecal persistence of EBV-infected memory B cells and ASCs, (iii) the prominence within CNS-resident cytotoxic T cells of multiple markers of sustained activation and innate-like function typical of responses to chronic viral infection²⁷³,

and (iv) the convergent action of multiple MS DMTs on memory B cells⁹⁶ and the B cell-dependent memory T cell phenotype identified in this thesis.

An alternative interpretation of these findings is that the observed cytotoxic T cell response reflects a protective antiviral mechanism that is attempting to control lytic EBV infection within the CNS, with tissue damage arising either as a direct consequence of uncontrolled viral infection or virally-assisted escape of autoreactive B cell clones from peripheral tolerance mechanisms²⁷⁴. While this possibility cannot be definitively excluded from the findings presented here, there is currently a lack of evidence for widespread neurotropic EBV infection within CNS tissue as a primary driver of MS pathology^{275,276}. Additionally, the absence of broadly prevalent disease-specific autoantibodies in MS additionally argues against an autoreactive B cell response as the major pathogenic mechanism^{48–50}. Moreover, the work in this thesis showing an association of the T_c1-TH phenotype with higher clinical disease activity and poor treatment response, and its enrichment in chronic lesion rims is more favourable to the proposed model that this cell state represents a maladaptive immune response that contributes to, rather than prevents, tissue injury.

6.3 Implications for therapeutic advances

The delineation of a peripherally-originating, CNS-homing, and B cell-dependent pathogenic T cell phenotype in pwMS reframes the mechanism of action of B cell depletion therapies as disruption of a dysregulated immunological niche, rather than eliminating an inherently pathogenic immune cell subset. Importantly, the sensitivity of the T_c1-TH phenotype to B cell depletion highlights a degree

of functional plasticity, suggesting that pathogenic immune states in MS may be actively maintained by persistent antigenic stimulation rather than durably established at the onset of disease. Interrupting the mechanisms that sustain the maladaptive immune circuit while preserving protective immune function may therefore represent a promising opportunity for the development of future therapeutic strategies, moving beyond indiscriminate immune cell ablation toward more selective immunomodulation¹¹.

Selectively disrupting the antigenic drive sustaining the pathogenic B–T cell axis may offer a more targeted alternative to indiscriminate ablation of memory B cells. If the cytotoxic T cell response is directed against latent EBV infection in memory B cells, then reducing the antigenic burden by promoting viral clearance or disrupting latent–lytic cycling with antiviral therapies could represent a viable therapeutic strategy in MS²⁷⁴. Historical trials of widely available anti-herpesvirus therapies in MS have shown modest and inconsistent benefit^{277,278}, likely due to their inability to target the latent phase of the EBV life cycle. In this regard, small molecule inhibitors and small interfering RNA (siRNA)-silencing of latent phase viral proteins^{279–281} may offer a more promising avenue for future therapeutic trials. Another strategy proposed by some authors is the use of therapeutic vaccination or infusion of *in vitro*-expanded autologous EBV-specific CD8⁺ T cells to boost antiviral immunity and promote clearance²⁸². However, this approach requires careful consideration in pwMS, where an augmented antiviral cytotoxic response could theoretically lead to paradoxical exacerbation of collateral tissue damage within the CNS, underscoring the need for a more nuanced understanding of the relationship between antiviral cellular immunity, antigen cross-reactivity, and

persistent compartmentalised inflammation in MS.

An alternative approach may be direct molecular targeting of the B–T cell interactions that sustain the dysregulated immune niche. Therapeutic targeting of the CD40–CD40L co-stimulatory pathway is already under active clinical investigation in MS. Anti-CD40L therapies (e.g. frexalimab) have demonstrated promising efficacy in early-phase trials²⁸³, and are thought to act by limiting the bidirectional amplification of T cell responses mediated by APCs. In the context of the disease model proposed in Section 6.2, anti-CD40L therapies may have additional specific beneficial effects, as CD40–CD40L signalling has been shown to be a key pathway mediating rescue of CD8⁺ T cells from functional exhaustion in the context of chronic viral infection²⁸⁴, which may contribute to the sustained cytotoxic T cell response within the CNS. Disrupting this pathway may therefore have the potential to selectively destabilise the maladaptive immune response while preserving broader immune competence.

Beyond targeting intercellular interactions, the transcriptional programs defining the T_c1-TH phenotype may reveal opportunities for more precise molecular intervention. For example, granzyme K, a serine protease stored within cytotoxic secretory granules in lymphocytes and highly expressed in T_c1-TH cells, has been shown to delineate a distinct subset of T cells that has been widely associated with chronic inflammatory states and autoimmunity^{285,286}. Multiple lines of evidence suggest that it may have pivotal roles in sustaining chronic tissue inflammation through diverse mechanisms, including stimulating proinflammatory signalling pathways via activation of the transmembrane G-protein-coupled receptor PAR-1^{287,288}, facilitating endothelial transmigration²⁸⁹ and activating the complement

cascade²⁹⁰. Similarly, the CCL5–CCR5 chemokine axis consistently emerges as an important signalling pathway for recruitment of both cytotoxic T cells and EBV-infected B cells to the CNS^{270,291}. Novel therapies targeting these pathways could therefore interfere with specific molecular mechanisms of immunological pathogenicity in MS.

Finally, the compartmentalised smouldering inflammation within chronic active lesion rims suggests that therapies capable of penetrating the BBB and modulating the functional state of proinflammatory T_{rm} cells and microglia may be important for the treatment of PIRA⁵⁴, particularly in patients with established disease. In future, delineation of the intercellular signalling pathways between these tissue-resident cell populations may support the development of CNS-penetrant small molecule immunomodulators targeting this niche to halt chronic inflammation and disability progression.

Taken together, emerging evidence endorses a shift in therapeutic strategies for MS from broad suppression or depletion of immune cell populations toward selective disruption of the pathogenic circuits that link peripheral immune activation to compartmentalised CNS inflammation. By identifying a B cell-dependent cytotoxic T cell state that bridges these distinct but interconnected elements of MS immunobiology, the work in this thesis supports this paradigm shift and provides an early framework for the rational development of therapies aimed at achieving durable disease control and restoring immune homeostasis.

6.4 Limitations

The work presented in this thesis has several important limitations. The discovery cohort in Chapter 3 comprised a relatively small number of patients, which may limit the generalisability of the findings to the broader population of pwMS. Additionally, the primary discovery analyses were performed in peripheral blood, justified by the lack of CNS penetration of anti-CD20 therapies. However, immunological changes in the blood may not fully reflect downstream changes within the CNS. To mitigate these limitations, the key finding from Chapter 3 was used to generate testable hypotheses which were validated in independent and orthogonal datasets in Chapter 4 and Chapter 5, including larger cohorts of pwMS and healthy control donors, as well as CSF and brain tissue samples. This provided support for the robustness and reproducibility of the identified disease-associated cell state.

All findings in this thesis are based on observational data derived from real-world clinical cohorts. As such, despite a considered study design, analytical approach, and multipronged validation, it lacks the experimental control and mechanistic rigour achievable in animal models or *in vitro* systems. Moreover, functional properties of identified cell states were inferred from prior biological knowledge of genes and surface proteins rather than directly measured, reflecting a general limitation of single-cell omics studies. However, animal models and other reductionist experimental systems commonly used in functional studies do not fully recapitulate human MS, nor capture the full complexity of immune network interactions in the disease–therapy context. Therefore, observational studies in pwMS represent an important paradigm for disease-relevant hypothesis generation,

which is complementary to and motivates future mechanistic experimental work.

The absence of paired TCR sequencing in the datasets analysed herein precluded direct assessment of the antigen specificity and clonality of T_c1-TH cells, limiting the ability to definitively link the identified transcriptional programs and surface protein phenotype to specific antigen-driven immune responses. Nevertheless, the gene programs and analytical approach developed in this thesis are readily portable and can be projected into newly generated or publicly available single-cell datasets that include TCR sequencing, providing a foundation for subsequent studies aimed at directly testing the mechanistic hypotheses generated in this thesis.

6.5 Future directions

The findings presented in this thesis motivate several directions for future investigation that would enable direct testing of the mechanistic hypotheses outlined here and facilitate translation into clinical settings.

A central next step is the integration of paired scRNA-seq and TCR sequencing to directly assess the clonality and antigen specificity of cells with a T_c1-TH phenotype. The first strategy is to apply this technique to FACS-isolated T cells from MS blood or CSF samples, followed by querying TCR sequences of clonally expanded populations against public databases for sequence similarity to known viral peptide-reactive TCRs. The paired scRNA-seq data could then be used to quantify the T_c1-TH signature per cell and compare this to predicted viral specificity to test the hypothesis that cells with this phenotype are enriched for clonally expanded EBV-reactive cells. However, significant technical challenges remain in accurate computational prediction of MHC-restricted peptide specificity from

TCR sequences alone¹²². This motivates an alternative strategy of labelling EBV-reactive T cells using DNA-barcoded peptide–MHC multimers²⁹² prior to single-cell sequencing, enabling direct high-throughput identification of MHC-restricted viral peptide-specific TCRs. Together, these approaches would facilitate evaluation of the hypothesised link between latent EBV infection and the T_c1-TH phenotype.

A second important direction is more detailed spatial characterisation of the identified pathogenic T cell state within the CNS. While the work in this thesis leveraged existing snRNA-seq datasets to demonstrate enrichment of the T_c1-TH signature within T cells extracted from chronic MS lesions, spatial transcriptomics and proteomics technologies offer the opportunity to resolve their precise localisation and cellular interactions *in situ*. This could be achieved either through integration of traditional spot-based spatial transcriptomics with computational deconvolution frameworks adapted to use cNMF-based weighted gene programs instead of discrete cell type-specific marker gene lists, or through application of single-cell resolution spatial transcriptomics platforms such as Xenium (10x Genomics) or MERSCOPE (Vizgen). These approaches would be particularly informative for delineating the colocalisation patterns and cellular interactions between pathogenic T_{rm} cells, EBV-infected B cells, and microglia within CNS-compartmentalised inflammatory niches.

From a clinical perspective, validation of the proposed pathogenic B–T cell axis in larger and more deeply phenotyped patient cohorts represents a key priority. In particular, extending longitudinal analyses to include greater numbers of anti-CD20 non-responders, as well as cohorts stratified by rate of PIRA, would enable more rigorous assessment of the relationship between the T_c1-TH signature, therapeutic

resistance, and long-term disability accumulation.

Finally, translation of these findings into scalable clinical assays will require approximation of the T_c1-TH transcriptional signature using combinatorial surface protein markers amenable to measurement using multiparametric flow or mass cytometry. Development of a minimal immunophenotyping panel to track circulating cells with this phenotype during therapy would facilitate broader validation across clinical cohorts and may eventually enable prospective monitoring of treatment response. Integration of cytometry data with CITE-seq profiles using recently developed modern analytical frameworks (e.g. CytoVI²⁹³) may facilitate translating high-dimensional systems immunology discoveries into clinically usable immunophenotypes.

