Macrophage-matrix interactions in the breast tumour microenvironment: the case of tenascin-C

Anís Nicholas Gammage

St. Cross College

University of Oxford

Supervisors

Professor Kim Midwood

Professor Irina Udalova

Thesis submitted for the Degree of

Doctor of Philosophy

Molecular and Cellular Medicine

Michaelmas, 2021

Table of Contents

Statement of Originality *5*
Abstract *6*
Abbreviations *8*
Acknowledgments *17*
Chapter 1 | Introduction *19*
1.1 | Overview **20**
1.2 | Cancer immunology at a glance **21**
1.3 | Breast cancer, immunity and immunotherapy **27**
1.4 | Macrophages and breast cancer: into the single-cell era **35**
1.5 | Extracellular matrix in breast cancer and inflammation **48**
1.6 | Tenascin-C in breast cancer **51**
1.7 | Thesis rationale and summary **55**
Chapter 2 | Materials and Methods *57*
2.1 | Establishment of NT193 cell lines **58**
2.2 | Culture of NT193 cell lines **59**
2.3 | Counting and viability testing of cells **60**
2.4 | Freezing and thawing of cells **60**
2.5 | Imaging of cells in culture **61**
2.6 | Use of animals **61**
2.7 | Orthotopic mammary tumour engraftment model **62**
2.8 | Dissociation of breast tumour tissue **63**
2.9 | Flow cytometry of breast tumour tissue **63**
2.9.1 | Sample preparation 63
2.9.2 | Antibodies and dyes 64
2.9.3 | Surface and intracellular analysis 66
2.9.4 | Sorting 67
2.9.5 | Compensation controls 68
2.9.6 | Gating strategies 69
2.9.7 | Software 69
2.10 | Droplet-based single-cell RNA sequencing **70**
2.11 | Computational analysis of mouse scRNAseq data **71**
2.11.1 | Generation of FASTQ files 71
2.11.2 | Data processing and quality control 71
2.11.3 | Feature selection and dimensionality reduction 72
2.11.4 | Unsupervised graph-based clustering 73
2.11.5 | Differential gene expression analysis 74
2.11.6 | Visualisation of gene expression distributions 74
2.11.7 | Geneset over-representation analysis 75
2.11.8 | Trajectory analysis with diffusion maps 75
2.11.9 | Pseudotemporal modelling of macrophage plasticity 76
2.11.10 | Triaging of marker genes for detection and isolation of macrophage subsets 76

2.12 | RT-qPCR **77**

2.13 | Differentiation and Isolation of BMDMs **78**

2.14 | ELISA **79**
2.14.1 | Kits 79
2.14.2 | BMDM culture supernatants 79
2.14.3 | FACS-sorted **Mac2**-like and **Mac3**-like culture supernatants 80
2.14.4 | Data acquisition and Analysis 80

2.15 | Pharmacological inhibition of KDM6B *in vivo* **81**

2.16 | Confocal immunofluorescence imaging **82**
2.16.1 | Tumour processing 82
2.16.2 | Antibodies 83
2.16.3 | Immuno-staining of tumour sections 83
2.16.4 | Data acquisition 84

2.17 | Mouse to human gene conversion **84**

2.18 | Human single-cell RNAseq conservation analysis **85**
2.18.1 | Azizi *et al.* dataset 85
2.18.2 | Wu *et al.* dataset 86

2.19 | Human bulk transcriptomic datasets **86**
2.19.1 | TCGA 86
2.19.2 | METABRIC 87

2.20 | Gene signature enrichments and quality control **87**

2.21 | Ecotype clustering **88**

2.22 | CIBERSORT **89**

2.23 | Survival analysis **89**

2.24 | Statistical analysis **89**

Chapter 3 | The myeloid landscape of TNC-regulated breast cancer at single-cell resolution ***92***

3.1 | Results **93**
3.1.1 | Introduction 93
3.1.2 | Tumour-infiltrating myeloid cell isolation for single-cell RNAseq 95
3.1.3 | Quality control 98
3.1.4 | Feature selection and dimensionality reduction 100
3.1.5 | Clustering and visualisation 103
3.1.6 | Characterisation of cluster phenotypes 108
3.1.7 | TNC-dependent effects 119
3.1.8 | Trajectory inference and pseudotemporal analysis of TAMs 123
3.1.9 | Summary of findings and hypotheses 129

3.2 | Technical Discussion **131**

3.3 | Biological Discussion **136**
3.3.1 | Introduction 136
3.3.2 | Heterogeneity and ontogeny of myeloid subpopulations 137
3.3.3 | Pathological relevance of myeloid subsets 140
3.3.4 | Co-existence of activatory and inhibitory phenotypes 142
3.3.5 | Tumour-specific phenotypic expansion 145
3.3.6 | Ribosome-mediated delegation of TAM effector functions 146
3.3.7 | TNC-dependent effects 148
3.3.8 | Summary & hypothetical pathogenic mechanism 150

Chapter 4 | Validation and exploration of macrophage phenotypes in vivo ***151***

4.1 | Results **152**

4.1.1 | Introduction 152
4.1.2 | Validation of the NT193 mammary tumour engraftment model 153
4.1.3 | Validation of functionally-delegated macrophage effector states 157
4.1.4 | Isolation of Mac3 and Mac2 by FACS-sorting 165
4.1.5 | Temporal dynamics of functionally-delegated macrophage states 172
4.1.6 | Temporal dynamics of monocyte-to-macrophage transition 176
4.1.7 | Spatial distribution of functionally delegated macrophage states 179
4.1.8 | *In vivo* alteration of functional-delegation by KDM6B: a Mac3 regulator 181

4.2 | Technical Discussion 185

4.3 | Biological Discussion 191
4.3.1 | Introduction 191
4.3.2 | Spatial insights into the pathological relevance of TAM effector functions 192
4.3.3 | TAM ontogeny, maturity, and the role of TNC 194
4.3.4 | Conserved & dichotomous TAM phenotypes 197
4.3.5 | KDM6B as a selective regulator of **Mac3** effector function 198
4.3.6 | Summary 199

Chapter 5 | Conservation and translational relevance of macrophage gene-signatures in human breast cancer 200

5.1 | Results 201
5.1.1 | Introduction 201
5.1.2 | Conversion of mouse genes to human homologs 202
5.1.3 | Conservation of mouse myeloid subsets in human single-cell RNAseq data 205
5.1.4 | Quality & refinement of gene signatures in human RNAseq data (TCGA) 214
5.1.5 | Gene signature enrichment & hierarchical clustering of TCGA data 219
5.1.6 | Survival analysis of patient tumour ecotypes in TCGA data 222
5.1.7 | Quality of myeloid gene signatures in human bulk array data (METABRIC) 226
5.1.8 | Gene signature enrichment & hierarchical clustering of METABRIC data 228
5.1.9 | Survival analysis of patient tumour ecotypes in METABRIC data 233

5.2 | Technical Discussion 236
5.2.1 | Introduction 236
5.2.2 | Conversion of mouse genes to human homologs 236
5.2.3 | Conservation of mouse myeloid subsets in human single-cell RNAseq data 238
5.2.4 | Gene signature approaches in bulk transcriptomic datasets 240
5.2.5 | Analysis of TNC in bulk transcriptomic datasets 242
5.2.6 | Interpretation of survival data 243

5.3 | Biological Discussion 245
5.3.1 | Introduction 245
5.3.2 | Conservation of mouse myeloid subsets in human single-cell RNAseq data 245
5.3.3 | Phenotypic expansion of myeloid subsets as a hallmark of pathogenesis 247
5.3.4 | Assessment of Mac3 regulators: *KDM6B* and *PIM1* 248
5.3.5 | The prognostic impact of myeloid ecotypes 249
5.3.6 | Elucidating a role for TNC 252

Chapter 6 | General Discussion 254

6.1 | Introduction 255

6.2 | The myeloid landscape in breast cancer 256

6.3 | The role of TNC 262

6.4 | Prognostic and therapeutic implications 265

6.5 | Future directions 267

Appendix 269

Bibliography 283

Statement of Originality

I declare that all data, words, figures, and results included herein are my own work and have not been submitted for another award, degree, or diploma at any other university or learning institution. I certify that this thesis is the result of my own work and all collaborations have been acknowledged at the appropriate location within the text.

This project was completed at the Kennedy Institute of Rheumatology, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford.

Abstract

Tumour-associated macrophages (TAMs) are key immune-suppressive constituents of the breast tumour microenvironment (TME) and correlate with poor prognosis. Tenascin-C (TNC) is an immunomodulatory extracellular matrix (ECM) protein which is upregulated in breast tumours, and is associated with increased metastasis and worsened prognosis. Our previous work showed that perturbation of the breast TME by tumour-derived TNC skews macrophages towards an immune-suppressive phenotype featuring downregulated antigen-presentation. Using TNC as an exemplar perturbant, this thesis sought to understand how macrophage effector functions such as antigen presentation become enhanced or suppressed within the TME. To achieve this, single-cell RNA sequencing (scRNAseq) was performed on infiltrating myeloid cells derived from syngeneic, orthotopically-engrafted, murine breast tumours that were engineered to express high or low levels of TNC. scRNAseq revealed marked heterogeneity in the intra-tumoural myeloid compartment with 7 distinct subsets observed, 5 of which were macrophages (**Mac1**-**5**). Notably, macrophages delegated key effector functions including antigen presentation (**Mac2**) and chemokine production (**Mac3**) to discrete subsets that were phenotypically, spatially and temporally distinct at the protein level by flow cytometry and confocal immunofluorescence imaging. The role of TNC was predominantly exerted at the level of subset abundance (rather than gene transcription), with 4 out of 7 myeloid subsets exhibiting TNC-dependent changes in abundance, including two macrophage subsets (**Mac1** & **Mac2**). Over time, TNC modestly weakened the segregation of **Mac2** and **Mac3** phenotypes, suggesting it may drive dysfunctional delegation of effector functions. Pseudotemporal analysis of scRNAseq data implicated the epigenetic regulator, *Kdm6b*, in governing the acquisition of **Mac3** effector function, and pharmacological inhibition of KDM6B *in vivo* shifted the proportion of **Mac3** via TAM reprogramming. Finally, mouse-derived myeloid subsets were well-conserved in the most comprehensive scRNAseq atlas of human breast cancer, and gene signature analyses in bulk clinical datasets demonstrated significant differences in prognosis for patients with distinct myeloid 'ecotype' profiles. Together, these

data demonstrate functionally divergent roles for TAM subsets in breast tumour progression and metastasis, while providing mechanistic insights into the impact of ECM-derived cues such as TNC on shaping the myeloid compartment, with potential therapeutic and prognostic implications for breast cancer patients.

Abbreviations

All abbreviations will be used in full form in the first instance, and in the abbreviated form thereafter, with the exception of gene names which will be used in abbreviated form on all occasions. Gene names that feature in the main text (not in figures) will be included in this abbreviations list. Where a gene name has a common prefix or suffix, for instance, CD4 and CD8, only the common part ('CD') will be listed as an abbreviation.

ACK	Ammonium Chloride Potassium
Actb	Beta Actin
AD	Additional Domain
ADCC	Antibody-Dependent Cellular Cytotoxicity
Adgre1	Adhesion G Protein-Coupled Receptor E1 (F4/80)
ADP	Adenosine Diphosphate
AF	AlexaFluor
AID	Activation-Induced Cytidine Deaminase
Aif1	Allograft Inflammatory Factor 1
ALL	Acute Lymphoblastic Leukaemia
ANOVA	Analysis of Variance
APC	Antigen-Presenting Cell
APC	Allophycocyanin
APOE	Apolipoprotein E
Arg1	Arginase 1
ASPA	Animals in Scientific Procedures Act
ATP	Adenosine Triphosphate
AUC	Area Under the Curve
B2m	Beta-2-Microglobulin
B7-H3	CD276
BCL	Base Call File
BH	Benjamini-Hochberg
BISCUIT	Bayesian Inference for Single-cell Clustering and Imputing Algorithm
BMDM	Bone Marrow-Derived Macrophage
Bmp2k	Bone Morphogenic Protein 2 Inducible Kinase
bp	Base Pair

BRCA	Breast Cancer Susceptibility Protein
BRCA	Breast Invasive Carcinoma Cohort (TCGA)
BSA	Bovine Serum Albumin
BV	Brilliant Violet
C1q	Complement C1q
CAF	Cancer-Associated Fibroblast
CAR	Chimeric Antigen Receptor
CAR-T	Chimeric Antigen Receptor T Cell
CCA	Canonical Correlation Analysis
CCD	Charge-Coupled Device
CCL	C-C Motif (Chemokine) Ligand
CCR	C-C Motif (Chemokine) Receptor
CD	Cluster of Differentiation
cDC	Conventional Dendritic Cell
CDH1	Cadherin-1 (E-Cadherin)
cDNA	Complementary DNA
CGAT	Computational Genomics Analysis and Training
Chil3	Chitinase-like 3
CHIT1	Chitinase 1
CIBERSORT	Cell-type Identification by Estimating Relative Subsets of RNA Transcripts
CITE-seq	Cellular Indexing of Transcriptomes and Epitopes by Sequencing
CLEC9A	C-Type Lectin Domain Containing 9A
CO_2	Carbon Dioxide
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSF1	Macrophage Colony Stimulating Factor 1 (see M-CSF)
CSF1R	Macrophage Colony Stimulating Factor 1 Receptor
CSF3R	Granulocyte Colony Stimulating Factor Receptor
CTC	Circulating Tumour Cells
CTL	Cytotoxic T Lymphocyte
CTLA-4	Cytotoxic T Lymphocyte Antigen 4
Cts	Cathepsin
CX3CR	CX3C Motif (Chemokine) Receptor
CXCL	C-X-C Motif (Chemokine) Ligand
CXCR	C-X-C Motif (Chemokine) Receptor
Cy	Cyanine
Cyba	Cytochrome B-245 Alpha Chain

CyTOF	Cytometry by Time of Flight
DAMP	Damage-Associated Molecular Pattern
DAPI	4'6-diamidino-2-phenylindole
DC	Dendritic Cell
DC	Diffusion Map Component
DEG	Differentially Expressed Gene
DMEM	Dulbecco's Modified Eagle Medium
dMMR	Mismatch-Repair-Deficient
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DNAse I	Deoxyribonuclease 1
dsDNA	Double-stranded DNA
DTR	Diptheria Toxin Receptor
ECM	Extracellular Matrix
EDTA	Ethylenediaminetetraacetic Acid
Eef	Eurkaryotic Elongation Factor
EGA	European Genome-Phenome Archive
EGF	Epidermal Growth Factor
EGR1	Early Growth Response 1
ELISA	Enzyme-Linked Immunosorbent Assay
EMT	Epithelial-to-Mesenchymal Transition
Epi	Epithelial Cell
ER	Estrogen Receptor (*ESR1*)
ERBB2	Erb-B2 Receptor Tyrosine Kinase 2
ESR1	Estrogen Receptor 1 (ER)
EZH2	Enhancer of Zeste Homolog 2
F4/80	*ADGRE1*
FABP5	Fatty Acid Binding Protein 5
FACS	Fluorescence-Activated Cell Sorting
FBG	Fibrinogen-like Globe
FBS	Fetal Bovine Serum
Fc	(Antibody) Fragment Crystallisable Region
FCγR	Fc Gamma Receptor
FD	Functional Delegation
FFPE	Formalin-Fixed Paraffin-Embedded
FITC	Fluorescein isothiocyanate

Flt3	FMS-like Tyrosine Kinase 3
FMO	Fluorescence Minus One
FNIII	Fibronectin Type III
FVB Mice	Friend Leukemia Virus B Mice
GAM	General Additive Model
GATA3	GATA Binding Protein 3
GEM	Gel Bead In Emulsion
GM-CSF	Granulocyte-Macrophage Colony Stimulating Factor
GO	Gene Ontology
GO.BP	GO Biological Processes
GO.BP	GO Molecular Function
GO.CC	GO Cellular Component
GSE	Gene Set Enrichment
GSEA	Gene-Set Enrichment Analysis
GSVA	Gene Set Variation Analysis
H3K27	Histone 3 Lysine 27
HDAC	Histone Deacetylase
HER2	Human Epidermal Growth Factor Receptor 2
HGNC	HUGO Gene Nomenclature Committee
Hist1h2ap	H2A Clustered Histone 11
HLA	Human Leukocyte Antigen
Hprt	Hypoxanthine-Guanine Phosphoribosyltransferase
HSC	Haematopoietic Stem Cells
HUGO	Human Genome Organisation
iCAF	Inflammatory-like CAF
ICB	Immune Checkpoint Blockade
IDO	Indoleamine 2,3-dioxygenase 1
Ifi	Interferon-Inducible Factor
IFNγ	Interferon Gamma
IgG	Immunoglobulin G
IHC	Immunohistochemistry
IL	Interleukin
iNOS	Inducible Nitric Oxide Synthase
IRES	Internal Ribosome Entry Site
IRF	Interferon Regulatory Factor
ITIM	Immunoreceptor Tyrosine-based Inhibitory Motif

KDM6B	Lysine Demethylase 6B
KEGG	Kyoto Encyclopaedia of Genes and Genomes
Ki-67	Marker of Proliferation Ki-67
KIR	Killer Cell Immunoglobulin-like Receptor
Klf13	Kruppel Like Factor 13
KMT2	Lysie Methyltransferase 2
KNN	k-Nearest Neighbour
L/D	Live/Dead Viability Dye
LAG3	Lymphocyte-Activation Gene 3
LAM	Lipid-Associated Macrophage
LAMP3	Lysosome-Associated Membrane Glycoprotein 3
Lgals9	Galectin 9
Lilrb4a	Leukocyte Immunoglobulin-like Receptor, Subfamily B, Member 4A
LM22	Leukocyte Signature Matrix 22
LN	Lymph Node
LPS	Lipopolysaccharide
Lrg1	Leucine-Rich Alpha-2-Glycoprotein
Ltb	Lymphotoxin Beta
Ly6	Lymphocyte Antigen 6
M-CSF	Macrophage Colony Stimulating Factor (see *CSF1*)
Mafb	V-Maf Musculoaponeurotic Fibrosarcoma Oncogene Homolog B
MAP3K1	Mitogen-Activated Protein Kinase Kinase Kinase 1
MARCO	Macrophage Receptor With Collagenous Structure
Mb	Megabase
MD-2	Myeloid Differentiation Factor 2
MDSC	Myeloid-Derived Suppressor Cell
METABRIC	Molecular Taxonomy of Breast Cancer International Consortium
MFI	Mean Fluorescence Intensity
MGI	Mouse Genome Informatics
MHC	Major Histocompatibility Complex
MIC	MHC Class I Polypeptide–Related Sequence
Mif	Macrophag Migration Inhibitory Factor
MMP	Matrix Metalloproteinase
MMTV	Mouse Mammary Tumour Virus
MNN	Mutual Nearest Neighbours
Mrc1	Mannose Receptor C-Type 1 (CD206)

mRNA	Messenger RNA
MS4A6A	Membrane-Spanning 4-Domains Subfamily A Member 6A
MSC	Mesenchymal Stem Cell
MSI-H	Microsatellite Instability - High
MSigDB	Molecular Signatures Database
MST	Minimum Spanning Tree
myCAF	Myofibroblast-like CAF
NADPH	Reduced Nicotinamide Adenine DinucleotidePhosphate
Near-IR	Near Infrared
NFkB	Nuclear Factor Kappa B Subunit
NK Cell	Natural Killer Cel
NKG2D	Natural Killer Group 2D
NLRP3	Nod-Like Receptor Family Pyrin Domain-containing 3
NO	Nitric Oxide
Notch2	Neurogenic Locus Notch Homolog Protein 2
NSCLC	Non-Small Cell Lung Cancer
OCT	Optimum Cutting Temperature Compound
OR	Odds Ratio
P2X7	Purinergic Receptor P2X 7
p53	Tumour Protein p53
PAGA	Partition-based Graph Abstraction
PAM50	Predictor Analysis of Microarray 50
PAMP	Pathogen-Associated Molecular Pattern
PARP	Poly (ADP-Ribose) Polymerase
PBS	Phosphate Buffered Saline
PC	Principal Component
PCA	Principal Component Analysis
Pcna	Proliferating Cell Nuclear Antigen
pCR	Pathological Complete Response
PCR	Polymerase Chain Reaction
PD-1	Programmed Cell Death 1
PD-L1	Programmed Cell Death Ligand 1
pDC	Plasmacytoid Dendritic cell
PDGF	Platelet-Derived Growth Factor
PE	Phycoerythrin
PerCP	Peridinin-Chlorophyll-Protein

PFA	Paraformaldehyde
PI3K	Phosphatidylinositol 4,5-Bisphosphate 3-Kinase
PIK3CA	PI3K Catalytic Subunit Alpha
Pim1	Pim-1 Proto-Oncogene, Serine/Threonine Kinase
PLAUR	Plasminogen Activator Urokinase Receptor
Pld3	Phospholipase D3
Plek	Pleckstrin
PMN	Polymorphonuclear Cell
Poly-dT	Poly-deoxythymine
PR	Progesterone Receptor
PRR	Pattern Recognition Receptors
PTEN	Phosphatase and tensin homolog
Ptprc	Protein Tyrosine Phosphatase Receptor Type C (CD45)
PVL	Perivascular-Like
PyMT	Polyoma Middle T Antigen
QC	Quality Control
Rag2	V(D)J Recombination-activating Protein 2
RNA	Ribonucleic Acid
RNS	Reactive Nitrogen Species
ROC	Receiver Operating Characteristic
ROS	Reactive Oxygen Species
RP	Ribosomal Protein
Rpl	Ribosomal Protein Large Subunit
RPMI	Roswell Park Memorial Institute Medium
Rps	Ribosomal Protein Small Subunit
RSEM	RNAseq by Expectation-Maximization Algorithm
RT	Room Temperature
RT-qPCR	Real-Time Quantitative Polymerase Chain Reaction
S100A	S100 Calcium Binding Protein A
scFv	Single-Chain Variable Fragment
scRNAseq	Single-cell RNA Sequencing
SD	Standard Deviation
shRNA	Small Hairpin RNA
SIGLEC1	Sialic Acid Binding Ig Like Lectin 1
SIRPα	Signal-Regulatory Protein Alpha
SLM	Smart Local Moving Algorithm

SNN	Shared Nearest Neighbour
Socs3	Suppressor Of Cytokine Signaling 3
SOX4	SRY-Box Transcription Factor 4
SPF	Specific Pathogen-free
SPP1	Secreted Phosphoprotein 1 (Osteopontin)
SSC	Side Scatter
STAT	Signal Transducer And Activator Of Transcription
TAM	Tumour-Associated Macrophage
TCGA	The Cancer Genome Atlas
TCR	T Cell Receptor
TGFBI	TGF Beta Induced
TGFβ	Transforming Growth Factor Beta
Tgm2	Transglutaminase 2
Th17	T Helper 17
Thbs1	Thrombospondin 1
TIE2	Tyrosine Kinase with Ig and EGF Homology Domains 2
TIGIT	T Cell immunoreceptor with Ig and ITIM Domains
TIL	Tumour-Infiltrating Lymphocyte
TIM3	T Cell Immunoglobulin And Mucin-Domain Containing-3
TLR	Toll-Like Receptor
TMB	Tumour Mutational Burden
TME	Tumour Microenvironment
Tmsb10	Thymosin Beta 10
TNBC	Triple-Negative Breast Cancer
TNC	Tenascin-C
Tnfsf9	TNF Superfamily Member 9 (41-BBL)
TNFα	Tumour Necrosis Factor Alpha
TPM	Transcripts Per Million
Treg	Regulatory T Cell
TREM2	Triggering Receptor Expressed on Myeloid Cells 2
tSNE	t-Distributed Stochastic Neighbour Embedding
Uba52	Ubiquitin A-52 Residue Ribosomal Protein Fusion
UMAP	Uniform Manifold Approximation and Projection
UMI	Unique Molecular Identifier
VEGF	Vascular Endothelial Growth Factor
Vim	Vimentin

VISTA	V-Set Immunoregulatory Receptor (*VSIR*)
Wfdc21	WAP Four-Disulfide Core Domain 21
XCL	X-C Motif (Chemokine) Ligand
XCR	X-C Motif (Chemokine) Receptor

Acknowledgments

It is truly a privilege to have been given the opportunity to undertake this research, to acquire new skills, to tackle meaningful problems, and to learn from world-class leaders in the field. I am therefore hugely grateful to the Clarendon Fund and NDORMS for providing me with the financial support to seize this opportunity.

Like the macrophages described in these pages, we are deeply influenced by our environment. As such, I feel supremely fortunate to have been surrounded by so many wonderfully talented and kind people on the long journey to, and through, Oxford. To thank everyone who helped me along this journey would no doubt exceed my word limit; however:

Firstly, I would like to thank my primary supervisor, Professor Kim Midwood, for her endless patience, compassion and support, especially during the most challenging periods of this DPhil. I am glad to say that I have learned a tremendous amount from her over the years, as a scientist and as a person. I would also like to thank my co-supervisor, Professor Irina Udalova, for helping guide the project, and for always making time to discuss any questions or issues I encountered along the way. In a scientific capacity, my gratitude extends to a great number of colleagues and scientists, for their advice, guidance and collaborative partnerships that enabled this work, including: Julie Mazet, Claire Deligne, Martha Gschwandter, Bill Chi, Maria Gomez-Vazquez, Moustafa Attar, Lion Uhl, Francesca Buffa, Gertraud Orend, Audrey Gerard, Stephen Sansom, Adrian Harris, Jon Webber, Members of the Kennedy Genomics Forum, Anja Schwenzer, Linda Mies, and the rest of the Midwood Group.

On a personal note, I would like to express my gratitude for the close and enduring friendships forged on Kingston Road, which were a huge source of support during this DPhil. Finally, I would like to thank my family – my parents, Jita and Mike, my brothers, Shamim and Payam, my whole extended family, and my partner, Talia, for being pillars of unwavering support, love and generosity throughout this time. This work would not have been possible without them.

For those loved and lost

Chapter 1 | Introduction

1.1 | Overview

This Chapter will be divided into the following subsections:

1.2 Cancer immunology at a glance

1.3 Breast cancer, immunity and immunotherapy

1.4 Macrophages and breast cancer: into the single-cell era

1.5 Extracellular matrix in breast cancer and inflammation

1.6 Tenascin-C in breast cancer

1.7 Thesis rationale and summary

1.2 | Cancer immunology at a glance

Cancer claimed the lives of ~10 million people in 2020 (1) and represents the second largest disease burden worldwide (2, 3). It develops through a process of somatic cell evolution (4-6), the driver mutations of which can precede the onset of disease by many years, if not decades (7). Yet uncontrolled cell division is not unique to humans nor mammals – it is a pervasive feature across Metazoans as well as certain plants, and is thought to represent a failure of multicellular systems to cooperate in suppressing somatic evolution (8-11). Thus, paradoxically, the origin of cancer is intimately linked with the foundations of most complex multicellular life.

While it is often conceptualised as a single phenomenon, cancer is better understood as a large collection of malignancies, with diverse presentations, that converge on shared phenotypic features – the hallmarks of cancer. Seminal work by Hanahan and Weinberg (12, 13) proposed eight distinct hallmarks: **(i)** sustaining proliferative signalling, **(ii)** evading growth suppressors, **(iii)** resisting cell death, **(iv)** enabling replicative immortality, **(v)** inducing angiogenesis, **(vi)** activating invasion and metastasis, **(vii)** deregulating cellular energetics, and **(viii)** avoiding immune destruction. The acquisition of these hallmarks is enabled by two further characteristics: **(i)** genomic instability leading to gene mutation, and **(ii)** tumour promoting inflammation. Importantly, genomic instability imparts cancer cells with a diverse and stochastic genetic repertoire, providing fertile ground for the Darwinian selection of mutations that confer a fitness advantage by enabling acquisition of further cancer hallmarks. With time, the iteration and parallelisation of this evolutionary process leads to profound clonal heterogeneity within an individual tumour, and is directly responsible for the resistance of tumours to various therapeutic interventions (14-16). Intra-tumour heterogeneity – beginning at the genetic level, but manifesting at all ‘omic, cellular and clinical levels – is therefore the fuel of cancer evolution and represents a unique therapeutic challenge in oncology.

While the aetiology of cancer is indisputably driven by genetic aberrations in somatic cells, tumour-promoting inflammation is a key facilitator of this process, and is known to contribute to all stages of tumourigenesis (17, 18). Indeed, rather than merely an outgrowth of malignant clones, cancer cells are now viewed more holistically as members of a dysfunctional tumour microenvironment (TME) – a diverse ecosystem of cellular, soluble and structural cues that dynamically coordinate to support tumour progression (19-21).

The composition of the TME varies greatly between different tumour types, but is generally composed of infiltrating immune cells (such as T cells, tumour-associated macrophages (TAMs), monocytes, Natural Killer (NK) cells, dendritic cells (DCs), or neutrophils) alongside stromal cells (such as endothelial cells, cancer-associated fibroblasts (CAFs), pericytes or adipocytes), embedded in a network of extracellular matrix (ECM). Each cell type may itself be comprised of numerous phenotypically distinct subsets which differ not only in their functionality, abundance, lifespan, and even antigen-specificity, but also in their migratory capacity, spatial localization and temporal dynamics (12, 14, 18, 19). The TME is therefore a source of tremendous functional complexity, and consequently, cancer cell phenotypes reflect the constraints imposed on them by their microenvironment, and vice versa. In addition to factors such as nutrient availability, acidity, and hypoxia, the TME dictates the inflammatory milieu of cytokines, chemokines and growth factors which can profoundly influence tumour pathogenesis (16, 19, 21, 22). Chronic exposure to these mediators contributes to the proliferation and survival of malignant cells, alongside angiogenesis, metastasis, response to therapy, and the subversion of adaptive immunity, thus reinforcing several cancer hallmarks (17). They also promote genetic instability in cancer cells either by: **(i)** directly generating reactive oxygen and nitrogen species (ROS/RNS) which induce DNA damage (23), or by **(ii)** secreting various cytokines including TNFα and IL-6 that can indirectly lead to inhibition of the tumour-suppressor gene, p53, or induce ectopic expression of AID (23) – a cytidine

deaminase which mediates somatic DNA hypermutation in B cells, and whose expression is otherwise tightly restricted to germinal centre B cells (24).

Despite its role in promoting cancer through inflammatory and tolerogenic mechanisms, the immune system is simultaneously capable of mounting potent anti-tumoural responses through various cytotoxic mechanisms, including via $CD8^+$ cytotoxic T lymphocytes (CTLs). To summarise, DCs phagocytose tumour cells to acquire tumour-specific (neo)antigens, undergo maturation, and migrate to the lymph node, where they cross-present these antigens via major histocompatibility complex (MHC) Class I to prime and activate naive $CD8^+$ T cells, generating CTLs that home to the tumour, encounter tumour-antigens, and induce cell death in cancer cells via perforin- and granzyme B- dependent mechanisms (25, 26). The cross-presentation of tumour antigens is also possible *in situ*, and has been described for both DCs and macrophages, another key professional antigen-presenting cell (APC) (27).

Indeed, the existence of such a cytotoxic anti-tumoural mechanism forms the basis of the 'cancer immunosurveillance' hypothesis, which posited that the immune system limits the emergence of tumours by eliminating nascently transformed cells (28) – and was later proven using immunodeficient *Rag2* knockout mice lacking T and B lymphocytes (29). Over time, the selective pressure imposed by CTLs sculpts the immunogenicity of tumours in a process of 'cancer immunoediting' (30, 31), leading to the selection of poorly immunogenic clones that diminish or escape CTL-mediated immunity, for instance, via the downregulation of tumour-associated antigens, MHC Class I, or antigen processing machinery (32).

However, these represent just a portion of the immune escape mechanisms and counter-mechanisms that co-evolve between cancer and immunes cells (33). Other examples of immune evasion strategies that tumours exploit include: **(i)** an immune-suppressive milieu of soluble mediators such as TGFβ, IL-10, IDO and Arginase (33), **(ii)** the manipulation of NK cell cytotoxicity by upregulating (decoy) ligands for the receptor, NKG2D (preventing NK cell

activation), or upregulating stress molecules such as MICA/MICB to engage KIRs (increasing NK cell inhibition) (34), **(iii)** the expansion of regulatory T cells, myeloid-derived suppressor cells (MDSCs) and pro-tumoural "M2-like" macrophages (18, 33), and **(iv)** the upregulation of immune checkpoint molecules such as CTLA-4, PD-1, LAG3, TIM3 and TIGIT that restrain T cell activation (35).

Immune checkpoints in particular have been the focus of intense investigation in recent years (36), and their therapeutic blockade has enjoyed remarkable success in the clinic, potently augmenting anti-tumour immunity, even in indications such as metastatic melanoma (37-39) and advanced non-small-cell lung cancer (NSCLC) (40, 41) that were once considered untreatable. Indeed, the success of immune checkpoint blockade (ICB) attests to the transformative potential of immunotherapy in treating cancer, and demonstrates the central importance of immune evasion in tumour pathogenesis. T cell-targeted immunomodulators, such as anti-PD-1/PD-L1 and anti-CTLA-4, are now routinely used as first or second line treatments for ~50 cancer types, either as a monotherapy, combined with one another, or combined with chemotherapy (36). The potential advantages of ICB over conventional therapies are numerous. First, CTLs can naturally recognise the diverse neoantigen landscape of tumour cells (42, 43), and can constantly adapt alongside the evolving tumour, limiting therapeutic resistance. Second, the immunogenic cell death of tumour cells promotes additional antigenicity via 'epitope spreading', and additional adjuvanticity via release of intracellular damage-associated molecules, together initiating a self-sustaining cycle of anti-tumour immunity (44, 45). Third, the positive correlation between ICB efficacy and tumour mutational burden (TMB) enables the stratification and treatment of patients with a TMB-high phenotype irrespective of their tumour type, substantially widening the patient population, as it did for anti-PD-1 therapy (46). Finally, the immune compartment is genetically stable and targeting it is therefore less likely to lead to evolved drug resistance due to underlying mutational heterogeneity, as occurs in tumour cells (14-16).

However, a number of issues remain. First, complete therapeutic response to ICB (with no detectable cancer present) are confined to only a small minority of patients (35, 36). Second, measuring TMB is time- and cost-intensive, and does not provide a robust prediction of response to ICB for all tumour types. The predictive value of TMB can also drop sharply in the presence of high neoantigen intra-tumour heterogeneity. Thus, accurately and efficiently predicting patients' response to ICB is challenging, and high quality predictive biomarkers are lacking (47). Third, inhibiting the physiological brake of immune activation via ICB often induces autoimmune-like responses that precipitate significant adverse events (36, 48) which can limit their usage. Fourth, tumours with low mutational burden, neoantigen load or T cell infiltration are unlikely to benefit from ICB therapy alone (36, 49). Fifth, acquired resistance to ICB is in fact still possible, and occurs in a reasonably large proportion of patients (approximately one third) (50).

In tandem with the rise of ICB, another promising approach has been adoptive cell transfer – a type of cell therapy. In particular, engineered T cell therapies known as chimeric antigen-receptor T cells (CAR-Ts) have enjoyed substantial clinical success in the treatment of haematological cancers (51-53). The chimeric receptor of CAR-T cells features an ectodomain composed of a single-chain variable fragment (scFv) responsible for the recognition of a specific tumour antigen, and is linked via a flexible hinge region to the transmembrane domain (typically derived from CD8α or CD28), which is in turn linked to an intracellular signalling domain (54, 55). The iteration of CAR-T technology over several generations has overcome numerous limitations, such as durability, persistence, and efficacy, which have been improved mostly through addition or modification of the intracellular signalling and costimulatory domain(s) (56). One of the attractions of CAR-T therapy is that it circumvents the issue of tumour cells downregulating antigen processing and presentation, as the scFv directly binds its target antigen on the tumour cell, bypassing the requirement for MHC (55). However, therein also lies its weakness: the specificity of CARs for only one antigen means they suffer from a counter-selective pressure which leads to 'antigen escape' and thus therapy resistance

(57). Thus, the application of CAR-T cells is intrinsically limited to cancers with a highly and uniformly expressed cancer-associated antigen, such as CD19 for B cell malignancies, and may yet still incur significant therapeutic resistance even in this case. For this reason, among others, the efficacy of CAR-T therapy in solid tumours has so far been poor (54, 56, 57).

Importantly, both ICB and CAR-T therapies share a major constraint which greatly limits their utility: they too often fail to sufficiently overcome the profound immune suppression imposed by the solid tumour microenvironment – in which the myeloid compartment and the ECM are key protagonists. In the following Sections, the relevance of immunity in breast cancer will be explored, alongside the role of myeloid cells and the ECM protein, tenascin-C (TNC), in contributing to immune-suppression in the breast TME.

1.3 | Breast cancer, immunity and immunotherapy

With ~2.3 million estimated new cases in 2020, breast cancer is the most commonly diagnosed cancer worldwide, and is also the leading cause of cancer-related mortality in women, accounting for ~685,000 deaths in the same year (1). It is a particularly heterogeneous disease comprised of several clinically distinct 'intrinsic' subtypes that bear substantial differences in their incidence, disease progression, prognosis, and response to therapy (58, 59). These subtypes are typically stratified by their expression of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and proliferation antigen Ki-67, leading to the following clinical assignments: Luminal A (ER^+ and/or PR^+, $HER2^-$, $Ki\text{-}67^-$), Luminal B (ER^+ and/or PR^+, $HER2^-$, $Ki\text{-}67^+$), HER2-enriched (ER^- PR^- $HER2^+$) and triple-negative (ER^- PR^- $HER2^-$) (59-61). These subtype-defining markers are typically assessed by immunohistochemical (IHC) staining and/or gene expression profiling using a 50-gene signature known as the Predictor Analysis of Microarray 50 (PAM50) (62). The PAM50-derived classification reveals that triple-negative breast cancer (TNBC) is in fact a subtype of a wider 'basal-like' phenotype, and has a strongly overlapping signature (63, 64). In fact, analysis of The Cancer Genome Atlas (TCGA) breast cancer dataset (65, 66) confirmed that that the great majority of basal-like tumours (78%) assigned by PAM50 were also triple-negative by IHC (67).

The incidence of subtypes varies by several factors such as age, but is generally highest for Luminal A (50-60%), followed by Luminal B (15-20%) and $HER2^+$ (15-20%), while estimation of basal-like is more variable (8-37%) as it additionally depends on tumour grade (68). It is worth noting, however, that there are modest differences in the methodology and classification of intrinsic subtypes between studies, meaning definitive prevalence information is challenging to obtain.

While a small proportion (5-10%) of breast cancers are hereditary, usually caused by highly penetrant mutations in the tumour suppressor genes, *BRCA1* or *BRCA2*, the vast majority of breast cancers are considered sporadic (69-72). The most common driver mutations across all breast cancers are observed in *PIK3CA* and *TP53*, each present in ~34% of samples, with other commonly mutated genes including *CDH1*, *GATA3*, *KMT2C*, *MAP3K1* and *PTEN*, in descending order of frequency (73). Mutation of *TP53* is especially frequent in TNBC (68%), and to a lesser extent HER2-enriched (39%) subtypes (74), indicating distinct disease mechanisms mediated by particular mutational profiles.

The prognosis of breast cancer varies markedly by subtype. The Luminal A subtype bears the most favourable prognosis with 4-year survival of ~93% while triple-negative disease bears the most unfavourable prognosis with only ~77% in the same period (75). In a large study which adjusted for confounding variables (such as age, tumour grade, disease stage, surgery, and adjuvant treatment), the hazard ratio of $HER2^+$ and triple-negative subtypes was 1.4 and 3.25, respectively, in comparison to Luminal A which had the most favourable prognosis (reference hazard ratio = 1.0) (76), indicating substantially poorer prognosis in these subtypes. The particularly poor prognosis of TNBC is partly due to its higher recurrence rate following surgery, which stands at 10-15% of cases per year, compared to 3-5% in hormone-receptor positive breast cancer (77, 78), though a lack of treatment options for TNBC has also contributed (79).

Ultimately, metastasis is responsible for the majority of cancer-associated mortality, with estimates ranging from 67-90% of all cancer deaths (80, 81). The process of metastasis is complex, involving the detachment of primary tumour cells, intravasation into systemic circulation, adherence to distal site vasculature, extravasation into the tissue, and finally colonisation – all while evading immune destruction, despite no longer being within the immune-suppressed TME (80). Breast tumours have a proclivity to metastasise to several specific sites – known as an organotropism – and the most commonly colonized tissues are

bone, liver, brain and lung, in descending order of frequency (82). Intrinsic subtypes of breast cancer also differ in their organotropisms, for instance triple-negative and HER2-enriched tumours both disproportionately metastasise to the lung and brain, whereas all subtypes are capable of bone metastasis (82). Interestingly, a meta-analysis of brain metastases and donor-matched primary breast tumours revealed a discordancy between the intrinsic subtype of primary versus metastatic tumours (83), indicating that some metastatic sites might exert subtype-dependent selective pressure. Interestingly, the mutational landscape of metastases varies substantially compared to primary tumours (74), with the frequency of mutated *TP53* (p53), *ESR1* (ER), and *KMT2D* being significantly higher in metastases, especially for ER^+ disease (74).

Understanding the molecular heterogeneity of breast cancer has enabled the development of targeted treatment strategies that have substantially improved patient outcomes, especially in $HER2^+$ and ER^+ disease (79), where anti-HER2 monoclonal antibody treatment (trastuzumab) (84) and estrogen blockers (e.g. tamoxifen) (85) have become first-line treatments. Other targeted treatments have also emerged, such as PI3Kα inhibition (e.g. alpelisib) for patients with a *PIK3CA* mutation (86), or poly (ADP-ribose) polymerase (PARP) inhibition (e.g. olaparib) for those with a *BRCA1* or *BRCA2* mutation (87).

As breast cancer has an intermediate level of mutational burden compared to other cancers where immunotherapies are widespread (88, 89), it has not received the same level of attention as TMB-high malignancies. Indeed, only ~5% of all breast cancer cases would fulfil a common definition of TMB-high, which refers to tumours bearing >10 mutations per megabase (Mb) (88, 90, 91). However, TMB varies significantly between intrinsic subtypes, with basal-like tumours carrying a substantially higher burden than ER^+ or $HER2^+$ subtypes (92, 93). Moreover, there is an elevated frequency of metastases bearing a TMB-high status (8-11%) compared with primary tumours (2-5%) (88, 94). Thus, there remains significant potential for immunotherapeutic interventions in distinct patient subpopulations. Indeed, the

proportion of tumour-infiltrating lymphocytes (TILs) is another feature contributing to the immunogenicity of breast cancer, and this similarly differs substantially between subtypes, with 20-30% of TNBC patients and 16-19% of $HER2^+$ patients exhibiting high TIL counts, compared to only 6-13% of hormone-receptor positive subtypes (95, 96). Indeed, studies collectively spanning thousands of patients have consistently demonstrated that increased infiltration of lymphocytes is associated with improved prognosis, especially in HER2-enriched and early stage TNBC populations (96-99). TIL levels in these same patient populations were also shown to be a predictor of response to chemotherapy (96, 99), thus reinforcing the importance of anti-tumour immune activation in positive clinical outcomes for these subtypes. Taken together, these studies suggest that TNBC and HER2 are the most immunogenic subtypes of breast cancer, and also carry the most unfavourable prognoses, indicating that they represent attractive candidates for immunotherapeutic interventions.

Beyond broad immune biomarkers such as lymphocyte infiltration, the granular composition of immune cells in the breast TME is highly complex, and reflects the interplay between components of innate immunity suppressing the microenvironment, and the concomitant suppression of adaptive immunity (100). As the predominant myeloid and lymphoid populations in the breast TME (101-103), TAMs and T cells are of particular interest for immuno-modulatory intervention. Crucially, both are capable of acquiring either highly anti-tumoural or pro-tumoural phenotypes, depending on their activation and polarization state (21). For instance, a high infiltration of regulatory T cells (Tregs) is correlated with HER2-overexpression, and this is associated with reduced progression-free and overall survival (104). In TNBC patients, a low infiltration of $CD8^+$ T cells was found to be negatively prognostic (105), which was reinforced by an analysis of 3837 patients, indicating that high infiltration of $CD8^+$ T cells was correlated with improved survival in TNBC, especially when there was high concomitant infiltration of $CD4^+$ memory T cells (106). Further work demonstrated that high expression of a gene signature representing a $CD8^+$ tissue-resident memory T cell subset was

associated with improved survival in early stage TNBC, providing greater prognostication than CD8 expression alone (107).

The idea that macrophages facilitate breast cancer progression has long been appreciated, since seminal genetic experiments in CSF1-deficient mice lacking macrophages confirmed their importance in breast malignancy (108, 109). Early clinical analyses also demonstrated a negatively prognostic role for breast tumours with high macrophage infiltration, in terms of relapse-free and overall survival, as well as a correlation with vascular grade – a marker of angiogenesis (110). These findings have since been reinforced by several other studies implicating macrophage infiltration with worsened prognosis in breast cancer (111-115), including a study of 1350 breast tumours which reported a macrophage-derived signature that was enriched in aggressive disease (characterized by higher tumor grade, decreased expression of ER and PR, higher mutation rate, and poorer prognosis) (114). Indeed, a meta-analysis of 16 studies totalling >4500 breast cancer patients demonstrated that a high density of TAMs was not only correlated with poor overall- and disease-free- survival, but was also associated with negative hormone-receptor status (ER^- PR^-), basal phenotype, tumour size, tumour grade and vascular invasion (116). Interestingly, TAMs localizing to the surrounding tumour stroma were associated with a poorer prognosis compared with TAMs infiltrating the tumour nest (116), highlighting the importance of the spatial organisation of TAMs in conferring their function and pathogenicity. Finally, analysis of primary breast tumours and matched metastases revealed that immune cell infiltration was significantly lower in metastases for all immune cell populations except immune-suppressive 'M2-like' macrophages (117), which were significantly enriched in metastases, alluding to a key role for TAMs in supporting metastasis.

As such, counteracting the pathogenic functions of TAMs is now recognised as a key emerging therapeutic approach in cancer, with potential to augment the efficacy of various other treatments, including T-cell modulatory therapies such as ICB (118-126). TAM-centric

targeting strategies in breast cancer can be broadly categorised as: **(i)** Inhibiting recruitment of monocytes, for example via anti-CCL2 treatment (127), **(ii)** Pan-macrophage depletion via anti-CSF1R treatment (128, 129) or liposome-encapsulated bisphosphonates such as clodronate (130-135), **(iii)** Phenotypic TAM reprogramming via PI3Kγ inhibitors (136, 137) or HDAC Class II inhibitors (138), **(iv)** Reactivating phagocytosis of tumour cells via the SIRPα-CD47 axis (139, 140), **(v)** Inhibiting angiogenesis via anti-TIE2 treatment (141, 142), or finally, **(vi)** Relieving immune-suppression via anti-MARCO treatment (143), IDO inhibition (144), or PD-L1 blockade (145-148). However, there are significant downsides to some of these approaches, such as the fact that pan-macrophage depletion prevents any of the potential benefits of macrophages in stimulating anti-tumour immunity (121), or that treatment-induced compensatory pathways can counter-productively worsen disease, as occurs upon cessation of anti-CCL2 therapy, which accelerates breast cancer metastasis (149).

The majority of these TAM-targeted agents are at early stages of clinical development for breast cancer, or have not yet reported efficacy data from clinical trials. Indeed, the only immunotherapies to so far receive regulatory FDA approval for breast cancer are those modulating the PD-1/PD-L1 immune checkpoint axis, and the targets of this pathway are predominantly PD-1$^+$ T cells and PD-L1$^+$ cancer cells, though PD-L1$^+$ TAMs also play a role. Remarkably, anti-PD-1 therapy (pembrolizumab) was approved in a cancer-agnostic manner for any unresectable or metastatic mismatch-repair-deficient (dMMR) or microsatellite-instability-high (MSI-H) solid tumours, irrespective of the primary tumour site (46, 150, 151). However, as the prevalence of MSI-H breast cancer is only <1% of all cases, while expression of PD-L1 is much higher at ~12% (152), PD-L1 blockade was expected to be more successful in widening the total treatable patient population. Though anti-PD-L1 therapy with atezolizumab did receive regulatory approval for unresectable locally advanced or metastatic PD-L1$^+$ TNBC in combination with the chemotherapy, nab-paclitaxel (145-147), this was subsequently withdrawn after it failed to meet its primary endpoint of superior progression-

free survival compared to paclitaxel in a later trial (148). Fortunately, anti-PD-1 (pembrolizumab) achieved success and gained regulatory approval in combination with various chemotherapy regimens for previously untreated locally recurrent inoperable or metastatic TNBC (153), and also for high-risk early-stage PD-L1$^+$ TNBC (154).

These clinical successes have widened the population of TNBC patients eligible for immunotherapeutic treatment, however, the trial of early-stage TNBC showed only a modest (13.6%) improvement in mean pathological complete response (pCR) between pembrolizumab-chemotherapy (64.8%) and placebo-chemotherapy (51.2%) treated patients (154). Together, these data indicate that TNBC – the most immunogenic breast cancer subtype – is amenable to immune checkpoint blockade with anti-PD-1 but not anti-PD-L1 therapy, suggesting that relieving the immune-suppressive effects of PD-L1$^+$ TAMs or PD-L1$^+$ tumour cells is not solely sufficient to reactivate anti-tumour immunity in TNBC patients. Instead, these data argue for direct T cell modulation (such as PD-1 blockade) as an essential cornerstone for TNBC immunotherapy, and hint that TAM modulatory approaches are likely to be effective only in combination with ICB, where the potential for synergy is high. Thus, while these clinical trials represent a major advance in the field, improving the immunogenicity and immune infiltration of the breast TME – for TNBC as well as other subtypes such as HER2$^+$ disease – will be critical for enhancing clinical efficacy and achieving the full potential of immunotherapy in breast cancer. Combination therapies which alleviate the immune-suppression and pro-tumoural activities imposed by the microenvironment, particularly from TAMs, are likely to help increase the sensitivity of breast tumours to PD-1 blockade or other immunotherapies. These TAM-targeted approaches therefore hold promise, but require deeper investigation.

The following subsection, **Section 1.4**, will explore the nature of TAMs in the breast TME, focussing on the key insights gleaned from recent advances in single-cell technologies, which

offer unprecedented insight into the function and phenotypes of these heterogeneous cell populations.

1.4 | Macrophages and breast cancer: into the single-cell era

Infiltrating myeloid cells of the innate immune system are central protagonists in the microenvironment of various cancers, where they contribute to fundamental pathological processes including metastasis, neo-angiogenesis, cancer cell proliferation, cancer stem cell survival, resistance to therapy, and the impairment of protective adaptive immunity (120-122, 155-158). As the predominant myeloid subpopulation in breast tumours, coupled to their key role in promoting each of these pro-tumoural functions to the detriment of patient prognosis, macrophages are of major interest as therapeutic targets in breast cancer, alongside a range of other solid tumours (120-122, 156). This section will explore the fundamental biology of macrophages in homeostasis and in breast cancer, while reviewing the novel insights gleaned in this field from the new era of high-dimensional single-cell technologies.

It is now widely appreciated that tumours co-opt macrophages for their own benefit. Seminal early work by Lin and colleagues showed that genetic ablation of macrophages using a homozygous null allele of macrophage-colony stimulating factor (*Csf1*) led to reduced tumour progression and metastasis in the Mouse Mammary Tumour Virus (MMTV) Polyoma Middle T antigen (PyMT) model of breast cancer (108), alongside a 50% reduction in vascular density (109). While macrophages are capable of anti-tumoural activity, this work constituted the first formal proof that the dominant phenotype of TAMs is pro-tumoural – engaging in a symbiotic relationship with cancer cells that supports disease progression (122, 156). Indeed, macrophages are critical co-ordinators of cancer pathogenesis, as they not only directly promote several of the necessary hallmarks of cancer, but they also facilitate the acquisition of further hallmarks through the induction of genetic instability (via ROS/RNS) and tumour-promoting inflammation (via recruitment of further immune cells) (12, 21, 120-122, 156, 159).

At a fundamental level, macrophages are professional phagocytes that act as immune sentinels – continuously patrolling, monitoring and responding to perturbations in their local tissue environment with exquisite sensitivity. They are present throughout all vertebrate tissues and sit at the nexus of innate and adaptive immunity – orchestrating the immune response in both infectious and sterile inflammatory environments via a multitude of effector functions. Owing to their ability to detect and phagocytose pathogens, apoptotic cells, cellular debris, and microbial products, macrophages (and other professional APCs such as DCs) play a key role in processing and (cross-) presenting peptide antigens on MHC molecules to inform adaptive immunity. Alongside phagocytosis and antigen presentation, macrophages can also participate in antibody-dependent cellular cytotoxicity (ADCC), scavenging of exogenous and endogenous ligands, and secretion of a diverse range of soluble factors that co-ordinate inflammatory and wound-healing processes. Beyond their immunological role, they also fulfil a range of trophic functions that are crucial for tissue homeostasis, development, remodelling, homeostasis, and the co-ordination of metabolic processes (160-164).

Macrophages achieve this functional versatility with an equally diverse spectrum of sensor and effector molecules (165). Pattern recognition receptors (PRRs) have the ability to detect exogenous pathogen-associated molecular patterns (PAMPs) as well as endogenous damage-associated molecular patterns (DAMPs) (166-168). Examples of PAMPs include lipoteichoic acid, double-stranded RNA, lipopolysaccharide (LPS), flagellin, and unmethylated CpG motifs, which are principally recognised by TLR2/TLR6, TLR3, TLR4, TLR5 and TLR9, respectively (167). Alongside many others, examples of DAMPs include uric acid, extracellular adenosine triphosphate (ATP), nuclear DNA and ECM-derived tenascin-C (TNC), which are recognised by NLRP3, P2X7, TLR9 and TLR4, respectively (166-168). A subtype of PRRs known as scavenger receptors, such as CD204, CD206, MARCO, CD163, and CD68 are upregulated in macrophages, where they bind and internalise a broad variety of ligands including oxidised or acetylated (high or low density) lipoproteins, phospholipids, polysaccharides, or cytoskeletal molecules (169, 170). In terms of effector molecules,

macrophages also express various chemokines (CCLs, CXCLs) cytokines (TGFβ, IL-1α, IL1β, IL-6, IL-8, IL-18, IL-23, TNFα, IFNγ), growth factors (VEGF, EGF, PDGF), co-stimulatory molecules for antigen-specific T/B-cell activation (CD80, CD86), and Fcγ receptors (FCGRs) which enable ADCC (160, 163-165, 171, 172).

Indeed, macrophages are typified by their phenotypic plasticity, occupying distinctive functional niches that are strongly defined by, and specialised for, their surrounding tissue environment – as evidenced by microglia, osteoclasts, Kupffer cells and Langerhans cells residing in the brain, bone, liver and skin, respectively (161, 162, 173). Remarkably, the physical transplantation of murine tissue-resident macrophages to a different tissue environment was shown to be sufficient for their full phenotypic reprogramming – acquiring the tissue-specific characteristics of macrophages in the recipient tissue (174), demonstrating their profound plasticity. In fact, the inherent plasticity of macrophages in detecting and adapting to external stimuli is thought to render them particularly vulnerable to hijacking and manipulation by tumours (156, 163, 175, 176).

Phenotypically, macrophages are often described according to their functional polarisation states, known as the M1/M2 paradigm, in which M1 and M2 represent the two extremes of macrophage activation. Classically activated pro-inflammatory (anti-tumoural) “M1” macrophages produce IL-12, IL-6, ROS and TNFα upon stimulation with LPS or IFNγ, to promote inflammation and host defence. By contrast, alternatively activated immune-suppressive (pro-tumoural) “M2” macrophages express IL-10 and TGFβ upon stimulation with IL-4 and IL-13, which mediates tissue-repair and angiogenic functions (177-180). However, this classification is greatly oversimplified – especially for macrophages in the sterile TME – and thus a more comprehensive framework that adequately explains the full phenotypic spectrum of macrophage plasticity is required (181, 182). Indeed, it is widely known that macrophage phenotypes are altered in the TME (176).

It is important to note the distinct ontogenies of macrophages, as these can also be a key determinant of phenotypic heterogeneity (121). Macrophages have three different origins. The most recognised source of macrophages is via the M-CSF or GM-CSF-dependent differentiation of blood-borne monocytes, derived from haematopoietic stem and progenitor cells in the bone marrow (120, 183, 184). Contrary to long-held belief, the majority of adult tissue macrophages are not derived from circulating monocytes (185), but instead originate during embryonic development, either in an initial myeloid-restricted haematogenic wave in the yolk sac (186), or in a subsequent wave derived from the foetal liver (187). These embryonically derived macrophages persist throughout life and are capable of self-renewal (185). The composition of embryonically-derived macrophages versus adult monocyte-derived macrophages differs depending on the tissue (121, 173). However, in the breast, an elegant study of normal mammary tissue versus the MMTV-PyMT model of breast cancer demonstrated that $Ly6C^{+}$ $CCR2^{+}$ monocytes seed the population of normal mammary tissue macrophages as well as the TAM population, and that TAMs had a higher capacity for subsequent proliferation (188, 189). Thus, the mammary gland is one of the few tissues where the replenishment of tissue macrophages is solely dependent on blood-borne monocytes, as opposed to (at least partial) local proliferation of tissue-resident macrophages seeded during foetal development, though this has only been established in mice and not yet in humans.

The complex phenotype of TAMs is therefore a reflection not just of their plasticity and exposure to their local environment, but also of their ontogeny (190). As previously described in **Section 1.3**, elevated TAM infiltration has been consistently associated with poor prognosis in breast cancer across a multitude of studies collectively spanning thousands of patients. While genome-wide profiling has transformed our understanding of cancer and immune biology in general, much of this understanding has previously been gleaned from expression profiling of ‘bulk’ populations of disparate cell types. The cell is the fundamental unit of life, yet the signal-averaging inherent to such bulk approaches obscures important biological insight

by failing to capture the true spectrum, diversity and distribution of individual cellular phenotypes (191). Early low-dimensional single-cell approaches such as flow cytometry, immunohistochemistry, fluorescent *in situ* hybridisation and immunofluorescence imaging provided an opportunity to improve the granularity of bulk phenotypic signatures – however, such methods have been highly constrained by the number of readouts that can be quantified simultaneously. In mouse models of breast cancer, bulk transcriptional profiling of flow-cytometrically sorted populations of TAMs confirmed that certain 'M2-like' genes such as *Arg1* (Arginase), and *Mrc1* (CD206) were enriched, alongside pro-angiogenic molecules such as *Vegfa* (VEGF) (192, 193). Mechanistically, these genes confer their anti-tumoural function in distinct ways. For instance, arginase is a prototypical 'M2-like' gene because it regulates the local availability of the amino acid, L-arginine, which is essential for the generation of the cytotoxic free radical, nitric oxide (NO) – a reaction catalysed by the 'M1-like' gene, inducible Nitric Oxide Synthase (iNOS/*Nos2*) (121, 180).

However, while a variety of TAM phenotypes have been described in breast cancer, which can be broadly categorised as pro-angiogenic, pro-metastatic or pro-tolerogenic, these phenotypes are typically defined by singular positivity of a limited number of effector molecules such as VEGF, TIE2, IL-10, IL-4, CD206, matrix metalloproteinases (MMPs) or chemokines (194). Yet, cellular identity is not defined by expression of a singular gene product, but by complex interacting molecular networks – operating at all levels of 'omic abstraction – which confer a specific function in space and time. The advent of single-cell 'omic technologies offers the opportunity to assay these levels of cellular identity in unprecedented detail – and with spatial and temporal resolution, owing to the emergence of spatial transcriptomics and pseudotemporal 'trajectory inference' algorithms. Together, these technologies herald a new era in biology, enabling us to unravel cellular, tissular and organismal complexity with extraordinary precision – revealing extensive heterogeneity, the presence of rare subpopulations, and an improved understanding of cellular phenotypes, localization, functionality, abundance, prevalence, ontogeny and plasticity (191, 195). While the scaling of

these technologies presents a range of technical challenges (196), their translational potential is profound, as they not only enable the identification of novel biomarkers but, importantly, enable the robust attribution of pathogenic signatures to specific cell subsets, thus aiding their evaluation as therapeutic targets. A list of key selected studies harnessing high-dimensional single-cell technologies to investigate myeloid cells in breast cancer is shown in **Table 1.1**.

Indeed, the full extent of immune cell heterogeneity and substructure within the tumour microenvironment had not been entirely grasped until the emergence of these highly parallelized single-cell technologies. This is especially true of the intra-tumoural myeloid compartment in human breast cancer, where unanticipated heterogeneity was observed in the monocytic precursors of TAMs, which segregated into as many transcriptionally distinct clusters as did T cells, in a scRNAseq study (102) that will be discussed in detail below.

In an early (2017) scRNAseq study of all cell types in the human breast tumour microenvironment, Chung *et al.* indicated the presence of a putative TAM population enriched with expression of *CD14*, *CD68* and phagocytic enzymes, and resembling an “M2” signature – expressing *CD163*, *MS4A6A* and *TGFBI*, alongside genes implicated in angiogenesis and tumour progression (IL-8 (*CXCL8*) and *PLAUR*) (197). However, this unbiased survey of the breast TME contained just 515 cells (fewer than 40 of which were macrophages) and was therefore not sufficiently powered to capture the full spectrum of phenotype or heterogeneity.

First Author	Year	Journal	Species	Technology used	Cells implicated	Samples analysed	State	Ref
An	2018	Cell Reports	Mouse	Mass cytometry (CyTOF, Fluidigm)	Broad panel (incl. T & B cells, neutrophils, monocytes, macrophages)	11,198 $CD45^+$ cells from 4 control and 4 tumour-bearing mice	Fresh	(198)
Azizi	2018	Cell	Human	Single-cell RNAseq (InDrop)	$CD45^+$ immune cells	47,016 cells from 8 treatment-naive tumours (ER^+, PR^+, $HER2^+$, TNBC), as well as matched normal breast, blood and lymph nodes	Fresh	(102)
Chung	2017	Nature Comms	Human	Single-cell RNAseq (Fluidigm C1/SMARTer)	Whole micro-environment	505 cells from 11 tumours (ER^+, PR^+, $HER2^+$, TNBC)	Fresh	(197)
Cheng	2021	Cell	Human	Single-cell RNAseq (mixed platforms: 10X Chromium, MARS-seq, inDrop, Smart-seq)	Pan-myeloid	131,161 cells from 210 patients across 15 cancer types. Derived from 380 samples of tumour, adjacent, blood and lymph node tissue.	Fresh	(199)
Janiszewska	2019	Nature Cell Biology	Mouse & Human (patient/cell-line-derived xenograft)	Single-cell RNAseq (10X Chromium)	$CD45^+$	7704 cells from blood, primary tumour and lung metastases of mice	Fresh	(200)
Keren	2018	Cell	Human	Multiplexed ion beam imaging by time-of-flight (MIBI-TOF)	Broad panel (incl. T & B cells, NK cells, dendritic cells, neutrophils, macrophages)	3000-10,000 cells per patient in 41 TNBC tumours	FFPE	(201)
Marques	2016	BMC Cancer	Human	Imaging flow cytometry (ImageStream, Amnis)	Epithelia, lymphocytes, macrophages	Unknown cell numbers, across 6 tumours	Frozen (OCT)	(202)
McGinnis	2019	Nature Methods	Human & Mouse (patient-derived xenograft)	Single-cell RNAseq (10X Chromium) with MULTI-SEQ	$hCD298^+$ human metastases & $mCD45^+$ mouse immune cells	10,427 cells from 1 control lung, 3 primary tumours and 5 lung metastases of various stages	Frozen (dissociated cell suspension)	(203)
Opzoomer	2021	Science Advances	Mouse	Single-cell RNAseq (10X Chromium)	Macrophages ($CD45^+CD11b^+Ly6G^-F4/80^{hi}$)	10,502 cells were derived from three MMTV-PyMT tumours	Fresh	(204)
Qiu	2019	bioRxiv	Human	Single-cell RNAseq (MIRALCS)	$CD45^+$	9683 cells from 14 treatment-naive TNBC tumours	Fresh	(205)
Schapiro	2017	Nature Methods	Human	Imaging mass cytometry (CyTOF, Fluidigm & ArF excimer GeoLas C laser system, Coherent)	$CD68^+$ macrophages, tumour cells, stromal cells	49 breast cancer samples and 9 normal breast samples (3 matched). Cell number unknown.	FFPE	(206)
Szczerba	2019	Nature	Human and Mouse	Single-cell RNAseq (SMART-seq2)	Circulating tumour cells (CTCs) & immune cells	749 CTCs from blood of 37 patients, including 33 CTC-immune cell clusters (exact number of cells not reported). 36 mice across 5 metastatic models of breast cancer assessed. Cell number unknown.	Fresh	(207)

Wagner	2019	Cell	Human	Mass cytometry (CyTOF2, Fluidigm)	Broad panel (whole tumour micro-environment)	26 million cells in 144 tumours, 50 non-tumour samples (including juxtatumour & mammoplasty)	Fresh	(103)
Wu	2021	Nature Genetics	Human	Single-cell RNAseq (10X Chromium) & CITE-seq (Biolegend)	Whole tumour micro-environment	130,246 cells from 26 tumours (TNBC, HER2$^+$, ER$^+$).	Fresh	(101)
				Spatial Transcriptomics (10X Visium)		Unknown cell numbers. 6 tumour sections analysed (TNBC & ER$^+$ only)	Frozen (OCT)	

Table 1.1 | Key selected studies harnessing high-dimensional single-cell technologies to investigate myeloid cells in human or mouse breast cancer. High-dimensionality was defined as studies simultaneously quantifying >30 parameters. Each study is listed in alphabetical order according to the first author's name. Metadata regarding cell numbers represent the number of cells that passed quality control after (bioinformatic) exclusion of low-quality cells. Note: cell numbers were not typically quantified for imaging based methods. Abbreviations: CTC = Circulating Tumour Cell; CITE-seq = Cellular Indexing of Transcriptomes and Epitopes by Sequencing; CyToF = Cytometry by Time of Flight; ER = Estrogen Receptor, FFPE = Formalin-Fixed Paraffin-Embedded; HER2 = Human Epidermal Growth Factor Receptor 2; MMTV-PyMT = Mouse Mammary Tumour Virus Polyoma Middle T Antigen; OCT = Optimum Cutting Temperature material; PR = Progesterone Receptor; TNBC = Triple-Negative Breast Cancer.

Comprehensive mass cytometric (CyToF) analysis of 144 human breast tumour samples defined three broad categories of intra-tumoural macrophages: **(i)** early immigrant macrophages (HLA-DR^{int}CD192$^+$), **(ii)** tissue-resident macrophages (CD206$^+$HLA-DR^{int}) and **(iii)** TAMs (CD64^{high}HLA-DR^{high}) (103), although it is unclear why tissue-resident macrophages were defined in this way. Aligning with a scRNAseq analysis (102), the phenotypic space of human breast macrophages (defined as the number of different phenotypes present) varied between tissues, with reduced tissue-resident macrophage abundance in juxtatumour tissue, and expansion of phenotypically heterogeneous TAM subsets in tumour tissue, expressing an assortment of canonical TAM markers including CD204, CD206, CD163 and CD169 (103).

In a pioneering approach incorporating *in situ* spatial information, an imaging mass cytometry study by Schapiro *et al.* demonstrated that TAMs surround a multitude of cellular phenotypes in breast cancer. This unsupervised network analysis showed TAMs making key interactions

with carcinoma cells (E-cadherin^{+}), stromal cells (vimentin^{+}), and, in higher grade tumours, carcinoma cells of a hypoxic (Carbonic anhydrase 9^{+}) and proliferative (Ki-67^{+}) phenotype (206). Together these data are suggestive of distinct roles for TAMs in their immediate 'nanoenvironment', as well as tumour-wide, in a manner that changes with disease stage.

Imaging flow cytometry analysis has also revealed 'iron donor' phenotypes of macrophages and lymphocytes infiltrating both primary human breast tumours and lymph node metastases, characterised by increased expression of ferroportin, ferritin, transferrin receptor 1, and hepcidin. Deregulated expression of iron-related proteins was additionally correlated with tumour size, negative hormone receptor status and worsened prognosis (202).

In the first large-scale (~47k cell) scRNAseq atlas capturing immune cells across multiple tissues of human breast cancer patients, Azizi *et al.* revealed 3 TAM, 7 monocyte, and 5 precursor of monocyte subpopulations, the proportions of which varied greatly between patients (102). These three TAM populations were relatively poorly discriminated, sharing most of their differentially expressed genes – and their gene covariance patterns were key to disambiguating their phenotype. For instance, while all three TAM clusters expressed high levels of the M2-like markers, *MARCO* and *B7-H3*, they positively co-varied in two clusters but negatively co-varied in the third cluster (102), thus highlighting the importance of understanding co-expression networks in single-cell datasets. Some myeloid clusters were present in both normal and tumour tissue (monocytic precursors: Cluster 27 and Cluster 24; monocytes: Cluster 40) however all three clusters of macrophages were tumour-specific, demonstrating the marked 'phenotypic expansion' induced by the tumour microenvironment, and reinforcing the importance of tissue-context (102). Supporting these findings, unusual tumour-specific phenotypes of macrophages and lymphocytes (compared to normal mammary gland) were observed in a CyToF study of mouse invasive lobular (breast) carcinoma, such as the presence of a novel CD11b^{+} CD11c^{high} CD49f^{high} macrophage subset, decreased infiltration of tissue-resident macrophages, and overexpression of MHC2, which is

usually associated with macrophage activation (198). Furthermore, a recent human pan-cancer scRNAseq analysis of tumour-infiltrating myeloid cells from 210 patients across 15 cancer types described the tumour-specific enrichment of pro-angiogenic $SPP1^{+}$ macrophages compared to normal tissue (199), corroborating the phenotypic expansion observed in other studies, and indicating that this phenomenon is conserved across solid tumours. Interestingly, scRNAseq of a human breast cancer cell line (MDA-MB-468)-derived xenograft model in mice revealed that a subset of M2-like macrophages (upregulating *Chil3*, *Marco*, *Mrc1* (CD206)) were enriched only in lung metastases and not in primary tumours (200). Likewise, single-nuclei RNAseq in a patient-derived xenograft mouse model revealed that the progressively metastatic lung accumulates pro-metastatic $Cd14^{high}$ monocytes (upregulating *Thbs1*) while reciprocally depleting anti-metastatic $Cd14^{low}$ monocytes (upregulating *Wfdc21*), interstitial macrophages and alveolar macrophages (203). Together, these data suggest that distinct local immune environments are critical to breast cancer progression, and provide an opportunity to therapeutically target specific TAMs and their monocytic precursors in advanced metastatic disease.

Remarkably, Azizi *et al.*, showed that while human TAMs were among the highest expressors of canonical “M2” gene signatures, they concomitantly expressed high levels of “M1” genes as well, resulting in a positive correlation. Diffusion maps, which infer the pseudotemporal ordering of cells, indicated that all three TAM clusters formed a single trajectory of TAM activation, along which M2 genes such as *MARCO* and *B7-H3*, alongside M1 genes such as *CCL3*, increased in their expression. Among the genes most highly correlated with the TAM activation trajectory were *APOE*, *CD68*, *TREM2* and *CHIT1*, likely indicating either the differentiation of recruited macrophages (from monocytic precursors) or the activation of tissue-resident macrophages. Importantly, however, the positive correlation of M1 and M2 signatures in the same cells (regardless of their ontogeny) suggests a high degree of plasticity in macrophages, and supports the idea that macrophage activation is markedly impacted by the tumour microenvironment in a manner that does not align with the polarization states

proposed by the classical M1/M2 paradigm (102) (**Figure 1.1**). Similar findings have since been reported in single-cell studies of larger breast cancer patient cohorts by Qiu *et al.* (205), Wagner *et al.* (103), and in the MMTV-PyMT mouse model of breast cancer (204) (**Table 1.1**), as well as in glioma (208) and non-small cell lung cancer (209), indicating a common phenomenon among TAMs.

However, in the largest, most recent and highest depth scRNAseq atlas of the breast tumour microenvironment by Wu *et al.*, which covered all major clinical subtypes of breast cancer, no analysis was performed on the correlation of M1 and M2 signatures in TAMs, nor on their pseudotemporal ordering with respect to monocytes (101). Thus, the most comprehensive study of the breast TME has not (yet) corroborated the 'continuous activation model' of TAMs proposed by Azizi *et al.* Notably, the Wu *et al.* study did corroborate other findings, such as the segregation of 13 heterogeneous myeloid subpopulations, including 3 which were assigned as conventional macrophages, 2 as lipid-associated macrophages, 3 as monocytes, and 1 as a general cycling (proliferating) myeloid population. This study also harnessed spatial transcriptomics, defining a lipid-associated macrophage subset preferentially localised to the invasive border, and additionally revealed novel heterotypic 'ecotype' interactions between various co-enriched cell types, which was able to prognostically stratify a large cohort of patients from a bulk clinical breast cancer dataset. Diverse distributions of immune checkpoint molecules were also observed in distinct macrophage and monocytes subpopulations, often depending on the intrinsic clinical subtype of the tumour, highlighting the differential immunogenicity, biomarker profiles and therapeutic opportunities for different strata of breast cancer patients.

Taken together, high-dimensional single-cell analyses of the myeloid compartment have revealed the presence of novel subpopulations, increased phenotypic expansion, extensive heterogeneity at the clinical, cellular and molecular levels which can change with tumour

progression, and finally, the discovery of positively correlating M1 and M2 gene signatures which challenge the paradigm of discrete macrophage polarisation states (**Figure 1.1**).

Finally, as previously described in **Sections 1.2**–**1.3**, the augmentation of adaptive immunity by ICB has achieved remarkable success in the clinic for certain cancer types by harnessing T-cell centric mechanisms of immune evasion. While modulating the PD-1/PD-L1 axis is approved in TNBC, the success of ICB has been less pronounced in breast cancer generally in comparison to other cancer types, and this may be partly explained by the pro-tumoural activities of the intra-tumoural myeloid compartment in breast cancer, particularly macrophages, which can regulate T cell activity to influence the outcome of the anti-tumoural immune response. Yet these myeloid-centric mechanisms of innate immune evasion have not yet enjoyed the same degree of clinical success as T-cell centric mechanisms, despite the high potential of synergistic efficacy when modulating both mechanisms simultaneously. The emergence of high-dimensional single-cell 'omic technologies provides an opportunity to understand the exact characteristics of these TAM populations and to define their pathogenicity. Rather than depleting all TAMs, novel opportunities are arising to specifically target pathogenic subsets, including TAM reprogramming strategies which can re-educate TAMs to an anti-tumoural phenotype. Thus, targeting the molecular switches underlying the acquisition of pathogenic macrophage phenotypes, rather than targeting the end-state phenotype itself, may be an attractive therapeutic strategy which rebalances macrophage function in favour of anti-tumoural activity (126, 210). Thus, in summary, macrophages can corrupt T cell targeted immunotherapies in many ways, and therefore, unravelling mechanisms of TAM-mediated immune suppression using single-cell technologies may pave the way for novel immunotherapeutic combinations with synergistic efficacy.