These avenues for future work build directly on the conceptual framework established in this thesis. By extending the current work to explore antigen specificity and spatial context, as well as potential for clinical stratification and development of a translational immunophenotyping assay, they offer a path toward more precise characterisation and measurement of dysregulated immune responses in MS and, ultimately, more targeted and durable therapeutic and monitoring strategies.

References

- 1 Walton C, King R, Rechtman L, Kaye W, Leray E, Marrie RA *et al.* Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition. *Multiple Sclerosis Journal* 2020; **26**: 1816–1821.
- 2 Jakimovski D, Bittner S, Zivadinov R, Morrow SA, Benedict RH, Zipp F *et al.* Multiple sclerosis. *The Lancet* 2024; **403**: 183–202.
- 3 Lublin FD, Reingold SC, National Multiple Sclerosis Society Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. Defining the clinical course of multiple sclerosis. *Neurology* 1996; **46**: 907.
- 4 Lublin FD, Reingold SC, Cohen JA, Cutter GR, Sørensen PS, Thompson AJ *et al.* Defining the clinical course of multiple sclerosis: The 2013 revisions. *Neurology* 2014; **83**: 278–286.
- 5 Ganjgahi H, Häring DA, Aarden P, Graham G, Sun Y, Gardiner S *et al.* AI-driven reclassification of multiple sclerosis progression. *Nature Medicine* 2025; **31**: 3414–3424.
- 6 Frischer JM, Bramow S, Dal-Bianco A, Lucchinetti CF, Rauschka H, Schmidbauer M *et al.* The relation between inflammation and neurodegeneration in multiple sclerosis brains. *Brain* 2009; **132**: 1175–1189.
- 7 Woo MS, Engler JB, Friese MA. The neuropathobiology of multiple sclerosis. *Nature Reviews Neuroscience* 2024; **25**: 493–513.
- 8 Lublin FD, Häring DA, Ganjgahi H, Ocampo A, Hatami F, Čuklina J *et al.* How patients with multiple sclerosis acquire disability. *Brain* 2022; **145**: 3147–3161.
- 9 Kuhlmann T, Moccia M, Coetzee T, Cohen JA, Correale J, Graves J *et al.* Multiple sclerosis progression: Time for a new mechanism-driven framework. *Lancet Neurol* 2023; **22**: 78–88.
- 10 Montalban X, Lebrun-Frénay C, Oh J, Arrambide G, Moccia M, Pia Amato M *et al.* Diagnosis of multiple sclerosis: 2024 revisions of the McDonald criteria. *The Lancet Neurology* 2025; **24**: 850–865.
- 11 Attfield KE, Jensen LT, Kaufmann M, Friese MA, Fugger L. The immunology of multiple sclerosis. *Nat Rev Immunol* 2022; **22**: 734–750.

- 12 International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* 2019; **365**: eaav7188.
- 13 Sospedra M, Martin R. IMMUNOLOGY OF MULTIPLE SCLEROSIS. *Annual Review of Immunology* 2005; **23**: 683–747.
- 14 Neefjes J, Jongsma MLM, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews Immunology* 2011; **11**: 823–836.
- 15 Holling TM, Schooten E, Van Den Elsen PJ. Function and regulation of MHC class II molecules in T-lymphocytes: Of mice and men. *Human Immunology* 2004; **65**: 282–290.
- 16 International Multiple Sclerosis Genetics Consortium, Harroud A, Stridh P, McCauley JL, Saarela J, Van Den Bosch AMR *et al.* Locus for severity implicates CNS resilience in progression of multiple sclerosis. *Nature* 2023; **619**: 323–331.
- 17 Bjornevik K, Münz C, Cohen JI, Ascherio A. Epstein–Barr virus as a leading cause of multiple sclerosis: Mechanisms and implications. *Nature Reviews Neurology* 2023. doi:10.1038/s41582-023-00775-5.
- 18 Bjornevik K, Cortese M, Healy BC, Kuhle J, Mina MJ, Leng Y *et al.* Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science* 2022; **375**: 296–301.
- 19 Handel AE, Williamson AJ, Disanto G, Handunnetthi L, Giovannoni G, Ramagopalan SV. An Updated Meta-Analysis of Risk of Multiple Sclerosis following Infectious Mononucleosis. *PLOS ONE* 2010; **5**: e12496.
- 20 Jacobs BM, Giovannoni G, Cuzick J, Dobson R. Systematic review and meta-analysis of the association between Epstein–Barr virus, multiple sclerosis and other risk factors. *Multiple Sclerosis Journal* 2020; **26**: 1281–1297.
- 21 Thorley-Lawson DA. Epstein-Barr virus: Exploiting the immune system. *Nature Reviews Immunology* 2001; **1**: 75–82.
- 22 Thorley-Lawson DA, Gross A. Persistence of the Epstein–Barr Virus and the Origins of Associated Lymphomas. *New England Journal of Medicine* 2004; **350**: 1328–1337.
- 23 Münz C. Latency and lytic replication in Epstein–Barr virus-associated oncogenesis. *Nature Reviews Microbiology* 2019; **17**: 691–700.

- 24 Levitskaya J, Coram M, Levitsky V, Imreh S, Steigerwald-Mullen PM, Klein G *et al.* Inhibition of antigen processing by the internal repeat region of the Epstein–Barr virus nuclear antigen-1. *Nature* 1995; **375**: 685–688.
- 25 Styles CT, Bazot Q, Parker GA, White RE, Paschos K, Allday MJ. EBV epigenetically suppresses the B cell-to-plasma cell differentiation pathway while establishing long-term latency. *PLOS Biology* 2017; **15**: e2001992.
- 26 Schneider-Hohendorf T, Gerdes LA, Pignolet B, Gittelman R, Ostkamp P, Rubelt F *et al.* Broader Epstein–Barr virus-specific T cell receptor repertoire in patients with multiple sclerosis. *Journal of Experimental Medicine* 2022; **219**: e20220650.
- 27 Schneider-Hohendorf T, Wünsch C, Falk S, Raposo C, Rubelt F, Mirebrahim H *et al.* Broader anti-EBV TCR repertoire in multiple sclerosis: Disease specificity and treatment modulation. *Brain* 2025; **148**: 933–940.
- 28 Lanz TV, Brewer RC, Ho PP, Moon J-S, Jude KM, Fernandez D *et al.* Clonally expanded B cells in multiple sclerosis bind EBV EBNA1 and GlialCAM. *Nature* 2022; **603**: 321–327.
- 29 Vietzen H, Kühner LM, Berger SM, Furlano PL, Bsteh G, Berger T *et al.* Accumulation of Epstein-Barr virus-induced cross-reactive immune responses is associated with multiple sclerosis. *Journal of Clinical Investigation* 2024; **134**: e184481.
- 30 Sattarnezhad N, Kockum I, Thomas OG, Liu Y, Ho PP, Barrett AK *et al.* Antibody reactivity against EBNA1 and GlialCAM differentiates multiple sclerosis patients from healthy controls. *Proceedings of the National Academy of Sciences* 2025; **122**: e2424986122.
- 31 Jelcic I, Al Nimer F, Wang J, Lentsch V, Planas R, Jelcic I *et al.* Memory B Cells Activate Brain-Homing, Autoreactive CD4+ T Cells in Multiple Sclerosis. *Cell* 2018; **175**: 85–100.e23.
- 32 Younis S, Moutusy SI, Rasouli S, Jahanbani S, Pandit M, Wu X *et al.* Epstein-Barr virus reprograms autoreactive B cells as antigen-presenting cells in systemic lupus erythematosus. *Science Translational Medicine* 2025; **17**: eady0210.
- 33 Leibovitch EC, Jacobson S. Evidence linking HHV-6 with multiple sclerosis: An update. *Current Opinion in Virology* 2014; **9**: 127–133.

- 34 Morandi E, Tanasescu R, Tarlinton RE, Constantinescu CS, Zhang W, Tench C *et al.* The association between human endogenous retroviruses and multiple sclerosis: A systematic review and meta-analysis. *PLOS ONE* 2017; **12**: e0172415.
- 35 Kunkl M, Frascolli S, Amormino C, Volpe E, Tuosto L. T Helper Cells: The Modulators of Inflammation in Multiple Sclerosis. *Cells* 2020; **9**: 482.
- 36 Lock C, Hermans G, Pedotti R, Brendolan A, Schadt E, Garren H *et al.* Gene-microarray analysis of multiple sclerosis lesions yields new targets validated in autoimmune encephalomyelitis. *Nature Medicine* 2002; **8**: 500–508.
- 37 Kebir H, Kreymborg K, Ifergan I, Dodelet-Devillers A, Cayrol R, Bernard M *et al.* Human TH17 lymphocytes promote blood-brain barrier disruption and central nervous system inflammation. *Nature Medicine* 2007; **13**: 1173–1175.
- 38 Viglietta V, Baecher-Allan C, Weiner HL, Hafler DA. Loss of Functional Suppression by CD4+CD25+ Regulatory T Cells in Patients with Multiple Sclerosis. *The Journal of Experimental Medicine* 2004; **199**: 971–979.
- 39 Martin R, Jaraquemada D, Flerlage M, Richert J, Whitaker J, Long EO *et al.* Fine specificity and HLA restriction of myelin basic protein-specific cytotoxic T cell lines from multiple sclerosis patients and healthy individuals. *The Journal of Immunology* 1990; **145**: 540–548.
- 40 Ota K, Matsui M, Milford EL, Mackin GA, Weiner HL, Hafler DA. T-cell recognition of an immuno-dominant myelin basic protein epitope in multiple sclerosis. *Nature* 1990; **346**: 183–187.
- 41 Von Essen MR, Ammitzbøll C, Börnsen L, Sellebjerg F. Assessment of commonly used methods to determine myelin-reactivity of T cells in multiple sclerosis. *Clinical Immunology* 2021; **230**: 108817.
- 42 Babbe H, Roers A, Waisman A, Lassmann H, Goebels N, Hohlfeld R *et al.* Clonal Expansions of Cd8+ T Cells Dominate the T Cell Infiltrate in Active Multiple Sclerosis Lesions as Shown by Micromanipulation and Single Cell Polymerase Chain Reaction. *Journal of Experimental Medicine* 2000; **192**: 393–404.
- 43 van Nierop GP, van Luijn MM, Michels SS, Melief M-J, Janssen M, Langerak AW *et al.* Phenotypic and functional characterization of T cells in white matter lesions of multiple sclerosis patients. *Acta Neuropathologica* 2017; **134**: 383–401.