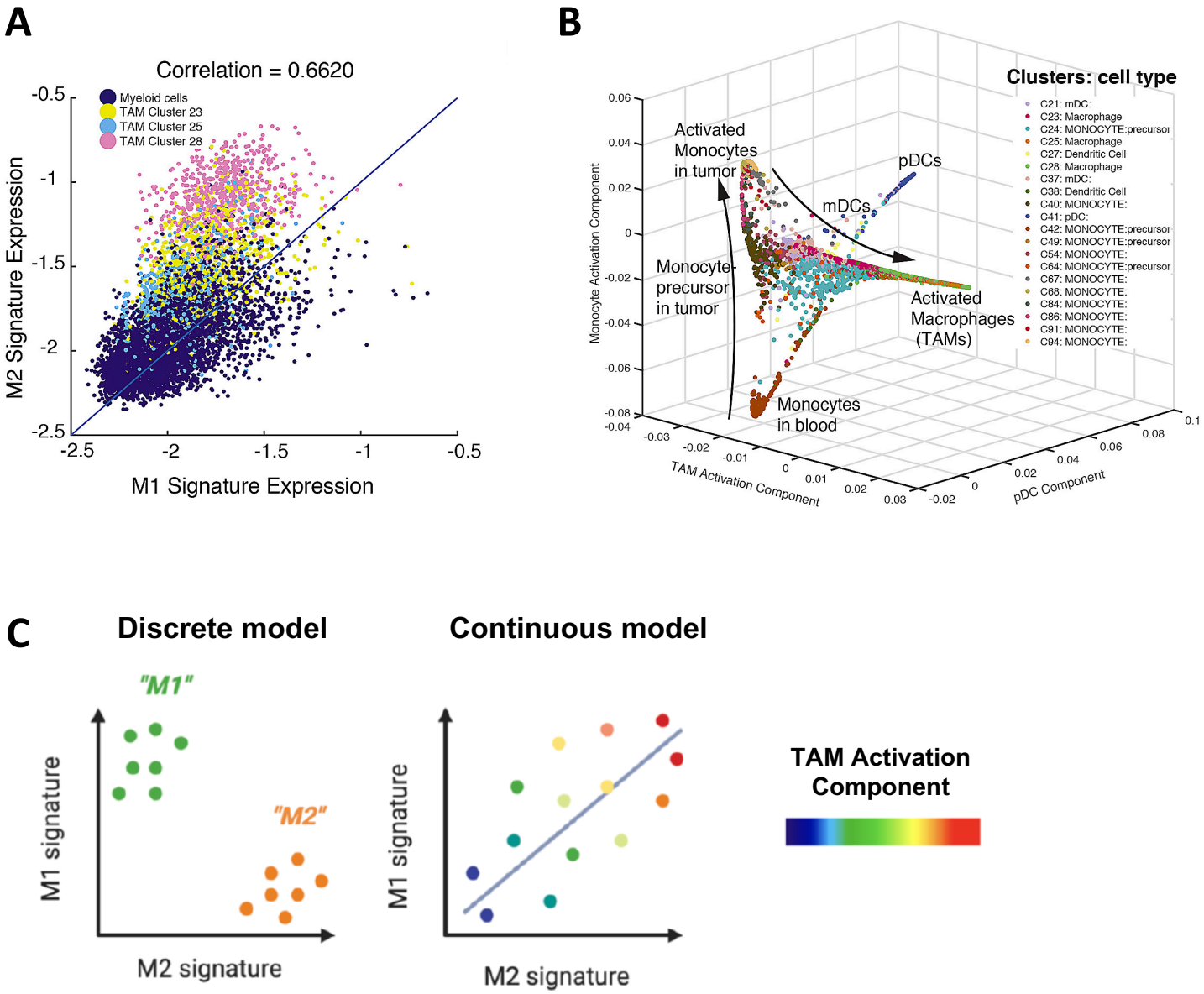


Figure 1.1 | Paradigms of macrophage activation. (A) Single-cell RNAseq of tumour-infiltrating immune cells from 8 human breast tumours by Azizi *et al.* (102) indicated a positive correlation between M1 and M2 signature gene expression in TAMs. Figure taken from Azizi *et al.* **(B)** Diffusion mapping is a non-linear dimensionality reduction method which establishes a pseudotemporal ordering of cells to visualise continuous biological processes, including cellular differentiation. Azizi *et al.* (102) generated a diffusion map of all monocytic and TAM clusters (the first three diffusion components of which are visualised on the X, Y, and Z axes), showing that TAMs form a single trajectory of activation, emanating from activated monocytic precursors. Figure taken from Azizi *et al.* **(C)** The classical M1/M2 paradigm of macrophage activation is a 'discrete' model in which pro-inflammatory (anti-tumoural) M1 macrophages or immune-suppressive (pro-tumoural) M2 macrophages represent the two polarized phenotypic extremes. However, the findings of Azizi *et al.* challenge this paradigm, instead suggesting that TAM activation is a continuous phenotypic spectrum, and that the M1/M2 genes correlate along this activation trajectory. Abbreviations: TAM = Tumour-Associated Macrophage

1.5 | Extracellular matrix in breast cancer and inflammation

The extracellular matrix (ECM) is the non-cellular component of tissues. It is composed of a complex assortment of (insoluble) macromolecules which not only provide a physical scaffold for the cellular constituents of the tissue (to ensure structural tissue integrity), but also provide biochemical and biomechanical cues that regulate cell growth, survival, adhesion, migration, and differentiation (211, 212). ECM genes were crucial for the evolution of multi-cellular organisms due to their key role in organising cells into multi-layer tissue structures, and the complexity of the ECM is thus reflective of the diversity of tissue structures required for Metazoan life (213). Interestingly, the lower incidence and lethality of tumours in plants may even be explained by their fixed cell-wall matrix, which limits the cellular motility necessary for metastasis (214). However, the importance of the ECM in supporting tumour progression has often been overlooked due to the widespread misconception of the matrix as an inert and static scaffold, rather than a dynamically remodelled regulator of tissue homeostasis which modulates almost every hallmark of cancer (215).

Indeed, the cancer-associated ECM is not only an integral characteristic of cancer but also a key driver of pathology and phenotype. In breast cancer specifically, high collagen density promotes tumour initiation and progression (216), matrix crosslinking enhances integrin signalling to promote tumour progression (217), and the presence of matrix metalloproteinases (MMPs) and mechanical stress are predictive of tumour formation (218). Prognostically, an increased mammographic density, which correlates with collagen deposition, is significantly associated with a higher risk of developing breast cancer (219), while high tissue stiffness was separately shown to be linked with malignant phenotypes (220, 221) such as high tumour grade, large tumour size, and negative ER status – with triple-negative tumours being the most stiffened (221). Thus, the pathologic role of the ECM in breast cancer progression is well established. It is worth noting, however, that healthy

mammary tissue is unique in its ability to structurally evolve throughout foetal development, puberty and pregnancy compared with other healthy tissues (222, 223), and therefore naturally harbours distinct ECM characteristics that guide mammary gland development, branching morphogenesis, lactation, and involution (224). Thus, the composition of the ECM is complex and dynamic, with roles reaching far beyond the mere provision of structural integrity, and the complex ECM requirements of the breast may render this tissue especially liable to ECM-mediated dysfunction.

As the accumulation of ECM proteins is intimately linked with the inflammatory response (225, 226), it is perhaps unsurprising that the ECM influences immune cell function. At a structural level, the ECM can provide migratory 'highways' along which T cells efficiently infiltrate tissues through the provision of chemotactic ECM fragments released by MMP-mediated cleavage (215, 226, 227). On the other hand, a physically stiffened tumour-associated ECM can suppress anti-tumoural T cell function (215). However, the ECM is not always working alone – these phenotypes are often achieved through co-ordination with other constituents of the TME, such as TAMs, which are able to suppress T cell recruitment by regulating ECM molecules, vascular structure and the chemokine milieu (121). Indeed, at a non-structural level, the ECM can indirectly modulate immune cells via regulating the availability of cytokines, chemokines and growth factors, as well as the hydration and pH of the tissue microenvironment. Alternatively, the ECM can directly alter immune cell phenotypes, for example, via engagement of inhibitory receptors (such as LAIR1 expressed by $CD8^+$ T cells), or via agonism of PRRs (215, 228, 229). A range of ECM proteins with PRR-stimulatory properties have been described, including tenascin C (TNC, hyaluronan, heparan sulfate, fibronectin, and fibrinogen, all of which are capable of binding TLR4, as well as decorin and biglycan which are additionally capable of binding TLR2 (168). The ECM therefore imparts specific signals that influence important early steps in the inflammatory cascade, such as immune cell differentiation and migration into inflamed tissue, as well as promoting immune cell activation and survival in chronically inflamed tissue (226).

The following subsection, **Section 1.6**, will explore in detail the functional and prognostic effects of one exemplar tumour-associated ECM molecule, TNC, in breast cancer, including current understanding on the role of TNC on altering macrophage phenotype.

1.6 | Tenascin-C in breast cancer

Tenascin-C (TNC) is a large, multimodular, hexameric glycoprotein of the extracellular matrix, which is induced specifically upon tissue damage, and whose expression is otherwise largely restricted to the developing embryo, stem cell niches, chronic inflammatory conditions and cancers (230-232). TNC has a diversity of pleiotropic functions, enabled by its multimodular structure, which features a series of epidermal growth factor (EGF)-like repeats, a series of fibronectin type III (FNIII)-like repeats, and a C-terminal fibrinogen-like globe (FBG) domain, each with a multitude of binding partners that include integrins, ECM proteins and TLR4 (232-234) (**Figure 1.2**). The tenascin gene family pre-dates the evolution of adaptive immunity, but TNC co-evolved with the appearance of immunoglobulin-based adaptive immunity, alongside C-C chemokines, MHC molecules, TCRs, and TLR4, suggesting that TNC plays an important and highly conserved role in the innate and adaptive immune response of jawed vertebrates (235).

As an endogenous DAMP, TNC promotes sterile inflammation through the (CD14- and MD-2-independent) activation of TLR4 on a variety of cell types including macrophages, dendritic cells and fibroblasts, as well as through interaction with α9β1 and αVβ3 integrins (236). Midwood *et al.* originally demonstrated that the FBG domain of TNC was responsible for binding TLR4 on synovial macrophages and fibroblasts, and that this drove chronic inflammation and joint damage in rheumatoid arthritis through the induction of pro-inflammatory cytokines, which could be mostly rescued in *Tnc* knockout mice (237). Later work in human tissue revealed that TNC-induced TLR4 stimulation generated divergent (though partially overlapping) phenotypes in monocyte-derived macrophages compared to the pathogen-associated TLR4 ligand, LPS, demonstrating distinct modes of TLR4 activation for DAMP- and PAMP-derived ligands (238).

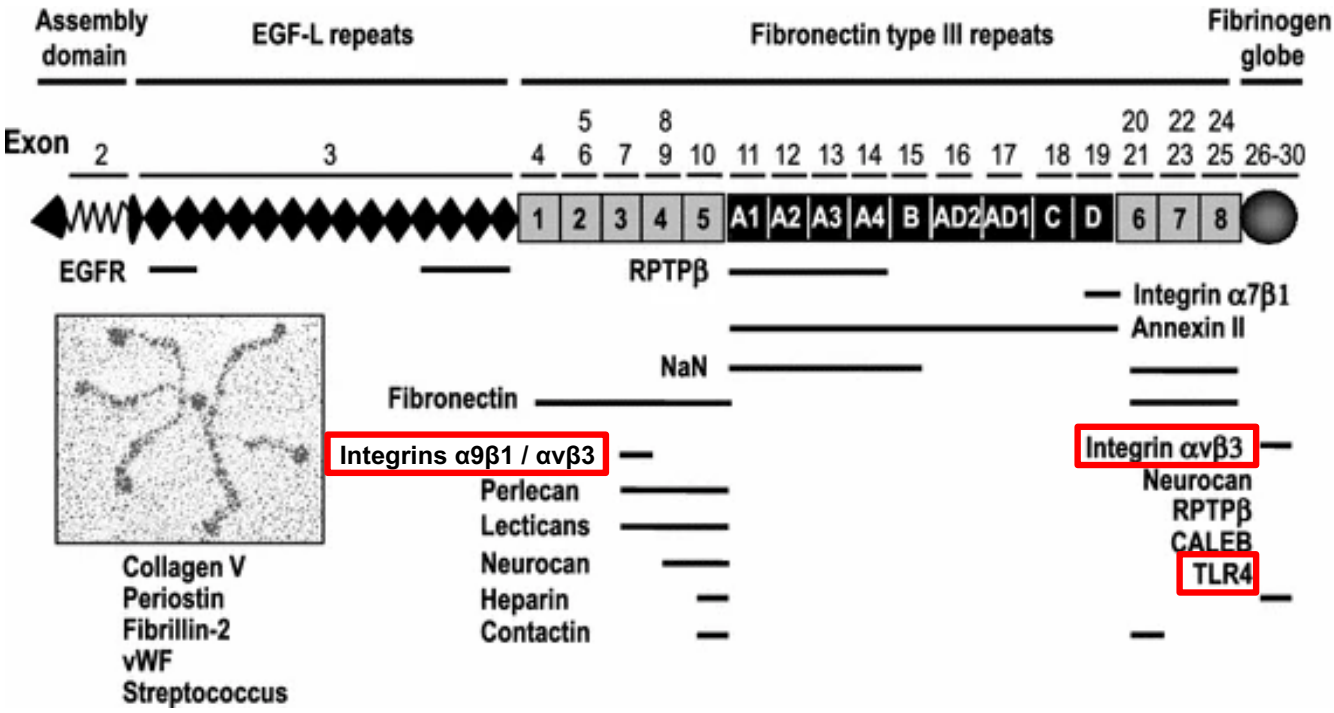


Figure 1.2 | Modular structure of TNC and its known binding partners. TNC is a multi-modular and hexameric extracellular matrix glycoprotein, as visualised by electron microscopy (left box), composed of the following four domains: **(i)** assembly domain, **(ii)** EGF-L repeats, **(iii)** Fibronectin Type III repeats, and **(iv)** Fibrinogen-like globe domain. TNC engages a wide variety of binding partners, three of which are responsible for mediating TNC's inflammatory effects (indicated in red): **(i)** TLR4, **(ii)** integrin $\alpha 9\beta 1$ and **(iii)** integrin $\alpha v\beta 3$ (236). Figure adapted from Midwood *et al.*, 2011 (233). Abbreviations: EGF-L = Epidermal Growth Factor-Like, TLR4 = Toll-like Receptor 4.

TNC is abundantly expressed in several solid tumours (231, 239), and numerous cellular sources have been described, including myeloid cells (240), fibroblasts (241), and tumour cells (239). Tumoural TNC often exhibits distinctive patterns of alternative splicing in its FNIII-like repeats compared with normal tissue, leading to divergent functional consequences (242). In breast cancer, early work described TNC as a stromal marker of malignancy in the rat mammary gland (243, 244), and this was followed by clinical data showing that total expression of TNC is profoundly increased in invasive and intraductal carcinomas compared with healthy breast tissue (245). Interestingly, the predominant splice variants observed within normal and tumour tissues were a small (180kDa) and large (330kDa) TNC isoform,

respectively (246), which aligned with later observations showing associations of various splice isoforms (varying in their AD1, AD2, B and D FNIII-like repeats) with invasiveness and proliferation (247-249). Importantly, elevated expression of TNC was associated with increased metastasis and worsened prognosis in breast cancer patients, and mechanistic investigation in immune-compromised mice showed that this was due to the role of TNC in promoting survival of circulating tumour cells which enabled their Notch/Wnt-dependent colonization of lung parenchyma (250). Supporting this, TNC has been implicated within gene expression signatures predicting distal metastasis in breast cancer (251-253), and associated with aggressive ER-negative tumours (254, 255) or local and distant recurrence (254, 256, 257), all of which are indicative of poor prognosis. Later work in the immunocompetent NT193 model of breast cancer was able to recapitulate key clinical phenotypes previously reported by Oskarsson *et al.* in immunodeficient mice (250), such as earlier tumour onset and larger lung metastases in TNC-high tumours (258). Various other cancers such as glioma, head and neck, colorectal and neuroendocrine tumours, also display TNC-dependent pro-invasive, pro-metastatic or negatively prognostic phenotypes (233, 259, 260).

TAMs can represent up to 50% of the tumour mass in breast cancer (261), and their interaction with tumour-associated ECM has long been appreciated since early *in vitro* work with monocytes cultured with collagen (262). More recent investigations have revealed deeper phenotypic complexity of the reciprocal interactions between TAM biology and ECM composition (263). Using the immunocompetent and orthotopic NT193 mouse model of $HER2^+$ breast cancer in which TNC had been engineered to express high or low levels of TNC in syngeneic tumours cells, either on a wildtype or TNC knockout background, Deligne *et al.* showed that tumour-derived TNC mediated a suppressive state in TAMs (downregulating MHC Class II) compared to host-stroma-derived TNC (264). Tumour-derived TNC correlated with macrophage infiltration, tumour size, tumour growth, and inversely correlated with infiltration of pro-inflammatory $CD86^+$ macrophage, while depletion of macrophages by intra-tumoural injection of clodronate liposomes ablated tumour growth (264). Blockade of the

TNC:TLR4 interaction with anti-FBG antibodies was able to decrease tumour growth and lung metastatic burden in a spontaneous model of mouse breast cancer, and this treatment synergised with anti-PD-L1 immune checkpoint inhibition (264). TNC-low murine tumours were additionally characterised by reduced Th17 differentiation and enhanced $CD8^+$ T cell proliferation, while human tumours stratified by low TNC and low "M2-like" macrophage infiltration were protected from mortality at 10 years post diagnosis (264). Together these data demonstrate a role for TNC in binding TLR4 on macrophages to promote breast tumour growth and metastasis, while coordinating pro-tumoural T cell phenotypes, highlighting the pathogenic roles of TNC in the tumour microenvironment. It is not known what factors define the differences in tumour-derived versus host-stroma-derived TNC in breast tumours, though it may be due to changes in alternative splicing, post-translational modification, spatiotemporal factors, or a combination thereof. Beyond the limited phenotypic information described above, the features of pathogenic TNC-associated macrophages – such as their functionality, plasticity, ontogeny, localization, prevalence, heterogeneity and therapeutic potential – remain mostly unknown.

The next subsection, **Section 1.7**, will pose the rationale, objectives, and hypotheses of this thesis, before summarising the key findings.

1.7 | Thesis rationale and summary

In some way, cancer can be conceptualised as a failure of the immune system to successfully identify and eliminate aberrant cells. Myeloid cells, particularly macrophages, are key immune-suppressive constituents of the breast tumour microenvironment, where they contribute to fundamental pathological processes and are consistently associated with poor prognosis. Yet immunotherapy in breast cancer has achieved only limited success, partly due to the immune-suppressive effects of the TME. TNC is an immunomodulatory ECM protein upregulated in breast cancer, where it is associated with metastasis and worsened prognosis. Previous work shows that this pathogenic phenotype is mediated in a TLR4- and macrophage- dependent manner. Understanding the mechanisms by which macrophages become co-opted by tumours and by TNC may aid the development of immunotherapeutic agents in breast cancer, and other TNC-regulated cancers, which have the potential to augment existing therapies. Deligne *et al.* showed that TNC polarises a pathogenic MHC2-low macrophage subset associated with increased tumour growth and metastasis in mice (264), highlighting the importance of acquiring or reinstating macrophage effector functions such as antigen presentation for anti-tumoural activity. However, detailed understanding of how macrophage effector functions transition from a homeostatic to pro-tumoural phenotype is limited, especially in the context of matrix-derived cues.

Given the above, the initial hypothesis of this thesis was that heterogeneous subpopulations of macrophages are responsible for TNC-dependent effects in breast cancer. If true, what are the precise roles of these heterogeneous subsets? What governs the acquisition and phenotype transition of polarised macrophage identities? Does TNC influence any other macrophage phenotypes aside from antigen presentation? Answers to these questions were sought using the immunocompetent, syngeneic, orthotopic and TNC-regulated NT193 mouse model of $HER2^{+}$ breast cancer, which recapitulates key clinical phenotypes including

metastasis and high TAM infiltration, and is well-validated across numerous publications (258, 264-267). The experimental strategy of this thesis was divided into three subsections. Firstly, **Chapter 3** focusses on discovering any potential heterogeneity and TNC-dependency of the myeloid infiltrate in breast tumours, using single-cell RNAseq. Secondly, **Chapter 4** focusses on exploring and validating macrophage phenotypes identified in the previous chapter, using flow cytometry, immunofluorescence imaging and *in vivo* pharmacological intervention. Thirdly, **Chapter 5** focusses on the conservation and translational relevance of mouse-derived macrophage phenotypes in human clinical breast cancer datasets.

To summarise, this thesis identified 7 heterogeneous myeloid subsets in breast tumours, 5 of which were macrophages (**Mac1**-**5**), alongside 1 **DC** and one **MDSC** subset. The role of TNC is predominantly exerted at the level of subset abundance rather than individual gene transcription, with 4 out of 7 subsets exhibiting TNC-dependent changes in abundance, including 2 macrophage subsets. Macrophages delegate key effector functions including antigen presentation (**Mac2**) and chemokine production (**Mac3**) to discrete subsets, which were not only validated at the protein level, but also shown to differ in their spatial and temporal dynamics. An epigenetic regulator, *Kdm6b*, potentially governing acquisition of the **Mac3** phenotype was targeted by pharmacological inhibition *in vivo,* and indeed shifted the proportion of **Mac3** via TAM reprogramming. Finally, mouse-derived myeloid subsets were well-conserved in the most comprehensive and recent single-cell RNAseq atlas of human breast cancer, and gene signature analyses in bulk clinical breast cancer datasets demonstrated significant differences in prognosis for patients with distinct myeloid 'ecotype' profiles.

Together these data demonstrate divergent roles for tumour-associated macrophages (TAMs) in breast tumour progression and metastasis, while providing mechanistic insights into the impact of TNC on TAMs, with potential therapeutic implications.

Chapter 2 | Materials and Methods

2.1 | Establishment of NT193 cell lines

The establishment of NT193 cells was performed by the laboratory of Professor Gertraud Orend at the University of Strasbourg. Briefly, NT193 cells were derived from the MMTV-NeuNT model of breast cancer (265), in which mammary tumours are induced in immunocompetent female FVB mice via mammary epithelia-restricted expression of a constitutively-active mutant of the breast proto-oncogene, *Erbb2* (*HER2*). To establish NT193 cells, primary breast tumours were harvested from MMTV-NeuNT mice, and their expression of *Tnc* was stably depleted by *ex vivo* transduction with a lentiviral shRNA construct targeting *Tnc* (TNC-low), or unchanged by a non-targeting control that was specific for a gene not found in mammals (TNC-high) (258). NT193 cells are thus composed of a co-culture of epithelial and mesenchymal cells derived from the primary tumour. The following lentiviral shRNA particles were used: Non-targeting control (Sigma-Aldrich, #SHC202V) and *Tnc*-targeted (Sigma-Aldrich, sh2, TRCN0000312138). Both constructs contained a Puromycin-selection cassette for positive selection in culture. The *Tnc*-targeting hairpin sequence was: 5'-CCGGGCATCAACACAACCAGTCTAACTCGAGTTAGACTGGTTGTGTTGATGCTTTTTG-3'.

Initial validation of TNC knockdown in NT193 cells was performed *in vitro* by Western blotting of NT193 cell lysates, and *in vivo* by immunofluorescence imaging of primary tumour sections from orthotopic NT193 mammary engraftments, both of which demonstrated potent TNC knockdown (258). The TNC-dependent phenotypes of the NT193 model have since been recapitulated in a further study (266), reinforcing its validity. NT193 cells were kindly provided on dry ice (-78°C) by the laboratory of Professor Gertraud Orend at the University of Strasbourg. In the Midwood group, knockdown of TNC in NT193 cells is validated periodically across different batches and passages by two methods: **(i)** real-time quantitative polymerase chain reaction (RT-qPCR) of NT193 cell lysates (see **Section 2.12** and **Figure 4.1A**), and **(ii)** Western blotting of NT193 cell lysates, which has been published in Deligne *et al.* (264).

The RT-qPCR validation of TNC knockdown shown in **Figure 4.1A** was performed alone. The Western blotting published in Deligne *et al.* is not shown in this thesis, but was entirely performed by Dr Claire Deligne at the Kennedy Institute of Rheumatology (University of Oxford).

2.2 | Culture of NT193 cell lines

NT193 cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM) containing glucose, GlutaMAX and pyruvate (Thermo Fisher, #10569010), supplemented with 10% heat-inactivated Fetal Bovine Serum (FBS) (Thermo Fisher, #10082147), 10,000 U/ml Penicillin and 10mg/ml Streptomycin (Thermo Fisher, #15140122), hereafter called 'NT193 culture medium'. To ensure continued positive-selection of lentiviral-transduced cells in culture, media were additionally supplemented with 10µg/ml Puromycin (Sigma-Aldrich, #P9620). Cells were maintained at 37°C in a humidified atmosphere of 5% CO_2 for a minimum of 2 weeks and a maximum of 6 weeks before use (passage 4-10). Cells were detached from tissue culture plates with Trypsin with 0.25% Ethylenediaminetetraacetic Acid (EDTA) (Thermo Fisher, #25200056) for 3 minutes at 37°C in a humidified atmosphere of 5% CO_2, and quenched with NT193 culture medium. Cell suspensions were then centrifuged at 1500rpm for 5 minutes, supernatant was discarded, and cells were resuspended in NT193 medium for plating on tissue-culture-treated dishes (Corning, #CLS430599-60EA). Cells were trypsinised once per 7 days, with 2 exchanges of fresh NT193 culture medium during this time.

Cells were periodically tested for the presence of *Mycoplasma* and consistently returned a negative result. This testing was performed by various members of the Midwood group at the Kennedy Institute of Rheumatology (University of Oxford).

2.3 | Counting and viability testing of cells

Cell viability was assessed using Trypan Blue solution (Sigma-Aldrich, #T8154-100ML) – a dye exclusively taken up by cells with a compromised and permeable cell membrane, indicative of cell death. Cells were diluted at 1:10 in Trypan Blue solution and visually confirmed for positivity of the dye using a haemocytometer (Hawksley, #AS1000+Z4) under a phase-contrast microscope (Olympus, ULWCD 0.30). The number of viable cells counted in a 16-square area of the haemocytometer was multiplied by 10^4 to provide the total number of viable cells per ml of volume, and further multiplied by the dilution factor of 10 to provide the total number of viable cells.

2.4 | Freezing and thawing of cells

Cells were cryopreserved in a solution of 10% heat-inactivated FBS (Thermo Fisher, #10082147) and 90% Dimethyl Sulfoxide (DMSO) (Thermo Fisher, #10387791) at a concentration of $5x10^6$ cells/ml in cryogenic storage vials (Thermo Fisher, #5000-0012). Cryovials were placed in a Mr Frosty container (Thermo Fisher, #5100-0001) and left overnight at -80°C to ensure gradual cooling, before transfer to long-term storage in vapour-phase liquid-nitrogen. To thaw the cells, the bottom-half of cryovials was immediately submerged in a 37°C water bath. Thawed cells were then centrifuged at 1200rpm for 5 minutes, supernatant was discarded, and cells resuspended in pre-warmed 37°C NT193 culture medium before plating in tissue culture dishes.

2.5 | Imaging of cells in culture

Imaging of cell culture plates was performed using a phase-contrast microscope (Olympus, CKX41) and a 12-bit Charge-Coupled Device (CCD) camera (QImaging, 74-0107-A0) at magnifications of 10X and 20X. Images were adjusted for automatic white-balancing prior to acquisition using the Teledyne Photometrics QCapture Pro software (version 7). For NT193 cells, multiple representative images were taken within 24 hours prior to engraftment to confirm appropriate formation of epithelial 'islands' and the expected ratio of epithelial versus mesenchymal cells. For FACS-sorted macrophage populations, multiple representative images were taken after either 2 or 24 hours in culture.

2.6 | Use of animals

Wild-type FVB mice were obtained from Charles River Laboratories (Strain: #207), bred in-house, and maintained under specific pathogen-free (SPF) conditions in accredited animal facilities at the University of Oxford. Mice were housed in independently ventilated cages at a constant temperature with water and food *ad libitum*. All procedures were conducted in accordance with the 1986 Animals in Scientific Procedures Act (ASPA), and with approval from the University of Oxford (Clinical Medicine) Ethical Committee.

Animal breeding and husbandry was performed by a Biological Services Unit at the University of Oxford.

2.7 | Orthotopic mammary tumour engraftment model

NT193 cells of either TNC-high or TNC-low status were cultured, trypsinised, and quenched as described in **Section 2.2**, then washed twice in 20ml cold (4°C) sterile autoclaved phosphate-buffered saline (PBS) (Thermo Fisher, #10649743) (previously diluted in Milli-Q de-ionized water). Cells were then resuspended to a concentration of $2x10^8$ cells/ml and kept on ice. Wild-type female FVB mice (Charles River Laboratories, Strain: #207), which are syngeneic with the NT193 & NeuNT models (258, 265), were selected for experiments only between the ages of 8-12 weeks. Mice were placed under anaesthesia via vapour-inhalation of isoflurane (Zoetis, #50019100) and then orthotopically injected in the left fourth mammary gland with $1x10^7$ cells in 50µl sterile PBS, using a 12.7mm long 29-Gauge insulin needle (Thermo Fisher, #13161931). The size of tumours was measured 3 times per week using calibrated Traceable™ Digital Calipers (Thermo Fisher, #15-077-958), and tumours volumes were calculated using the formula: (length x width x height)/2. After 18 days post-engraftment (and additionally at Day 8 and Day 13 for the time-course analyses), mice were sacrificed using a rising concentration of CO_2 (a Schedule 1 method), and tumours were harvested for subsequent analysis.

All tumour engraftments were performed alone with no assistance, with the exception of the initial engraftment for the single-cell RNAseq experiment, which was performed entirely by Dr Claire Deligne at the Kennedy Institute of Rheumatology (University of Oxford).

2.8 | Dissociation of breast tumour tissue

Upon dissection, harvested tumour tissue was briefly placed into sterile PBS before it was cut into small pieces (<5mm^3) with a sterile scalpel (Swann Morton, #0504), and digested in RPMI-1640 medium (Thermo Fisher, # 21875034) supplemented with 5% heat-inactivated Fetal Bovine Serum (FBS) (Thermo Fisher, #10082147), 10,000 U/ml Penicillin and 10mg/ml Streptomycin (Thermo Fisher, #15140122), 500ug/ml Liberase TM (Roche, #05401127001) and 100µg/ml DNAse I (Sigma-Aldrich, #11284932001) for 30 minutes at 37°C under agitation. Liberase TM was composed of highly purified Collagenase I, Collagenase II and a medium concentration of Thermolysin (a non-clostridial neutral protease) to dissociate tissue while limiting cell death and preventing cleavage of cell surface-bound proteins which would otherwise confound surface-targeted flow cytometry. Following digestion, tumour cell suspensions were kept on ice or at 4°C at all times for all subsequent steps.

2.9 | Flow cytometry of breast tumour tissue

2.9.1 | Sample preparation

Digested tumour cell suspensions were counted using a haemocytometer as previously described (**Section 2.3**). For non-sorting experiments, a total of 2x10^6 viable cells were seeded per well into a V-bottom 96-well plate (Thermo Fisher, #651101). After centrifugation at 1600rpm for 2 minutes, cells were re-suspended for 15 minutes in a blocking buffer comprised of sterile PBS with 5% heat-inactivated FBS (Thermo Fisher, #10082147) and 2.5µg/ml TruStain FcX™ (Biolegend, #101319). Cells were then centrifuged and washed one further time with sterile PBS before subsequent staining, fixation and acquisition. For sorting experiments, a variable number of viable cells were transferred to a sterile capped

polypropylene tube (Thermo Fisher, #10314791), and were then blocked and washed in proportionally equal volumes, using the same protocol as non-sorting experiments.

2.9.2 | Antibodies and dyes

A list of all antibodies and viability dyes used for flow cytometry experiments is provided in **Table 2.1** below.

Target	Fluorochrome	Clone	Host & Isotype	Type	Supplier	ID	Dilution
CCRL2	AF594	498321	Rat IgG1	Surface	RnD Systems	FAB5519 T-100UG	1:20
CD11b	PerCP-Cy5.5	M1/70	Rat IgG2b, κ	Surface	Biolegend	101227	1:200
CD11c	PE	N418	Armenian Hamster IgG	Surface	Biolegend	117307	1:200
CD206	BV605	C068C2	Rat IgG2a, κ	Surface	Biolegend	141721	1:200
CD3	BV785	17A2	Rat IgG2b, κ	Surface	Biolegend	100231	1:200
CD4	PE-TexasRed	GK1.5	Rat IgG2b, κ	Surface	Abcam	Ab51467	1:200
CD45	PE-Cy7	30-F11	Rat IgG2b, κ	Surface	Biolegend	103127	1:200
CD45	AF700	30-F11	Rat IgG2b, κ	Surface	Biolegend	103127	1:200
CD8	Pacific Blue	53-6.7	Rat IgG2a, κ	Surface	BD Biosciences	558106	1:200
CD83	PE-Cy7	Michel-19	Rat IgG1, κ	Surface	Biolegend	121517	1:200
CD86	BV650	GL-1	Rat IgG2a, κ	Surface	Biolegend	105035	1:200
CD86	BV785	GL-1	Rat IgG2a, κ	Surface	Biolegend	105043	1:200
CD86	PE	GL-1	Rat IgG2a, κ	Surface	Biolegend	105007	1:200
CX3CR1	PE	SA011F11	Mouse IgG2a, κ	Surface	Biolegend	149005	1:200
CXCL1	AF647	1174A	Rabbit IgG	Intracellular	Novus Biologicals	IC4532R-100UG	1:10
F4/80	Pacific Blue	BM8	Rat IgG2a, κ	Surface	Biolegend	123123	1:200
H-2K^{d}/D^{d} (MHC1)	FITC	34-1-2S	Mouse IgG2a, κ	Surface	Biolegend	114706	1:200
IA/IE (MHC2)	FITC	2G9	Rat IgG2a, κ	Surface	BD Biosciences	562009	1:200
IA/IE (MHC2)	BV510	M5/114.15.2	Rat IgG2b, κ	Surface	Biolegend	107635	1:200
Isotype	AF594	43414	Rat IgG1	As required	RnD Systems	IC005T	As required
Isotype	AF647	60024B	Rabbit IgG	As required	Novus Biologicals	IC1051R	As required
Isotype	APC	RTK2758	Rat IgG2a, κ	As required	Biolegend	400512	As required
Isotype	BV605	R3-34	Rat IgG1, κ	As required	BD Biosciences	562993	As required
Isotype	PE	RTK2758	Rat IgG2a, κ	As required	Biolegend	400508	As required
Ki67	PE	16A8	Rat IgG2a, κ	Intracellular	Biolegend	652403	1:200
Live/Dead	Yellow (AmCyan)	N/A	N/A	Viability Dye	Thermo Fisher	L34959	1:200
Live/Dead	Near-IR (APC-Cy7)	N/A	N/A	Viability Dye	Thermo Fisher	34967	1:200
Ly-6C	BV785	HK1.4	Rat IgG2c, κ	Surface	Biolegend	128041	1:200
Ly-6G	BV650	1A8	Rat IgG2a, κ	Surface	Biolegend	127641	1:200
NK1.1	APC	PK136	Rat IgG2a, κ	Surface	Biolegend	108709	1:200
PLAUR	AF405	109801	Rat IgG2a	Surface	RnD Systems	FAB531V -100UG	1:200
TLR2	PE	CB225	Rat IgG2a, κ	Surface	Biolegend	148603	1:200
TLR4	APC	SA15-21	Rat IgG2a, κ	Surface	Biolegend	145405	1:200
TNF	BV605	MP6-XT22	Rat IgG1, κ	Intracellular	Biolegend	506329	1:10

Table 2.1 | Antibodies and viability dyes used for flow cytometry. A list of all antibodies, isotye controls and viability dyes used for surface, intracellular and sorting flow cytometry, presented in alphabetical order. For viability dyes, equivalent fluorochromes used for compensation setup are provided in parentheses. Fluorochrome Abbreviations: AF = AlexaFluor®, APC = Allophycocyanin, BV = Brilliant Violet™, Cy = Cyanine, FITC = Fluorescein isothiocyanate, IR = Infrared, PE = Phycoerythrin, PerCP = Peridinin-Chlorophyll-Protein.

2.9.3 | Surface and intracellular analysis

After following the procedure in **Section 2.9.1**, cells were re-suspended for 25 minutes in PBS-SE Buffer – composed of sterile PBS with 2% heat-inactivated FBS (Thermo Fisher, #10082147) and 2mM EDTA (Sigma-Aldrich, #03690) – supplemented with a panel of surface-targeted primary antibodies and a viability dye, the details of which are stated in **Table 2.1**. All surface antibodies and viability dyes were diluted at 1:200. Samples were protected from light during staining and at all times thereafter until acquisition. After surface staining, cells were centrifuged and washed two further times with sterile PBS.

For non-sorting experiments, surface-only stained cells were fixed with 4% Paraformaldehyde Solution (PFA) (Thermo Fisher, #15670799) for 10 minutes, centrifuged and re-suspended in PBS with 2% Bovine Serum Albumin (BSA) (Sigma-Aldrich, #A7906), before being transferred to a sterile polypropylene tube (Thermo Fisher, #10655411) prior to acquisition. Surface-stained cells additionally undergoing intracellular staining were simultaneously fixed and permeabilized with Cytofix/Cytoperm™ Solution (BD Biosciences, #554722) for 1 hour, centrifuged and washed once with Perm/Wash Buffer (BD Biosciences, #554723), and then stained with intracellular-targeted antibodies diluted in Perm/Wash Buffer for 30 minutes. Both intracellular antibodies were stained at a dilution of 1:10, which was the top end of the range recommended by the manufacturer in each case. Cells were then centrifuged and washed two further times with Perm/Wash buffer, and re-suspended in PBS with 2% BSA (Sigma-Aldrich, #A7906) before being transferred to a sterile polypropylene tube (Thermo Fisher, #10655411) prior to acquisition. Intracellular- and/or surface-stained samples were immediately acquired on either a LSRII (BD Biosciences) or LSRFortessa X-20 (BD Biosciences) analytical flow cytometer.

All flow cytometry experiments were performed alone with no assistance, with the exception of the initial experiment where tumours undergoing single-cell RNA-sequencing were also analysed by flow cytometry (**Figure 3.1**), which was entirely performed by Dr Claire Deligne at the Kennedy Institute of Rheumatology (University of Oxford).

2.9.4 | Sorting

Cells were processed and surface stained according to **Sections 2.9.1** & **2.9.2** but were not fixed. Instead, surface-stained cells were re-suspended in Calcium/Magnesium-free PBS (Thermo Fisher, #14190144) supplemented with 2% BSA (Sigma-Aldrich, #A7906) and 10µg/ml DNase I (Sigma-Aldrich, #11284932001), and then filtered through a 70µm cell strainer (Thermo Fisher, #10788201) into a sterile capped polypropylene tube (Thermo Fisher, #10314791). Sorting was performed using a FACSAria III Sorter (BD Biosciences) operating at low pressure (20psi) and with a 100µm nozzle to preserve cell viability by reducing the hydrodynamic forces exerted upon large immune cells such as macrophages. Sorted cells were collected in sterile capped polypropylene tubes (Thermo Fisher, #10314791) that had previously been vortexed for 20 seconds with sterile heat-inactivated FBS (Thermo Fisher, #10082147). All sorted populations were gated for doublet exclusion and for negativity of a Live/Dead (L/D) viability dye. After sorting, cells were kept on ice at all times, and the purity of isolated populations was later checked on the sorter where cell numbers permitted doing so, however this was not possible for low abundance populations (<100k cells) which were instead assessed by subsequent intracellular flow cytometry using an analytical flow cytometer.

For the sorting and single-cell RNA-sequencing of intra-tumoural myeloid cells in **Chapter 3**, three TNC-high and three TNC-low tumours were harvested and pooled per TNC condition, prior to sorting. To ensure purification of $CD11b^+$ myeloid cells and exclusion of $CD3^+$ T cells, the following gating strategy was used: $CD11b^+ CD3^- L/D^-$.

For the sorting of intra-tumoural macrophages for RT-qPCR in **Chapter 4**, six TNC-high and eight TNC-low tumours were harvested and processed individually (unpooled) and then sorted according to the following gating strategy: $CD11b^{+}$ $F4/80^{+}$ L/D^{-}.

For the sorting of **Mac2**-like and **Mac3**-like intra-tumoural macrophages in **Chapter 4**, four TNC-high and five TNC-low tumours were harvested and pooled together, prior to sorting. The following gating strategies were applied: **Mac2**-like ($CD45^{+}$ $CD11b^{int/hi}$ $F4/80^{+}$$MHC1^{hi}$ $MHC2^{hi}$ L/D^{-}) and **Mac3**-like ($CD45^{+}$ $CD11b^{hi}$ $F4/80^{+}$ $MHC1^{lo}$ $MHC2^{lo}$ SSC^{hi} L/D^{-}). For the Side-Scatter (SSC)-high gate, approximately the top 30% of the SSC distribution was taken.

All operation of the FACS Aria III Sorter was performed on each occasion by Dr Jonathan Webber at the Kennedy Institute of Rheumatology (University of Oxford).

2.9.5 | Compensation controls

Compensation controls were generated immediately before acquisition by mixing 1μl of antibody with 1 drop of pre-vortexed Compensation Beads (Thermo Fisher, #01-2222-42) in a polypropylene tube containing 300μl PBS with 2% BSA (Sigma-Aldrich, #A7906). A compensation control was generated for every antibody-fluorochrome in a given panel, and incubated for 5 minutes at room temperature, protected from light. The only antibody which could not bind the compensation beads was anti-CXCL1-AF647 (Biotechne, #IC4532R), as this antibody was raised in rabbit, which is not recognized by the secondary antibodies on the compensation beads. Therefore, a surrogate rat anti-mouse antibody with the same fluorochrome was used instead.

2.9.6 | Gating strategies

The details of each gating strategy are shown in the corresponding Results chapter for each experiment.

2.9.7 | Software

Flow cytometry data was acquired using BD FACS DIVA (version 8.0.1) and analysed using Treestar FlowJo (version 10.4.2).

2.10 | Droplet-based single-cell RNA sequencing

Following flow cytometric sorting described in **Section 2.9.4**, purified $CD11b^+$ intra-tumoural myeloid cells were kept on ice, centrifuged at 1300rpm for 5 minutes at 4°C, and re-suspended to a concentration of $1x10^6$ cells/ml in RPMI-1640 medium (Thermo Fisher, # 21875034) supplemented with 10% heat-inactivated Fetal Bovine Serum (FBS) (Thermo Fisher, #10082147). Cells were then processed for droplet-based single-cell RNA-sequencing (scRNAseq) with the Chromium 3' v2 platform (10X Genomics), using the Single Cell 3' Library & Gel Bead Kit v2 (10X Genomics, #PN-120237) according to manufacturer instructions. Briefly, single-cells were combined at a limiting dilution with reverse-transcription reagents, gel beads containing barcoded oligonucleotides, and oil on a microfluidic chip, together forming a gel bead in emulsion (GEM) which serves as a reaction vesicle for reverse transcription. The single cell within each GEM was lysed, the gel bead was dissolved, and reverse transcription of polyadenylated RNA (representing most mRNA in eukaryotes) was performed via a poly-deoxythymine (poly-dT) primer, followed by amplification. As each bead has a unique 10X barcode, all complementary DNAs (cDNAs) detected from a single-cell possess the same barcode, and because each individual primer molecule includes a unique molecular identifier (UMI), every individual RNA molecule can be identified, enabling computational exclusion of duplicates. Following amplification, cDNAs were randomly enzymatically fragmented and ligated with an Illumina Read 2 adaptor prior to the final Sample Index PCR step, which provided final 10X single-cell libraries. These libraries then underwent paired-end Illumina sequencing on a single sequencing lane of a HiSeq 4000 instrument (Illumina). Paired end reads of 26bp and 98bp in length were used for Read 1 and Read 2, respectively. Approximately 10,000 single cells were loaded and a total of 5452 cells were recovered, with a mean read depth of ~61k and ~71k reads per cell for TNC-high and TNC-low samples, respectively.

Purified single-cell suspensions of $CD11b^+$ myeloid cells were given to Dr Moustafa Attar and Dr Hubert Slawinski from the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (University of Oxford), who performed all subsequent steps independently.

2.11 | Computational analysis of mouse scRNAseq data

2.11.1 | Generation of FASTQ files

Raw base call (BCL) files generated by the Illumina sequencer were demultiplexed to generate FASTQ files using the 'cellranger mkfastq' pipeline within the Cellranger toolkit (version 3) (10X Genomics) which itself wraps around 'bcl2fastq' software (Illumina). This was performed by Santiago Revale from the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (University of Oxford).

2.11.2 | Data processing and quality control

FASTQ sequencing files were processed and analysed within the CGAT-core virtual environment (version 0.5.13) (268) using alpha release 'tenx' software from the 'sansomlab/tenx' GitHub repository (https://github.com/sansomlab/tenx), accessed on 11th September 2019. This software was composed of two Python3 pipelines and principally based on the 10X Genomics Cell Ranger software (version 3.0.0) and Seurat (version 3.1.1) (269). Firstly, pipeline_cellranger performed mapping and quantitation of reads via alignment to the *Mus musculus* reference genome assembly, mm10, using 'cellranger count' across Aggregation of data from both TNC-high and TNC-high conditions was then performed using

'cellranger aggr', including normalization for sequencing depth by random downsampling of UMI counts. Second, pipeline_seurat was configured to run a range of quality control steps by excluding cells expressing <500 genes or >3500 genes, cells with greater than 10% of their UMIs assigned to mitochondrial genes, or cells with >15,000 UMIs. Likewise, genes expressed in fewer than 3 cells were excluded. Differences in the number of UMIs and the percentage of mitochondrial genes between TNC-high and TNC-low conditions were regressed out according to a linear model prior to further analysis (as recommended in (270)). Cells that passed quality control were then randomly downsampled to equalise the number of cells per TNC condition using the seurat_begin.R script in pipeline_seurat. Finally, pipeline_seurat performed a range of downstream analyses including feature selection, dimensionality reduction, visualization, clustering, differential expression analysis and geneset over-representation analysis, which are described in **Sections 2.11.3**, **2.11.4**, **2.11.5**, and **2.11.7**. All parameters were specified for each step using the pipeline.yml configuration file. Python (version 3.6.2) and R (version 3.6.0) were used for all above and subsequent steps.

2.11.3 | Feature selection and dimensionality reduction

Data normalization and variance stabilization was performed with the SCTransform method (271), using a regularized negative binomial regression, which is called by the SCTransform() function from Seurat. The top 3000 genes exhibiting the highest variation across cells (the most highly variable features) were used as the input for linear dimensionality reduction by principal component analysis (PCA) using the RunPCA() function from Seurat with a default of 50 PCs. These PCs were then randomly permuted for 100 replicate samplings with the JackStraw procedure (272, 273) using the JackStraw() function from Seurat, specifying 'num.replicate=100'. All statistically significant PCs with a Benjamini-Hochberg (BH) adjusted P value of <0.05 were then used as an input for unsupervised clustering, described in

Section 2.11.4, or for visualisation by non-linear dimensionality reduction methods: t-distributed Stochastic Neighbour Embedding (tSNE) (274, 275) and Uniform Manifold Approximation and Projection (UMAP) (276, 277). These were performed with the RunTSNE() and RunUMAP() functions from Seurat with default settings, except for the tSNE parameters 'perplexity=30' and 'max.iter=10000'. All of the above procedures were performed using the seurat_begin, seurat_tsne, and seurat_umap scripts in pipeline_seurat.

2.11.4 | Unsupervised graph-based clustering

Using the embeddings of all the significant PCs defined by JackStraw analysis in **Section 2.11.3**, the FindNeighbors() function from Seurat generated a *k*-nearest neighbor (*k*-NN) graph of all cells (from both TNC-high and TNC-low conditions) based on their Euclidean distance in PCA space. The default 'k.param=20' was used. The shared overlap in local neighbourhoods (Jaccard similarity) was used to refine edge weights between connected cells, resulting in a shared-nearest neighbor (SNN) graph. All nearest neighbor analysis in FindNeighbors() was implemented using the R package 'RANN' (version 2.6.1). Unsupervised clustering of the SNN graph was then performed using the FindClusters() function from Seurat, specifying the Smart Local Moving (SLM) method 'algorithm=3' at 'resolution=0.3'. A series of resolution parameters (0.1, 0.2, 0.3, 0.4, 0,5) were assayed in parallel using the cluster_resolutions argument in pipeline_seurat, which in turn visualized changes in cluster composition at each resolution using the clustree() function from the R package clustree (version 0.4.4) (278). All of the above procedures were performed using the seurat_cluster.R, seurat_clustree.R, seurat_FindMarkers.R and seurat_jackstraw.R scripts in pipeline_seurat.

2.11.5 | Differential gene expression analysis

Differential gene expression analysis between clusters was performed with the FindMarkers() function from Seurat, specifying the non-parametric Wilcoxon rank sum test (test.use='wilcox'). Testing was limited to genes detected in at least 10% of cells (min.pct=0.1), and which showed, on average, at least a ±0.25 log2 fold-change (logfc.threshold=0.25). Only markers that were conserved between both TNC-high and TNC-low conditions were selected, using the 'conservedfactor' argument in pipeline_seurat. P values were adjusted with the BH correction. The heatmap of the top 20 differentially expressed marker genes per cluster was generated by the 'markerComplexHeatmap' function in the seurat_summariseMarkers.R script of pipeline_seurat, via the (now deprecated) 'tenxutils' package (version 0.1) on GitHub (https://github.com/sansomlab/tenx/tree/master/tenxutils). Differential gene expression analysis between TNC-high and TNC-low conditions within the same cluster was performed using the 'seurat_summariseMarkersBetween.R' script in pipeline_seurat.

2.11.6 | Visualisation of gene expression distributions

The expression distribution of genes was visualized in four ways: **(i)** violin plot, **(ii)** bubble plot, **(iii)** superimposition onto a UMAP, or **(iv)** heatmap. Violin plots were generated using the plotHorizontalViolins() function from the (now deprecated) 'tenxutils' (version 0.1) repository on GitHub (https://github.com/sansomlab/tenx/tree/master/tenxutils). Bubble plots were generated using the DotPlot() function, expression-coloured UMAP plots were generated using the FeaturePlot() function, and heatmaps visualizing the expression of genes within PCA space were generated using the DimHeatmap() function, all in Seurat (version 3.1.1).

2.11.7 | Geneset over-representation analysis

Significantly differentially expressed genes (BH-adjusted P <0.05) from each cluster were tested for their over-representation in the following genesets from the Molecular Signatures Database (MSigDB) Version 6.1 (279, 280): Gene Ontology (GO) Biological Processes (GO.BP), Molecular Function (GO.MF), Cellular Components (GO.CC), Kyoto Encyclopaedia of Genes and Genomes (KEGG), Canonical Pathways, Reactome, Biocarta, Immunological Signatures, and Transcription Factor Motifs. The odds ratio and statistical significance of over-represented genesets was determined with a hypergeometric Fisher's exact test (BH-adjusted P <0.05) using the analyseGenesets() function from the 'gsfisher' package (version 0.2) at the GitHub repository: https://github.com/sansomlab/gsfisher. All geneset analysis was performed with the following pipeline_seurat scripts: summariseGenesets.R, genesetAnalysis.R and fetch_geneset_annotations.R.

2.11.8 | Trajectory analysis with diffusion maps

Diffusion maps are a non-linear dimensionality reduction method that are able to capture biologically relevant changes in continuous gene expression (281), and have been used extensively to model cellular differentiation. Using the embeddings of all the significant PCs defined by JackStraw analysis in **Section 2.11.3**, diffusion maps were generated using the DiffusionMap() function in the R package 'destiny' (version 3.7.1). Only macrophage clusters were selected for inclusion in the diffusion map. The first three diffusion map components (eigenvectors) were visualized in three-dimensional plots generated with the scatter3D() function in the R package 'plot3D' (version 1.4).

2.11.9 | Pseudotemporal modelling of macrophage plasticity

Using the Slingshot algorithm (282) from the R package 'slingshot' (version 2.0.0), the getLineages() function was used to create a minimum spanning tree (MST) defining the three macrophage trajectories of interest (**Mac1**>**Mac2**, **Mac1**>**Mac3**, **Mac1**>**Mac4**>**Mac5**). This was achieved by manual specification of the root cluster as naive-like **Mac1** (start.clus=0) and the ending clusters as the putative terminal macrophage effector states of **Mac2**, **Mac3** and **Mac5** (end.clus=c(1,2,5)). Numerical values represent cluster numbers. MST coordinates were then passed to the getCurves() function to generate smooth principal curves and compute pseudotime values. The statistical significance of gene expression changes over pseudotime were determined by regressing every gene on each pseudotime trajectory with a general additive model (GAM), using the gam() function in the R package 'gam' (version 1.16.1). Selected significant pseudotemporal changes in gene expression were then plotted using ComplexHeatmap (version 2.0).

R code for fitting the GAM was provided by Maria Gomez Vazquez from the Kennedy Institute of Rheumatology (University of Oxford).

2.11.10 | Triaging of marker genes for detection and isolation of macrophage subsets

To determine macrophage marker genes that optimally represent the terminal effector states of enriched antigen-presentation (**Mac2**) and chemokine production (**Mac3**), differential expression analysis was re-performed on macrophage clusters alone (**Mac1-5**), using the receiver operating characteristic (ROC) method (test.use='roc') in the FindAllMarkers() function of Seurat (version 4.0.1). Testing was limited to genes detected in at least 10% of cells (min.pct=0.1), and which showed, on average, at least a ±0.25 log2 fold-change

(logfc.threshold=0.25). The ROC method generated area under the curve (AUC) values which represent the predictive statistical power of a marker to classify two groups of cells. These AUC values were used to rank candidate marker genes, alongside additional characteristics such as marker gene expression level, frequency of expression within a cluster, and the concordance of the marker with the overarching biological function of the cluster.

2.12 | RT-qPCR

For the validation of *Tnc* knockdown in NT193 cells, a total of $5x10^6$ cells at passage 6 were lysed in RNeasy lysis buffer, RLT (Qiagen, #79216), supplemented with 1% β-Mercaptoethanol (Sigma-Aldrich, #444203), for both TNC-high and TNC-low conditions. RNA was isolated using the RNeasy Mini Kit (Qiagen, #74104) according to manufacturer instructions.

For the analysis of FACS-sorted tumour macrophages (as described in **Section 2.9.4**), RNA was isolated using the RNeasy Micro Kit (Qiagen, #74004) according to manufacturer instructions. The purity and concentration of isolated RNA was calculated using a NanoDrop One Spectrophotometer (ThermoFisher, #ND-ONE-W), and a total of 100ng of RNA was reverse-transcribed per sample using the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher, #10400745) and a Programmable Thermal Controller (MJ Research Inc., #PTC-100) with the following settings: 25°C for 10 minutes, 37°C for 120 minutes, 85°C for 5 minutes, 4°C thereafter. RT-qPCR was then performed in a 384-well plate on a Viaa7 RT-PCR system (Thermo Fisher, #4453545) using the TaqMan Universal PCR Master Mix (Thermo Fisher, #4305719) alongside FAM-conjugated TaqMan primers targeting the following genes: *Arg1* (Thermo Fisher, Mm00475988_m1), *Cd86* (Thermo Fisher, Mm00444540_m1), *Hprt* (Thermo Fisher, Mm00446968_m1) *Mrc1* (Thermo Fisher,

Mm01329362_m1), *Tnc* (Thermo Fisher, Mm00495662_m1), and *Tnf* (Thermo Fisher, Mm99999068_m1). Each biological replicate was run in technical triplicates, and data were analysed using the Delta-Delta Ct method (283), presented as a percentage of the 'housekeeping' gene, *Hprt* (for FACS-sorted tumour macrophage samples) or as a fold-change relative to the percentage of *Hprt* (for NT193 samples).

For FACS-sorted tumour macrophage samples, RNA isolation and cDNA generation was performed jointly with Dr Martha Gschwandtner (MG) at the Kennedy Institute of Rheumatology (University of Oxford) on one out of two occasions, and solely performed by MG on the remaining occasion. The RT-qPCR was performed entirely by MG on both occasions.

Alongside the validation of TNC knockdown in NT193 cells presented in this thesis (**Figure 4.1A**), additional validation of other NT193 batches and passages (not shown in this thesis) was periodically performed by various members of the Midwood group at the Kennedy Institute of Rheumatology (University of Oxford). Please refer to **Section 2.1** for further details of NT193 cell validation.

2.13 | Differentiation and Isolation of BMDMs

Bone marrow-derived macrophages (BMDMs) were isolated by flushing the femur of naive wild-type female FVB mice and lysing erythrocytes with 1ml ice-cold Ammonium Chloride Potassium (ACK) Lysing Buffer (Thermo Fisher, #A1049201). Differentiation was performed by culturing cells for 7 days with 10ng/ml Macrophage-Colony Stimulating Factor (M-CSF) (Peprotech, #315-02) in 'BMDM culture medium' composed of DMEM containing glucose, GlutaMAX and pyruvate (Thermo Fisher, #10569010), 10% heat-inactivated FBS (Thermo

Fisher, #10082147), 1X Antibiotic Antimycotic Solution (Sigma-Aldrich, #A5955) and 50μM β-Mercaptoethanol (Thermo Fisher, #21985023). Media was exchanged on Days 3 and 5. On Day 7, BMDMs were washed with FBS-free media and plated in a 96-well flat-bottom plate (Corning, #353072) at a concentration of $1x10^5$ cells per well.

BMDM differentiation was performed entirely by Dr Anja Schwenzer at the Kennedy Institute of Rheumatology (University of Oxford).

2.14 | ELISA

2.14.1 | Kits

The concentration of CXCL1 and TNF protein was quantified using the Mouse CXCL1 ELISA Kit (RnD Systems, #DY453) and Mouse TNFα ELISA Kit (RnD Systems, #DY410), respectively, both of which use a streptavidin-conjugated Horse Radish Peroxidase (HRP)-based readout. All steps were carried out according to the manufacturer's instructions.

2.14.2 | BMDM culture supernatants

BMDMs from four wild-type female FVB mice were seeded into 96-well flat-bottom plates at a concentration of $1x10^5$ cells per well, and stimulated with 10ng/ml lipopolysaccharide (LPS) (Sigma-Aldrich, #L4524) for 24 hours in BMDM culture medium, or left unstimulated. Supernatants were assessed for the presence of soluble CXCL1 and TNF, compared to unstimulated BMDMs. Supernatants from the four mouse BMDM samples were pooled and analysed as technical duplicates for the CXCL1 ELISA, or run as four independent biological

replicates for the TNF ELISA. Supernatants of all samples were diluted at 1:8 in reagent diluent (provided in each ELISA kit).

BMDM stimulation was performed by Dr Anja Schwenzer at the Kennedy Institute of Rheumatology (University of Oxford). All subsequent ELISAs of BMDM culture supernatants for both CXCL1 and TNF were performed alone with no assistance.

2.14.3 | FACS-sorted **Mac2**-like and **Mac3**-like culture supernatants

Macrophages with phenotypes resembling **Mac2** and **Mac3** were FACS-sorted from orthotopic mouse NT193 breast tumours as described in **Section 2.9.4**. Sorted cells were seeded into a 96-well flat-bottom plate (Corning, #353072) in RPMI-1640 medium (Thermo Fisher, # 21875034) supplemented with 10% heat-inactivated FBS (Thermo Fisher, #10082147), and incubated for either 2 or 24 hours at 37°C in a humidified atmosphere of 5% CO_2. Cells that were incubated for 2 hours were seeded at a density of $6.2x10^4$ cells per well and were not stimulated. Cells that were incubated for 24 hours were seeded at a density of $1.1x10^5$ cells per well and were stimulated with 10ng/ml LPS (Sigma-Aldrich, #L4524) for the final 2 hours of culture, or left unstimulated. All culture supernatants were analysed without any dilution, and performed in technical duplicates.

2.14.4 | Data acquisition and Analysis

The optical density at a wavelength of 450nm was measured per well using a FLUOstar Omega Microplate Reader (BMG Labtech), and then normalized to a “blank” well that had been treated with all reagents but contained no culture supernatant samples. Using the reference standards provided for CXCL1 and TNF in each kit, a 4-parameter standard curve

model was fitted based on blank-normalised values at 450nm. This yielded concentration values per well in pg/ml that were multiplied by 8 (the dilution factor) to provide actual concentration values. Data normalization and standard curve model fitting was performed using the MARS Data Analysis Software (version 4.01), and statistical analysis was performed in GraphPad Prism (version 8).

2.15 | Pharmacological inhibition of KDM6B *in vivo*

Mice were orthotopically engrafted with NT193 breast tumours according to **Section 2.7** and treated with a small molecule inhibitor of KDM6B, GSKJ4 (Selleckchem, #S7070), or a vehicle control. The number of vehicle-treated mice was 3 for TNC-high and 3 for TNC-low conditions. The number of GSKJ4-treated mice was 6 for TNC-high and 5 for TNC-low conditions. Mice were dosed on days 1, 4, 6, 8, 11, 13, and 15 post-engraftment. GSKJ4 was administered at a dose of 1mg per kilogram of mouse body weight (mg/kg). The route of administration was an intra-peritoneal injection at all timepoints except day 8, where an intra-tumoural injection was performed. GSKJ4 was dissolved in sterile Milli-Q de-ionized water and then diluted to the appropriate concentration in sterile PBS (Thermo Fisher, #14190144), as per manufacturer instructions. Vehicle control was composed of sterile PBS (Thermo Fisher, #14190144) supplemented with sterile Milli-Q de-ionized water (using the same volume of diluent as for the 1mg/kg solution). Mice were weighed on days 0, 1, 4, 6, 8, 11, 13, 15 and 18, while measurements of tumour size (see **Section 2.7**) were performed on all days except days 0 and 1, when tumours were not yet palpable. Mouse weights and tumour sizes were presented as raw values, and tumour sizes were also separately normalised to the average of the vehicle-control value for each timepoint.

2.16 | Confocal immunofluorescence imaging

2.16.1 | Tumour processing

Mouse breast tumours were dissected, briefly placed into sterile PBS, and then immediately fixed with 'Antigenfix' PFA solution (Diapath, #P0016) overnight under gentle agitation at 4°C, protected from light. Fixed tumours were then washed with PBS for 1 hour under gentle agitation at 4°C, and incubated overnight in PBS supplemented with 30% sucrose (Sigma-Aldrich, #S9578) at 4°C, embedded in Optimal Cutting Temperature (OCT) compound (Sakura Finetek, #4583). OCT-embedded tumours were stored at -80°C prior to tissue sectioning using a cryostat microtome (Leica, #CM1900UV) to obtain 10µm cryosections for staining.

All steps were performed solely by Julie Mazet at the Kennedy Institute of Rheumatology (University of Oxford), with the exception of the PFA fixation step which was performed alone on one out of two occasions.

2.16.2 | Antibodies

A list of all antibodies used for immunofluorescence experiments is provided in **Table 2.2** below.

Target	Fluorochrome	Clone	Host & Isotype	Type	Supplier	ID	Dilution
CXCL1	N/A	Polyclonal	Rabbit IgG	Primary	Proteintech	12335-1-AP	1:50
TNC	N/A	MTn-12	Rat IgG1	Primary	Sigma-Aldrich	T3413	1:150
F4/80	APC	BM8	Rat IgG2a, κ	Primary	Biolegend	123115	1:150
F4/80	FITC	BM8	Rat IgG2a, κ	Primary	Biolegend	123107	1:150
IA/IE (MHC2)	AF488	M5/114.1 5.2	Rat IgG2b, κ	Primary	Biolegend	107616	1:150
Rabbit IgG	AF594	Polyclonal	Rabbit IgG	Secondary	Abcam	Ab150080	1:400
Rat IgG	AF555	Polyclonal	Goat IgG	Secondary	Thermo Fisher	A-21434	1:400
Nuclei (dsDNA)	DAPI	N/A	N/A	Dye	Sigma-Aldrich	D9542	0.5µg/ml

Table 2.2 | Antibodies used for immunofluorescence imaging. A list of all primary and secondary antibodies used for immunofluorescence imaging experiments. Fluorochrome Abbreviations: AF = AlexaFluor®, APC = Allophycocyanin, DAPI = 4'6-diamidino-2-phenylindole, FITC = Fluorescein isothiocyanate.

2.16.3 | Immuno-staining of tumour sections

Cryosections were brought to room temperature (RT) and rehydrated in PBS for 5 minutes. After removing excess OCT, cryosections were incubated for 4 hours at RT in a wet chamber, protected from light, in a blocking solution of PBS supplemented with 2% heat-inactivated FBS (Thermo Fisher, #10082147), 10% Goat Serum (Sigma-Aldrich, #G9023), 2.5µg/ml Fc-receptor blocking antibody targeting CD16/CD32 (Biolegend, #101302) and 0.1% Triton X-100 (Thermo Fisher, #327371000). Staining with primary antibodies was performed in the blocking solution overnight at 4°C, staining with secondary antibodies was performed for 4 hours at RT, and nuclei staining with DAPI in PBS was performed for 15 minutes at RT.

All steps were performed solely by Julie Mazet at the Kennedy Institute of Rheumatology (University of Oxford).

2.16.4 | Data acquisition

Slides were mounted using Fluoromount-G® (SouthernBiotech, #0100-01), and imaged with a Zeiss LSM880 confocal microscope. Image analysis was performed with Imaris software (version 9.6, Bitplane), and cell quantification was performed using the 'colocalization' function.

All steps were performed solely by Julie Mazet at the Kennedy Institute of Rheumatology (University of Oxford).

2.17 | Mouse to human gene conversion

The top positive differentially-expressed marker genes defining each mouse myeloid cluster in the single-cell RNAseq dataset, as described in **Section 2.11.5**, were selected for gene conversion. Mouse genes were provided as Mouse Genome Informatics (MGI) symbols, and were converted to human gene symbols from the Human Genome Organisation (HUGO) Gene Nomenclature Committee (HGNC) (284). Two methods were employed to convert genes. First, the 'bioMart' R package (version 2.48.3) (285) was used to retrieve *Homo sapiens* and *Mus musculus* gene identifier information from the Ensembl database (which performs homology mapping) with the useMart() function followed by the getLDS() function. Second, an MGI text database 'report' of all human and mouse gene homology was retrieved from the following URL, accessed on 27th May 2021: http://www.informatics.jax.org/downloads/reports/HOM_MouseHumanSequence.rpt

This homology database originally sourced its data from the Alliance of Genome Resources (286). To convert genes, the universal “HomoloGene.ID” of each mouse gene was used to subset a list of human genes that shared the same “HomoloGene.ID”, yielding a list of equivalent human homologs for every mouse gene.

2.18 | Human single-cell RNAseq conservation analysis

2.18.1 | Azizi *et al.* dataset

The Azizi *et al.* single-cell RNAseq dataset of FACS-sorted $CD45^+$ immune cells from 8 human breast tumours (102) was retrieved with the R package ‘GEOquery’ (version 2.60) (287), using the getGEO() function, specifying the dataset ‘GSE114724’. This dataset also included matched normal breast tissue, blood and lymph node samples. The original clustering of cells described by the authors was recapitulated using their Bayesian Inference for Single-cell Clustering and Imputing (BISCUIT) algorithm and appropriate parameter choice (102). Imputed expression values are shown for gene expression analysis. Only clusters assigned by the authors as monocytes (C36, C40, C54, C67, C68, C86, C91, C94), monocyte precursors (C24, C27, C42, C49, C64) and macrophages (C23, C35, C28) were taken forward for analysis. Only clusters that were majority (>65%) derived from tumour samples were considered, and two clusters of monocytes (C36) and monocyte precursors (C42) were therefore excluded due to their disproportionately high abundance in blood samples. The average scaled expression of a given marker gene and the percentage of cells expressing it per cluster was calculated using the DotPlot() function in the R package ‘Seurat’ (version 4.0.5), and plotted using the R package ‘ggplot2’.

The retrieval of the Azizi *et al.* dataset and performance of the BISCUIT algorithm with the author's settings was performed entirely by Yuechan Chi from the Department of Oncology (University of Oxford). All remaining analysis was performed alone with no assistance.

2.18.2 | Wu *et al.* dataset

The Wu *et al.* single-cell RNAseq dataset of the whole microenvironment of 28 human breast tumours (101) was accessed via the interactive Single-Cell Portal hosted by the Broad Institute, accessed at the following URL on the 30th September 2021: https://singlecell.broadinstitute.org/single_cell. The average expression of a given gene and the percentage of cells expressing it per cluster was calculated by Single-Cell Portal, exported, and plotted using the R package 'ggplot2'. Expression values were scaled against all other plotted values using the scale() function in base R.

2.19 | Human bulk transcriptomic datasets

2.19.1 | TCGA

Bulk RNA-sequencing data for 1045 human breast tumours from The Cancer Genome Atlas (TCGA) (65, 66) was retrieved from the Genomic Data Commons via the R package 'TCGAbiolinks' (version 2.9.4) (288). Transcripts were quantified and normalised by the authors (65, 66) using the RNA-seq by Expectation-Maximization (RSEM) algorithm (289), and these values were multiplied by 10^6 to provide transcripts per million (TPM). TPM expression values were then log2(x+1) transformed prior to analysis. Updated clinical subtype assignments for 1021 breast tumours using the Prediction Analysis of Microarray 50 (PAM50)

classifier (41) were kindly provided by the authors, Professor Charles M. Perou and Associate Professor Joel S. Parker, from the University of North Carolina.

2.19.2 | METABRIC

Bulk microarray data for 1991 human breast tumours from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohort (290) were accessed at EGAS00000000083 with permission from the European Genome-Phenome Archive (EGA) following approval from the METABRIC Data Access Rights committee. Permission was granted to Professor Francesca Buffa at the Department of Oncology (University of Oxford), with whom a collaboration was established. Clinical metadata was retrieved from cBioPortal (291) at the following URL: https://www.cbioportal.org/study/summary?id=brca_metabric. The cleaning and amalgamation of clinical metadata was performed by Yuechan Chi from the Department of Oncology (University of Oxford). Expression values were log2-normalised by the authors (290).

2.20 | Gene signature enrichments and quality control

For the TCGA analysis, mouse gene lists were converted to human homologs as described in **Section 2.15**, and were then refined by excluding genes that were assigned as pan-cancer or breast-specific 'fitness genes' in a genome-wide CRISPR knockout screen of 324 human cancer cell lines (25 of which were breast cancer cell lines) (292). Refined signatures were also applied to the METABRIC dataset. The R package 'sigQC' (version 0.1.22) (293) was then employed to perform quality control of mouse-derived myeloid gene signatures in both human bulk transcriptomic datasets (TCGA and METABRIC). mRNA expression matrices (samples versus genes) for both datasets, alongside gene signature lists, were inputted into the make_all_plots() function to perform all quality control analyses, generate all plots, and

calculate enrichment scores. Specification of the argument 'doNegativeControl=TRUE' ensured random resampling of gene signatures of an equal length to the original gene signature to create an appropriate null distribution (negative control). The Gene Set Variation Analysis (GSVA) algorithm (294) was called by sigQC, and GSVA enrichment scores for every tumour samples were extracted and left unscaled.

2.21 | Ecotype clustering

Defining cellular ecotypes by hierarchical clustering of bulk transcriptomic samples using single-cell inferred marker genes was first described in breast cancer by Wu *et al.* (101), albeit using CIBERSORTx (295) rather than a gene signature enrichment approach. Here, the GSVA enrichment score of every tumour for each myeloid gene-signature was hierarchically clustered using the Heatmap() function in the R package 'ComplexHeatmap' (version 2.8.0) (296). This function in turn calls the dist() and hclust() functions in the base R package 'stats' (version 3.6.2) which computes the Euclidean distance between observations via the complete linkage of pairwise inter-cluster distances. Dendrogram clusters of distinct myeloid enrichment profiles (hereafter called myeloid ecotypes) were defined by performance of consensus k-means clustering on columns (tumours), specifying a total of 10 ecotypes with the argument 'column_km=10', and specifying 1000 iterations of the k-means routine by stating 'column_km_repeats=1000'. The column order of each myeloid ecotype was extracted (rather than the dendrogram order) and named in chronological order (left to right) prior to survival analysis. Expression of TNC, CIBERSORT abundances, and tumour sizes were all scaled prior to k-means clustering.

2.22 | CIBERSORT

To serve as a positive control in the myeloid ecotype analysis of the TCGA dataset, the relative abundance of tumour-infiltrating immune cells was estimated by computational deconvolution of 22 immune cell types using the CIBERSORT method and LM22 gene signature matrix (29), available at: https://cibersort.stanford.edu/runcibersort.php
[accessed 21st March 2018]. As is recommended for RNAseq data, quantile normalization was disabled. Only the relative abundance values of "naive" (M0), "pro-inflammatory" (M1) and "anti-inflammatory" (M2) macrophages are shown for simplicity. All CIBERSORT data were scaled.

2.23 | Survival analysis

Following the ecotype clustering described in **Section 2.19**, Kaplan-Meier survival analysis was performed using the kmTCGA() function from the R package 'RTCGA' (version 1.22). P values were calculated by performance of a log-rank test. For the TCGA analysis, the x axis was set to a 4000 day period since diagnosis (~11 years) while for the METABRIC analysis, this was set to 5000 days (~13.7 years). All data represent overall survival (OS).

2.24 | Statistical analysis

For all bioinformatic analyses, details of statistical testing and modelling are incorporated into the relevant subsections of this Chapter. For all wet-lab analyses, details of statistical testing are described below, and were performed using GraphPad Prism (version 8).

To ascertain a mean (parametric) difference between 2 unpaired groups, an unpaired two-tailed Student's t-test was performed. Welch's correction was applied to avoid assuming consistent standard deviations between the two groups (homoscedasticity). Where multiple unpaired t-tests were required, P values were adjusted for multiple comparisons using the Holm-Sidak method, which provides improved statistical power compared to other methods. Consistent standard deviations were similarly not assumed.

To ascertain a mean (parametric) difference between >2 unpaired groups, an Analysis of Variance (ANOVA) was performed. As the effect of only one independent variable (eg cell subsets) was being assessed, a one-way ANOVA was performed. In each case, consistent standard deviations between groups were not assumed, and thus the Brown-Forsythe and Welch test was performed. P values were adjusted for multiple comparisons using the Dunnett T3 method.