- 44 Machado-Santos J, Saji E, Tröscher AR, Paunovic M, Liblau R, Gabriely G *et al.* The compartmentalized inflammatory response in the multiple sclerosis brain is composed of tissue-resident CD8+ T lymphocytes and B cells. *Brain* 2018; **141**: 2066–2082.
- 45 Fransen NL, Hsiao C-C, van der Poel M, Engelenburg HJ, Verdaasdonk K, Vincenten MCJ *et al.* Tissue-resident memory T cells invade the brain parenchyma in multiple sclerosis white matter lesions. *Brain* 2020; **143**: 1714–1730.
- 46 Wingerchuk DM, Banwell B, Bennett JL, Cabre P, Carroll W, Chitnis T *et al.* International consensus diagnostic criteria for neuromyelitis optica spectrum disorders. *Neurology* 2015; **85**: 177–189.
- 47 Attfield KE, Armen AP, Kuttikkatte SB, Farooq R, Francis AG, Dendrou CA *et al.* Identification of genetic risk loci associated with aquaporin 4-positive neuromyelitis optica spectrum disorder: A genome-wide association study. *The Lancet Neurology* 2026; **25**: 482–491.
- 48 de Sèze J, Maillart E, Gueguen A, Laplaud DA, Michel L, Thouvenot E *et al.* Anti-CD20 therapies in multiple sclerosis: From pathology to the clinic. *Frontiers in Immunology* 2023; **Volume 14 - 2023**.
- 49 Cencioni MT, Mattosio M, Magliozzi R, Bar-Or A, Muraro PA. B cells in multiple sclerosis — from targeted depletion to immune reconstitution therapies. *Nature Reviews Neurology* 2021; **17**: 399–414.
- 50 Höftberger R, Lassmann H, Berger T, Reindl M. Pathogenic autoantibodies in multiple sclerosis — from a simple idea to a complex concept. *Nature Reviews Neurology* 2022; **18**: 681–688.
- 51 Laroni A, Armentani E, Kerlero De Rosbo N, Ivaldi F, Marcenaro E, Sivori S *et al.* Dysregulation of regulatory CD56bright NK cells/T cells interactions in multiple sclerosis. *Journal of Autoimmunity* 2016; **72**: 8–18.
- 52 Vietzen H, Berger SM, Kühner LM, Furlano PL, Bsteh G, Berger T *et al.* Ineffective control of Epstein-Barr-virus-induced autoimmunity increases the risk for multiple sclerosis. *Cell* 2023; **186**: 5705–5718.e13.
- 53 Van Horssen J, Bö L, Vos CMP, Virtanen I, De Vries HE. Basement Membrane Proteins in Multiple Sclerosis-Associated Inflammatory Cuffs: Potential Role in Influx and Transport of Leukocytes. *Journal of Neuropathology and Experimental Neurology* 2005; **64**: 722–729.

- 54 Smolders J, Fransen NL, Hsiao C-C, Hamann J, Huitinga I. Perivascular tissue resident memory T cells as therapeutic target in multiple sclerosis. *Expert Review of Neurotherapeutics* 2020; **20**: 835–848.
- 55 Smolders J, van Luijn MM, Hsiao C-C, Hamann J. T-cell surveillance of the human brain in health and multiple sclerosis. *Seminars in Immunopathology* 2022; **44**: 855–867.
- 56 Serafini B, Rosicarelli B, Magliozzi R, Stigliano E, Aloisi F. Detection of Ectopic B-cell Follicles with Germinal Centers in the Meninges of Patients with Secondary Progressive Multiple Sclerosis. *Brain Pathology* 2004; **14**: 164–174.
- 57 Magliozzi R, Howell O, Vora A, Serafini B, Nicholas R, Puopolo M *et al.* Meningeal B-cell follicles in secondary progressive multiple sclerosis associate with early onset of disease and severe cortical pathology. *Brain* 2006; **130**: 1089–1104.
- 58 Magliozzi R, Howell OW, Reeves C, Roncaroli F, Nicholas R, Serafini B *et al.* A Gradient of neuronal loss and meningeal inflammation in multiple sclerosis. *Annals of Neurology* 2010; **68**: 477–493.
- 59 Howell OW, Reeves CA, Nicholas R, Carassiti D, Radotra B, Gentleman SM *et al.* Meningeal inflammation is widespread and linked to cortical pathology in multiple sclerosis. *Brain* 2011; **134**: 2755–2771.
- 60 Serafini B, Rosicarelli B, Veroni C, Mazzola GA, Aloisi F. Epstein-Barr Virus-Specific CD8 T Cells Selectively Infiltrate the Brain in Multiple Sclerosis and Interact Locally with Virus-Infected Cells: Clue for a Virus-Driven Immunopathological Mechanism. *Journal of Virology* 2019; **93**: 10.1128/jvi.00980–19.
- 61 Luchetti S, Fransen NL, Van Eden CG, Ramaglia V, Mason M, Huitinga I. Progressive multiple sclerosis patients show substantial lesion activity that correlates with clinical disease severity and sex: A retrospective autopsy cohort analysis. *Acta Neuropathologica* 2018; **135**: 511–528.
- 62 Kuhlmann T, Ludwin S, Prat A, Antel J, Brück W, Lassmann H. An updated histological classification system for multiple sclerosis lesions. *Acta Neuropathologica* 2017; **133**: 13–24.
- 63 Absinta M, Sati P, Masuzzo F, Nair G, Sethi V, Kolb H *et al.* Association of Chronic Active Multiple Sclerosis Lesions With Disability In Vivo. *JAMA Neurology* 2019; **76**: 1474.

- 64 Serafini B, Rosicarelli B, Veroni C, Aloisi F. Tissue-resident memory T cells in the multiple sclerosis brain and their relationship to Epstein-Barr virus infected B cells. *Journal of Neuroimmunology* 2023; **376**: 578036.
- 65 Yong VW. Microglia in multiple sclerosis: Protectors turn destroyers. *Neuron* 2022; **110**: 3534–3548.
- 66 Absinta M, Maric D, Gharagozloo M, Garton T, Smith MD, Jin J *et al.* A lymphocyte–microglia–astrocyte axis in chronic active multiple sclerosis. *Nature* 2021; **597**: 709–714.
- 67 Lerma-Martin C, Badia-i-Mompel P, Ramirez Flores RO, Sekol P, Schäfer PSL, Riedl CJ *et al.* Cell type mapping reveals tissue niches and interactions in subcortical multiple sclerosis lesions. *Nature Neuroscience* 2024; **27**: 2354–2365.
- 68 Feng R, Spieth L, Liu L, Berghoff S, Franz J, Liu Q *et al.* Single-cell spatial transcriptomic profiling defines a pathogenic inflammatory niche in chronic active multiple sclerosis lesions. *Immunity* 2025. doi:10.1016/j.immuni.2025.10.003.
- 69 Olejnik P, Roszkowska Z, Adamus S, Kasarek K. Multiple sclerosis: A narrative overview of current pharmacotherapies and emerging treatment prospects. *Pharmacological Reports* 2024; **76**: 926–943.
- 70 The IFNB Multiple Sclerosis Study Group. Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. *Neurology* 1993; **43**: 655–655.
- 71 Johnson KP, Brooks BR, Cohen JA, Ford CC, Goldstein J, Lisak RP *et al.* Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis. *Neurology* 1995; **45**: 1268–1276.
- 72 Gold R, Kappos L, Arnold DL, Bar-Or A, Giovannoni G, Selmaj K *et al.* Placebo-Controlled Phase 3 Study of Oral BG-12 for Relapsing Multiple Sclerosis. *New England Journal of Medicine* 2012; **367**: 1098–1107.
- 73 Fox RJ, Miller DH, Phillips JT, Hutchinson M, Havrdova E, Kita M *et al.* Placebo-Controlled Phase 3 Study of Oral BG-12 or Glatiramer in Multiple Sclerosis. *New England Journal of Medicine* 2012; **367**: 1087–1097.
- 74 The Lenercept Multiple Sclerosis Study Group and The University of British Columbia MS/MRI Analysis Group. TNF neutralization in MS. *Neurology* 1999; **53**: 457–457.

- 75 Selmaj K, Raine CS, Cross AH. Anti—tumor necrosis factor therapy abrogates autoimmune demyelination. *Annals of Neurology* 1991; **30**: 694–700.
- 76 Sharief MK, Hentges R. Association between Tumor Necrosis Factor- α and Disease Progression in Patients with Multiple Sclerosis. *New England Journal of Medicine* 1991; **325**: 467–472.
- 77 Gregory AP, Dendrou CA, Attfield KE, Haghikia A, Xifara DK, Butter F *et al.* TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis. *Nature* 2012; **488**: 508–511.
- 78 Cohan SL, Lucassen EB, Romba MC, Linch SN. Daclizumab: Mechanisms of Action, Therapeutic Efficacy, Adverse Events and Its Uncovering the Potential Role of Innate Immune System Recruitment as a Treatment Strategy for Relapsing Multiple Sclerosis. *Biomedicines* 2019; **7**: 18.
- 79 Bielekova B, Catalfamo M, Reichert-Scrivner S, Packer A, Cerna M, Waldmann TA *et al.* Regulatory CD56bright natural killer cells mediate immunomodulatory effects of IL-2R α -targeted therapy (daclizumab) in multiple sclerosis. *Proceedings of the National Academy of Sciences* 2006; **103**: 5941–5946.
- 80 Gold R, Giovannoni G, Selmaj K, Havrdova E, Montalban X, Radue E-W *et al.* Daclizumab high-yield process in relapsing-remitting multiple sclerosis (SELECT): A randomised, double-blind, placebo-controlled trial. *The Lancet* 2013; **381**: 2167–2175.
- 81 Kappos L, Wiendl H, Selmaj K, Arnold DL, Havrdova E, Boyko A *et al.* Daclizumab HYP versus Interferon Beta-1a in Relapsing Multiple Sclerosis. *New England Journal of Medicine* 2015; **373**: 1418–1428.
- 82 Polman CH, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH *et al.* A Randomized, Placebo-Controlled Trial of Natalizumab for Relapsing Multiple Sclerosis. *New England Journal of Medicine* 2006; **354**: 899–910.
- 83 Kappos L, Radue E-W, O'Connor P, Polman C, Hohlfeld R, Calabresi P *et al.* A Placebo-Controlled Trial of Oral Fingolimod in Relapsing Multiple Sclerosis. *New England Journal of Medicine* 2010; **362**: 387–401.
- 84 Kappos L, Bar-Or A, Cree BAC, Fox RJ, Giovannoni G, Gold R *et al.* Siponimod versus placebo in secondary progressive multiple sclerosis (EXPAND): A double-blind, randomised, phase 3 study. *The Lancet* 2018; **391**: 1263–1273.