To ascertain a mean (parametric) difference between >2 paired groups, a two-way ANOVA was performed, because the paired nature of each group added a second independent variable to assess. In every case, this second independent variable represented 'time', and therefore each test was a repeated-measures two-way ANOVA (for example, assessing the effects of TNC-status over time in tumour sizes). In each case, equal variability of differences between groups was not assumed (ie assuming no sphericity), and thus the Geisser-Greenhouse correction was applied. Following this, P values were adjusted for multiple comparisons using several methods (Dunnett, Tukey or Sidak), depending on the question and comparison being made. The Tukey test was used when the mean of one group was being compared to all other groups individually, at a given timepoint (eg effects of TNC-high/TNC-low with GSKJ4/vehicle treatment, at each timepoint). The Dunnett test was used when the mean of one group was being compared to a control group (eg an unstimulated control, or the earliest timepoint in a time-course). Finally, the Sidak test was used when the mean of one group was being compared to only one other group (ie a pairwise comparison

between TNC-high/TNC-low tumours at each timepoint in a time-course), due to its moderately increased power compared to a Bonferroni correction. (Adjusted) P values below a threshold of 0.05 were considered statistically significant. Error bars represent the standard error of the mean.

Chapter 3 | The myeloid landscape of TNC-regulated breast cancer at single-cell resolution

3.1 | Results

3.1.1 | Introduction

Infiltrating myeloid cells are key protagonists in the breast tumour microenvironment, where they contribute to an array of fundamental pathological processes (79, 120, 155). Using a well-validated, immunocompetent, syngeneic and orthotopic model of murine breast cancer which recapitulates key clinical phenotypes (265, 267), previous work has demonstrated the central importance of tolerogenic macrophages in orchestrating TNC-dependent tumour progression (264). However, precise phenotypic information on the features of pathogenic TNC-associated macrophages – such as their functionality, plasticity, ontogeny, localization, prevalence, heterogeneity and therapeutic potential – remain mostly unknown.

To understand how these pathogenic macrophages are influenced by exposure to tumour-derived TNC, single-cell RNA-sequencing was performed on infiltrating myeloid cells, isolated by $CD11b^+$ FACS sorting, from syngeneic breast tumours engineered to produce either high or low levels of TNC via lentiviral shRNA knockdown. This approach led to the identification of novel intra-tumoural myeloid subsets and states associated with this model of breast cancer, and revealed the impact of tumour-derived TNC on their relative abundance. Together these data provide a unique insight into the multifaceted roles of myeloid subpopulations, their interaction with the extracellular matrix, and their contribution to breast cancer pathogenesis.

This chapter will be divided into the following subsections:

3.1.2 Tumour-infiltrating myeloid cell isolation for single-cell RNAseq

3.1.3 Quality control

3.1.4 Feature selection and dimensionality reduction

3.1.5 Clustering and visualisation

3.1.6 Characterisation of cluster phenotypes

3.1.7 TNC-dependent effects

3.1.8 Trajectory inference and pseudotemporal analysis of TAMs

3.1.9 Summary of findings and hypotheses

3.1.2 | Tumour-infiltrating myeloid cell isolation for single-cell RNAseq

To investigate the impact of tumour-derived TNC on the intra-tumoural myeloid compartment, the NT193 model of murine breast cancer was exploited (258). Briefly, this model is derived from the MMTV-NeuNT model of breast cancer, in which tumours are genetically induced in female FVB mice via mammary epithelia-restricted expression of a constitutively-active mutant of the breast proto-oncogene, *Erbb2* (*HER2/Neu*) (265). Using harvested primary breast tumours from MMTV-NeuNT mice, expression of TNC was stably depleted by *ex vivo* transduction with a lentiviral shRNA construct targeting TNC (TNC-low), or unchanged by a non-targeting control (TNC-high). The resultant NT193 cells can then be injected into the mammary gland of wildtype female FVB mice – ensuring a syngeneic, orthotopic and immune-competent model of breast cancer in which only tumour-derived TNC is altered (258).

Previous investigation of this model revealed that the presence of tumour-derived TNC significantly accelerated tumour growth, positively correlated with macrophage infiltration, and reduced the abundance of anti-tumoural $CD86^{hi}$ or $MHC2^{hi}$ macrophages (264). Recapitulating the pathological relevance of macrophages in breast cancer (108), pan-depletion of macrophages by intra-tumoural injection of clodronate liposomes was able to ablate tumour growth. Confirming the role of TNC, antibody-mediated blockade of the TNC:TLR4 interaction significantly reduced tumour growth and was accompanied by increased MHC2 expression by macrophages (264). Together, these observations provided a strong rationale for deeper phenotypic characterization of the intra-tumoural myeloid compartment, and its response to tumour-derived TNC. To achieve this, single-cell RNA-sequencing was performed on intra-tumoural myeloid cells that had been isolated by $CD11b^{+}$ FACS sorting. The scope of this investigation was restricted specifically to $CD11b^{+}$ myeloid cells, rather than the entire $CD45^{+}$ immune compartment, to increase the average read depth per sequenced cell, thereby improving the detection and resolution of myeloid subpopulations.

Consistent with previous reports of the NT193 engraftment model (258, 264), the growth of TNC-high tumours was significantly accelerated compared to TNC-low tumours (**Figure 3.1A**). At 21 days post-engraftment, three representative tumours of either TNC-status were harvested and pooled separately (with stoichiometrically equal numbers of cells from each tumour) prior to isolation of $CD11b^+$ myeloid cells by FACS sorting. To ensure a single-cell suspension of viable myeloid cells for single-cell RNAseq, and to exclude contaminating T cells, a $CD11b^+CD3^-LiveDead^-$ gating strategy was applied, and the purity of sorted $CD11b^+$ populations was >95% in both TNC conditions (**Figure 3.1B**). Flow cytometric analysis of sorted versus unsorted samples confirmed that myeloid cells were successfully enriched by $CD11b^+$ FACS sorting (**Figure 3.1C**), and that macrophages displayed a similar TNC-dependent phenotype as previously observed (264) – namely, the expansion of $CD86^+$ macrophages in TNC-low tumours (**Figure 3.1D**). Purified $CD11b^+$ cells were then processed for droplet-based single-cell RNA-seq using the 10X Genomics Chromium system, resulting in the capture of ~4800 myeloid cells across both TNC conditions combined, with a mean read depth of 61-71k reads per cell (**Appendix Figure 3.1**).

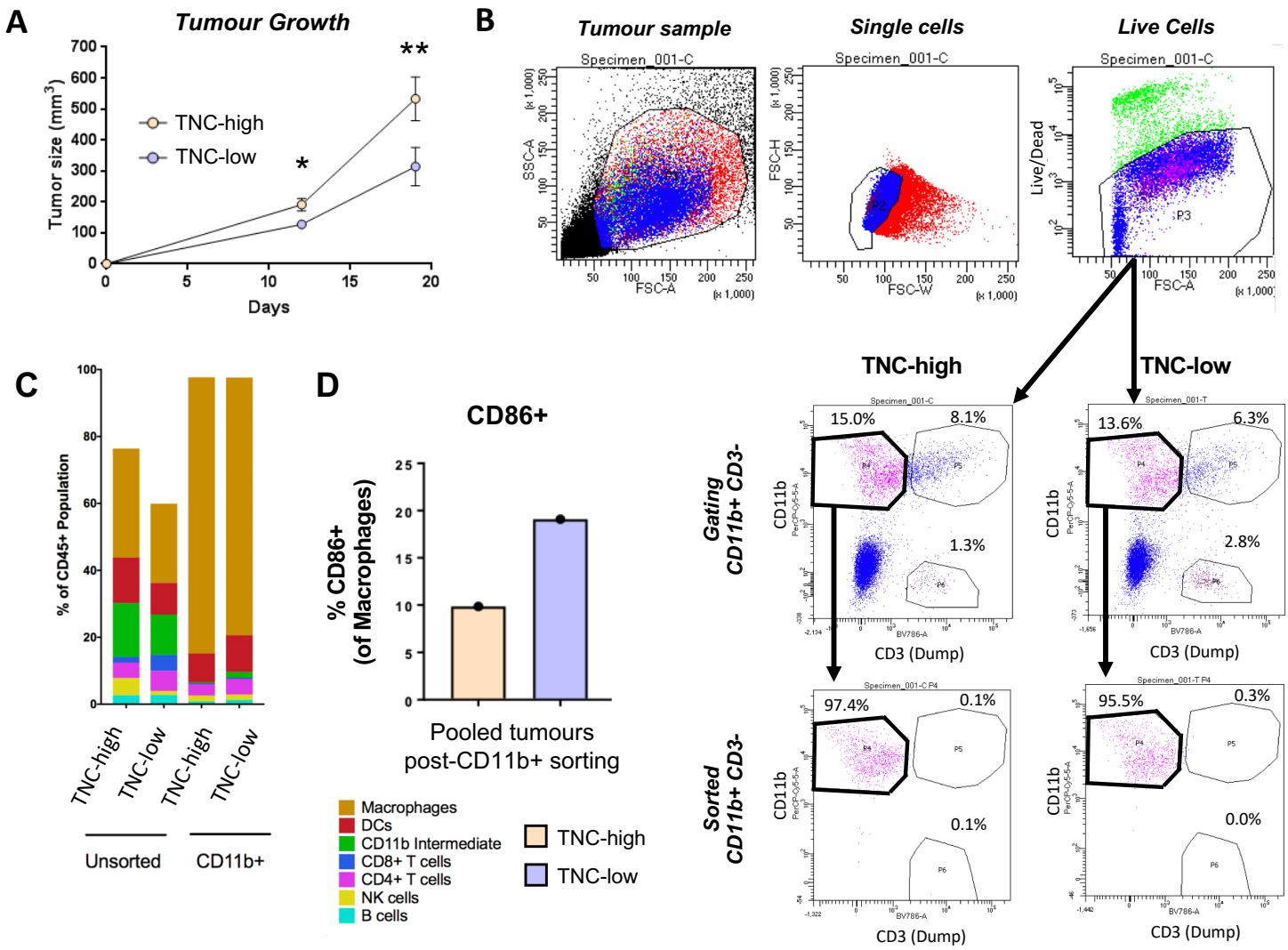


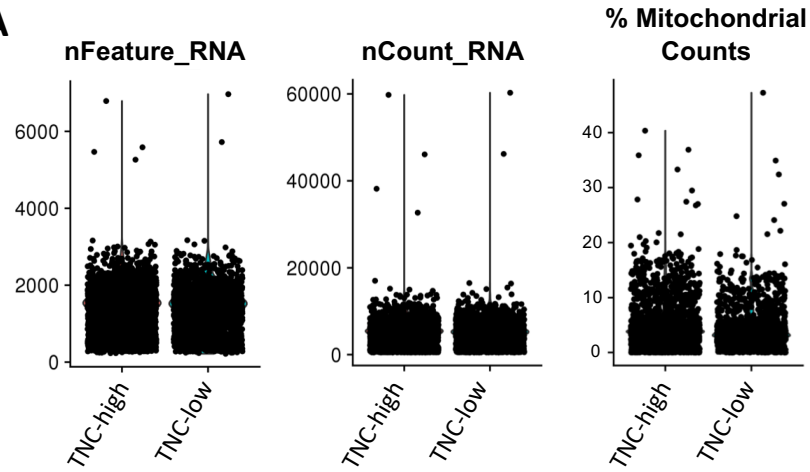
Figure 3.1 | Flow cytometric analysis of immune cells in the microenvironment of TNC-high or TNC-low mouse breast tumours processed for CD11b$^+$ single-cell RNAseq. A total of 6 TNC-high and 6 TNC-low NT193 mammary tumours were orthotopically engrafted into syngeneic immunocompetent wild-type female FVB mice. A subset of 3 TNC-high and 3 TNC-low representative tumours were selected based on the quality of their engraftment, harvested and pooled with stoichiometrically equal numbers of cells per tumour, prior to myeloid cell isolation by CD11b$^+$ FACS-sorting. **(A)** Growth curves of TNC-high and TNC-low tumours (n=3 for each TNC condition, pooled). Error bars represent the standard error of the mean. Statistical significance was determined by performance of a two-way repeated-measures ANOVA on the mean of each TNC condition per timepoint, assuming no sphericity (applying the Geisser-Greenhouse correction) and adjusting for multiple comparisons using the Sidak method. * $P < 0.05$; ** $P < 0.01$. **(B)** Gating strategy for isolation of tumour-infiltrating myeloid cells by CD11b$^+$ FACS sorting, including a CD3 dump channel to exclude contaminating T cells. **(C)** Flow cytometric analysis of the immune infiltrate in TNC-high versus TNC-low tumours before and after CD11b$^+$ FACS sorting. Data is presented as a percentage of the CD45$^+$ population. **(D)** Flow cytometric validation of TNC-dependent CD86$^+$ macrophage phenotype within these samples. Acknowledgment: All FACS sorting was performed by Dr. Jonathan Webber, and all remaining procedures and analysis were performed by Dr. Claire Deligne, both at the Kennedy Institute of Rheumatology (University of Oxford). Abbreviations: FACS = Fluorescence-Activated Cell Sorting.

3.1.3 | Quality control

The quality of sequenced cells was assessed using three commonly used quality-control metrics: the number of genes (features), unique molecular identifiers (UMIs, or Counts), and the percentage of mitochondrial reads, per cell (**Figure 3.2A**). These metrics were used to inform a filtering strategy indicated by dashed lines for the removal of low-quality cells or doublets prior to further analysis – namely, excluding cells: (i) without 500-3500 genes expressed, (ii) with more than 15,000 UMIs, or (iii) with more than 10% of reads mapping to mitochondrial genes (**Figure 3.2B**). The number of features (genes) and counts (UMIs) was consistent between each TNC condition; however, the percentage of mitochondrial reads expressed within TNC-low tumours was considerably lower than TNC-high tumours. To minimise any downstream bias, differences in the number of features and the percentage of mitochondrial reads were both regressed out according to a linear model prior to further analysis. To visualise how these quality-control metrics potentially affected the global structure of the data, a non-linear dimensionality reduction method known as UMAP was employed, which represents arguably the best approximation of the underlying topology, preserving both local similarity and global structure (276, 277). UMAP analysis demonstrated marked heterogeneity in the number of genes, UMIs and the percentage of mitochondrial counts between distinct groups of cells (**Figure 3.2C**), potentially owing to either technical or biological factors. After applying the quality control filters, myeloid cells from TNC-high tumours were randomly downsampled to equal the number of cells from TNC-low tumours (2012 cells per TNC condition) (**Figure 3.2D**), in order to simplify downstream comparisons.

From this point onwards, analytical procedures were performed on aggregated data from both TNC conditions combined, with the exception of **Section 3.1.7** which considers the TNC-dependent effects.

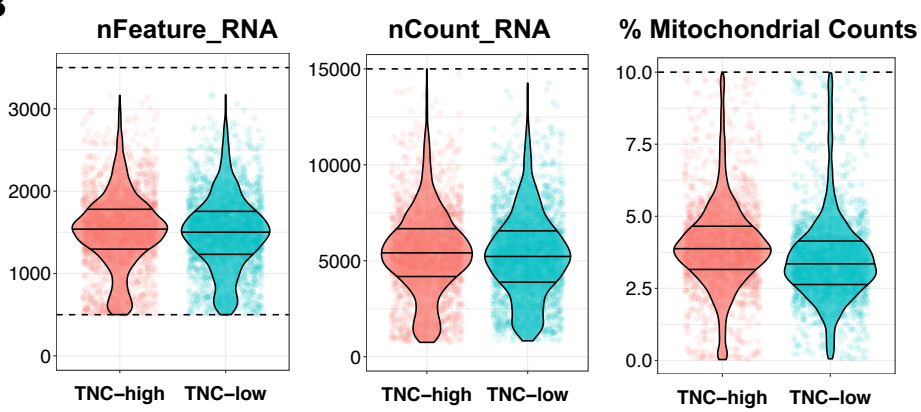
A



D

	TNC-high	TNC-low
Input cell number	3148	2304
After QC filtering	2566	2012
After downsampling	2012	2012

B



Filtering parameter	Value
% of mitochondrial RNA counts permitted	0-10%
Number of genes per cell permitted	500-3500
Minimum number of cells a gene must be expressed in	3
Latent variables regressed out	nCount_RNA %MitoCounts

C

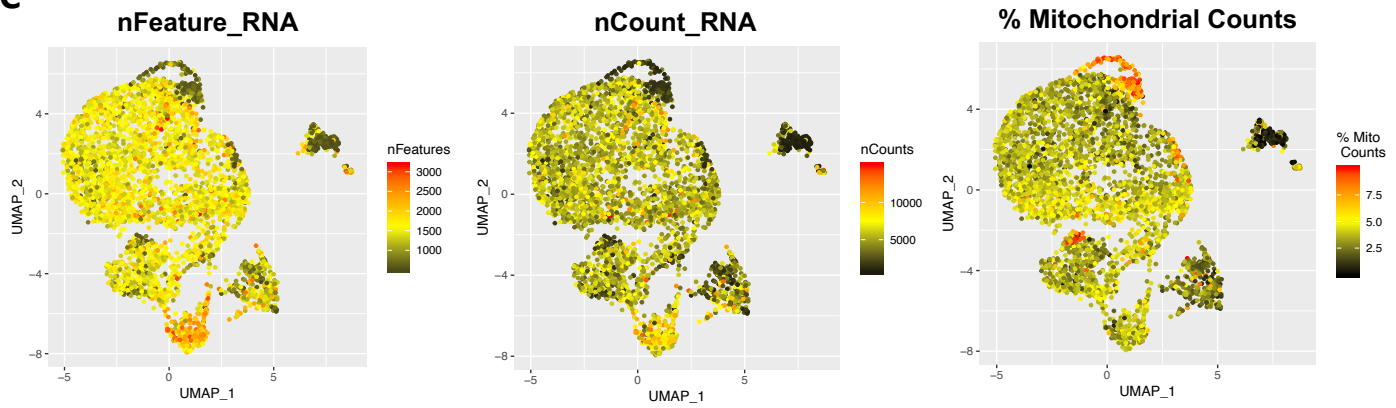


Figure 3.2 | Quality control & data filtering of myeloid cells captured by scRNAseq (A) Initial cell quality assessment using recommended QC metrics, showing the number of features (genes), number of counts (unique molecular identifiers, UMIs) and the percentage of reads mapping to mitochondrial genes. **(B)** Cell quality assessment after application of a filtering strategy to remove low quality cells or doublets. Horizontal lines on violin plots represent the upper quartile, median, and lower quartile, in descending order, respectively **(C)** UMAP projection, representing the similarity of cells within a global structure, superimposed with quality metrics to visualize their distribution. This plot represents all cells aggregated from both TNC conditions. For visualisation purposes, outlier cells expressing greater than 3500 genes or 15,000 counts were excluded. **(D)** Summary of cell numbers per TNC condition, after successive rounds of quality control using the aforementioned filtering strategy, followed by random downsampling.

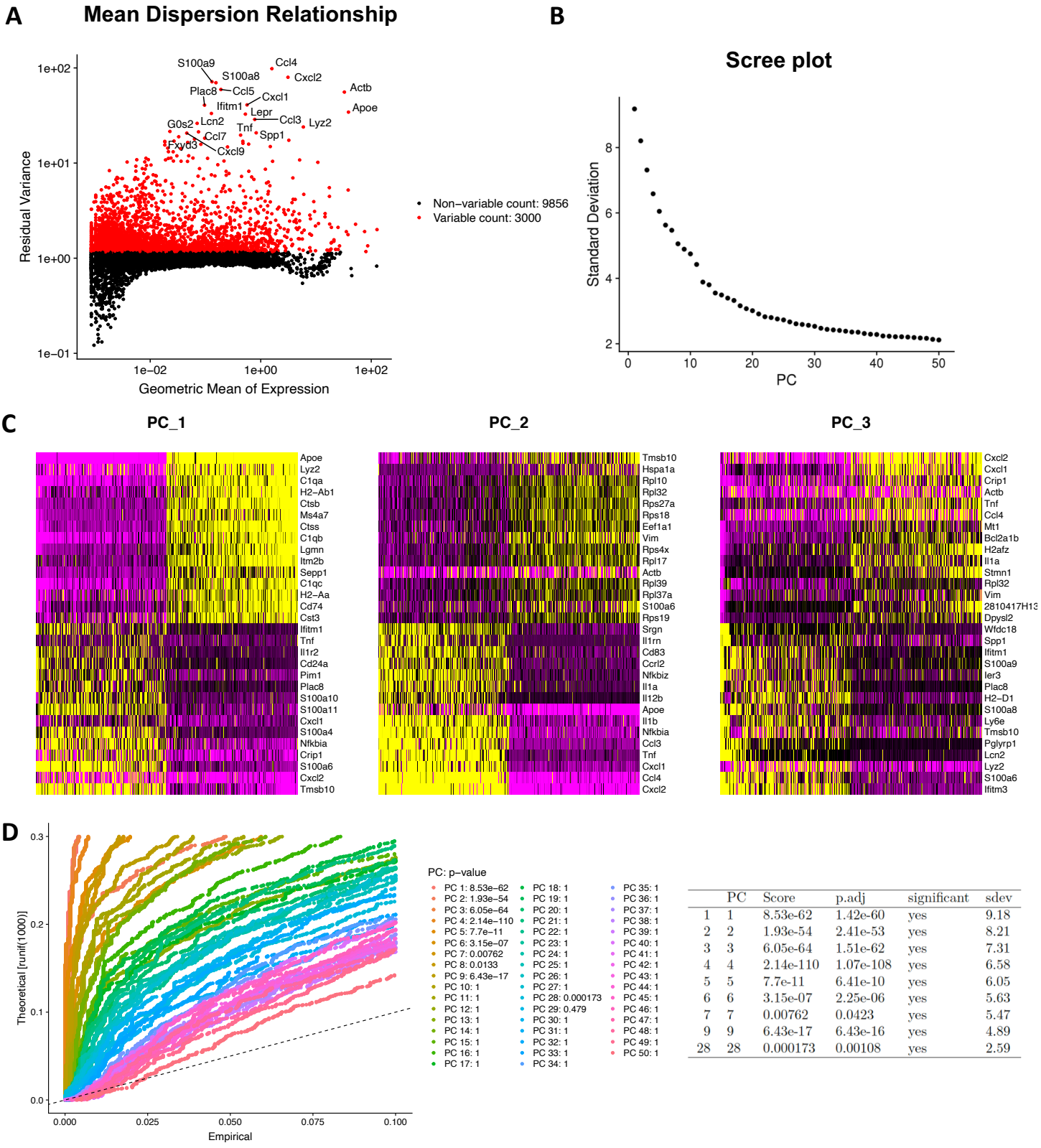
3.1.4 | Feature selection and dimensionality reduction

High-dimensional data, such as single-cell RNAseq data, suffers from the "curse of dimensionality", coined by Richard Bellman in 1957 (297), which describes the computational challenge of extracting representative information from high-dimensional space. With the large numbers of zero counts inherent to scRNAseq data (ie a ‘sparse matrix’ where zero counts represent missing expression values as well as truly negative values), the genes detected will not necessarily be informative of the overall data structure. Two main methods were employed to mitigate this.

First, data normalisation and variance stabilisation were performed using the SCTransform method (271), which optimises the removal of technical variation while retaining biological heterogeneity. To improve biological signal in downstream analyses (298), only the top 3000 genes exhibiting the highest cell-to-cell variation were selected (ie the most highly variable genes) (**Figure 3.3A**). This analysis also enables visualisation of the expression and dispersion characteristics of the genes with highest residual variance, highlighting the extremely high expression of genes such as *Apoe* and *Actb*.

Second, to further improve the signal-to-noise ratio and ease the computational burden on downstream analyses, linear dimensionality reduction was performed using principal component analysis (PCA), seeking to capture the underlying structure of the data in a minimal number of dimensions, using the top 3000 highly variable genes as an input. The top principal components (PCs), explaining the majority of (linear) variation in the data, thus represent a robust compression of the dataset. However, selection of an appropriate number of PCs requires further assessment. To determine the dimensionality of the data, a scree plot was initially used to (subjectively) identify a reasonable number of dimensions by selection of the "elbow" of the curve, where the marginal variance explained by PCs begins to diminish

considerably (**Figure 3.3B**). Following this, an exploration of the biological relevance of these sources of heterogeneity was achieved by plotting the top cells and genes according to their PCA scores. For example, the first 3 PCs contain strong enrichment of cytokines, chemokines, MHC molecules, and complement, aligning with biological expectations of myeloid cell transcriptomes (**Figure 3.3C, Appendix Figure 3.2**). Finally, this information was then compared with a quantitative unsupervised analysis using the JackStraw random permutation procedure (272, 273) which identifies 'significant' PCs, strongly enriching low *P* value genes (**Figure 3.3D**). A total of 9 PCs were considered significant ($P \leq 0.05$) after multiple-testing correction, and the top cells and genes of these PCs are visualised in **Appendix Figure 3.2**.



	PC	Score	p.adj	significant	sdev
1	1	8.53e-62	1.42e-60	yes	9.18
2	2	1.93e-54	2.41e-53	yes	8.21
3	3	6.05e-64	1.51e-62	yes	7.31
4	4	2.14e-110	1.07e-108	yes	6.58
5	5	7.7e-11	6.41e-10	yes	6.05
6	6	3.15e-07	2.25e-06	yes	5.63
7	7	0.00762	0.0423	yes	5.47
9	9	6.43e-17	6.43e-16	yes	4.89
28	28	0.000173	0.00108	yes	2.59

Figure 3.3 | Feature selection, linear dimensionality reduction, & Jackstraw analysis. (A) The mean dispersion relationship between gene expression and residual variance across all cells, calculated with the SCTransform method (271), highlighting the top 3000 most highly variable genes, the top 20 of which are labelled. **(B)** Scree plot visualizing the eigenvalues (percentage of variance explained, by standard deviation) for each principal component (PC). **(C)** Heatmaps of the first three PCs, visualizing the top 15 highly and top 15 lowly expressed genes per PC. Only the top 500 cells are shown per PC. **(D)** The Jackstraw permutation procedure (272, 273) was used to quantify the statistical significance of the first 50 PCs. PCs with an adjusted P value of ≤0.05 after Benjamini-Hochberg correction were considered significant.

3.1.5 | Clustering and visualisation

Graph-based clustering methods and 'modularity optimisation' algorithms are routinely used to effectively partition single-cell transcriptome profiles into meaningfully connected communities of similar cells, known as clusters.

First, a *k*-nearest neighbour (*k*-NN) graph (299, 300) was constructed on all cells from both TNC conditions, based on the Euclidean distance in PCA space of all significant PCs, with edges drawn between cells with similar gene expression patterns. The edge weights were then refined between connected cells based on the shared overlap in their local neighbourhoods (the Jaccard similarity), resulting in a Shared-Nearest Neighbours (SNN) (301) representation of the data (**Figure 3.4A**). While this SNN graph is visualised with a tSNE projection, it is important to note that tSNE embeddings were not used to cluster the data. Next, densely connected regions in the SNN graph were detected using a modularity optimisation algorithm called the Smart Local Moving (SLM) algorithm (302, 303), which iteratively groups cells together to partition highly interconnected clusters. For optimisation purposes, a resolution parameter can be set to alter the granularity of clustering. To aid the assessment of clustering quality over multiple resolutions, a cluster resolution analysis was performed using Clustree (278) (**Figure 3.4B**), alongside inspection of cluster-defining marker heatmaps at each resolution (**Appendix Figure 3.3**). This demonstrated an optimal segregation of clusters at a resolution of 0.3 – as lower resolutions either generated clusters with a low frequency of marker gene expression (eg cluster 0 at 0.1 resolution), or failed to discriminate a highly differential but small cluster (eg cluster 7 at 0.3 resolution is missing at 0.2 resolution), while higher resolutions consistently over-clustered the data into non-exclusive overlapping populations.

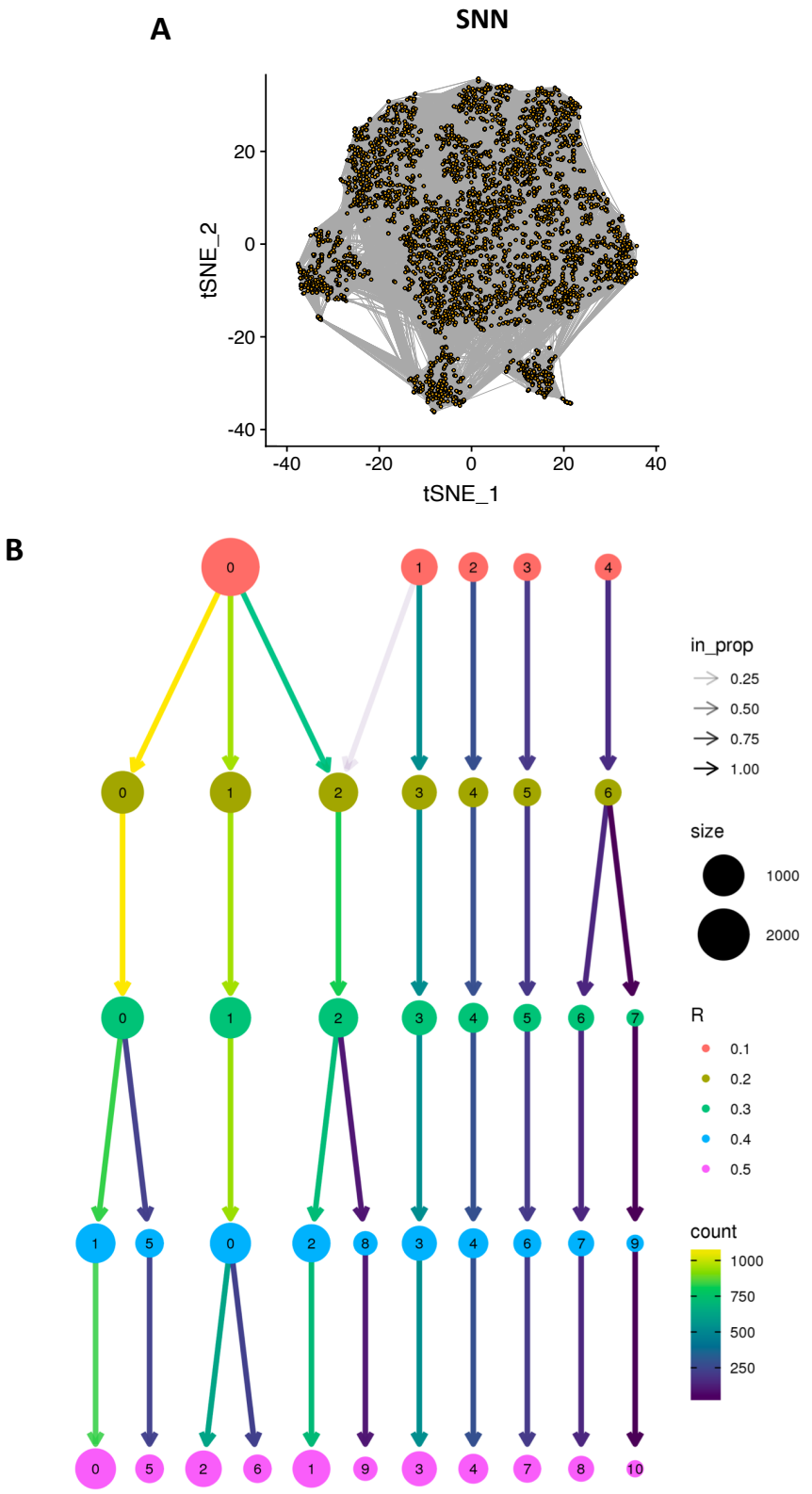


Figure 3.4 | Community detection and clustering optimization (A) tSNE projection of a Shared Nearest Neighbour graph, with edges drawn between cells based on the shared overlap in their local neighbourhoods from a K-nearest neighbour graph (KNN). The KNN (k=20) was constructed based on the Euclidean distance between cells in PCA space, using all significant PCs (n=9) determined by the JackStraw procedure (**Figure 3.3**). For the KNN graph, this plot represents all cells aggregated from both conditions. tSNE perplexity was set to 30, and the maximum number of iterations was 10,000. tSNE was used only for visualization and not for clustering. **(B)** Cluster resolution analysis with Clustree (278), visualising how the composition of clusters changes at different levels of resolution ('R') using a phylogenetic tree. This tree relates the 'average' cell from each cluster in gene expression space, and shows the total size of clusters ('size'), the proportion of cells from the incoming cluster node ('in_prop'), and their absolute cell number ('count'). Clustering was performed using the Smart Local Moving (SLM) algorithm. Abbreviations: PCA = Principal Components Analysis; tSNE = t-distributed Stochastic Neighbour Embedding.

Cluster assignments were additionally visualised by overlaying onto two non-linear dimensionality reduction graphs: tSNE (**Figure 3.5A**) and UMAP (**Figure 3.5B**). Both visualisation tools successfully partitioned the SLM clusters in low-dimensional space, with the exception of Cluster 0 in the tSNE projection, which is divided into two separate communities. tSNE captures local similarity at the expense of global structure, unlike UMAP which is superior at preserving both, and thus the UMAP representation is likely to be a more representative approximation of the underlying topology (276). This UMAP data indicates relatively high levels of similarity between clusters 0, 1 and 2, whereas 3, 4 and 5 are increasingly segregated, and clusters 6 and 7 are highly dissimilar from the rest of the population (**Figure 3.5B**).

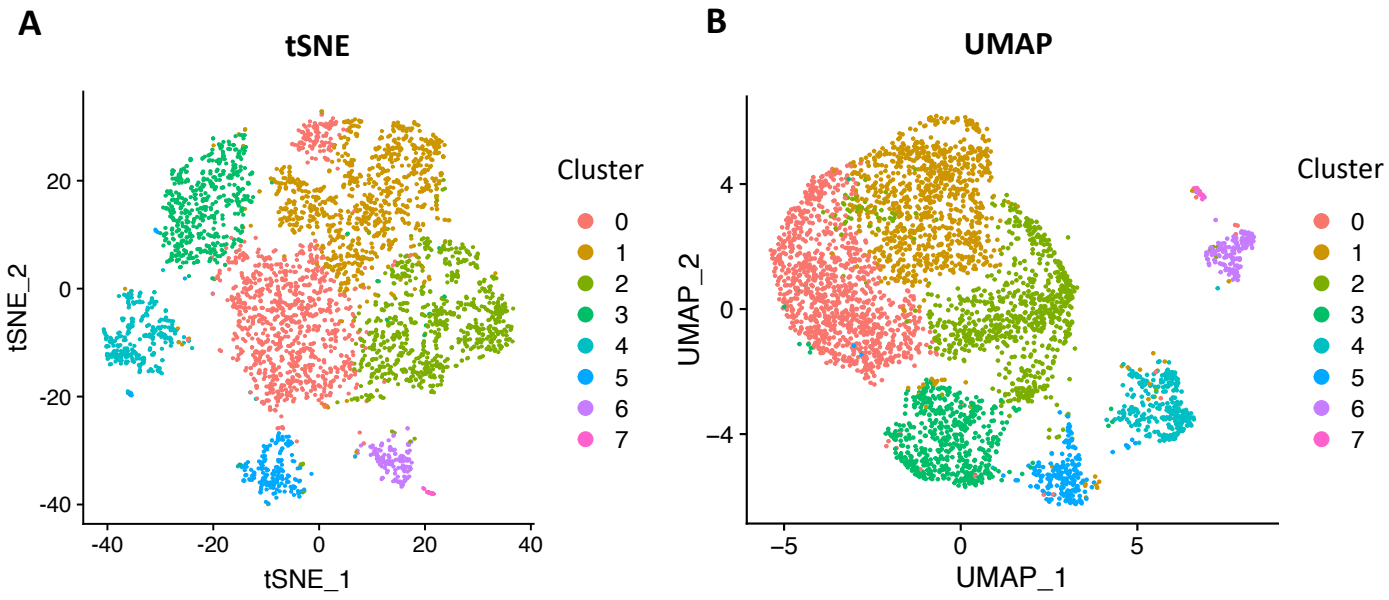


Figure 3.5 | Cluster visualisation by non-linear dimensionality reduction. (A) tSNE and **(B)** UMAP projections of Shared-Nearest-Neighbour graph clustering, performed using the SLM algorithm at an optimised resolution of 0.3, determined in part by prior Clustree resolution analysis. Each plot represents an aggregation of all cells from both TNC-high and TNC-low tumours. tSNE perplexity was set to 30, and the maximum number of iterations was 10,000. Cluster labels are assigned in decreasing order of cluster size, with the first cluster (cluster 0) being the largest. Abbreviations: tSNE = t-distributed Stochastic Neighbour Embedding; UMAP = Uniform Manifold Approximation and Projection.

Confirming successful partitioning using SLM at a resolution of 0.3, differential expression analysis indicated clear demarcation of eight clusters (labelled in order of the number of cells per cluster, with the largest cluster first). These clusters featured many highly discriminatory and non-overlapping cluster-defining marker genes (**Figure 3.6**). Cluster 0 was unusual in being defined by more modestly upregulated genes compared to other clusters.

Finally, the absolute number and proportion of cells in each cluster was enumerated (**Appendix Figure 3.4A-B**), alongside the distribution of QC metrics per cluster, which indicated that cluster 6 was of a generally lower quality with fewer detected genes and UMIs, though it still had many positively differentially-expressed genes (DEGs), possibly indicating genuine biological variation (**Appendix Figure 3.4C-E**).

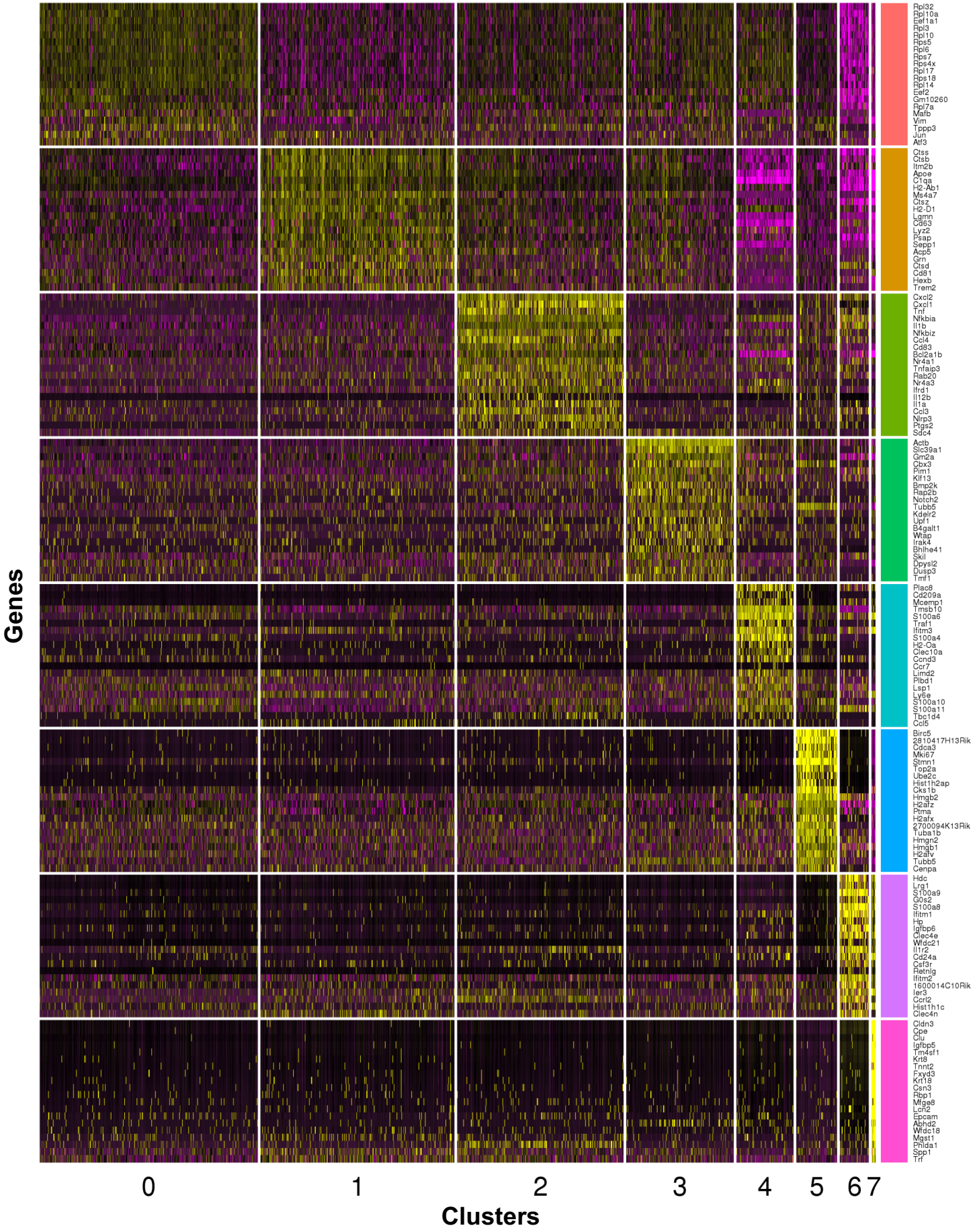
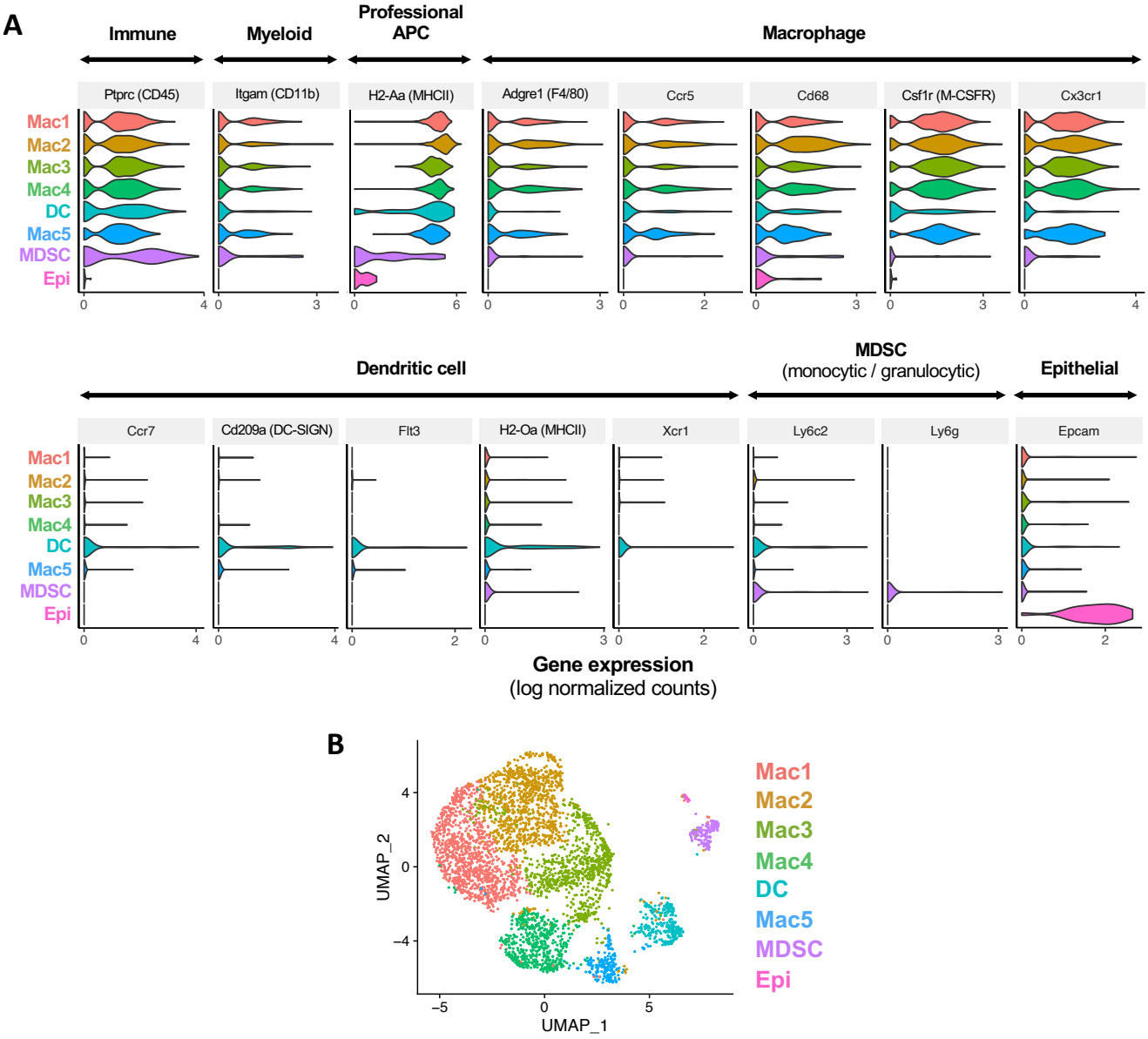


Figure 3.6 | Identification of cluster-discriminating marker genes. Differential expression analysis between clusters was performed on aggregated data (both TNC-high and TNC-low) using the Wilcox test. Testing was limited to genes: with a log fold-change of >0.25, detected in at least 10% of cells, and conserved between TNC-high and TNC-low conditions. The top 20 differentially expressed marker genes per cluster (ie rows), defined by their adjusted P value (Benjamini-Hochberg correction) are displayed by heatmap. A total of 4024 cells (ie columns) selected after QC and downsampling were included (2012 from each condition).

3.1.6 | Characterisation of cluster phenotypes

To assign top-level cell type annotations to each of the eight clusters identified, a panel of well-validated immune differentiation markers was assessed with violin plots (**Figure 3.7A**), revealing the presence of five macrophage clusters (**Mac1-5)**, one dendritic cell cluster (**DC**), one putative myeloid-derived suppressor cell cluster (**MDSC**), as well as one (very small) cluster of contaminating epithelial cells (**Epi**) which were excluded from subsequent analyses. Macrophage populations were identified by their positivity for five commonly described macrophage markers (*Adgre1*, *Csf1r*, *Cx3cr1*, *Ccr5* and *Cd68*), which were all in reasonable agreement. This indicated that macrophages were the predominant myeloid subset within breast tumours, corroborating earlier flow cytometry data (**Figure 3.1C)** and previous work (264). The annotation of **DC** was made on the basis of their low but consistent expression of **DC**-specific or highly **DC**-enriched markers (including *Ccr7*, *Cd209a*, *Flt3*, *H2-Oa*, *Xcr1*), all of which localised to discrete sub-cluster locations when visualised by UMAP (**Figure 3.7A, 3.7C**), indicating considerable **DC** heterogeneity. It should also be noted that low *Ly6c2* and *Csf1r* expression was observed in the **DC** cluster, alluding to their potential monocytic origin, and perhaps indicating the presence of a small number of monocytes (**Figure 3.7A, 3.7C**). The putative assignment of **MDSC** was made on the basis of (albeit low) expression of *Ly6g* (granulocytic) and *Ly6c2* (monocytic) markers, and striking downregulation of MHC class II (*H2-Aa*) (**Figure 3.7A**) – a suppressive phenotype that MDSC have long been known to adopt (304).



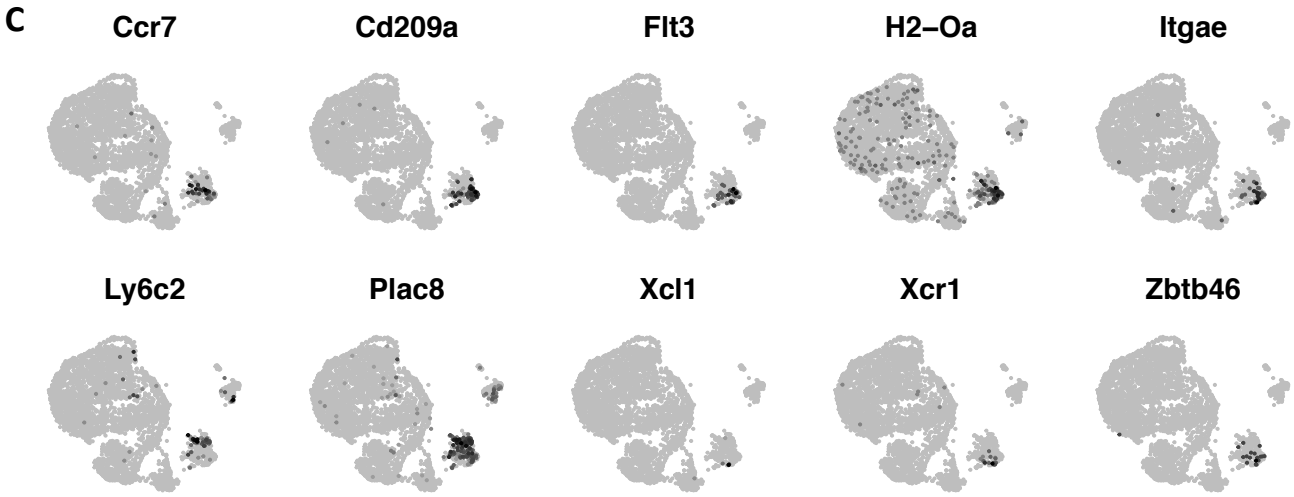


Figure 3.7 | Cell type assignment. (A) Violin plots visualizing the expression of well-described immune differentiation markers were used to assign top-level cell type annotations to all clusters. Gene names with a better-known protein alias are shown in parentheses. **(B)** UMAP projection of all cells, coloured by their cell type annotation. (C) UMAPs visualizing the expression distribution of common Dendritic Cell (DC) marker genes, and the monocytic marker, *Ly6c2*. Abbreviations: APC = antigen-presenting cell; MDSC = myeloid-derived suppressor cell, UMAP = Uniform Manifold Approximation and Projection.

To characterize the biological functions and phenotypes of each cluster, several strategies were undertaken. First, to initially understand some of the key genes driving the differential biology of each cluster, the distribution of top cluster-discriminating markers was visualized by violin plot and UMAP projection, shown in **Figure 3.8**. Second, to quantitate the biological functions of all clusters in an unbiased way, geneset over-representation analyses were performed using geneset annotation databases including KEGG, GO, Canonical Pathways and others (**Figure 3.9, Appendix Figures 3.4-3.5**). Third, to compare cluster signatures with previously established immune phenotypes, expression of well-described immune phenotypic markers was assessed (**Figure 3.10**).

Mac1 modestly but significantly upregulates a range of genes relating to ribosomal translation (range of average log-fold change: 0.26 to 0.37, adjusted *P* values: 1e-17 to 2e-55), including ribosomal protein genes (*Rpl10a*, *Rpl3*, *Rpl6*, *Rpl32*), as well as translation elongation factors (*Eef2*, *Eef1a1*) (**Figures 3.6, 3.8A & 3.9**) **Mac1** also upregulates the transcription factor *Mafb*, which might support terminal macrophage differentiation (305), and promote an anti-inflammatory phenotype (306). However, given that **Mac1** is the largest TAM subset by cell number (**Appendix Figure 3.4A-B**), has the fewest positively-discriminating marker genes (**Appendix Figure 3.4E**), and downregulates MHC class I (*H2-K1*) (**Figure 3.10**), this cluster might possibly represent a reservoir of naïve-like ‘unpolarised’ macrophages.

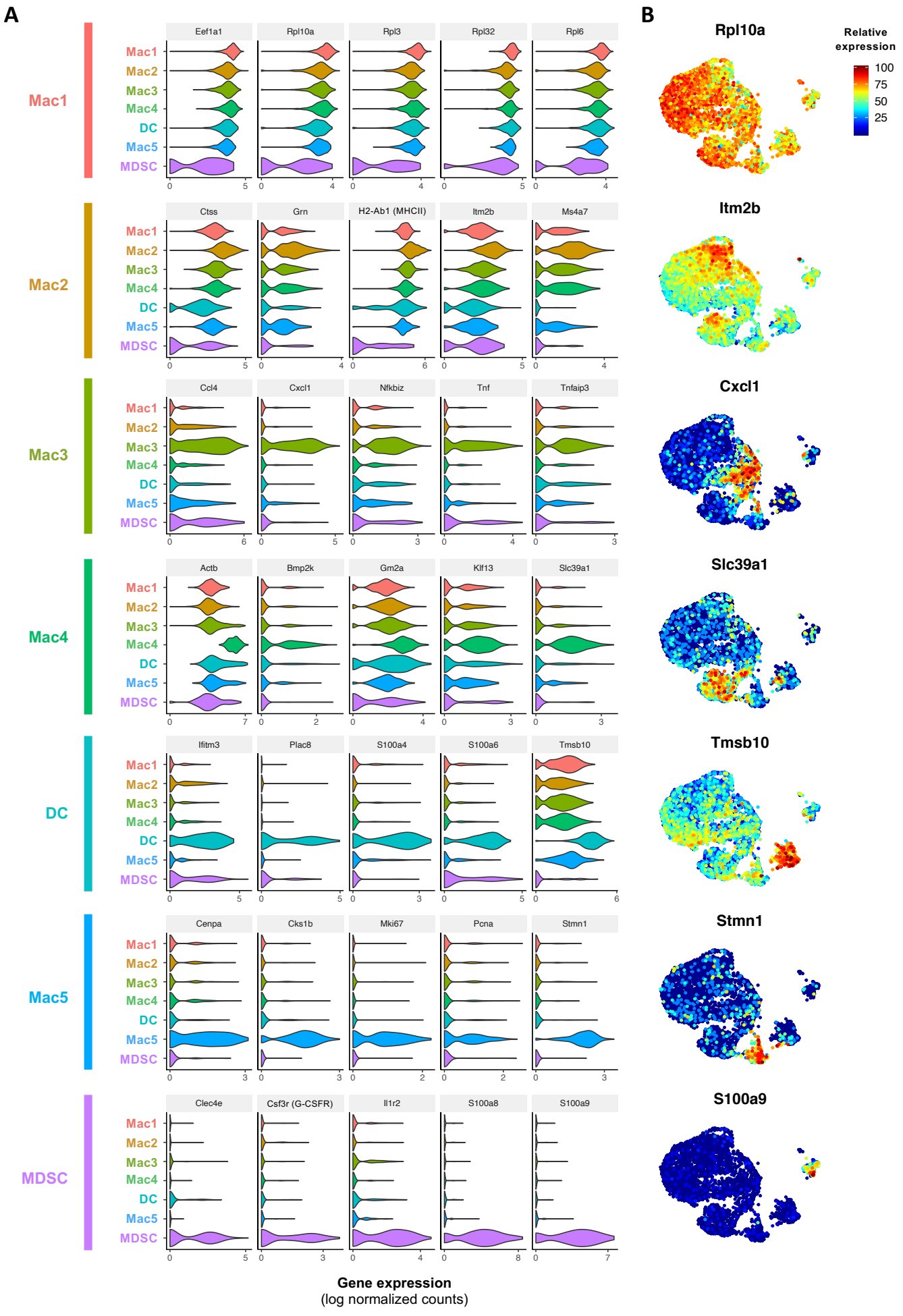


Figure 3.8 | Visualization of cluster-discriminating marker genes. The distribution of selected cluster-discriminating marker genes was visualised by violin plot **(A)** or UMAP projection **(B)** for all clusters except contaminating epithelial cells, which were excluded. Data are aggregated from both TNC-high & TNC-low conditions.

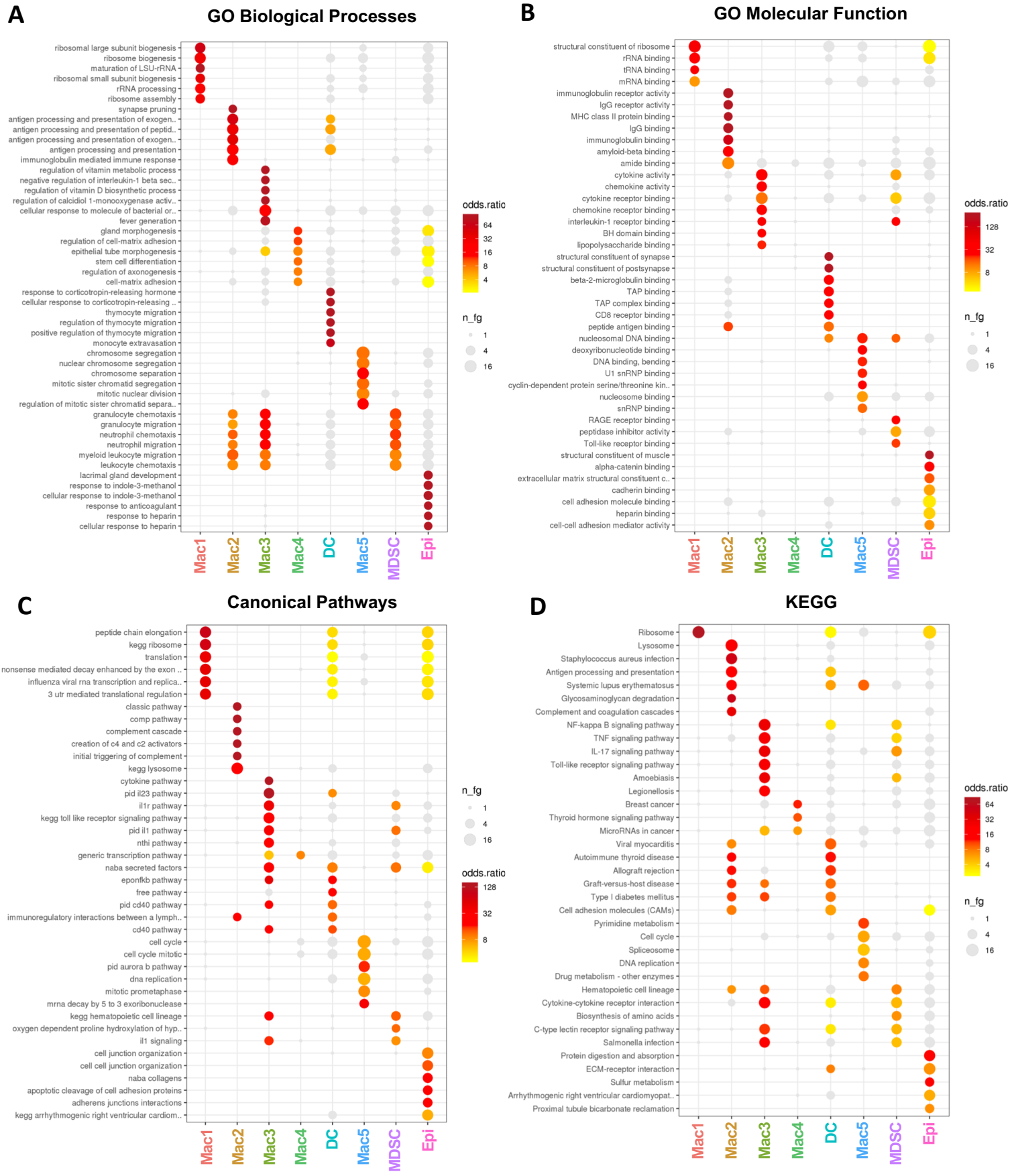


Figure 3.9 | Phenotypic characterization of clusters by geneset over-representation analysis. All genes significantly differentially expressed between clusters (BH–adjusted *P* value < 0.05) were assessed for their over-representation of the following functionally annotated genesets: **(A)** GO Biological processes **(B)** GO Molecular Function **(C)** Canonical Pathways **(D)** KEGG pathways. The minimum odds ratio for inclusion of a geneset was 1.5, and the minimum number of over-represented foreground genes ('n_fg') was 3. Abbreviations: BH = Benjamini-Hochberg correction; GO = Gene ontology; KEGG = Kyoto Encyclopaedia of Genes and Genomes; MSigDB = Molecular Signatures Database.

Mac2 are principally characterized by enhanced antigen presentation (**Figure 3.9A**), through their upregulation molecules related to MHC Class I (*H2-D1*, *H2-K1*, *B2m*), MHC Class II (*H2-Aa*, *H2-Ab1*, *Cd74*), and cathepsin proteases for antigen processing (*Ctsb*, *Ctss*, *Ctsz*) (**Figures 3.8A & 3.10**). Significant enrichment of complement-related pathways due to upregulation of *C1qa*, *C1qb* and *C1qc* was also observed (**Figure 3.9C, Appendix Figure 3.4B**), alongside modest upregulation of pro-phagocytic genes such as the high affinity Fc-gamma receptor (*Fcgr1*), a subunit of NADPH oxidase involved in ROS generation (*Cyba*), and a key myeloid receptor that regulates phagocytosis (*Trem2*). Together, these enrichments indicate that **Mac2** is of a predominantly anti-tumoural phenotype. However, the potent anti-inflammatory molecule, *Apoe*, is also elevated in **Mac2**, perhaps indicating a counterbalance to an otherwise highly inflammatory state.

Mac3 highly upregulates a range of chemokines (*Ccl3*, *Ccl4*, *Cxcl1*, *Cxcl2*) and cytokines (*Tnf*, *Il1a*, *Il1b*), for which it is a strong, primary or even sole source within the myeloid compartment (**Figure 3.10**), enriching pathways related to TLR, NFkB, IL1, TNF, IL17 and IL23 signalling (**Figure 3.9, Appendix Figure 3.5**). The chemotactic functions of *Cxcl1*, *Cxcl2* and *Ccl3* converge on the recruitment of polymorphonuclear cells (PMN) such as neutrophils, and, together with *Ccl4*, these genes all have well described roles in promoting breast tumour invasion or metastasis (251, 307-311). **Mac3** also strongly upregulates anti-inflammatory mediators including *Tnfaip3*, an inhibitor of TLR4 signaling in macrophages (312) and *Nfkbiz*, which is essential for macrophage expression of *Il10* (313) (**Figure 3.10**), together potentially indicating a degree of regulatory feedback within **Mac3** to dampen inflammatory signaling. Specific induction of a large number of genes implicated in pattern recognition receptor (PRR) signalling was also observed (**Appendix Figure 3.7**) – for instance, the inflammasome subunit, *Nlrp3*, is likely responsible for the increased expression of *Il1b* in **Mac3** (**Figure 3.10**), which is a cytokine crucial for pathogenic intra-tumoural Th17 differentiation (314).



Figure 3.10 | Phenotypic characterization of myeloid subsets by expression analysis of known immune genes. Violin plot expression distributions of well-described inflammatory mediators, arranged by shared function, ontology or gene family. Gene names with a better-known protein alias are shown in parentheses. APC = Antigen-Presenting Cell.

Mac4, which is partially segregated from other macrophage subsets in a UMAP visualisation (**Figure 3.7B**), strongly upregulates the cytoskeletal protein β-actin (*Actb*) and the transcription factor, *Klf13*, while also moderately upregulating morphogenesis-promoting gene, *Notch2*, and a kinase involved in positive regulation of notch signalling, *Bmp2k* (**Figures 3.6 & 3.8A, Appendix Figure 3.7**). Like **Mac1,** *Mafb* expression is similarly high in **Mac4**, indicating a differentiated macrophage state. However, **Mac4** has the fewest combined discriminatory marker genes of all myeloid cells (**Appendix Figure 3.4E**), only modestly enriched a handful of biological functions such as morphogenesis, cell-matrix adhesion, and generic transcription (**Figure 3.9**, **Appendix Figure 3.5**). Together, this phenotype is suggestive of a subset undergoing a potentially substantial change in morphology, but with few other biological phenotypes, perhaps suggesting that **Mac4** represents a transitionary state between **Mac1** and other TAM phenotypes.

Mac5 strongly upregulates proliferation genes (*Mki67*, *Pcna*) (**Figure 3.8A**), enriches cell cycle processes (**Figure 3.9, Appendix Figure 3.5**), and despite being the smallest TAM cluster by number, has the largest number of differentially expressed genes within all TAMs (**Appendix Figure 3.4E**), while being even further segregated from other myeloid subsets in a UMAP visualisation (**Figure 3.7B**). **Mac5** also upregulates the pleiotropic cytokine, osteopontin (*Spp1*), which has prognostic and diagnostic value for high-grade and disseminated breast cancer (315). Moreover, a receptor for SPP1, *Cd44*, is expressed in **Mac3** and **Mac5**, but is particularly downregulated in **Mac2** and **Mac4**, suggesting potentially differential susceptibilities to SPP1 signaling between macrophage subsets (**Figure 3.10**).

DC upregulates specific MHC Class II molecules such as *H2-Oa* and *H2-DMb2*, and to a lesser extent MHC Class I (*H2-D1*), whose upregulation was also shared by **Mac2** (**Figures 3.8 & 3.10)**. While **DC** enriched processes and functions relating to antigen presentation **(Figure 3.9A-B)**, this was less significant compared to **Mac2**, perhaps indicating decreased potency as an APC. **DC** might also mediate a pro-tumoural role through its high

and specific upregulation of *S100a4*, a gene known to promote metastasis in breast cancer (316), which can be suppressed by anti-S100A4 blockade (317). Conversely, modest induction of the cytokine, lymphotoxin beta (*Ltb*) (**Figure 3.10**) might indicate an anti-tumoural phenotype of **DC**, given that *Ltb*-producing DCs have previously been correlated with a clinically favourable outcome in breast cancer patients (318). Finally, **DC** strongly upregulates thymosin beta 10 (*Tmsb10*) (**Figure 3.8**), a regulator of actin-polymerization and cytoskeletal organization, perhaps reflecting a specific morphological phenotype of **DC**.

MDSC is the least abundant myeloid subset, and possesses the largest number of differentially expressed genes across all myeloid clusters (**Appendix Figure 3.4E**), consistent with its marked segregation from all other myeloid cells in a UMAP visualisation (**Figure 3.7B**). **MDSC** exclusively expresses very high quantities of the TLR4 ligand, calprotectin (*S100a8*/*S100a9*) (**Figure 3.8**), while strongly enriching pathways related to granulocyte chemotaxis, PRR signaling, and IL1 signalling (**Figure 3.9**). These upregulated genes represent a diverse repertoire of immune suppressive and pro-tumoural genes, some of which have well-described roles in breast cancer pathogenesis. These include potent upregulation of genes that contribute to angiogenesis (*Cxcr2*, *Lrg1* and *Il1b*), breast cancer metastasis (*Ccl3*, *Ccl4*, *Cxcl2*), APC co-inhibition (*Vsir*/VISTA), and regulation of IL1 signalling via high expression of the decoy receptor, *Il1r2*, and the receptor antagonist, *Il1rn*. The pleiotropic cytokine, *Mif*, whose overexpression in breast cancer cells is beneficial but whose presence extracellularly within the stroma is possibly pro-oncogenic (319), is expressed highly in all myeloid cells, but is particularly enriched in **MDSC** (**Figure 3.10**). Another key feature of the suppressive **MDSC** phenotype observed in this study is the specific downregulation of almost all classical MHC class I and II molecules (*H2-Aa*, *H2-Ab1*, *H2-DMb2*, *H2-DMa*, *H2-K1*), with the exception of *H2-D1* (Class I) whose expression was retained (**Figure 3.10**), possibly to prevent 'missing-self' destruction by natural killer (NK) cells. The **MDSC** cluster also displays universally decreased expression of ribosomal protein genes in comparison to other myeloid subsets (**Figure 3.6 & 3.8**), yet very high expression of many immune effector molecules

(**Figure 3.10**), perhaps indicating differences in ribosomal translation and energetic demand between cell types. Finally, a small number of highly upregulated genes in **MDSC** were also upregulated in **Mac3**, such as *Mif*, *Il1b, Ccl3*, *Ccl4*, and *Cxcl2* (**Figure 3.10**), alluding to some degree of functional overlap between these subsets.

While TAM populations exhibited significant heterogeneity, as evidenced by their clustering and differential signatures, they concomitantly shared a number of common features. For instance, TAMs universally expressed: *Tgfbr2* (which promotes breast cancer metastases when expressed in myeloid cells (320)), *Il10rb* (conferring susceptibility to tolerogenic IL-10 signalling), *Cxcl16* (which provides a strong chemotactic stimulus for breast cancer metastases (321)), *Tnfr2*, (an inhibitory Tnf receptor implicated in TNBC metastasis (322)), and extremely high levels of *Apoe* (which has potent anti-inflammatory effects in macrophages (323)) (**Figure 3.10**), thus together facilitating tumour progression. Furthermore, no myeloid subsets expressed appreciable quantities of the (typically) anti-tumoural cytokine, *Ifng*, though did express the receptor, *Ifngr*. Elevated expression of *Cd52*, an inhibitor of TLR/TNFR signalling in mononuclear phagocytes (324), indicated a pan-suppressive phenotype preserved across all myeloid cells (**Figure 3.10**). Interestingly, all TAM clusters also expressed low levels of *Ccr2*, most frequently in **Mac1** and **Mac5**, and least frequently in **Mac2** (**Figure 3.10**), alluding to the monocytic origin of TAMs in this model, with the differences in their relative frequencies perhaps explained by the recency of their differentiation.

The immunosuppressive phenotype common to all TAMs appeared to be mediated at multiple levels. TAMs upregulate canonical 'M2-like' genes such as *Cd204* and *Tgm2*, while also upregulating co-inhibitory checkpoint molecules such as TIM-3 (*Havcr2*), VISTA (*Vsir*), and Galectin-9 (*Lgals9*) (**Figure 3.10)**, all but one of which have already been associated with worsened prognosis or unresponsiveness to immunotherapy when expressed on TAMs across various cancers (325-328). Interestingly, however, these TAM subsets simultaneously expressed canonical M1-like genes such as *Irf5* and *Socs3*, alongside the co-stimulatory gene

Cd86, thus challenging the conventional separation of these phenotypes within the M1/M2 paradigm. Supporting this finding, pro-phagocytic (*Cyba*, *Fcgr1*, *Fcgr3*, *Trem2*, *Aif1*, *Pld3*) and anti-phagocytic (*Sirpa*, *Fcgr2b*, *Lilrb4a*) genes similarly co-existed within the same subsets. Taken together, these data demonstrate widespread mechanisms of immune tolerance within TAMs, alongside co-expression of conventional activation markers, yielding novel combinations of TAM phenotypes in the breast tumour microenvironment.

3.1.7 | TNC-dependent effects

Unsupervised clustering and differential expression analysis of all cells, aggregated across both TNC conditions, revealed 7 functionally distinct myeloid clusters, including 5 macrophage states, 1 DC cluster and 1 MDSC cluster. To understand the impact of a TNC-high versus TNC-low tumour microenvironment on infiltrating myeloid cells, a TNC-dependent analysis of population structure and transcriptional landscape was undertaken.

At the level of cellular abundance, superimposition of TNC-status onto a UMAP of all cells indicated a modest segregation of macrophages according to TNC-status, however every cluster had considerable representation in either TNC condition, indicating that this was not a problematic batch effect (**Figure 3.11A-B**). Quantifying the proportions of cells in each cluster according to TNC-status revealed several myeloid populations with notable TNC-dependent changes in abundance: **Mac1**, **Mac2**, **DC,** and **MDSC** (**Figure 3.11C**). In TNC-low tumours, expansion of **Mac2** and **DC** subsets was observed, both of which are highly enriched for antigen presentation genes (**Figures 3.9 & 3.10; Appendix Figure 3.6**) and might thus be considered to be of an anti-tumoural phenotype. In contrast, TNC-high tumours exhibited an expansion of both **Mac1** – a subset which upregulates ribosomal genes and downregulates MHC Class I (*H2-K1*) antigen presentation – alongside putative **MDSC**, which strongly enriches genes related to granulocyte chemotaxis (**Figure 3.9A**) such as *Cxcr2*, calprotectin (*S100a8*/*S100a9*), *Il1b*, and IL1 receptor antagonist (*Il1rn*) (**Figures 3.8 & 3.10**). A potential trend towards a modest ~11% expansion of cytokine/chemokine-producing **Mac3** was also observed in TNC-high tumours, alongside a reciprocally modest ~13% decrease in morphogenic/transitionary **Mac4** abundance (**Figure 3.11C**), though the significance of these effects cannot be established without independent (unpooled) biological replicates.

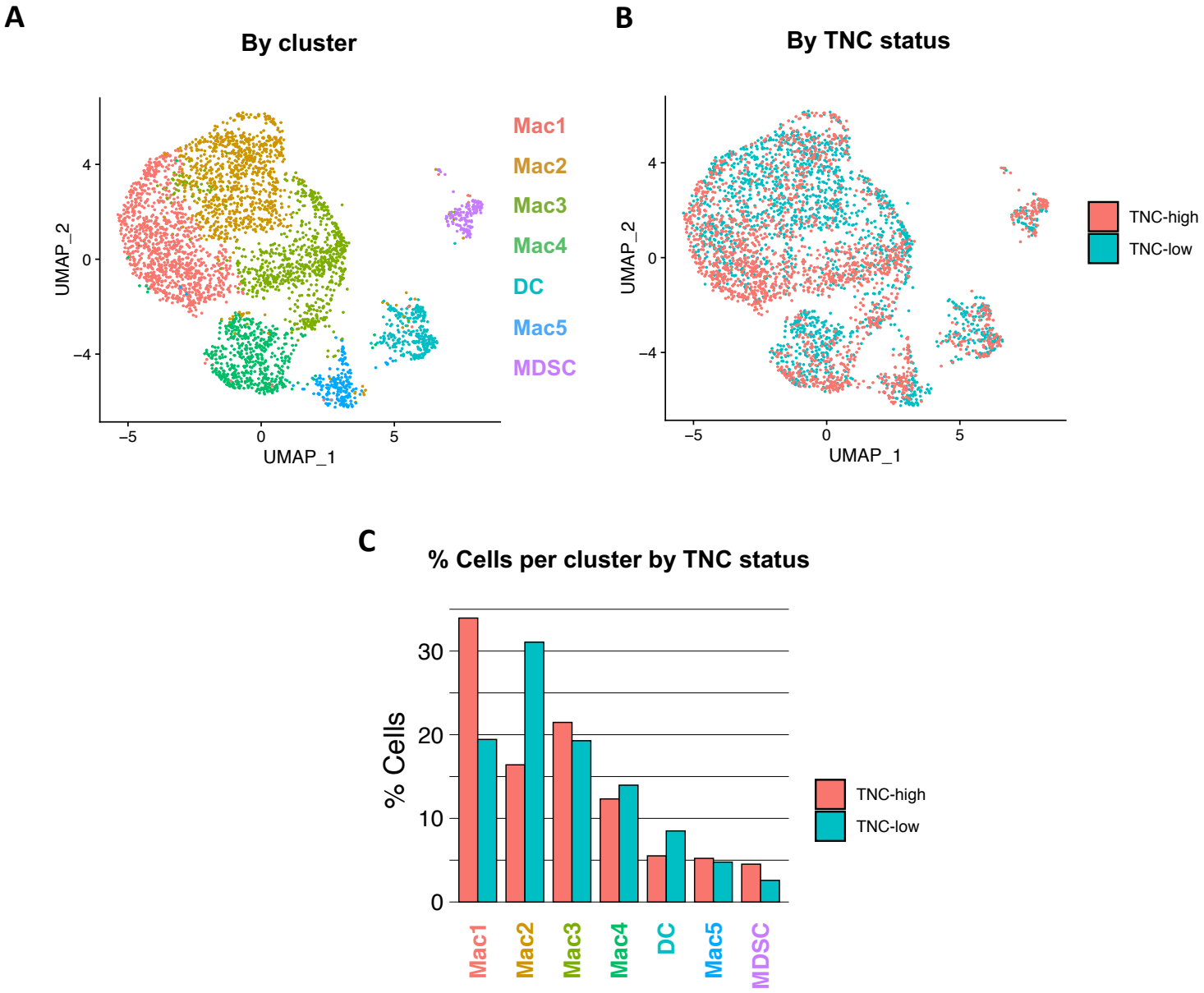


Figure 3.11 | TNC-dependent effects on myeloid population structure. UMAP projections of all cells (except epithelia) aggregated from both TNC conditions, coloured by either **(A)** cluster assignment or **(B)** TNC-status. **(C)** The percentage of cells within each cluster per TNC status.