- 85 Bloomgren G, Richman S, Hotermans C, Subramanyam M, Goelz S, Natarajan A *et al.* Risk of Natalizumab-Associated Progressive Multifocal Leukoencephalopathy. *New England Journal of Medicine* 2012; **366**: 1870–1880.
- 86 McGuigan C, Craner M, Guadagno J, Kapoor R, Mazibrada G, Molyneux P *et al.* Stratification and monitoring of natalizumab-associated progressive multifocal leukoencephalopathy risk: Recommendations from an expert group. *Journal of Neurology, Neurosurgery & Psychiatry* 2015; : jnnp-2015-311100.
- 87 Hartung H-P, Gonsette R, König N, Kwiecinski H, Guseo A, Morrissey SP *et al.* Mitoxantrone in progressive multiple sclerosis: A placebo-controlled, double-blind, randomised, multicentre trial. *The Lancet* 2002; **360**: 2018–2025.
- 88 Confavreux C, O'Connor P, Comi G, Freedman MS, Miller AE, Olsson TP *et al.* Oral teriflunomide for patients with relapsing multiple sclerosis (TOWER): A randomised, double-blind, placebo-controlled, phase 3 trial. *The Lancet Neurology* 2014; **13**: 247–256.
- 89 Vermersch P, Czlonkowska A, Grimaldi LM, Confavreux C, Comi G, Kappos L *et al.* Teriflunomide versus subcutaneous interferon beta-1a in patients with relapsing multiple sclerosis: A randomised, controlled phase 3 trial. *Multiple Sclerosis Journal* 2014; **20**: 705–716.
- 90 Giovannoni G, Comi G, Cook S, Rammohan K, Rieckmann P, Sørensen PS *et al.* A Placebo-Controlled Trial of Oral Cladribine for Relapsing Multiple Sclerosis. *New England Journal of Medicine* 2010; **362**: 416–426.
- 91 Hauser SL, Waubant E, Arnold DL, Vollmer T, Antel J, Fox RJ *et al.* B-Cell Depletion with Rituximab in Relapsing–Remitting Multiple Sclerosis. *New England Journal of Medicine* 2008; **358**: 676–688.
- 92 Hauser SL, Bar-Or A, Comi G, Giovannoni G, Hartung HP, Hemmer B *et al.* Ocrelizumab versus Interferon Beta-1a in Relapsing Multiple Sclerosis. *N Engl J Med* 2017; **376**: 221–234.
- 93 Montalban X, Hauser SL, Kappos L, Arnold DL, Bar-Or A, Comi G *et al.* Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis. *N Engl J Med* 2017; **376**: 209–220.
- 94 Steinman L, Fox E, Hartung H-P, Alvarez E, Qian P, Wray S *et al.* Ublituximab versus Teriflunomide in Relapsing Multiple Sclerosis. *New England Journal of Medicine* 2022; **387**: 704–714.

- 95 Hauser SL, Bar-Or A, Cohen JA, Comi G, Correale J, Coyle PK *et al.* Ofatumumab versus Teriflunomide in Multiple Sclerosis. *New England Journal of Medicine* 2020; **383**: 546–557.
- 96 Baker D, Marta M, Pryce G, Giovannoni G, Schmierer K. Memory B Cells are Major Targets for Effective Immunotherapy in Relapsing Multiple Sclerosis. *EBioMedicine* 2017; **16**: 41–50.
- 97 Kappos L, Hartung H-P, Freedman MS, Boyko A, Radü EW, Mikol DD *et al.* Atacicept in multiple sclerosis (ATAMS): A randomised, placebo-controlled, double-blind, phase 2 trial. *The Lancet Neurology* 2014; **13**: 353–363.
- 98 Hartung H-P, Kieseier BC. Atacicept: Targeting B cells in multiple sclerosis. *Therapeutic Advances in Neurological Disorders* 2010; **3**: 205–216.
- 99 van Oosten BW, Lai M, Hodgkinson S, Barkhof F, Miller DH, Moseley IF *et al.* Treatment of multiple sclerosis with the monoclonal anti-CD4 antibody cM-T412: Results of a randomized, double-blind, placebo-controlled MR-monitored phase II trial. *Neurology* 1997; **49**: 351–357.
- 100 Abdul-Majid K-B, Wefer J, Stadelmann C, Stefferl A, Lassmann H, Olsson T *et al.* Comparing the pathogenesis of experimental autoimmune encephalomyelitis in CD4^{-/-} and CD8^{-/-} DBA/1 mice defines qualitative roles of different T cell subsets. *Journal of Neuroimmunology* 2003; **141**: 10–19.
- 101 Coles AJ, Twyman CL, Arnold DL, Cohen JA, Confavreux C, Fox EJ *et al.* Alemtuzumab for patients with relapsing multiple sclerosis after disease-modifying therapy: A randomised controlled phase 3 trial. *The Lancet* 2012; **380**: 1829–1839.
- 102 Cohen JA, Coles AJ, Arnold DL, Confavreux C, Fox EJ, Hartung H-P *et al.* Alemtuzumab versus interferon beta 1a as first-line treatment for patients with relapsing-remitting multiple sclerosis: A randomised controlled phase 3 trial. *The Lancet* 2012; **380**: 1819–1828.
- 103 Havrdova E, Arnold DL, Cohen JA, Hartung H-P, Fox EJ, Giovannoni G *et al.* Alemtuzumab CARE-MS I 5-year follow-up. *Neurology* 2017; **89**: 1107–1116.
- 104 Comi G, Alroughani R, Boster AL, Bass AD, Berkovich R, Fernández Ó *et al.* Efficacy of alemtuzumab in relapsing-remitting MS patients who received additional courses after the initial two courses: Pooled analysis of the CARE-MS, extension, and TOPAZ studies. *Multiple Sclerosis Journal* 2020; **26**: 1866–1876.

- 105 Jimenez-Sanchez S, Maksoud R, Eaton-Fitch N, Marshall-Gradisnik S, Broadley SA. The role of alemtuzumab in the development of secondary autoimmunity in multiple sclerosis: A systematic review. *Journal of Neuroinflammation* 2024; **21**: 281.
- 106 Muraro PA, Pasquini M, Atkins HL, Bowen JD, Farge D, Fassas A *et al.* Long-term Outcomes After Autologous Hematopoietic Stem Cell Transplantation for Multiple Sclerosis. *JAMA Neurology* 2017; **74**: 459–469.
- 107 Kalincik T, Sharmin S, Roos I, Freedman MS, Atkins H, Burman J *et al.* Comparative Effectiveness of Autologous Hematopoietic Stem Cell Transplant vs Fingolimod, Natalizumab, and Ocrelizumab in Highly Active Relapsing-Relapsing Multiple Sclerosis. *JAMA Neurology* 2023; **80**: 702–713.
- 108 Burt RK, Balabanov R, Burman J, Sharrack B, Snowden JA, Oliveira MC *et al.* Effect of Nonmyeloablative Hematopoietic Stem Cell Transplantation vs Continued Disease-Modifying Therapy on Disease Progression in Patients With Relapsing-Relapsing Multiple Sclerosis: A Randomized Clinical Trial. *JAMA* 2019; **321**: 165–174.
- 109 Mariottini A, De Matteis E, Cencioni MT, Muraro PA. Haematopoietic Stem Cell Transplantation for the Treatment of Multiple Sclerosis: Recent Advances. *Current Neurology and Neuroscience Reports* 2023; **23**: 507–520.
- 110 Krämer J, Bar-Or A, Turner TJ, Wiendl H. Bruton tyrosine kinase inhibitors for multiple sclerosis. *Nature Reviews Neurology* 2023; **19**: 289–304.
- 111 Montalban X, Vermersch P, Arnold DL, Bar-Or A, Cree BAC, Cross AH *et al.* Safety and efficacy of evobrutinib in relapsing multiple sclerosis (evolutionRMS1 and evolutionRMS2): Two multicentre, randomised, double-blind, active-controlled, phase 3 trials. *The Lancet Neurology* 2024; **23**: 1119–1132.
- 112 Oh J, Arnold DL, Cree BAC, Ionete C, Kim HJ, Sormani MP *et al.* Tolebrutinib versus Teriflunomide in Relapsing Multiple Sclerosis. *New England Journal of Medicine* 2025; **392**: 1893–1904.
- 113 Fox RJ, Bar-Or A, Traboulsee A, Oreja-Guevara C, Giovannoni G, Vermersch P *et al.* Tolebrutinib in Nonrelapsing Secondary Progressive Multiple Sclerosis. *New England Journal of Medicine* 2025; **392**: 1883–1892.
- 114 Fischbach F, Richter J, Pfeffer LK, Fehse B, Berger SC, Reinhardt S *et al.* CD19-targeted chimeric antigen receptor T cell therapy in two patients with multiple sclerosis. *Med* 2024; **5**: 550–558.e2.

- 115 Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nature Reviews Immunology* 2018; **18**: 35–45.
- 116 Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R *et al.* Massively parallel digital transcriptional profiling of single cells. *Nature Communications* 2017; **8**: 14049.
- 117 Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R *et al.* Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* 2017; **357**: 661–667.
- 118 Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics* 2012; **13**: 227–232.
- 119 Srivastava H, Lippincott MJ, Currie J, Canfield R, Lam MPY, Lau E. Protein prediction models support widespread post-transcriptional regulation of protein abundance by interacting partners. *PLOS Computational Biology* 2022; **18**: e1010702.
- 120 Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nature Methods* 2017; **14**: 865–868.
- 121 Stuart T, Satija R. Integrative single-cell analysis. *Nature Reviews Genetics* 2019; **20**: 257–272.
- 122 Heumos L, Schaar AC, Lance C, Litinetskaya A, Drost F, Zappia L *et al.* Best practices for single-cell analysis across modalities. *Nature Reviews Genetics* 2023; **24**: 550–572.
- 123 Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol* 2018; **36**: 421–427.
- 124 Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* 2019; **16**: 1289–1296.
- 125 Gayoso A, Lopez R, Xing G, Boyeau P, Valiollah Pour Amiri V, Hong J *et al.* A Python library for probabilistic analysis of single-cell omics data. *Nature Biotechnology* 2022; **40**: 163–166.
- 126 Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. *Nature Methods* 2018; **15**: 1053–1058.

- 127 Luecken MD, Büttner M, Chaichoompu K, Danese A, Interlandi M, Mueller MF *et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nature Methods* 2022; **19**: 41–50.
- 128 Stoeckius M, Zheng S, Houck-Loomis B, Hao S, Yeung BZ, Mauck WM *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biology* 2018; **19**: 224.
- 129 Tran HTN, Ang KS, Chevrier M, Zhang X, Lee NYS, Goh M *et al.* A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biology* 2020; **21**: 12.
- 130 Antonsson SE, Melsted P. Batch correction methods used in single-cell RNA sequencing analyses are often poorly calibrated. *Genome Research* 2025; **35**: 1832–1841.
- 131 Mulè MP, Martins AJ, Tsang JS. Normalizing and denoising protein expression data from droplet-based single cell profiling. *Nature Communications* 2022; **13**: 2099.
- 132 Gayoso A, Steier Z, Lopez R, Regier J, Nazon KL, Streets A *et al.* Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nature Methods* 2021; **18**: 272–282.
- 133 Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, Marioni JC *et al.* MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology* 2020; **21**: 111.
- 134 Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A *et al.* Integrated analysis of multimodal single-cell data. *Cell* 2021; **184**: 3573–3587.e29.
- 135 Zheng Y, Caron DP, Kim JY, Jun S-H, Tian Y, Mair F *et al.* ADThorm: Robust integration of single-cell protein measurement across CITE-seq datasets. *Nature Communications* 2025; **16**: 5852.
- 136 Büttner F, Pratanwanich N, McCarthy DJ, Marioni JC, Stegle O. F-scLVM: Scalable and versatile factor analysis for single-cell RNA-seq. *Genome Biology* 2017; **18**: 212.
- 137 Kotliar D, Veres A, Nagy MA, Tabrizi S, Hodis E, Melton DA *et al.* Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *eLife* 2019; **8**: e43803.