To determine whether gene expression within clusters was altered by the presence of TNC, differential expression analysis between cells from TNC-high and TNC-low tumours was performed for each cluster (**Figure 3.12A**). The largest numbers of TNC-dependent differentially expressed genes were observed in **DC** and proliferating **Mac5**, with >100 in each, while other TAM subsets displayed markedly fewer, the least being **Mac1** with 32 genes, and the least overall in the myeloid compartment being **MDSC** with just 2 genes (**Figure 3.12B**). Among the genes regulated in a TNC-dependent manner, some have important roles in coordinating adaptive immunity, such as classical (*H2-K1*, *H2-D1*) and non-classical (*H2-Q4*, *H2-T23*) MHC class I molecules, and associated β2-microglobulin (*B2m*), which were upregulated in the majority of all macrophage clusters from TNC-low tumours (**Figure 3.12C**). Other key immune genes expanded in TNC-low tumours include *Ccl3* (in **Mac3**), *Ccl4* (in **Mac5**) and *Apoe* (in all macrophage clusters). Conversely, macrophage clusters from TNC-high tumours more frequently expressed *Ccr2*, though the actual level of *Ccr2* expression was broadly similar between TNC conditions for cells that did express it (**Figure 3.12C**). The most strong, consistent and significant TNC-dependent change in gene expression was displayed by *Uba52*, an unusual fusion protein of ubiquitin attached to ribosomal protein L40, which was downregulated in TNC-high tumours (**Figure 3.12A & 3.12C**).

However, in general, the extent of TNC-dependent differential gene expression was often modest, and comprised a large proportion of transcripts corresponding to pseudogenes (*Gm10260*, *Rps18-ps3*), long non-coding RNAs (*Gm10076*, *Gm2000*, *Gm26917*), and nucleus-sequestered proteins (*Ifi27l2a*, *Hist1h2ap*) (**Figure 3.12A**) some of which lacked conservation in humans. Together this limited the translational relevance and feasibility of TNC-dependent gene signatures for further investigation, and suggested that TNC exerts its effects predominantly at the level of subset abundance, with only minor changes in individual gene expression.

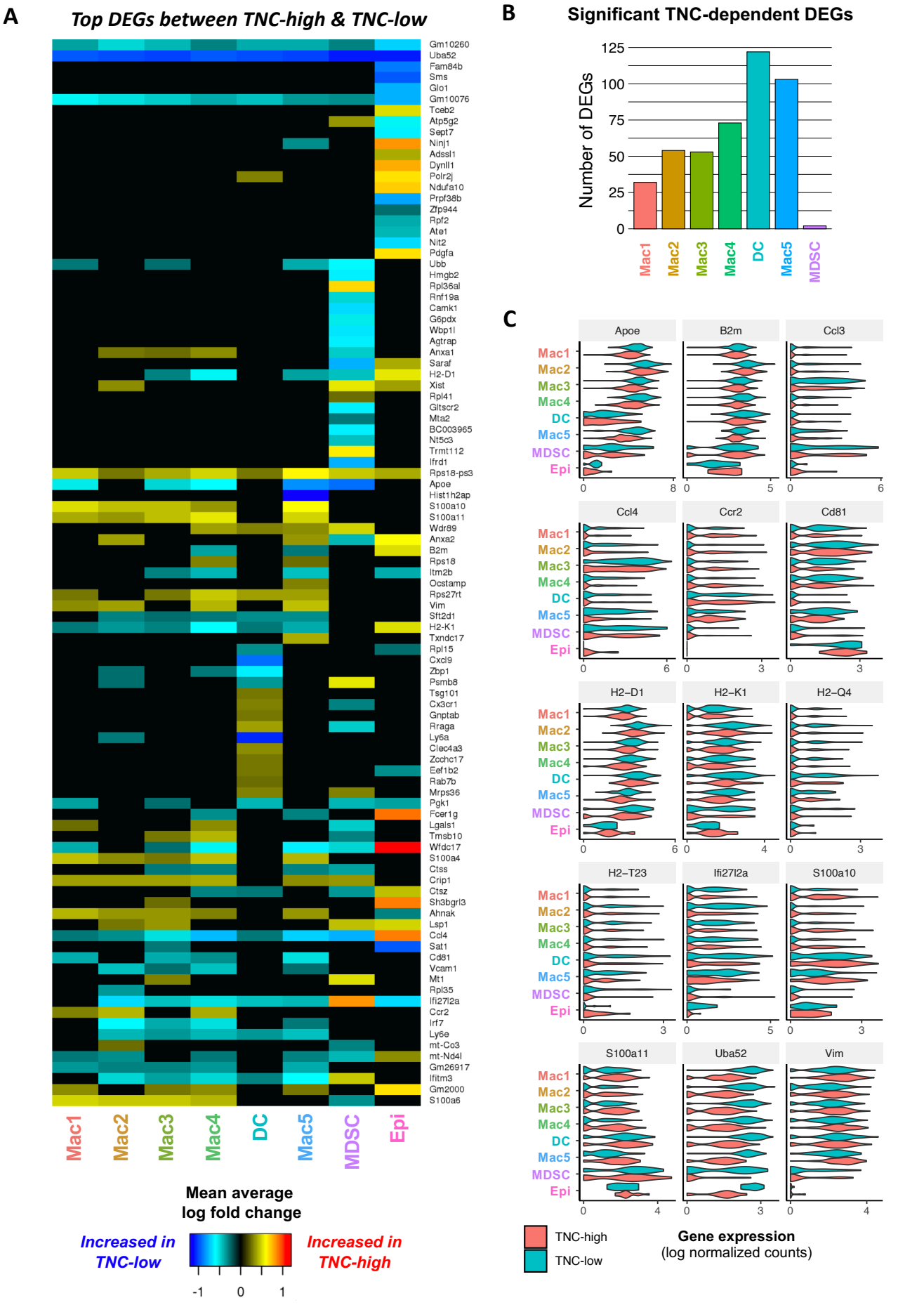


Figure 3.12 | TNC-dependent effects on myeloid transcriptional landscape. (A) Heatmap of the top differentially expressed genes per cluster between TNC-high and TNC-low conditions. Genes were included on the basis of their average log2 fold-change. **(B)** The number of genes significantly differentially expressed (Benjamini-Hochberg-adjusted $P < 0.05$) between TNC-high and TNC-low conditions per cluster. **(C)** Violin plots of key selected TNC-dependent genes. Abbreviations: DEG = Differentially Expressed Gene.

3.1.8 | Trajectory inference and pseudotemporal analysis of TAMs

Clustering based approaches aid the classification of cells into discrete states, however, they neglect the fact that biological processes are often continuous – such as cellular activation, plasticity and differentiation (329). Cellular transcriptional states therefore lie along a phenotypic continuum that can be reconstructed using dynamic models of continuous gene expression, known as trajectory inference algorithms.

As TAMs are the largest myeloid subset by number (**Figure 3.11C, Appendix Figure 3.4**), and exhibit TNC-dependent changes at both the level of population structure and immune gene expression (**Figures 3.11-3.12**), these cells were selected for further analysis. To explore the phenotypic plasticity of functionally-distinct TAM subsets, the diffusion map algorithm (281) was employed, establishing a pseudotemporal ordering of cells based on continuous changes in gene expression (**Figure 3.13A**). Visualisation of the first three diffusion map components (DC1-3) indicated that putative 'naïve-like' **Mac1**, the largest macrophage subset by cell number, form a natural 'root' cluster from which other macrophage subsets emanate, potentially supporting the notion of **Mac1** as a naïve-like state. Three distinct and opposing trajectories were observed for **Mac2**, **Mac3** and **Mac5**, representing the great degree of phenotypic heterogeneity between their respective roles in antigen-presentation, cytokine/chemokine production, and proliferation. Finally, the morphogenic subset, **Mac4**, lies between all other TAM subsets, thus reinforcing its characterisation as a transitionary state between other macrophage states.

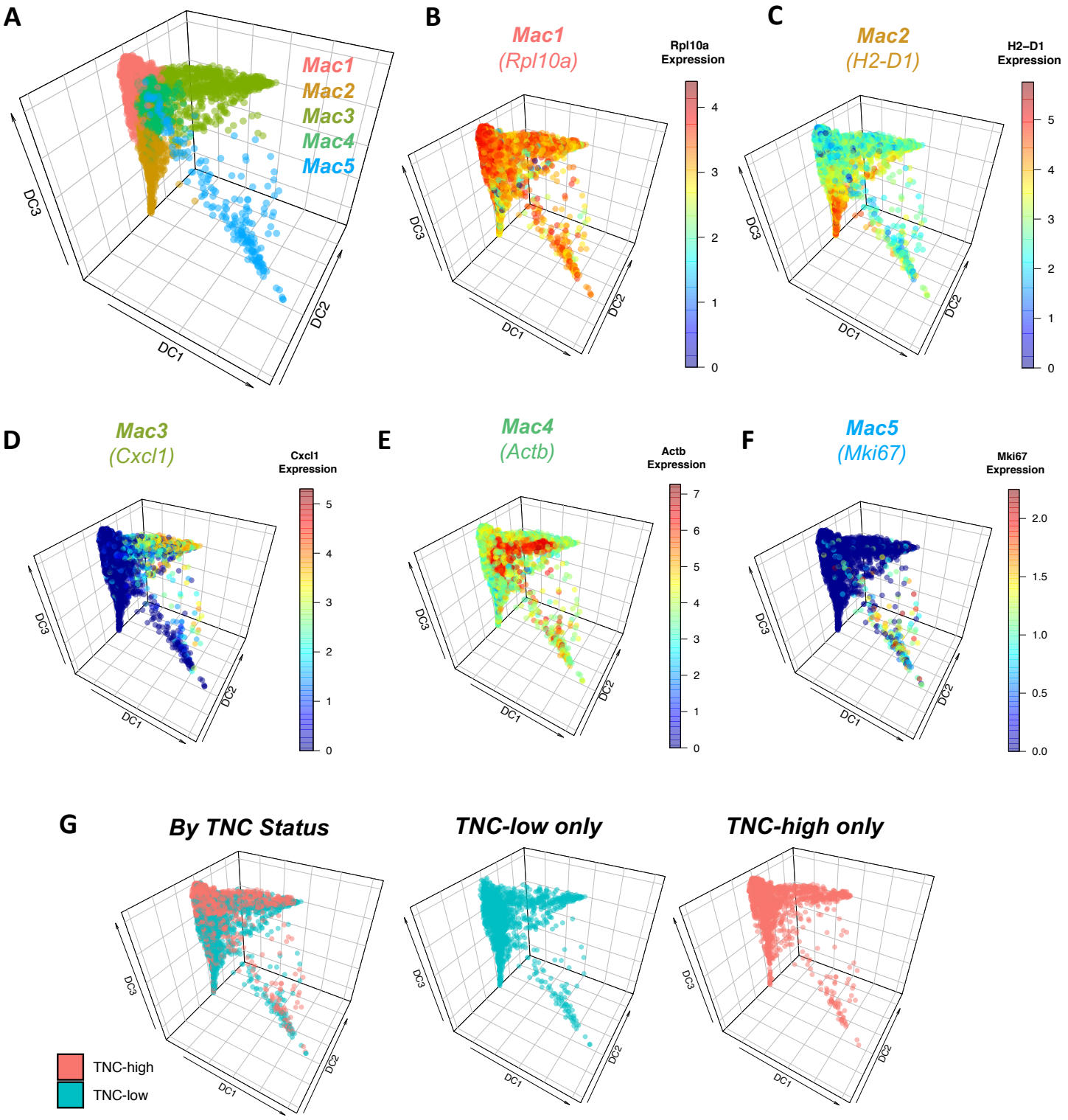


Figure 3.13 | Trajectory inference of macrophage subsets with diffusion maps. Diffusion map analysis was performed to establish a pseudotemporal ordering of all macrophages using the first three diffusion map components (DCs). Trajectories are visualised in a three-dimensional plot according to: **(A)** Cluster assignment, **(B-F)** Cluster-discriminating genes for all macrophage clusters, and **(G)** TNC-status. Expression data is represented as log normalized counts.

Visualisation of selected discriminatory marker genes of each macrophage cluster indicated differences in their distribution along given trajectories – for instance, *Cxcl1* (**Mac3**) and *H2-D1* (**Mac2**) expression both increased along their respective trajectories, whereas *Rpl10a* (**Mac1**), *Actb* (**Mac4**) and *Mki67* (**Mac5**) did not (**Figure 3.13B-F**). Corroborating previous data (**Figure 3.10B-C**), superimposition of TNC-status onto diffusion maps confirmed the preference of macrophages from TNC-low microenvironments to skew towards the antigen-presentation-enriched **Mac2**, and vice versa for ribosome-enriched **Mac1** in TNC-high tumours (**Figure 3.13G**).

To statistically assess gene expression dynamics along diffusion map trajectories, the Slingshot pseudotime algorithm (282) was employed to fit a minimum spanning tree, using the first three diffusion components and cluster labels as an input. Given the absence of an explicit monocytic cluster, the presence of the proliferating **Mac5** cluster (which might replenish the TAM population), and the evidence that **Mac1** may represent a naïve-like state, either **Mac1** or **Mac5** were selected as the initial root clusters for Slingshot analysis in two independent runs. In both cases, part or whole of the trajectory inferred was inconsistent with the topology of the diffusion map – for instance, an unusual trajectory was produced that passed sequentially from **Mac1**>**Mac2**>**Mac3**, thus connecting two poorly connected communities on the diffusion map: **Mac2**>**Mac3** (data not shown).

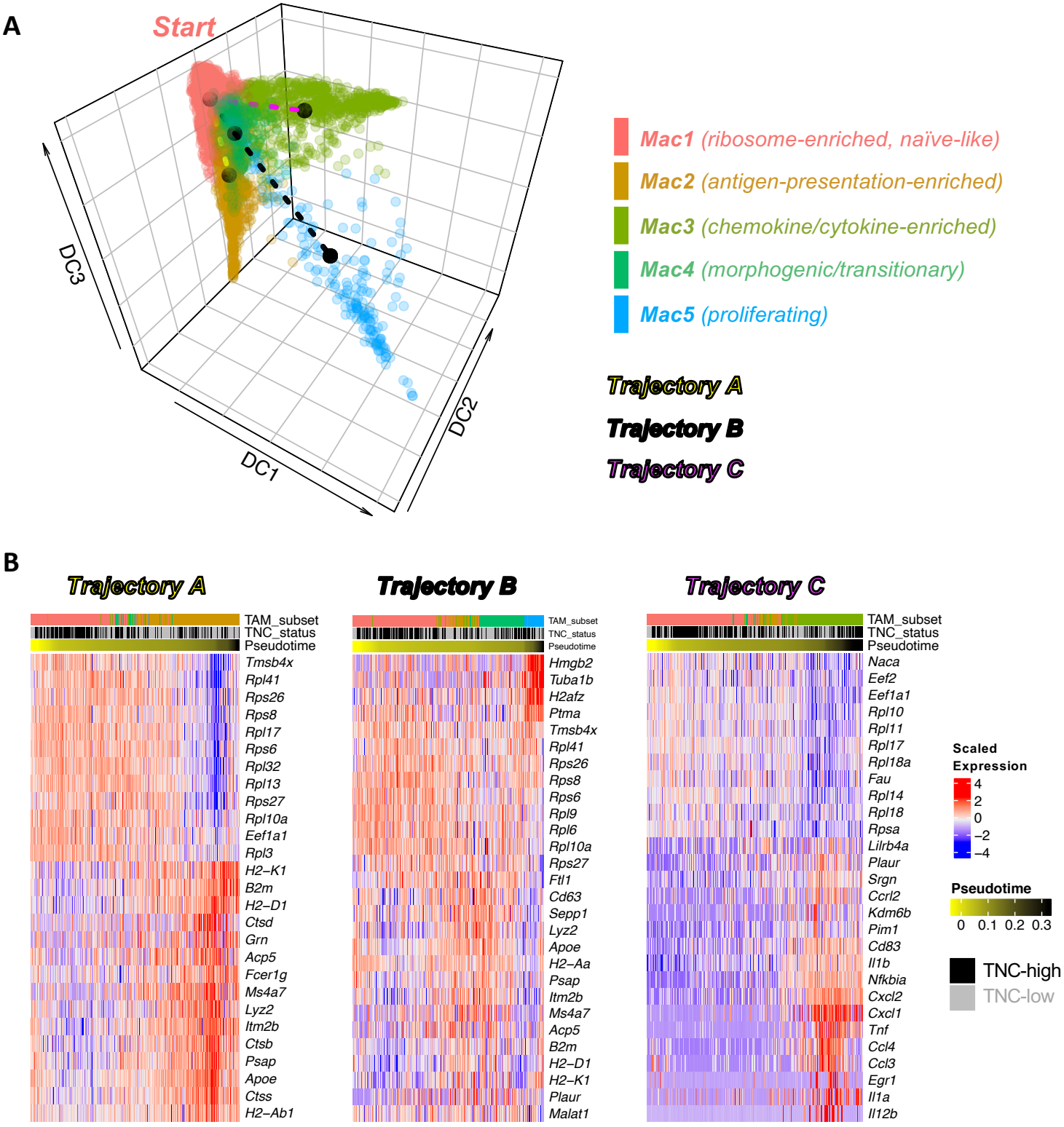


Figure 3.14 | Diffusion pseudotime analysis of macrophage plasticity. (A) Using the Slingshot algorithm (282), a minimum-spanning-tree was fitted for the three clear diffusion map trajectories emanating from **Mac1**, a putative naïve-like subset (specified as the root cluster), which respectively extend towards **Mac2**, **Mac3** and **Mac5** (specified as the leaf clusters). **(B)** Heatmaps showing selected genes found to have significant changes in expression along a smooth curve fitted to each pseudotime trajectory. This was computed using a generalize additive model (GAM), and BH-adjusted *P* values < 0.05 were considered significant. The number of genes significantly differentially expressed in trajectories A, B and C was 3004, 3739 and 2479, respectively. Cells (columns) in each trajectory are ordered by pseudotime. Acknowledgment: R code for fitting the GAM was provided by Maria Gomez Vazquez at the Kennedy Institute of Rheumatology (University of Oxford).

As recommended (330), *a priori* biological knowledge was used to better inform the analysis and improve the directionality of inferred trajectories. Earlier data in **Figure 3.11C** indicated that **Mac1** and **Mac2** are the only macrophage subsets to considerably change in abundance according to TNC-status. Previous flow cytometric analysis in this disease model established that an antigen-presentation-enriched macrophage subset (ie analogous to **Mac2**) is expanded in TNC-low tumours by approximately the same proportion as observed by scRNAseq in **Figure 3.11C** (264). Thus, with the knowledge that the **Mac1**>**Mac2** trajectory is consistent with well-validated phenotypes observed at the FACS level in a statistically robust number of replicates, this trajectory was 'defined' by supervised specification of start (**Mac1**) and end (**Mac2**) clusters using Slingshot, which fitted a minimum-spanning tree, designated 'trajectory A'. This approach was then replicated for the two clear remaining topological diffusion map trajectories (**Figure 3.14A**), each emanating from 'naïve-like' **Mac1.** This produced the **Mac1**>**Mac4**>**Mac5** trajectory (trajectory B), and the **Mac1**>**Mac3** trajectory (trajectory C) (**Figure 3.14A**).

This supervised analysis yielded several key observations. First, there appears to be a co-ordinated downregulation of ribosomal protein genes as macrophages acquire effector functions such as antigen presentation (trajectory A) and chemokine/cytokine production (trajectory C). The precise and reciprocal nature of this effect suggests that these processes might be inter-dependent. The same phenotype was not observed in Trajectory B, however, indicating that proliferation or self-renewal may not be regulated in the same manner as explicit macrophage effector functions. Second the expression dynamics of trajectory A indicate the presence of a brief and intense phase of protease induction (*Ctss*, *Ctsb*, *Ctsd*, *Lyz2*), followed by upregulation of membrane-bound MHC Class I machinery (*B2m*, *H2-K1*, *H2-D1*), perhaps indicating that the former is necessary for the latter. Interestingly, MHC class II machinery (*H2-Ab1*) did not follow the same dynamics. Third, trajectory B revealed biphasic and reciprocal changes in the expression of ribosomal proteins and MHC molecules, though these changes were more modest. It should also be noted that ribosomal protein genes appear to be modestly

differentially regulated through pseudotime, with some becoming downregulated at the end of trajectory B (*Rps27*, *Rpl10a*) and others not (*Rps26*, *Rps8*). In general, however, trajectory B did not identify any genes smoothly and linearly changing with pseudotime in one direction. Fourth, the expression dynamics of cytokines/chemokines differed markedly along trajectory C, with some favouring gradual progression (*Il1a*, *Il1b*, *Cxcl2*), while others favoured sharp specific induction (*Ccl3*, *Ccl4*, *Tnf*), perhaps hinting at differential regulation and function. Of particular interest is the finding that the lysine demethylase, *Kdm6b*, and the kinase, *Pim1*, are specifically induced immediately preceding the onset of chemokine and cytokine upregulation in **Mac3**, perhaps alluding to an epigenetic program controlling this response (**Figure 3.14B**). Given the role of these aforementioned molecules in promoting breast tumour growth and metastasis, this finding may have therapeutic relevance.

3.1.9 | Summary of findings and hypotheses

This study aimed to identify the full spectrum of pathogenic myeloid cell states in the breast tumour microenvironment, and to understand how these states change upon exposure to tumour-derived TNC. By performing scRNAseq on $CD11b^+$ myeloid cells that had been FACS-sorted from TNC-high or TNC-low breast tumours, this work shows that the intratumoural myeloid compartment is dominated by functionally distinct macrophage subsets with specialised effector activities. TNC appears to exert its effects principally at the level of altering myeloid subset abundances, although modest changes in the expression of individual genes was also observed. Finally, pseudotemporal analysis indicated multiple trajectories of phenotypic plasticity between different macrophage states, and proposed putative regulators of this process.

To prioritise the experimental validation of the most prominent, interesting, novel and tractable findings from the scRNAseq study, a summary of all key observations was generated (**Table 3.1**). From this, it was apparent that the most novel and tractable finding with the largest effect size was Finding 2: the functional delegation of macrophage effector processes to discrete specialised subsets, such as antigen-presentation-enriched (**Mac2**), chemokine/cytokine-enriched (**Mac3**) and proliferating (**Mac5**) macrophages, with the hypothesis that these effector functions are not simultaneously co-existing in tumour macrophages. This finding additionally involved TNC-dependent mechanisms at the level of both population structure (eg changes in **Mac1** and **Mac2** abundance) and the transcriptional landscape (eg modest changes in MHC or chemokine gene expression). This finding was therefore selected for validation and further assessment in **Chapter 4**.

(Next page) **Table 3.1 | Summary of findings from scRNAseq data.** To prioritize follow-up validation of key findings, all principal observations were listed alongside their concomitant effect-sizes and TNC-dependency. Associated hypotheses were then assessed for their tractability and novelty. The average logFC is derived by comparison of a given cluster to all other clusters. Abbreviations: FC = Fold-change, OR = Odds Ratio.

Finding	Observation	Effect Size	TNC dependent?	Hypothesis	Tractability	Novelty
1	The largest TAM cluster, **Mac1**, significantly upregulates many ribosomal-translation genes	**Small** (average logFC = 0.29)	No (NB: see **Hypothesis 6**)	Ribosomal enrichment is a pathogenic biomarker of naïve-like macrophages (**Mac1**) that have failed to polarize towards an antigen presentation phenotype (**Mac2**)	**Low:** no adequate **Mac1** surface markers exist for isolation methods (eg MACS, FACS), and ribosomal enrichment is relatively modest.	High
	Mac1 is expanded ~2-fold in TNC-high tumours.	**Medium** (FC = 2)	Yes, at cellular level			
2	Discrete macrophage subsets strongly enrich chemokine production, antigen presentation, or proliferation.	**Large** (OR = 26, 36, 7, respective)	Partly, at the cellular level (Eg **Mac1** / **Mac2**)	Macrophage effector functions are delegated to discrete specialized subsets. For instance, highly antigen presenting TAMs (**Mac2**) do not produce *Cxcl1* or *Tnf* (**Mac3**), and vice versa. This is partly regulated by TNC, at both the transcriptional and cellular levels.	**High:** surface markers identified and validated antibody reagents commercially available.	High
	Several MHC genes are specifically upregulated in macrophages from TNC-low tumours.	**Medium** (average logFC = 0.35)	Yes, at transcriptional level			
3	Large quantities of endogenous TLR4 ligands are produced by the myeloid compartment, such as calprotectin (*S100a8/S100a9*) in **MDSC**, *Hmgb1* in **Mac5**, alongside others. *Tlr4* levels are very low, perhaps indicating specific downregulation.	**Large** (S100a8/a9 average logFC = 5.5 & 5.1	Partly, at cellular level (Eg **MDSC** abundance)	Co-ordinated myeloid expression of endogenous TLR4 ligands results in a pathogenic, multi-stimulus, pan-TLR4 signaling in the tumour microenvironment. Inhibition of TLR4 signaling with available small molecules would confer anti-tumoural activity.	**Medium:** surface markers of relevant subsets identified, plus validated antibody reagents and small molecule inhibitors are commercially available. TLR4 FACS staining is challenging to use due to rapid receptor internalisation upon binding.	Medium
4	**MDSC** (comprising 3.5% of myeloid cells) are expanded ~2-fold in TNC-high tumours and possess a strong suppressive phenotype.	**Medium** (FC= ~2)	Yes, at cellular level	**MDSC** recruitment is an important feature of TNC-dependent disease, and **MDSC** possess a number of soluble and surface-bound therapeutic targets, depletion of which will ameliorate disease. **MDSC** depletion will also reduce suppressive TAM infiltration (provided that **MDSC** can indeed differentiate into TAMs).	**Medium: MDSC** is a small subset, however does have highly discriminatory surface markers, alongside validated antibody reagents, and potential therapeutic targets. A limitation is the potentially low quality of the captured **MDSC** transcriptomes, necessitating further validation.	High
5	**DC** are expanded ~2-fold in TNC-low tumours, and possess the highest number of TNC-dependent differentially expressed genes (122).	**Medium** (FC= ~2)	Yes, at both cellular and transcriptional levels	**DC** are the principal target of tumour-derived TNC, as they are regulated at both cellular and transcriptional levels. Their potent antigen presentation phenotype, and ability to migrate to lymph nodes to inform adaptive immunity (unlike macrophages), render them prime therapeutic targets.	**Medium: DC** is a relatively small subset (only 7% of myeloid cells) and exhibits great heterogeneity – however several reasonable surface markers are available, and some validated antibodies.	High
6	*Uba52*, encoding a fusion protein of ubiquitin and RPL40, is highly and specifically upregulated in all myeloid cells from TNC-low tumours. It is the most significantly upregulated gene across all clusters.	**Large** (average logFC across all clusters = 1.0)	Yes, at transcriptional level	*Uba52* expression increases ribosome ubiquitination. In order to efficiently polarize between states (eg from **Mac1** to **Mac2**), macrophages downregulate ribosomal proteins while upregulating effector molecules (eg chemokine or MHC molecules). Ubiquitinated ribosomes in *Uba52*-high cells are more readily downregulated and degraded, enabling more efficient polarization towards effector functions. This explains why naïve-like ribosome-enriched **Mac1** accumulate in TNC-high tumours (which are *Uba52*-low), as they have failed to polarize to other subsets. Thus UBA52 and ribosome enrichment mediate an arrest of polarization. Finally, as TLR4-stimulation can promote iron storage, perhaps the TNC-high environment leads to low iron bioavailability which also limits **Mac1**'s ability to polarize.	**Low:** relatively poorly described gene which is exclusively intracellular, not cluster-specific, and has very limited reagents available.	High
7	*Pim1* and *Kdm6b* are upregulated immediately preceding cytokine/chemokine induction, and this signal is enhanced in TNC-high tumours, while the kinetics of *Cxcl1* and *Ccl3* induction also change.	**Medium** (N/A)	Yes	*Pim1* and *Kdm6b* represent an epigenetic transcriptional program in macrophages immediately preceding cytokine/chemokine induction. The strength and kinetics of this epigenetic program is TNC-dependent.	**Medium:** surface markers of relevant subsets identified, but inhibiting kinases/demethylases often has large off-target effects	High
8	No explicit monocytic precursors to TAMs were detected, yet **Mac5** strongly enriched proliferation.	**Medium** (N/A & OR = 7)	No	Tissue-resident **Mac5** replenishes the entire TAM population by proliferation.	**Medium:** lineage tracing experiments are challenging and time-consuming	Medium
9	A smaller proportion of mitochondrial reads were observed in myeloid cells captured from TNC-low tumours. Mitochondrial genes were TNC-dependently downregulated (*Mrps36*, *mt-Co3*, *mt-Nd4l*)	**Small** (% change mito reads = 10%; All mito gene average logFC = 0.29)	Yes, at both cellular and transcriptional levels	Tumour-derived TNC induces the accumulation of dysfunctional mitochondria in myeloid cells, leading to dysregulated metabolism and impaired function.	**Medium:** reagents and expertise likely only available through collaboration.	High

(Table legend on previous page)

3.2 | Technical Discussion

As with any experimental method or design, there are a number of nuances and caveats worth noting. First, and perhaps most importantly, there are limits to what can be interpreted from this dataset, given that it represents the pooling of n=3 tumours per TNC condition, thus a biological n=1 scenario. While statistically significant differences can be established between the >4000 cells passing quality control, which enables the examination of *intra*-tumour heterogeneity, it is not possible to examine the contribution of *inter*-tumour heterogeneity, since individual biological replicates have been merged without a means of later deconvolving them. This, for example, means that the significance of TNC-dependent changes in cell subset abundance (**Figure 3.10C**) must be interpreted with caution, and necessitates the performance of extensive follow-up experiments in a statistically robust number of biological replicates, as determined by a power calculation, to validate these observations (see **Chapter 4**).

Second, it should be acknowledged that the gating strategy for $CD11b^+$ myeloid cell isolation by FACS-sorting, using CD3 as a negative marker to exclude T cells (based on established protocols (331)), may have inadvertently also excluded a small subpopulation of CD3+ and/or TCR^+ macrophages, whose role in tumour biology is poorly understood (332).

Third, it remains possible that minor batch-dependent misalignment of TNC-high and TNC-low conditions has occurred. Using one of the leading batch-correction methods available at the time, a canonical correlation analysis (CCA) (333) was performed, and while agreement between both TNC conditions was reasonable, a modest misalignment was observed (data not shown). Importantly, however, an attempt to realign the data based on CCA components was overly punitive – abolishing genuine biological variation for which there is robust evidence from previous flow cytometric analyses. For example, previous work in this same mouse

model of breast cancer described expansion of MHC2^{hi} macrophages (analogous to **Mac2**) in TNC-low tumours (264), yet this effect was abolished upon CCA alignment (data not shown). Popular alternative batch-correction methods include an iterative 'soft' k-means-based approach, Harmony (334), and a Bayesian approach, ComBat (335), both of which yielded similar results to CCA alignment (data not shown). Further approaches based on mutual-nearest-neighbours (MNN) (336), neural networks (337), or a recent combination of CCA and MNN (269) are also available, and might potentially provide superior results. However, given the reasonable agreement of the two TNC conditions, the overly punitive adjustment methods, and the internal validation provided through consistency with previous well-validated data (from statistically powered analyses at the protein level), there was sufficient confidence in the native unadjusted data to proceed with downstream analysis. To minimise any potential batch-specific bias, differential expression analysis between (and within) clusters was performed using only marker genes conserved between both TNC conditions, and therefore cluster-specific gene expression and concomitant downstream interpretations should not be unduly affected by batch-dependency.

Fourth, while the average read depth per sequenced cell is reasonable for 10X data, at 61-71k reads per cell (**Appendix Figure 3.1**), extrapolating insights from lowly expressed genes should be performed with caution, as there is potentially considerable technical noise in scRNAseq data, even despite the extensive implementation of quality-control procedures to mitigate it. For example, the majority of immune genes relating to APC co-inhibition and co-stimulation are lowly expressed genes (**Figure 3.9**) and, consequently, only top-level conclusions are drawn from these. Yet, in spite of this point, it should be noted that a number of seemingly lowly expressed genes at the mRNA level may still express sufficient protein to be physiologically pertinent and detectable with antibody-based methods. A prime example of this would be CD11b (*Itgam*), whose mRNA expression in myeloid clusters is low and, in some cases, almost absent (**Figure 3.6**), despite the fact that these cells were isolated based on surface detection of CD11b protein (ITGAM) by FACS sorting (**Figure 3.1**). This might suggest

a number of possibilities, which may not necessarily be mutually exclusive: **(i)** the current sequencing depth is insufficient to capture such genes, **(ii)** certain genes might be periodically rather than constitutively induced, meaning only a subset of cells express transcript at any one time, or **(iii)** the well-known discrepancy between mRNA and protein levels (338) in such cases might be explainable by long protein half-life, preferential IRES-mediated translation, or endocytic recycling of surface proteins, meaning that very low transcript levels can be maintained indefinitely without impeding biological function. Thus, while caution is advised when interpreting lowly expressed genes, their potential biological importance should not be neglected.

Fifth, while it is recommended best-practise (339) that any inter-sample differences in the percentage of mitochondrial reads are regressed out prior to further analysis, it remains possible that this significant difference between TNC-high and TNC-low tumours (**Figure 3.2B**) might represent genuine biological variation, given the potential roles for TNC in regulating metabolism. However, these differences were relatively modest and thus seem unlikely to significantly impact the biological interpretation of the analysis. In fact, using data that had not been regressed for the percentage of mitochondrial reads resulted in consistently suboptimal clustering (data not shown), and thus best-practise was followed, whereby these differences between TNC conditions were regressed out prior to further analysis.

Sixth, geneset over-representation analysis is a powerful tool for understanding the biological functions of lists of differentially expressed genes (340), however it relies on a hypergeometric (Fisher) test in which a list of differentially expressed gene names (the foreground genes) are compared to that of a biologically-defined geneset annotation (the background genes). This analysis is not weighted depending on the strength of differential expression between genes in the same foreground geneset (ie they are all considered equal), and thus potential biological signal is overlooked. A possible improvement on this would be to perform: Gene-set enrichment analysis (GSEA) which defines an enrichment score based on the ordered rank of

differentially expressed genes (279); LEGO, a method which incorporates network-based gene weights (341); or Gscluster, another network-weighted approach which harnesses protein-protein interaction networks, and clusters genesets to aid biological interpretation (342). One way to extract further biological meaning, especially from the immune gene expression analysis, would be to assess receptor-ligand interaction networks using NicheNet (343) or CellPhoneDB (344), to more intimately understand the degree of inter-cluster communication and signalling.

Seventh, the **MDSC** cluster has a consistently lower number of genes, counts and percentage of mitochondrial reads than other clusters (**Appendix Figure 3.4C**), reduced ribosomal protein expression (**Figure 3.6 & 3.8A**), and a particularly high number of negatively differentially expressed genes (**Appendix Figure 3.4E**). Together, these could be indicative of lower quality transcriptomes, meaning this data should be interpreted with caution. In addition, the small number of cells in the **MDSC** cluster meant there was insufficient statistical power to make robust inferences on TNC-dependent changes in gene expression, so these trends must also not be over-interpreted. However, indicating that these issues are not prohibitive to interpretation, there was sufficient read depth to demonstrate that **MDSC** transcriptomes had a high number of unique positively differentially expressed genes (**Appendix Figure 3.4E**), including marked upregulation of key immune effector genes (*S100a8*, *S100a9*, *Il1b*, *Ptprc*) even more highly than other myeloid subsets (**Figures 3.8 & 3.10**), suggesting that these differences may represent genuine and unique biology.

Eighth, strong and highly significant upregulation of β-actin (*Actb*) – the 2nd most highly variable gene across all myeloid transcriptomes – was identified in **Mac4**, indicating that the widespread use of actin as a 'housekeeping' normalisation control in other studies is ill-advised, aligning with a comparative analysis of housekeeping genes in macrophages (345).

Finally, while our approach to trajectory analysis does not enable inference of developmental lineages, this supervised analysis does enable visualisation of the phenotypic plasticity of macrophage subsets, and the determination of markers that significantly drive continuous changes in gene expression along such phenotypic trajectories. Thus, pseudotime data should only be interpreted within this framework. The Slingshot algorithm was originally selected due to its consistently strong performance compared to other methods in terms of accuracy, stability, scalability and usability (346). Unsupervised analysis of the data using an alternative trajectory inference tool, especially one better optimised for cyclic topologies, such as PAGA (347), might also be advisable to complement this analysis. With respect to pseudotime interpretation: assuming that RNA content is equal between all cells (which scRNAseq workflows do assume by within-cell scaling), the reduced expression of ribosomal protein genes (which are very highly and constitutively expressed) coinciding with high upregulation of other genes (MHC molecules, cytokines, and chemokines) in **Mac2**, **Mac3** and **Mac5** (**Figure 3.10**) might be reflective not of a ribosomal protein downregulation *per se*, but merely of a proportional decrease in detection (ie a dilution effect caused by a reciprocal increase of other genes). It is also important to emphasise that diffusion map trajectories do not necessarily have to represent biological processes – they merely represent similarity at the transcriptional level, and few trajectory inference methods include an evaluation of uncertainty in their model to account for processes at other 'omic levels (339, 348). Similarly, the lack of a clear monocytic cluster in this dataset represents a significant level of uncertainty in the ontogeny of TAMs which is not being accounted for. Thus, extensive and time-resolved follow-up experiments at the protein level, with a robust number of biological replicates, must be performed to validate these findings.