- 138 Svensson V, Gayoso A, Yosef N, Pachter L. Interpretable factor models of single-cell RNA-seq via variational autoencoders. *Bioinformatics* 2020; **36**: 3418–3421.
- 139 Wang W, Tan H, Sun M, Han Y, Chen W, Qiu S *et al.* Independent component analysis based gene co-expression network inference (ICAnet) to decipher functional modules for better single-cell clustering and batch integration. *Nucleic Acids Research* 2021; **49**: e54–e54.
- 140 Kunes RZ, Walle T, Land M, Nawy T, Pe’er D. Supervised discovery of interpretable gene programs from single-cell data. *Nature Biotechnology* 2024; **42**: 1084–1095.
- 141 Pouyabahar D, Andrews T, Bader GD. Interpretable single-cell factor decomposition using sciRED. *Nature Communications* 2025; **16**: 1878.
- 142 Kotliar D, Curtis M, Agnew R, Weinand K, Nathan A, Baglaenko Y *et al.* Reproducible single-cell annotation of programs underlying T cell subsets, activation states and functions. *Nature Methods* 2025; **22**: 1964–1980.
- 143 Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nature Methods* 2017; **14**: 297–301.
- 144 Srivatsan SR, McFaline-Figueroa JL, Ramani V, Saunders L, Cao J, Packer J *et al.* Massively multiplex chemical transcriptomics at single-cell resolution. *Science* 2020; **367**: 45–51.
- 145 Cao Y, Lin Y, Ormerod JT, Yang P, Yang JYH, Lo KK. scDC: Single cell differential composition analysis. *BMC Bioinformatics* 2019; **20**: 721.
- 146 Büttner M, Ostner J, Müller CL, Theis FJ, Schubert B. scCODA is a Bayesian model for compositional single-cell data analysis. *Nature Communications* 2021; **12**: 6876.
- 147 Otto DJ, Arriaga-Gomez E, Thieme E, Yang R, Lee SC, Setty M. Comparing phenotypic manifolds with Kompot: Detecting differential abundance and gene expression at single-cell resolution. 2025. doi:10.1101/2025.06.03.657769.
- 148 Boyeau P, Hong J, Gayoso A, Kim M, McFaline-Figueroa JL, Jordan MI *et al.* Deep generative modeling of sample-level heterogeneity in single-cell genomics. *Nature Methods* 2025; **22**: 2264–2274.

- 149 Zhao J, Jaffe A, Li H, Lindenbaum O, Sefik E, Jackson R *et al.* Detection of differentially abundant cell subpopulations in scRNA-seq data. *Proceedings of the National Academy of Sciences* 2021; **118**: e2100293118.
- 150 Burkhardt DB, Stanley JS, Tong A, Perdigoto AL, Gigante SA, Herold KC *et al.* Quantifying the effect of experimental perturbations at single-cell resolution. *Nature Biotechnology* 2021; **39**: 619–629.
- 151 Dann E, Henderson NC, Teichmann SA, Morgan MD, Marioni JC. Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nature Biotechnology* 2022; **40**: 245–253.
- 152 Yi H, Plotkin A, Stanley N. Benchmarking differential abundance methods for finding condition-specific prototypical cells in multi-sample single-cell datasets. *Genome Biology* 2024; **25**: 9.
- 153 Breiman L. Random Forests. *Machine Learning* 2001; **45**: 5–32.
- 154 Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B *et al.* From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence* 2020; **2**: 56–67.
- 155 Cantoni C, Smirnov RA, Firulyova M, Andhey PS, Bradstreet TR, Esaulova E *et al.* A single-cell compendium of human cerebrospinal fluid identifies disease-associated immune cell populations. *The Journal of Clinical Investigation* 2025; **135**. doi:10.1172/JCI177793.
- 156 Esaulova E, Cantoni C, Shchukina I, Zaitsev K, Bucelli RC, Wu GF *et al.* Single-cell RNA-seq analysis of human CSF microglia and myeloid cells in neuroinflammation. *Neurology Neuroimmunology & Neuroinflammation* 2020; **7**: e732.
- 157 Pappalardo JL, Zhang L, Pecsok MK, Perlman K, Zografou C, Raddassi K *et al.* Transcriptomic and clonal characterization of T cells in the human central nervous system. *Science Immunology* 2020; **5**: eabb8786.
- 158 Ramesh A, Schubert RD, Greenfield AL, Dandekar R, Loudermilk R, Sabatino JJ Jr. *et al.* A pathogenic and clonally expanded B cell transcriptome in active multiple sclerosis. *Proc Natl Acad Sci U S A* 2020; **117**: 22932–22943.
- 159 Schafflick D, Xu CA, Hartlehnert M, Cole M, Schulte-Mecklenbeck A, Lautwein T *et al.* Integrated single cell analysis of blood and cerebrospinal fluid leukocytes in multiple sclerosis. *Nat Commun* 2020; **11**: 247.

- 160 Kaufmann M, Evans H, Schaupp AL, Engler JB, Kaur G, Willing A *et al.* Identifying CNS-colonizing T cells as potential therapeutic targets to prevent progression of multiple sclerosis. *Med* 2021; **2**: 296–312 e8.
- 161 Mölder F, Jablonski K, Letcher B, Hall M, Tomkins-Tinch C, Sochat V *et al.* Sustainable data analysis with Snakemake [version 2; peer review: 2 approved]. *F1000Research* 2021; **10**. doi:10.12688/f1000research.29032.2.
- 162 Benjamin K, Dinar Y, Alexander D. STARsolo: Accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. *bioRxiv* 2021; : 2021.05.05.442755.
- 163 Swanson E, Reading J, Graybuck LT, Skene PJ. BarWare: Efficient software tools for barcoded single-cell genomics. *BMC Bioinformatics* 2022; **23**: 106.
- 164 Lun ATL, Riesenfeld S, Andrews T, Dao TP, Gomes T, Marioni JC *et al.* EmptyDrops: Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biology* 2019; **20**: 63.
- 165 Fleming SJ, Chaffin MD, Arduini A, Akkad A-D, Banks E, Marioni JC *et al.* Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nature Methods* 2023; **20**: 1323–1335.
- 166 Klein H-U. Demuxmix: Demultiplexing oligonucleotide-barcoded single-cell RNA sequencing data with regression mixture models. *Bioinformatics* 2023; **39**: btad481.
- 167 Germain P, Lun A, Garcia Meixide C, Macnair W, Robinson M. Doublet identification in single-cell sequencing data using scDblFinder [version 2; peer review: 2 approved]. *F1000Research* 2022; **10**. doi:10.12688/f1000research.73600.2.
- 168 McCarthy DJ, Campbell KR, Lun ATL, Wills QF. Scater: Pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* 2017; **33**: 1179–1186.
- 169 Rainer J, Gatto L, Weichenberger CX. EnsemblDb: An R package to create and use Ensembl-based annotation resources. *Bioinformatics* 2019; **35**: 3151–3153.
- 170 Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology* 2024; **42**: 293–304.

- 171 Domínguez Conde C, Xu C, Jarvis LB, Rainbow DB, Wells SB, Gomes T *et al.* Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* 2022; **376**: eabl5197.
- 172 Gong Q, Sharma M, Glass MC, Kuan EL, Chander A, Singh M *et al.* Multi-omic profiling reveals age-related immune dynamics in healthy adults. *Nature* 2025. doi:10.1038/s41586-025-09686-5.
- 173 Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature Immunology* 2019; **20**: 163–172.
- 174 Monaco G, Lee B, Xu W, Mustafah S, Hwang YY, Carré C *et al.* RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Reports* 2019; **26**: 1627–1640.e7.
- 175 Wolf FA, Angerer P, Theis FJ. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biology* 2018; **19**: 15.
- 176 Li G, Song B, Singh H, Surya Prasath VB, Leighton Grimes H, Salomonis N. Decision level integration of unimodal and multimodal single cell data with scTriangulate. *Nature Communications* 2023; **14**: 406.
- 177 Genovese CR, Roeder K, Wasserman L. False discovery control with p-value weighting. *Biometrika* 2006; **93**: 509–524.
- 178 Fang Z, Liu X, Peltz G. GSEAPy: A comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* 2023; **39**: btac757.
- 179 Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O *et al.* Scikit-learn: Machine learning in python. *Journal of Machine Learning Research* 2011; **12**: 2825–2830.
- 180 Wei J, Moon J, Yasumizu Y, Zhang L, Radassi K, Buitrago-Pocasangre N *et al.* Transcriptomic profiling after B cell depletion reveals central and peripheral immune cell changes in multiple sclerosis. *The Journal of Clinical Investigation* 2025; **135**. doi:10.1172/JCI182790.
- 181 Waede M, Kingo C, Damsbo K, Joergensen MU, Oubounyt M, Gjerstorff M *et al.* Phenotypic Changes in a Monocyte Cluster with High Interleukin-1 Beta Expression during Long-Term Anti-CD20 Therapy. *Annals of Neurology* 2025; **n/a**. doi:10.1002/ana.78004.
- 182 Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using Lme4. *Journal of Statistical Software* 2015; **67**: 1–48.

- 183 Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 2017; **82**: 1–26.
- 184 Lenth RV. Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software* 2016; **69**: 1–33.
- 185 Kappos L, Wolinsky JS, Giovannoni G, Arnold DL, Wang Q, Bernasconi C *et al.* Contribution of Relapse-Independent Progression vs Relapse-Associated Worsening to Overall Confirmed Disability Accumulation in Typical Relapsing Multiple Sclerosis in a Pooled Analysis of 2 Randomized Clinical Trials. *JAMA Neurology* 2020; **77**: 1132.
- 186 Obermeier B, Mentele R, Malotka J, Kellermann J, Kümpfel T, Wekerle H *et al.* Matching of oligoclonal immunoglobulin transcriptomes and proteomes of cerebrospinal fluid in multiple sclerosis. *Nature Medicine* 2008; **14**: 688–693.
- 187 Brändle SM, Obermeier B, Senel M, Bruder J, Mentele R, Khademi M *et al.* Distinct oligoclonal band antibodies in multiple sclerosis recognize ubiquitous self-proteins. *Proceedings of the National Academy of Sciences* 2016; **113**: 7864–7869.
- 188 Baker D, MacDougall A, Kang AS, Schmierer K, Giovannoni G, Dobson R. CD19 B cell repopulation after ocrelizumab, alemtuzumab and cladribine: Implications for SARS-CoV-2 vaccinations in multiple sclerosis. *Multiple Sclerosis and Related Disorders* 2022; **57**: 103448.
- 189 Nissimov N, Hajiyeva Z, Torke S, Grondey K, Brück W, Häusser-Kinzel S *et al.* B cells reappear less mature and more activated after their anti-CD20-mediated depletion in multiple sclerosis. *Proceedings of the National Academy of Sciences* 2020; **117**: 25690–25699.
- 190 Rodriguez-Mogeda C, van Lierop ZYGJ, van der Pol SMA, Coenen L, Hogenboom L, Kamermans A *et al.* Extended interval dosing of ocrelizumab modifies the repopulation of B cells without altering the clinical efficacy in multiple sclerosis. *Journal of Neuroinflammation* 2023; **20**: 215.
- 191 Wang M, Otto C, Fernández Zapata C, Dehlinger A, Gallaccio G, Diekmann L-M *et al.* Comprehensive analysis of B cell repopulation in ocrelizumab-treated patients with multiple sclerosis by mass cytometry and proteomics. *iScience* 2025; **28**: 112383.
- 192 Schneider-Hohendorf T, Wunsch C, Schulte-Mecklenbeck A, Revie L, Raposo C, Strauli N *et al.* B- and T cell receptor sequencing elucidates characteristics of lymphocyte depletion by ocrelizumab. *iScience* 2025; **28**. doi:10.1016/j.isci.2025.113068.

- 193 Beckers L, Baeten P, Popescu V, Swinnen D, Cardilli A, Hamad I *et al.* Alterations in the innate and adaptive immune system in a real-world cohort of multiple sclerosis patients treated with ocrelizumab. *Clinical Immunology* 2024; **259**: 109894.
- 194 Yang S, Zhang T, Feng B, Zhang C. Effects of Ofatumumab Treatment on Immune Cells Landscape in Patients With Relapsing-remitting Multiple Sclerosis (P7-6.014). *Neurology* 2024; **102**: 3417.
- 195 Feige J, Moser T, Akgün K, Schwenker K, Hitzl W, Haschke-Becher E *et al.* Repeated iv ANTI-CD20 treatment in multiple sclerosis: Long-term effects on peripheral immune cell subsets. *Annals of Clinical and Translational Neurology* 2024; **11**: 450–465.
- 196 Björnsson GS, Sigurgrímsdóttir H, Maggadóttir SM, Einarsdóttir BÓ, Sveinsson ÓÁ, Hjaltason H *et al.* Long-term use of rituximab increases T cell count in MS patients. *Frontiers in Immunology* 2024; **15**: 1412668.
- 197 Willison AG, Hagler R, Weise M, Elben S, Huntemann N, Masanneck L *et al.* Effects of Anti-CD20 Antibody Therapy on Immune Cell Dynamics in Relapsing-Remitting Multiple Sclerosis. *Cells* 2025; **14**: 552.
- 198 Garcia A, Dugast E, Shah S, Morille J, Lebrun-Frenay C, Thouvenot E *et al.* Immune Profiling Reveals the T-Cell Effect of Ocrelizumab in Early Relapsing-Remitting Multiple Sclerosis. *Neurol Neuroimmunol Neuroinflamm* 2023; **10**: e200091.
- 199 Palanichamy A, Jahn S, Nickles D, Derstine M, Abounasr A, Hauser SL *et al.* Rituximab Efficiently Depletes Increased CD20-Expressing T Cells in Multiple Sclerosis Patients. *The Journal of Immunology* 2014; **193**: 580–586.
- 200 Wilk E, Witte T, Marquardt N, Horvath T, Kalippke K, Scholz K *et al.* Depletion of functionally active CD20+ T cells by rituximab treatment. *Arthritis & Rheumatism* 2009; **60**: 3563–3571.
- 201 Gingele S, Jacobus TL, Konen FF, Hümmert MW, Sühs K-W, Schwenkenbecher P *et al.* Ocrelizumab Depletes CD20+ T Cells in Multiple Sclerosis Patients. *Cells* 2019; **8**. doi:10.3390/cells8010012.
- 202 Shinoda K, Li R, Rezk A, Mexhitaj I, Patterson KR, Kakara M *et al.* Differential effects of anti-CD20 therapy on CD4 and CD8 T cells and implication of CD20-expressing CD8 T cells in MS disease activity. *Proceedings of the National Academy of Sciences* 2023; **120**: e2207291120.

References

- 203 Mathias A, Pantazou V, Perriot S, Canales M, Jones S, Oberholster L *et al.* Ocrelizumab Impairs the Phenotype and Function of Memory CD8+ T Cells. *Neurology Neuroimmunology & Neuroinflammation* 2023; **10**: e200084.
- 204 Abbadessa G, Lepore MT, Bruzzaniti S, Piemonte E, Miele G, Signoriello E *et al.* Ocrelizumab Alters Cytotoxic Lymphocyte Function While Reducing EBV-Specific CD8+ T-Cell Proliferation in Patients With Multiple Sclerosis. *Neurology Neuroimmunology & Neuroinflammation* 2024; **11**: e200250.
- 205 Sabatino JJ, Wilson MR, Calabresi PA, Hauser SL, Schneck JP, Zamvil SS. Anti-CD20 therapy depletes activated myelin-specific CD8+ T cells in multiple sclerosis. *Proceedings of the National Academy of Sciences* 2019; **116**: 25800–25807.
- 206 Lehmann-Horn K, Schleich E, Hertenberg D, Hapfelmeier A, Kämpfel T, Von Bubnoff N *et al.* Anti-CD20 B-cell depletion enhances monocyte reactivity in neuroimmunological disorders. *Journal of Neuroinflammation* 2011; **8**: 146.
- 207 Ochs J, Nissimov N, Torke S, Freier M, Grondy K, Koch J *et al.* Proinflammatory CD20+ T cells contribute to CNS-directed autoimmunity. *Science Translational Medicine* 2022; **14**: eabi4632.
- 208 Schuh E, Berer K, Mulazzani M, Feil K, Meinel I, Lahm H *et al.* Features of Human CD3+CD20+ T Cells. *The Journal of Immunology* 2016; **197**: 1111–1117.
- 209 von Essen MR, Stolpe LE, Bach Søndergaard H, Sellebjerg F. The origin of human CD20+ T cells: A stolen identity? *Frontiers in Immunology* 2024; **Volume 15 - 2024**.
- 210 Butzkueven H, Farr P, Ozakbas S, Boz C, Kalincik T, Taylor L *et al.* Real-World Effectiveness of Ocrelizumab in Relapsing Multiple Sclerosis: An MSBase Registry Sub-Study. *Multiple Sclerosis and Related Disorders* 2025. doi:10.1016/j.msard.2025.106885.
- 211 Smyth LCD, Kipnis J. Redefining CNS immune privilege. *Nature Reviews Immunology* 2025; **25**: 766–775.
- 212 Ostkamp P, Deffner M, Schulte-Mecklenbeck A, Wunsch C, Lu IN, Wu GF *et al.* A single-cell analysis framework allows for characterization of CSF leukocytes and their tissue of origin in multiple sclerosis. *Sci Transl Med* 2022; **14**: eadc9778.

- 213 Holm Hansen R, Talbot J, Højsgaard Chow H, Bredahl Hansen M, Buhelt S, Herich S *et al.* Increased Intrathecal Activity of Follicular Helper T Cells in Patients With Relapsing-Remitting Multiple Sclerosis. *Neurology Neuroimmunology & Neuroinflammation* 2022; **9**: e200009.
- 214 Feger U, Luther C, Poeschel S, Melms A, Tolosa E, Wiendl H. Increased frequency of CD4+ CD25+ regulatory T cells in the cerebrospinal fluid but not in the blood of multiple sclerosis patients. *Clinical and Experimental Immunology* 2007; **147**: 412–418.
- 215 Beltrán E, Gerdes LA, Hansen J, Flierl-Hecht A, Krebs S, Blum H *et al.* Early adaptive immune activation detected in monozygotic twins with prodromal multiple sclerosis. *The Journal of Clinical Investigation* 2019; **129**: 4758–4768.
- 216 Hayashi F, Mittl K, Dandekar R, Gerdts J, Hassan E, Schubert RD *et al.* Antigen specificity of clonally enriched CD8+ T cells in multiple sclerosis. *Nature Immunology* 2026; **27**: 490–502.
- 217 Cepok S, Rosche B, Grummel V, Vogel F, Zhou D, Sayn J *et al.* Short-lived plasma blasts are the main B cell effector subset during the course of multiple sclerosis. *Brain* 2005; **128**: 1667–1676.
- 218 Eggers EL, Michel BA, Wu H, Wang S, Bevan CJ, Abounasr A *et al.* Clonal relationships of CSF B cells in treatment-naïve multiple sclerosis patients. *JCI Insight* 2017; **2**: e92724.
- 219 Kister I, Lotan I, Wallach A, Bacon T, Cutter G, Arbini A. CSF plasmablasts differentiate MS from other neurologic disorders. *Multiple Sclerosis and Related Disorders* 2021; **48**: 102719.
- 220 Jacobs BM, Gasperi C, Kalluri SR, Al-Najjar R, McKeon MO, Else J *et al.* Single-cell analysis of cerebrospinal fluid reveals common features of neuroinflammation. *Cell Reports Medicine* 2025; **6**: 101733.
- 221 Lucchinetti C, Brck W, Parisi J, Scheithauer B, Rodriguez M, Lassmann H. Heterogeneity of multiple sclerosis lesions: Implications for the pathogenesis of demyelination. *Annals of Neurology* 2000; **47**: 707–717.
- 222 Dal-Bianco A, Grabner G, Kronnerwetter C, Weber M, Höftberger R, Berger T *et al.* Slow expansion of multiple sclerosis iron rim lesions: Pathology and 7 T magnetic resonance imaging. *Acta Neuropathologica* 2017; **133**: 25–42.
- 223 Smolders J, Heutinck KM, Fransen NL, Remmerswaal EBM, Hombrink P, Ten Berge IJM *et al.* Tissue-resident memory T cells populate the human brain. *Nature Communications* 2018; **9**: 4593.

References

- 224 Louveau A, Herz J, Alme MN, Salvador AF, Dong MQ, Viar KE *et al.* CNS lymphatic drainage and neuroinflammation are regulated by meningeal lymphatic vasculature. *Nature Neuroscience* 2018; **21**: 1380–1391.
- 225 Papadopoulos Z, Smyth LCD, Smirnov I, Gibson DA, Herz J, Kipnis J. Differential impact of lymphatic outflow pathways on cerebrospinal fluid homeostasis. *Journal of Experimental Medicine* 2025; **222**: e20241752.
- 226 Meng H, Zhao H, Cao X, Hao J, Zhang H, Liu Y *et al.* Double-negative T cells remarkably promote neuroinflammation after ischemic stroke. *Proceedings of the National Academy of Sciences* 2019; **116**: 5558–5563.
- 227 Mackay LK, Minnich M, Kragten NAM, Liao Y, Nota B, Seillet C *et al.* Hobit and Blimp1 instruct a universal transcriptional program of tissue residency in lymphocytes. *Science* 2016; **352**: 459–463.
- 228 Parga-Vidal L, Behr FM, Kragten NAM, Nota B, Wesselink TH, Kavazović I *et al.* Hobit identifies tissue-resident memory T cell precursors that are regulated by Eomes. *Science Immunology* 2021; **6**: eabg3533.
- 229 Veroni C, Serafini B, Rosicarelli B, Fagnani C, Aloisi F. Transcriptional profile and Epstein-Barr virus infection status of laser-cut immune infiltrates from the brain of patients with progressive multiple sclerosis. *Journal of Neuroinflammation* 2018; **15**: 18.
- 230 Colombo M, Dono M, Gazzola P, Roncella S, Valetto A, Chiorazzi N *et al.* Accumulation of Clonally Related B Lymphocytes in the Cerebrospinal Fluid of Multiple Sclerosis Patients. *The Journal of Immunology* 2000; **164**: 2782–2789.
- 231 Owens GP, Bennett JL, Lassmann H, O'Connor KC, Ritchie AM, Shearer A *et al.* Antibodies produced by clonally expanded plasma cells in multiple sclerosis cerebrospinal fluid. *Annals of Neurology* 2009; **65**: 639–649.
- 232 Laichalk LL, Thorley-Lawson DA. Terminal Differentiation into Plasma Cells Initiates the Replicative Cycle of Epstein-Barr Virus In Vivo. *Journal of Virology* 2005; **79**: 1296–1307.
- 233 Butzkueven H, Kappos L, Wiendl H, Trojano M, Spelman T, Chang I *et al.* Long-term safety and effectiveness of natalizumab treatment in clinical practice: 10 years of real-world data from the Tysabri Observational Program (TOP). *Journal of Neurology, Neurosurgery & Psychiatry* 2020; **91**: 660–668.

- 234 Rollot F, Couturier J, Casey R, Wiertlewski S, Debouverie M, Pelletier J *et al.* Comparative Effectiveness of Natalizumab Versus Anti-CD20 in Highly Active Relapsing–Remitting Multiple Sclerosis After Fingolimod Withdrawal. *Neurotherapeutics* 2022; **19**: 476–490.
- 235 Barbuti E, Castiello A, Pozzilli V, Carotenuto A, Tomasso I, Moccia M *et al.* Comparative effectiveness, safety and persistence of ocrelizumab versus natalizumab in multiple sclerosis: A real-world, multi-center, propensity score-matched study. *Neurotherapeutics* 2025; **22**: e00537.
- 236 Kappos L, O'Connor PW, Polman CH, Vermersch P, Wiendl H, Pace A *et al.* Clinical effects of natalizumab on multiple sclerosis appear early in treatment course. *Journal of Neurology* 2013; **260**: 1388–1395.
- 237 Plavina T, Muralidharan KK, Kuesters G, Mikol D, Evans K, Subramanyam M *et al.* Reversibility of the effects of natalizumab on peripheral immune cell dynamics in MS patients. *Neurology* 2017; **89**: 1584–1593.
- 238 Klotz L, Berger T, Brownlee WJ, Chan A, Lycke J, Oreja-Guevara C *et al.* Twenty years of natalizumab in multiple sclerosis: Lessons learned and future outlook. *Therapeutic Advances in Neurological Disorders* 2025; **18**: 17562864251372752.
- 239 Ryerson LZ, Foley J, Chang I, Kister I, Cutter G, Metzger RR *et al.* Risk of natalizumab-associated PML in patients with MS is reduced with extended interval dosing. *Neurology* 2019; **93**. doi:10.1212/WNL.00000000000008243.
- 240 Ryerson LZ, Foley JF, Defer G, Cohen JA, Arnold DL, Butzkueven H *et al.* Exploratory clinical efficacy and patient-reported outcomes from NOVA: A randomized controlled study of intravenous natalizumab 6-week dosing versus continued 4-week dosing for relapsing-remitting multiple sclerosis. *Multiple Sclerosis and Related Disorders* 2023; **72**: 104561.
- 241 Stüve O, Marra CM, Jerome KR, Cook L, Cravens PD, Cepok S *et al.* Immune surveillance in multiple sclerosis patients treated with natalizumab. *Annals of Neurology* 2006; **59**: 743–747.
- 242 Planas R, Jelčić I, Schippling S, Martin R, Sospedra M. Natalizumab treatment perturbs memory- and marginal zone-like B-cell homing in secondary lymphoid organs in multiple sclerosis. *European Journal of Immunology* 2012; **42**: 790–798.
- 243 Putzki N, Baranwal MK, Tettenborn B, Limmroth V, Kreuzfelder E. Effects of Natalizumab on Circulating B Cells, T Regulatory Cells and Natural Killer Cells. *European Neurology* 2010; **63**: 311–317.

- 244 Koudriavtseva T, Sbardella E, Trento E, Bordignon V, D'Agosto G, Cordiali-Fei P. Long-term follow-up of peripheral lymphocyte subsets in a cohort of multiple sclerosis patients treated with natalizumab. *Clinical and Experimental Immunology* 2014; **176**: 320–326.
- 245 Kaufmann M, Haase R, Proschmann U, Ziemssen T, Akgün K. Real-World Lab Data in Natalizumab Treated Multiple Sclerosis Patients Up to 6 Years Long-Term Follow Up. *Frontiers in Neurology* 2018; **9**: 1071.
- 246 Kivisäkk P, Healy BC, Viglietta V, Quintana FJ, Hootstein MA, Weiner HL *et al.* Natalizumab treatment is associated with peripheral sequestration of proinflammatory T cells. *Neurology* 2009; **72**: 1922–1930.
- 247 Kimura K, Nakamura M, Sato W, Okamoto T, Araki M, Lin Y *et al.* Disrupted balance of T cells under natalizumab treatment in multiple sclerosis. *Neurology Neuroimmunology & Neuroinflammation* 2016; **3**: e210.
- 248 Van Langelaar J, Van Der Vuurst De Vries RM, Janssen M, Wierenga-Wolf AF, Spilt IM, Siepman TA *et al.* T helper 17.1 cells associate with multiple sclerosis disease activity: Perspectives for early intervention. *Brain* 2018; **141**: 1334–1349.
- 249 Börnsen L, Christensen JR, Ratzer R, Oturai AB, Sørensen PS, Søndergaard HB *et al.* Effect of Natalizumab on Circulating CD4+ T-Cells in Multiple Sclerosis. *PLoS ONE* 2012; **7**: e47578.
- 250 Tanasescu R, Frakich N, Chou I-J, Filippini P, Podda G, Xin G *et al.* Natalizumab Treatment of Relapsing Remitting Multiple Sclerosis Has No Long-Term Effects on the Proportion of Circulating Regulatory T Cells. *Neurology and Therapy* 2023; **12**: 2041–2052.
- 251 Traub JW, Pellkofer HL, Grondey K, Seeger I, Rowold C, Brück W *et al.* Natalizumab promotes activation and pro-inflammatory differentiation of peripheral B cells in multiple sclerosis patients. *Journal of Neuroinflammation* 2019; **16**: 228.
- 252 Cuculiza Henriksen A, Ammitzbøll C, Petersen ER, McWilliam O, Sellebjerg F, Von Essen MR *et al.* Natalizumab differentially affects plasmablasts and B cells in multiple sclerosis. *Multiple Sclerosis and Related Disorders* 2021; **52**: 102987.
- 253 Liampas A, Tseriotis V-S, Mavridis T, Vavougiou GD, Zis P, Hadjigeorgiou GM *et al.* Effects of natalizumab on oligoclonal bands in the cerebrospinal fluid of patients with multiple sclerosis: A systematic review and meta-analysis. *Neurological Sciences* 2025; **46**: 1541–1553.

- 254 Frisch ES, Häusler D, Weber MS. Natalizumab Promotes Activation of Peripheral Monocytes in Patients With Multiple Sclerosis. *Neurology Neuroimmunology & Neuroinflammation* 2023; **10**: e200114.
- 255 Xu W, Zhao X, Wang X, Feng H, Gou M, Jin W *et al.* The Transcription Factor Tox2 Drives T Follicular Helper Cell Development via Regulating Chromatin Accessibility. *Immunity* 2019; **51**: 826–839.e5.
- 256 Horiuchi S, Wu H, Liu W-C, Schmitt N, Provot J, Liu Y *et al.* Tox2 is required for the maintenance of GC T_{FH} cells and the generation of memory T_{FH} cells. *Science Advances* 2021; **7**: eabj1249.
- 257 Schmitt N. Role of T Follicular Helper cells in Multiple Sclerosis. *Journal of Nature and Science* 2015; **1**: e139.
- 258 Quinn JL, Axtell RC. Emerging Role of Follicular T Helper Cells in Multiple Sclerosis and Experimental Autoimmune Encephalomyelitis. *International Journal of Molecular Sciences* 2018; **19**: 3233.
- 259 Morille J, Mandon M, Rodriguez S, Roulois D, Leonard S, Garcia A *et al.* Multiple Sclerosis CSF Is Enriched With Follicular T Cells Displaying a Th1/Eomes Signature. *Neurology Neuroimmunology & Neuroinflammation* 2022; **9**: e200033.
- 260 Lassmann H. Pathogenic Mechanisms Associated With Different Clinical Courses of Multiple Sclerosis. *Frontiers in Immunology* 2019; **9**: 3116.
- 261 Merkler D, Vincenti I, Masson F, Liblau RS. Tissue-resident CD8 T cells in central nervous system inflammatory diseases: Present at the crime scene and ...guilty. *Current Opinion in Immunology* 2022; **77**: 102211.
- 262 Gottlieb A, Pham HPT, Saltarrelli JG, Lindsey JW. Expanded T lymphocytes in the cerebrospinal fluid of multiple sclerosis patients are specific for Epstein-Barr-virus-infected B cells. *Proceedings of the National Academy of Sciences* 2024; **121**: e2315857121.
- 263 Serafini B, Rosicarelli B, Franciotta D, Magliozzi R, Reynolds R, Cinque P *et al.* Dysregulated Epstein-Barr virus infection in the multiple sclerosis brain. *The Journal of Experimental Medicine* 2007; **204**: 2899–2912.
- 264 Veroni C, Aloisi F. The CD8 T Cell-Epstein-Barr Virus-B Cell Dialogue: A Central Issue in Multiple Sclerosis Pathogenesis. *Frontiers in Immunology* 2021; **Volume 12 - 2021**.

- 265 Giovannoni G, Butzkueven H, Dhib-Jalbut S, Hobart J, Kobelt G, Pepper G *et al.* Brain health: Time matters in multiple sclerosis. *Multiple Sclerosis and Related Disorders* 2016; **9**: S5–S48.
- 266 Brown JWL, Coles A, Horakova D, Havrdova E, Izquierdo G, Prat A *et al.* Association of Initial Disease-Modifying Therapy With Later Conversion to Secondary Progressive Multiple Sclerosis. *JAMA* 2019; **321**: 175.
- 267 Harding K, Williams O, Willis M, Hrastelj J, Rimmer A, Joseph F *et al.* Clinical Outcomes of Escalation vs Early Intensive Disease-Modifying Therapy in Patients With Multiple Sclerosis. *JAMA Neurology* 2019; **76**: 536.
- 268 Filippi M, Amato MP, Centonze D, Gallo P, Gasperini C, Inglese M *et al.* Early use of high-efficacy disease-modifying therapies makes the difference in people with multiple sclerosis: An expert opinion. *Journal of Neurology* 2022; **269**: 5382–5394.
- 269 Giovannoni G. EBV-specific T-cell responses are telling us something important about multiple sclerosis. *Brain* 2025; **148**: 692–694.
- 270 Lindsey JW. Epstein-Barr virus and multiple sclerosis: A mechanism comes into focus. *Multiple Sclerosis and Related Disorders* 2026; **105**: 106884.
- 271 Pender MP. The Essential Role of Epstein-Barr Virus in the Pathogenesis of Multiple Sclerosis. *The Neuroscientist* 2011; **17**: 351–367.
- 272 Pender MP, Csurhes PA, Pfluger CMM, Burrows SR. CD8 T cell deficiency impairs control of Epstein–Barr virus and worsens with age in multiple sclerosis: Figure 1. *Journal of Neurology, Neurosurgery & Psychiatry* 2012; **83**: 353–354.
- 273 McLane LM, Abdel-Hakeem MS, Wherry EJ. CD8 T Cell Exhaustion During Chronic Viral Infection and Cancer. *Annual Review of Immunology* 2019; **37**: 457–495.
- 274 Giovannoni G, James L, Adeniran AA, Gold J, Young LS, Selwood DL *et al.* The case for targeting latent and lytic Epstein-Barr virus infection in multiple sclerosis. *Brain* 2025; **148**: 3057–3071.
- 275 Willis SN, Stadelmann C, Rodig SJ, Caron T, Gattenloehner S, Mallozzi SS *et al.* Epstein–Barr virus infection is not a characteristic feature of multiple sclerosis brain. *Brain* 2009; **132**: 3318–3328.

- 276 Sargsyan SA, Shearer AJ, Ritchie AM, Burgoon MP, Anderson S, Hemmer B *et al.* Absence of Epstein-Barr virus in the brain and CSF of patients with multiple sclerosis. *Neurology* 2010; **74**: 1127–1135.
- 277 Lycke J, Svennerholm B, Hjelmquist E, Frisn L, Badr G, Andersson M *et al.* Acyclovir treatment of relapsing-remitting multiple sclerosis: A randomized, placebo-controlled, double-blind study. *Journal of Neurology* 1996; **243**: 214–224.
- 278 Bech E, Lycke J, Gadeberg P, Hansen HJ, Malmeström C, Andersen O *et al.* A randomized, double-blind, placebo-controlled MRI study of anti-herpes virus therapy in MS. *Neurology* 2002; **58**: 31–36.
- 279 Li N, Thompson S, Schultz DC, Zhu W, Jiang H, Luo C *et al.* Discovery of Selective Inhibitors Against EBNA1 via High Throughput In Silico Virtual Screening. *PLoS ONE* 2010; **5**: e10126.
- 280 Messick TE, Smith GR, Soldan SS, McDonnell ME, Deakyne JS, Malecka KA *et al.* Structure-based design of small-molecule inhibitors of EBNA1 DNA binding blocks Epstein-Barr virus latent infection and tumor growth. *Science Translational Medicine* 2019; **11**: eaau5612.
- 281 Mei Y-P, Zhu X-F, Zhou J-M, Huang H, Deng R, Zeng Y-X. siRNA targeting LMP1-induced apoptosis in EBV-positive lymphoma cells is associated with inhibition of telomerase activity and expression. *Cancer Letters* 2006; **232**: 189–198.
- 282 Pender MP, Burrows SR. Epstein–Barr virus and multiple sclerosis: Potential opportunities for immunotherapy. *Clinical & Translational Immunology* 2014; **3**: e27.
- 283 Vermersch P, Granziera C, Mao-Draayer Y, Cutter G, Kalbus O, Staikov I *et al.* Inhibition of CD40L with Frexalimab in Multiple Sclerosis. *New England Journal of Medicine* 2024; **390**: 589–600.
- 284 Bhadra R, Cobb DA, Khan IA. CD40 Signaling to the Rescue: A CD8 Exhaustion Perspective in Chronic Infectious Diseases. *Critical Reviews in Immunology* 2013; **33**: 361–378.
- 285 Guo C-L, Wang C-S, Wang X-H, Yu D, Liu Z. GZMK+CD8+ T cells: Multifaceted roles beyond cytotoxicity. *Trends in Immunology* 2025; **46**: 562–572.
- 286 Xin C, Liu P, Zhan Q, Cao W. Gzmk+ CD8 T cells in inflammatory diseases. *Frontiers in Immunology* 2025; **16**: 1661755.

References

- 287 Cooper DM, Pechkovsky DV, Hackett TL, Knight DA, Granville DJ. Granzyme K Activates Protease-Activated Receptor-1. *PLoS ONE* 2011; **6**: e21484.
- 288 Sharma M, Merkulova Y, Raithatha S, Parkinson LG, Shen Y, Cooper D *et al.* Extracellular granzyme K mediates endothelial activation through the cleavage of protease-activated receptor-1. *The FEBS Journal* 2016; **283**: 1734–1747.
- 289 Herich S, Schneider-Hohendorf T, Rohlmann A, Khaleghi Ghadiri M, Schulte-Mecklenbeck A, Zondler L *et al.* Human CCR5^{high} effector memory cells perform CNS parenchymal immune surveillance via GZMK-mediated transendothelial diapedesis. *Brain* 2019; **142**: 3411–3427.
- 290 Donado CA, Theisen E, Zhang F, Nathan A, Fairfield ML, Rupani KV *et al.* Granzyme K activates the entire complement cascade. *Nature* 2025; **641**: 211–221.
- 291 Mahad D, Ransohoff R. CCR5 as a potential therapeutic target in multiple sclerosis. *Drugs of the Future* 2008; **33**: 705.
- 292 Bentzen AK, Marquard AM, Lyngaa R, Saini SK, Ramskov S, Donia M *et al.* Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nature Biotechnology* 2016; **34**: 1037–1045.
- 293 Ingelfinger F, Levy N, Ergen C, Bakulin A, Becker A, Boyeau P *et al.* CytoVI: Deep generative modeling of antibody-based single cell technologies. 2025. doi:10.1101/2025.09.07.674699.

Supplementary files

Supplementary File 1

Marker genes for each coarse (level 2) cell type identified in the discovery cohort dataset.

Supplementary File 2

Marker genes for each fine (level 3) cell type identified in the discovery cohort dataset.

Supplementary File 3

Top 100 weighted genes for each cNMF-derived gene program from B cells, T cells, and monocytes in the discovery cohort dataset.

Supplementary File 4

GO enrichment analysis results for cNMF-derived gene programs for B cells, T cells, and monocytes in the discovery cohort dataset.

Supplementary File 5

Differential expression testing results between T_c1-TH high versus low memory T cells in the discovery cohort dataset.

Software packages

Table B.1. Key software packages.

Software	Version	Repository
bcl2fastq	2.20.0.422	–
STAR	2.7.11a	https://github.com/alexdobin/STAR
BarCounter	1.0.0	https://github.com/AllenInstitute/BarCounter-release
R	4.3.3	https://github.com/wch/r-source
Python	3.11.10	https://github.com/python/cpython
Snakemake	7.31.1	https://github.com/snakemake/snakemake
Seurat	5.1.0	https://github.com/satijalab/seurat
Scanpy	1.10.3	https://github.com/scverse/scanpy
DropletUtils	1.22.0	https://github.com/MarioniLab/DropletUtils
CellBender	0.3.2	https://github.com/broadinstitute/CellBender
demuxmix	1.4.0	https://github.com/huklein/demuxmix
scDbtFinder	1.16.0	https://github.com/plger/scDbtFinder
scater	1.30.1	https://github.com/alanocallaghan/scater
ensembladb	2.26.0	https://github.com/jorainer/ensembladb
harmony	1.2.1	https://github.com/immunogenomics/harmony
harmonypy	0.0.10	https://github.com/slowkow/harmonypy
CellTypist	1.7.1	https://github.com/Teichlab/celltypist
SingleR	2.4.0	https://github.com/dviraran/SingleR
scTriangulate	0.13.0	https://github.com/franklign/scTriangulate
MELD	1.0.2	https://github.com/KrishnaswamyLab/MELD
cNMF	1.7.0	https://github.com/dylkot/cNMF
GSEAPy	0.10.4	https://github.com/zqfang/GSEAPy
scikit-learn	1.5.2	https://github.com/scikit-learn/scikit-learn
SHAP	0.50.0	https://github.com/shap/shap
scvi-tools	1.2.0	https://github.com/scverse/scvi-tools
starCAT	1.0.9	https://github.com/immunogenomics/starCAT
lme4	1.1_35.5	https://github.com/lme4/lme4
lmerTest	3.1_3	https://github.com/runehaubo/lmerTestR
emmeans	1.11.2_8	https://github.com/rvlenth/emmeans
tidyverse	2.0.0	https://github.com/tidyverse
ggpubr	0.6.0	https://github.com/kassambara/ggpubr
patchwork	1.3.0	https://github.com/thomasp85/patchwork
cowplot	1.1.3	https://github.com/wilkelab/cowplot
ComplexHeatmap	2.18.0	https://github.com/jokergoo/ComplexHeatmap
Matplotlib	3.9.2	https://github.com/matplotlib/matplotlib

Publications and presentations

Publications:

1. Attfield KE, Armen AP, Kuttikkatte SB, **Farooq R**, Francis AG, Dendrou CA, Leite MI, Waters P, International NMOSD Genetics Consortium, Weinshenker BG, Karczewski KJ, Neale BM, Jensen LT, Rossjohn J, Gold R, Korn T, McVean G, Palace J, Fugger L. Identification of genetic risk loci associated with aquaporin 4-positive neuromyelitis optica spectrum disorder: a genome-wide association study. *The Lancet Neurology* 2026; **25**: 482–491.
2. **Farooq R**, Cutler B, Pisa M, Kuttikkatte SB, Palace J, DeLuca GC, Gold R, Theis FJ, Fugger L, Attfield KE. Multimodal discovery of a pathogenic B cell-dependent T cell state in multiple sclerosis. Submitted to *Brain* (under peer review).

Presentations:

1. **Farooq R**. High-throughput single cell multiomic analysis of multiple sclerosis patient blood to identify causal immune cell subsets. Presented at *ABN Annual Meeting*, Liverpool, UK, May 2025 [Oral Presentation].