3.3 | Biological Discussion

3.3.1 | Introduction

Myeloid cells are key protagonists in the microenvironment of many solid cancers, including breast cancer, where they contribute to a diversity of fundamental pathological processes (120-122). The extracellular matrix protein, TNC, is associated with worsened prognosis in breast cancer patients (250), and high levels of tumour-derived TNC promote a tolerogenic macrophage phenotype in murine breast cancer (264). To understand the pathogenic interaction between myeloid cells and tumour-derived TNC in breast cancer, single-cell RNAseq was performed on intra-tumoural myeloid cells exposed to either high or low levels of tumour-derived TNC. This approach enabled high-resolution mapping of heterogeneous myeloid states, and revealed a multitude of TNC-dependent and TNC-independent phenotypes at both cellular and transcriptional levels.

This section will be divided into the following subsections:

3.3.2 Heterogeneity and ontogeny of myeloid subpopulations

3.3.3 Pathological relevance of myeloid states

3.3.4 Co-existence of activatory and inhibitory phenotypes

3.3.5 Tumour-specific phenotypic expansion

3.3.6 Ribosome-mediated delegation of TAM effector functions

3.3.7 TNC-dependent effects

3.3.8 Summary & hypothetical pathogenic mechanism

3.3.2 | Heterogeneity and ontogeny of myeloid subpopulations

Myeloid cells are members of innate immunity with key roles in immune defence and tissue homeostasis. As they become co-opted by tumours to aid pathogenesis, myeloid cells can have a profound impact on cancer prognosis and the response of patients to immunotherapy (157, 158, 349). The phenotype of intra-tumoural myeloid cells has been extensively characterised with low-dimensional technologies, yet the advent of high-dimensional single-cell technologies revealed unexpected heterogeneity, tumour-specific phenotypes and the co-existence of functional expression signatures once thought to be dichotomous (102).

Aligning with previous single-cell surveys of the immune landscape in human breast tumours (101-103, 350), scRNAseq revealed extensive myeloid cell heterogeneity in the NT193 model of breast cancer. In total, seven myeloid subpopulations were identified, comprising one **DC** cluster, one **MDSC** cluster and five macrophage clusters, the latter partitioning into distinct functional states according to their capacity for ribosomal protein upregulation (**Mac1**), enhanced antigen presentation (**Mac2**), chemokine/cytokine production (**Mac3**), cellular morphogenesis (**Mac4**) and proliferation (**Mac5**) (**Figures 3.4-3.6**). This precise segregation of macrophage effector functions has not previously been described, though several aspects of these findings are supported by previous studies. For example, CXCL1-expressing macrophages (analogous to **Mac3**) that recruit MDSCs have previously been observed in breast cancer (308), however the existence of a discrete macrophage population responsible for the majority of chemokine production has not been described in breast cancer – though it was in an orthotopic immune-competent model of lung adenocarcinoma (351). Likewise, previous work in the Midwood group demonstrated that exposure to high levels of tumour-derived TNC mediates a shift towards immune-suppressive MHC-downregulation in macrophages (264) (analogous to **Mac2**). However, scRNAseq revealed deeper features of this **Mac2** phenotypic state – for instance, showing that these MHC changes occur at both the

level of cellular abundance (**Figure 3.10C**), and via the transcription of specific classical and non-classical MHC genes (**Figure 3.11C**).

scRNAseq analysis of nine human breast tumours by Azizi and colleagues described considerable and unexpected heterogeneity in the monocyte compartment, showing that monocytes segregated into an even greater number of phenotypic clusters than T cells (102) – a cellular compartment notorious for its extensive heterogeneity. While a small number of Ly6C^{hi} monocytic-like cells were identified in the present work (as part of the **DC** and **MDSC** clusters) (**Figure 3.7A**), no explicit monocytic cluster(s) were observed. Instead, a greater degree of heterogeneity was observed among macrophages, which segregated into 5 functionally-distinct states, rather than the 3 largely overlapping macrophage states reported by Azizi *et al.*, which differed only in their gene covariance patterns rather than their differential expression. This could potentially reflect: (i) species-specific differences between human and mouse breast cancer, (ii) the lower sequencing depth of the Azizi *et al.* dataset compared to the increased resolution afforded by CD11b^{+} FACS-sorting, or (iii) differences between a HER2-driven breast cancer model (whose pathogenesis is more immune-dependent) as opposed to other molecular breast cancer subtypes, given that only 1 out of 9 human samples from Azizi *et al.* was HER2-positive.

In terms of macrophage ontogeny, the absence of an explicit monocytic cluster might suggest that TAMs in the NT193 murine model are predominantly derived from proliferating tissue-resident macrophages rather than recruited monocytes, or that monocytes were not captured by this analysis for technical reasons (see Technical Discussion, **Section 3.2**). While all macrophage clusters expressed variable quantities of *Ccr2* (**Figure 3.10**), indicating a monocytic origin, it is important to note that the normal mammary gland relies entirely on blood-borne monocytes to seed its tissue-resident macrophage population (188) (rather than being derived from a foetal progenitor), and that this resident population is capable of self-renewal. Thus, any TAM population detected by scRNAseq may well express *Ccr2*

irrespective of whether it was derived from a recent monocyte or a proliferating resident macrophage that was originally derived from a monocyte. To exclude this possibility, comparison of *Ccr2* expression in tissue-resident macrophages should be assessed in public datasets, preferably from normal mammary gland tissue. Importantly, expression of the terminal-differentiation gene, *Mafb*, across all TAMs gives confidence that **Mac1** is indeed a macrophage population rather than a Ly6C^{lo} non-classical monocyte population. However, given that monocytic MDSCs can differentiate into TAMs in hypoxic contexts via *STAT3* downregulation (352) it remains possible that monocytic cells in the **MDSC** cluster might at least partially seed or replenish the TAM population, as *Stat3* expression is absent in **MDSC** yet present in all other myeloid clusters (**Appendix Figure 3.7**). Overall, the ontogeny of TAMs in this dataset remains unclear, and necessitates further investigation.

3.3.3 | Pathological relevance of myeloid subsets

scRNAseq revealed a wealth of phenotypic information on myeloid subpopulations. The pathogenicity of these novel phenotypes is therefore often speculative; however, many phenotypes were also supported by previous breast cancer literature, thus reinforcing their validity. For example, the pathogenic relevance of novel subsets such as ribosome-enriched **Mac1** or morphogenic/transitionary **Mac4** have not yet been described, while chemokine-enriched **Mac3** highly upregulates *Ccl3*, *Cxcl1* and *Cxcl2*, all of which have previously been implicated as tumour-promoting or metastasis-promoting factors in the tumour microenvironment. However, establishing the pathogenic relevance of these molecules can be challenging, as it is confounded by the pleiotropy and context-dependence of immune signaling processes. Highlighting this issue is the **Mac3**-derived chemokine, *Ccl3*, whose high expression in a subset of macrophages promotes pulmonary metastasis in murine breast tumours (310), while impaired *Ccl3* expression promotes bone metastasis of breast tumours in another context (353). Together this alludes to the pleiotropy, complexity and context-dependence of CCL3 activity, and additionally suggests that therapeutic modulation of the chemokine signaling network is likely to be challenging. Indeed, this crosstalk between TAMs and tumours mediates similar mechanisms via CXCL1 – another chemokine strongly (and solely) expressed by **Mac3**. TAM-derived CXCL1 has previously been implicated in promoting tumour growth and metastatic burden in mouse breast cancer via a NFkB/SOX4-dependent mechanism, and was associated with lymph node metastasis, poor survival and a basal-like subtype (mostly triple-negative) in a human clinical dataset (308). Other work has defined a network of paracrine signals between myeloid, endothelial and breast cancer cells that together drives metastasis and chemotherapy resistance (307), proposing pathogenic roles for *Cxcl1*, *Cxcl2*, *Tnf* and calprotectin (*S100a8*/*S100a9*), all of which are highly expressed by **Mac3** and/or **MDSC** (**Figures 3.7A & 3.9**). In fact, the extremely high expression of calprotectin – an endogenous TLR4 ligand – is a key feature of the **MDSC** cluster, and S100A9 is likely to be strictly required for MDSC recruitment, as MDSCs are ablated in *S100a9*-null

mice with lymphoma (354). Furthermore, in breast cancer, calprotectin is associated with high-grade, basal-like tumours bearing poor prognosis (355), and calprotectin-high $Gr1^+$ myeloid cells recruit further cells of the same phenotype via TLR4 signalling to promote brain metastasis (356). Thus, MDSC-induced TLR4 signalling appears to drive a pathogenic positive feedback loop in breast cancer, with potential relevance to (and perhaps synergy with) TNC-dependent TLR4 signalling. Given that calprotectin-high **MDSC** are expanded ~2-fold in TNC-high tumours, this subset may potentially represent an attractive therapeutic target, helping to break the cycle of pathogenic TLR4 signaling while perhaps also halting the possible differentiation of tolerogenic mMDSC-derived macrophages

3.3.4 | Co-existence of activatory and inhibitory phenotypes

One of the peculiarities of omic-wide single-cell analyses of intra-tumoural immune cells is the discovery of co-existing activatory and inhibitory markers on the same individual cells or clusters (102, 103), meaning that the distinction between protective versus pathogenic cell states is difficult to disentangle. While this co-existence might be partly explained by the lack of a temporal dimension in such analyses (ie a cell switching phenotype will temporarily have both activatory and inhibitory gene transcripts present), the finding that canonical M1 and M2 macrophage markers genes strongly positively correlate (102) remains highly unusual and at odds with the paradigm of discrete macrophage polarization states (172, 179, 180). The present work similarly identified multiple examples of unexpected co-existing phenotypes that challenge the conventional polarization model. For instance, macrophage subsets frequently expressed co-inhibitory or "M2-like" molecules such as *Tgm2*, *Cd204*, *Havcr2* (TIM3), *Vsir* (VISTA), and *Lgals9* (Galectin-9), while also expressing canonical "M1-like" or co-stimulatory genes such as *Irf5*, *Socs3* and *Cd86*. A similar observation was made for co-expression of pro-phagocytic (*Cyba*, *Fcgr1*, *Fcgr3*, *Trem2*, *Aif1*, *Pld3*) and anti-phagocytic (*Sirpa*, *Fcgr2b*, *Lilrb4a*) genes in macrophages. Together, these findings suggest a deeper phenotypic complexity of TAMs in the breast tumour microenvironment than has been conventionally appreciated, aligning with other single-cell studies of breast cancer (101-103). Single-cell studies of macrophages in various other cancers including glioma (208) and non-small cell lung cancer (209) reported similar findings, indicating that this may be a common phenomenon among TAMs, independent of tissue origin.

However, in contrast to previous single-cell studies of the human breast TME which suggested that macrophages lie along a single phenotypic continuum (the 'TAM activation component') (102), the present work proposes the existence of distinct and opposing activation trajectories of macrophages (defined by effector functions) (**Figure 3.14**), that could not be fully explained by a single pseudotemporal 'activation component', nor the classic macrophage polarisation

model. The present work additionally argues for no single state of anti-tumour immune activation – rather, there appears to be a universally compromised myeloid compartment, where even highly inflammatory macrophages (**Mac3**) are pleiotropic, likely contributing to pathogenic recruitment of gMDSC, angiogenesis and remodelling of the tumour vasculature.

Another example of a complex co-existing phenotype can be found in *Apoe* – a multi-faceted gene with pathogenic roles in metabolism (357), anti-inflammatory effects in macrophages (323), and now considered an inhibitory checkpoint of the classical complement pathway due to its engagement of activated C1q (358). *Apoe* is extremely highly expressed in all TAM subsets (**Figure 3.9**), in fact representing one of the most highly expressed genes across the (polyadenylated) transcriptomes of all tumour-infiltrating myeloid cells. Paradoxically, *Apoe* levels are highest in the MHC-enriched **Mac2** subset, whose cellular phenotype is otherwise strongly anti-tumoural, and whose abundance as MHC-enriched macrophages correlates with reduced tumour size in previous work (264). **Mac2** is simultaneously the highest expressor of *C1qa* and *C1qb* (**Appendix Figure 3.7**) – which are targets of APOE as a checkpoint molecule – potentially alluding to an intriguing *cis*-acting compensatory mechanism wherein *Apoe* is induced to dampen anti-tumoural macrophage activation by MHC-dependent and complement-dependent processes. Interestingly, *Apoe* expression is TNC-dependently regulated in macrophages, but not in **MDSC** or **DC** (**Figure 3.11B**). Counterintuitively, *Apoe* expression is higher in all TAM subsets from TNC-low tumours compared with TNC-high tumours, though this effect is more modest for **Mac2**. This might either suggest a pleiotropic role for *Apoe* in tumours (perhaps enhancing macrophage activation in certain cases through an unknown mechanism), or more likely, indicating that the anti-tumoural effects of TNC deficiency (skewing TAMs towards an MHC-enriched phenotype) are partly counteracted by the anti-inflammatory effects of APOE. This latter point could perhaps explain why the role of TNC on tumour progression is only partial. In either case, proliferating **Mac5** exhibit the largest TNC-dependent upregulation of *Apoe* across all TAMs, alluding to a unique role for *Apoe* and TNC in potentially regulating TAM self-renewal in the tumour microenvironment. Furthermore,

a recent single-cell (spatial) transcriptomic study of the human breast TME described a lipid-associated macrophage subset expressing high levels of APOE (101), perhaps indicating conservation of the **Mac2** subset, and its co-existence of activatory and inhibitory phenotypes. Taken together, these data suggest that APOE might have pleiotropic effects in the breast tumour microenvironment, or may indeed be induced to dampen macrophage activation, perhaps especially in the TNC-low microenvironment.

3.3.5 | Tumour-specific phenotypic expansion

The complex heterogeneity and co-existing phenotypes of myeloid cells highlights a limitation of this data – the fact that tumour tissue was analysed with no 'normal tissue' comparators means it is not possible to directly discern tumour-macrophage phenotypes from normal-macrophage phenotypes. It therefore remains possible that some of these macrophage states are tumour-specific or tumour-induced. Indeed, Azizi and colleagues offer compelling evidence of a marked 'phenotypic expansion' in the immune compartment of human breast tumours compared to normal mammary tissue, with many tumour-specific immune phenotypes observed, suggesting a prominent role for the tumour microenvironment in shaping immune function (102). In particular, monocytic cells derived from tumours represented the highest 'phenotypic volume' occupied by any immune cell type by several orders of magnitude, suggesting that monocytic cells (and perhaps macrophages derived from them) are uniquely liable to the impact of the (breast) tumour microenvironment. Interestingly, the genes with the largest variance between normal and tumour tissue enriched signalling pathways activated in the TME, including TNFα, TGFβ, type I (IFNα) and type II (IFNγ) interferon responses, alongside hypoxia (102). Together, this suggests that the high production of **Mac3** and **MDSC**-derived *Tnf*, and to a lesser extent, pan-macrophage- and **DC**-derived *Tgfb1*, might be a tumour-specific myeloid phenotype.

3.3.6 | Ribosome-mediated delegation of TAM effector functions

This study identified that, **Mac1**, a macrophage cluster that modestly enriches a large range of ribosomal proteins and is the largest TAM subset, accounting for ~26% of the macrophage population on average. **Mac1** is also markedly expanded ~2-fold in TNC-high tumours, while antigen-presentation-enriched **Mac2** is reciprocally expanded in TNC-low tumours. Ribosome-enriched immune cells have previously been observed in single-cell studies of acute lymphoblastic leukaemia (ALL) where increased ribosomal protein (RP) expression is associated with developmental immaturity (359), and in zebrafish haematopoietic stem cells (HSCs), where cell fate decisions are followed by coordinated downregulation of RPs and upregulation of lineage-specific factors to control lineage differentiation (360). This not only alludes to the deep conservation of this RP-related developmental mechanism in cellular differentiation, but also reinforces the assignment of the RP-enriched cluster (**Mac1**) as a naïve-like root state for macrophage pseudotime analysis. It remains possible, however, that tissue-resident (non-monocyte derived) macrophages may not be governed by the same RP-related mechanism. Pseudotime data indicated that the downregulation of ribosomal protein genes is indeed a concerted and coordinated mechanism that coincides with the upregulation of gene programs required for polarization towards antigen-presentation enriched (**Mac2**) or chemokine/cytokine-producing (**Mac3**) phenotypes, but, crucially, not for proliferating (**Mac5**) or morphogenic (**Mac4**) subsets. This suggests that downregulation of ribosomal proteins may enable (or be required for) acquisition of specific macrophage effector functions, but not other cellular processes. Taken together, it seems that the presence of TNC appears to arrest tumour macrophages in a naïve-like state of developmental immaturity (marked by elevated ribosomal protein expression (**Mac1**)), which limits their capacity to polarise towards a more anti-tumoural antigen-presenting phenotype (**Mac2**).

Given the enormous energetic expense of ribosome biogenesis in eukaryotes (361), this functional prioritisation of macrophage effector functions above ribosome production is

perhaps an energetic necessity. Furthermore, as these myeloid cells are residing within the hostile growth conditions of the tumour microenvironment, it might be that the requirement for ribosome downregulation is even more pronounced here. Thus, the tumour microenvironment might directly induce macrophage differentiation via nutrient starvation that necessitates ribosomal protein downregulation to maintain energetic homeostasis. Alternatively, as the ribosome-enriched pool of **Mac1** may represent a non-starved phenotype, this could perhaps suggest that these TAMs do not reside within the centre of the tumour, and thus have greater nutrient availability. If this is true, and **Mac1** resides preferentially on the outside of the tumours, this might explain their increased presence in TNC-high tumours, as TNC is predominantly concentrated in outer matrix tracks. Thus, it might be that TNC arrests **Mac1** on the tumour periphery and prevents their migration and differentiation in the tumour core where they can present tumour antigens (**Mac2**). These speculative possibilities underscore the importance of resolving the spatial dimension of these macrophage phenotypes via imaging approaches in a follow up study.

Finally, in terms of TNC-independent findings, the pseudotime trajectory of ribosome-enriched **Mac1** moving towards chemokine/cytokine-producing **Mac3** (**Figure 3.10B**) revealed a sequential induction of factors such as the epigenetic histone demethylase, *Kdm6b*, and the kinase, *Pim1*, which immediately preceded (pathogenic) chemokine and cytokine upregulation. This potentially suggests the presence of a kinase-inhibitor-targetable (epigenetic) program regulating pathogenic **Mac3** polarization, with potential therapeutic relevance.

3.3.7 | TNC-dependent effects

There are several potential mechanisms by which tumour-derived TNC exerts its effects in the breast tumour microenvironment, and these can be divided into changes of either cellular abundance or individual gene expression.

At the cellular level, tumour-derived TNC appears to either induce or maintain a state of naivety in macrophages (**Mac1**), preventing their functional transition towards effector functions with an anti-tumoural phenotype, such as enhanced antigen presentation (**Mac2**), while not affecting the acquisition of likely pro-tumoural effector functions such as chemokine/cytokine production (**Mac3**). Previous work found that MHC2^{hi} macrophages (resembling **Mac2**) are modulated by the TNC-TLR4 axis (via the FBG-C domain), as blockade of the TNC-TLR4 interaction increased the abundance of **Mac2** in breast tumours (264). However, it remains possible that TNC's many other domains and binding sites (233) may also be involved in mediating TNC-dependent effects.

From a mechanistic perspective, there may also be parallels in the way that tumour-derived TNC alters the abundance of myeloid populations: **Mac1**, **Mac2**, **DC** and **MDSC**. For instance, there may be a conserved pathogenic mechanism by which tumour-derived TNC induces or maintains a naive-like phenotype in myeloid cells, which would explain the ~2-fold expansion of naive-like **Mac1** and inherently undifferentiated **MDSC** in tumours which have high levels of tumour-derived TNC. This notion is supported by the role of TNC in embryonic development, where it is known to generate a niche orchestrating neural stem cell development, while inhibiting cellular differentiation into neurons (362). Equally, given that **Mac2** and **DC** are both professional APCs with a highly antigen-presenting phenotype, their expansion in tumours with low levels of tumour-derived TNC may hint at a conserved mechanism through which TNC suppresses professional APCs. Indeed, APCs are also modulated by TNC at the level of individual gene expression, with marked downregulation of both classical and non-classical

MHC genes observed in TNC-high tumours. Thus, the immunologic roles of TNC in breast cancer converge on the inhibition of antigen presentation, at both the cellular and transcriptional level, implying that this mechanism is of key importance to TNC's pathogenic function.

Other genes regulated at the cellular level by tumour-derived TNC include the Epithelial-to-Mesenchymal transition (EMT) marker, vimentin (*Vim*), whose expression is upregulated by myeloid cells in TNC-high tumours (**Figure 3.12C**), aligning with the role of TNC in promoting breast tumour metastasis (250, 363). In addition, pan-myeloid downregulation of *Uba52* (a ubiquitin-fused ribosomal protein gene) in TNC-high tumours suggests that TNC may confer its pathogenic function by altering ribosome availability. Indeed, UBA52 deficiency is embryonic-lethal in mice, and UBA52-deficient cells display decreased protein synthesis and cell cycle arrest (364), perhaps similar to the arrest of naive-like **Mac1** observed in TNC-high tumours, which fail to acquire **Mac2** effector function. Finally, it is also worth noting that an increased frequency of *Ccr2*-expressing myeloid cells was observed in TNC-high tumours (**Figure 3.10**), perhaps suggesting that tumour-derived TNC may promote increased recruitment of $CCR2^{+}$ monocytes.

Given the range of novel and tractable findings arising from this scRNAseq analysis, the TNC-dependency of macrophage phenotypes was no longer the primary investigative focus. Instead, a combination of TNC-independent and TNC-dependent phenotypes will be taken forward for deeper evaluation in latter chapters.

3.3.8 | Summary & hypothetical pathogenic mechanism

In this chapter, scRNAseq revealed that TAMs delegate their effector functions to distinct subsets, and that tumour-derived TNC is able to modulate the abundance and transcriptional programs of myeloid subpopulations. The insights from this chapter hint at a potentially unifying pathogenic mechanism for myeloid cells in this model of breast cancer. First, a repertoire of chemokines produced by **Mac3** not only directly promote metastasis of tumour cells, but also facilitate the influx of **MDSC** into breast tumours, which then increases net TLR4 stimulation in the microenvironment due to secretion of high quantities of calprotectin, leading to chronic inflammation and induction of anergy/tolerance. In a possibly additive or synergistic effect, increased tumour-derived TNC in the microenvironment then exacerbates TLR4 stimulation, leading to further immune tolerance. Meanwhile, tumour-derived TNC induces or maintains a state of naivety in macrophages (**Mac1**), possibly via downregulation of *Uba52*, which prevents their functional transition towards a more anti-tumoural phenotype featuring enhanced antigen presentation (**Mac2**).

Considering the novelty and magnitude of the phenotypic effects described in this chapter, and their tractability for follow-up experiments, the most suitable finding to initially investigate was the functional delegation of macrophage effector functions to distinct subsets, which will be the focus of **Chapter 4**.

Chapter 4 | Validation and exploration of macrophage phenotypes *in vivo*

4.1 | Results

4.1.1 | Introduction

In the previous chapter, single-cell RNAseq analysis identified five phenotypically-distinct clusters of macrophages, comprising the majority of the intra-tumoural myeloid compartment. These data indicated that macrophage effector functions such as antigen-presentation and chemokine/cytokine production are delegated to functionally-specialised subsets (**Mac2** and **Mac3**, respectively)**.** The present chapter aims to validate and characterise this observation in greater depth by: **(a)** assessing functional delegation of macrophages at the protein level with improved statistical power, **(b)** isolating subset-specific macrophage populations for functional *in vitro* assays, **(c)** resolving macrophage subsets in spatial and temporal dimensions, and **(d)** disrupting the acquisition of macrophage effector functions via pharmacological intervention *in vivo*. To achieve this, a combination of flow cytometry, FACS-sorting, ELISA and confocal immunofluorescence imaging was employed.

This chapter will be divided into the following subsections:

4.1.2 Validation of the NT193 mammary tumour engraftment model

4.1.3 Validation of functionally-delegated macrophage effector states

4.1.4 Isolation of **Mac3** by FACS sorting

4.1.5 Temporal dynamics of functionally-delegated macrophage states

4.1.6 Temporal dynamics of monocyte-to-macrophage transition

4.1.7 Spatial distribution of functionally delegated macrophage states

4.1.8 *In vivo* alteration of functional-delegation by KDM6B: a **Mac3** regulator

4.1.2 | Validation of the NT193 mammary tumour engraftment model

To validate key phenotypes of the mammary tumour engraftment model previously published in the Midwood group (264), NT193 tumour cell co-cultures were first assessed for their expression of *Tnc* mRNA by RT-qPCR. Consistent with previous findings, strong knockdown of *Tnc* was observed in cells transduced with lentiviral shRNA constructs targeting *Tnc*, compared to a non-targeting control, indicating a retention of knockdown efficacy (**Figure 4.1A**). Light microscopy imaging of NT193 cells in culture confirmed a consistent morphology in both TNC-high and TNC-low conditions, characterized by large 'islands' of epithelial cells surrounded by dense strata of fibrous mesenchymal cells (**Figure 4.1B**).

To validate the *in vivo* phenotypes of the model, NT193 cells were orthotopically engrafted into the mammary gland of wildtype FVB mice and resultant tumours were assessed for their growth over time. Corroborating previous data, TNC-low tumours displayed significantly reduced growth compared to TNC-high tumours (**Figure 4.2A**). At Day 18 post-engraftment, tumours were harvested and analysed by flow cytometry to assess the immune-phenotype of infiltrating myeloid cells. The abundance of macrophages, defined as $CD45^{+}CD11b^{+}F4/80^{+}$ cells, were significantly reduced in TNC-low tumours, where they displayed increased expression of antigen-presentation (MHC1) and co-stimulation (CD86) machinery compared to TNC-high tumours, alongside a concomitant decrease in the immune-suppressive marker, CD206 (**Figure 4.2B**). Finally, to verify their phenotype at the transcriptional level, macrophages were enriched to a purity of >97% by FACS-sorting (**Figure 4.2C**) and assessed for their mRNA expression of various immune genes by RT-qPCR. Consistent with previous analyses at the protein level (**Figure 4.2B**), macrophages from TNC-low tumours expressed decreased immune-suppressive molecules, such as arginase (*Arg1*) and CD206 (*Mrc1*) compared to TNC-high tumours, while increasing expression of immune-stimulatory molecules such as *Cd86* and the inflammatory cytokine, *Tnf* (**Figure 4.2D**). Taken together,

these data corroborate the phenotypes previously observed in this model (264), thus validating it for subsequent use.

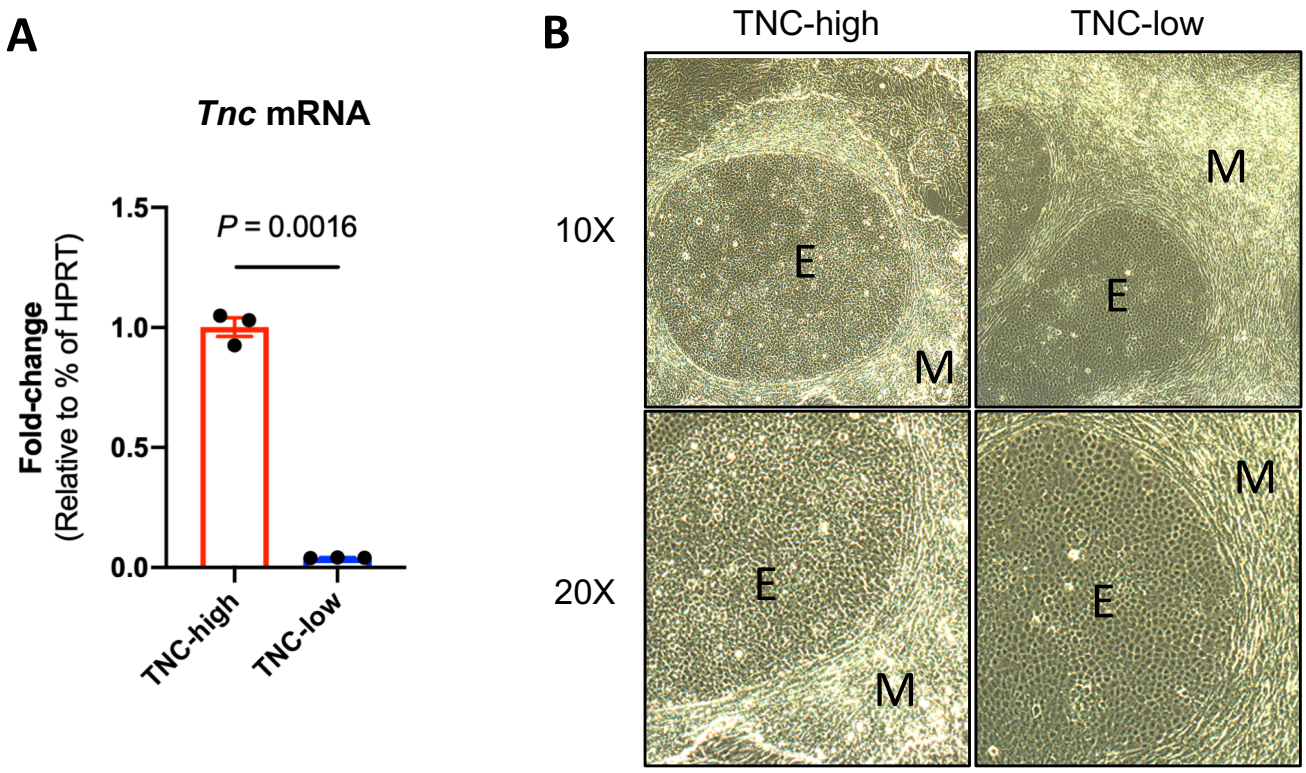


Figure 4.1 | Validation of TNC knockdown and morphology in NT193 co-cultures. (A) NT193 tumour cells transduced with lentiviral shRNA targeting either TNC (TNC-low) or a control targeting a gene not found in mammals (TNC-high) were assessed for *Tnc* gene transcription by RT-qPCR. Individual culture samples were run as technical triplicates for RT-qPCR, data were normalized to the housekeeping gene, *Hprt*, and statistical significance was assessed by performance of an unpaired two-tailed T test with Welch's correction (not assuming equivalent standard deviations between conditions). Error bars represent the standard deviation. **(B)** Morphology of NT193 tumour cell co-cultures at 10X or 20X magnification, demonstrating their characteristic formation of epithelial 'islands' (E) surrounded by dense strata of mesenchyme (M). Note: establishment of the NT193 model, including lentiviral transduction, was performed by the laboratory of Professor Gertraud Orend at the University of Strasbourg. Abbreviations: RT-qPCR = Real-Time Quantitative Polymerase Chain Reaction.

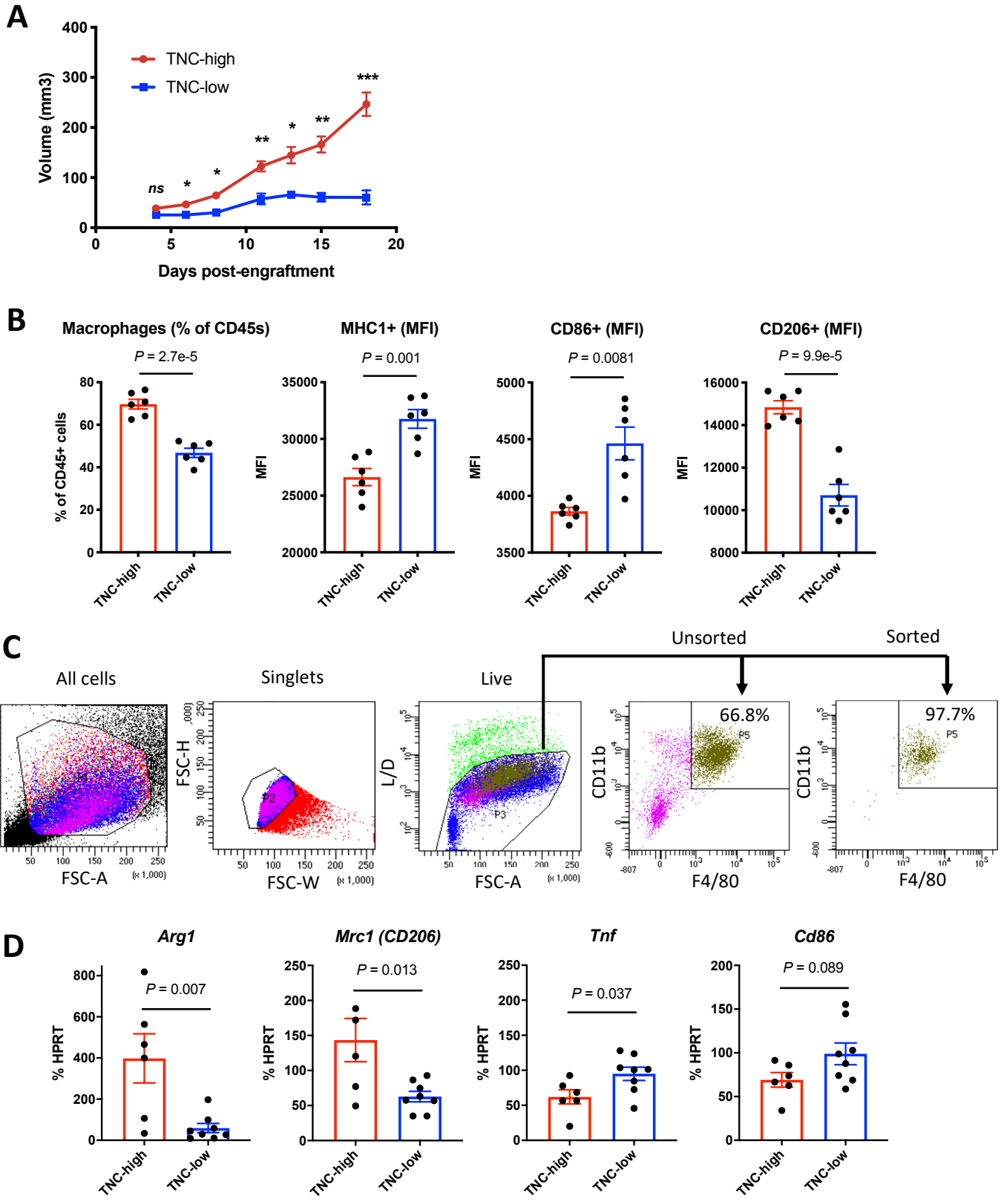


Figure 4.2 | Re-validation of orthotopic engraftment model phenotypes. Syngeneic NT193 tumour cells transduced with lentiviral shRNA targeting either *Tnc* (TNC-low, n=6) or a non-mammalian control (TNC-high, n=6) were orthotopically engrafted into the mammary gland of wildtype FVB mice for 18 days. **(A)** Tumour growth was monitored over time and a two-way repeated measures ANOVA was performed on the mean of each TNC condition per timepoint, assuming no sphericity (applying the Geisser-Greenhouse correction) and adjusting for multiple comparisons using the Sidak method.

(*continued on next page*).

Figure 4.2 (continued). * P < 0.05; ** P < 0.01; *** P < 0.001; ns = not significant. **(B)** Expression of phenotypic immune markers was performed by multi-parameter flow cytometry. A single TNC-low tumour regressed prior to Day 18 and was not included in subsequent FACS analysis. An unpaired T test with Welch's correction was performed for all conditions. **(C-D)** Intra-tumoural macrophages were isolated by FACS sorting $CD45^+CD11b^+CD3^-F4/80^+$ cells from TNC-high (n=6) and TNC-low (n=8) engrafted NT193 tumours. **(C)** The gating strategy is shown with a single representative tumour as an example, and percentages are presented as a proportion of the parent $CD45^+$ population. **(D)** Immune gene transcription was assessed by RT-qPCR on FACS-sorted macrophages, and data was normalized to the housekeeping gene, *Hprt*, followed by performance of an unpaired two-tailed T test with Welch's correction (not assuming equivalent standard deviations between conditions). Error bars represent the standard error of the mean. Acknowledgment: RT-qPCR was performed by Dr Martha Gschwandtner, and FACS-sorting was performed by Dr. Jonathan Webber, both at the Kennedy Institute of Rheumatology (University of Oxford). Abbreviations: FACS = Fluorescence-Activated Cell Sorting, RT-qPCR = Real-Time Quantitative Polymerase Chain Reaction.

4.1.3 | Validation of functionally-delegated macrophage effector states

To confirm the existence of functionally-delegated TAM effector states at the protein level, I reasoned that the most extreme phenotypes of any given effector state – represented by the effector genes most highly expressed at the terminus of its diffusion map trajectory in the scRNAseq analysis (**Figures 4.3A-B & 3.13-3.14**) – should be mutually exclusive to any other effector state phenotype. Two opposing macrophage effector states were selected to test this – antigen-presentation-enriched (**Mac2**) and cytokine/chemokine-enriched (**Mac3**) – as the former was implicated in mediating an anti-tumoural effect in our previous investigation of the NT193 model (264), and the latter produces a milieu of chemokines with known pro-metastatic roles in breast cancer (251, 307-311). The hypothesis was therefore that these opposing macrophage phenotypes would be non-overlapping, leading to single-positive **Mac2** or single-positive **Mac3** populations, as illustrated in **Figure 4.3C**. To confirm that the **Mac2** and **Mac3** marker genes selected by diffusion map analysis were indeed discriminatory markers, differential expression and Receiver Operating Characteristic (ROC) analyses were performed across all macrophage clusters, demonstrating sufficiently high cluster-specific upregulation and Area Under the Curve (AUC) values for each marker (AUC > 0.7) (**Figure 4.3D**).

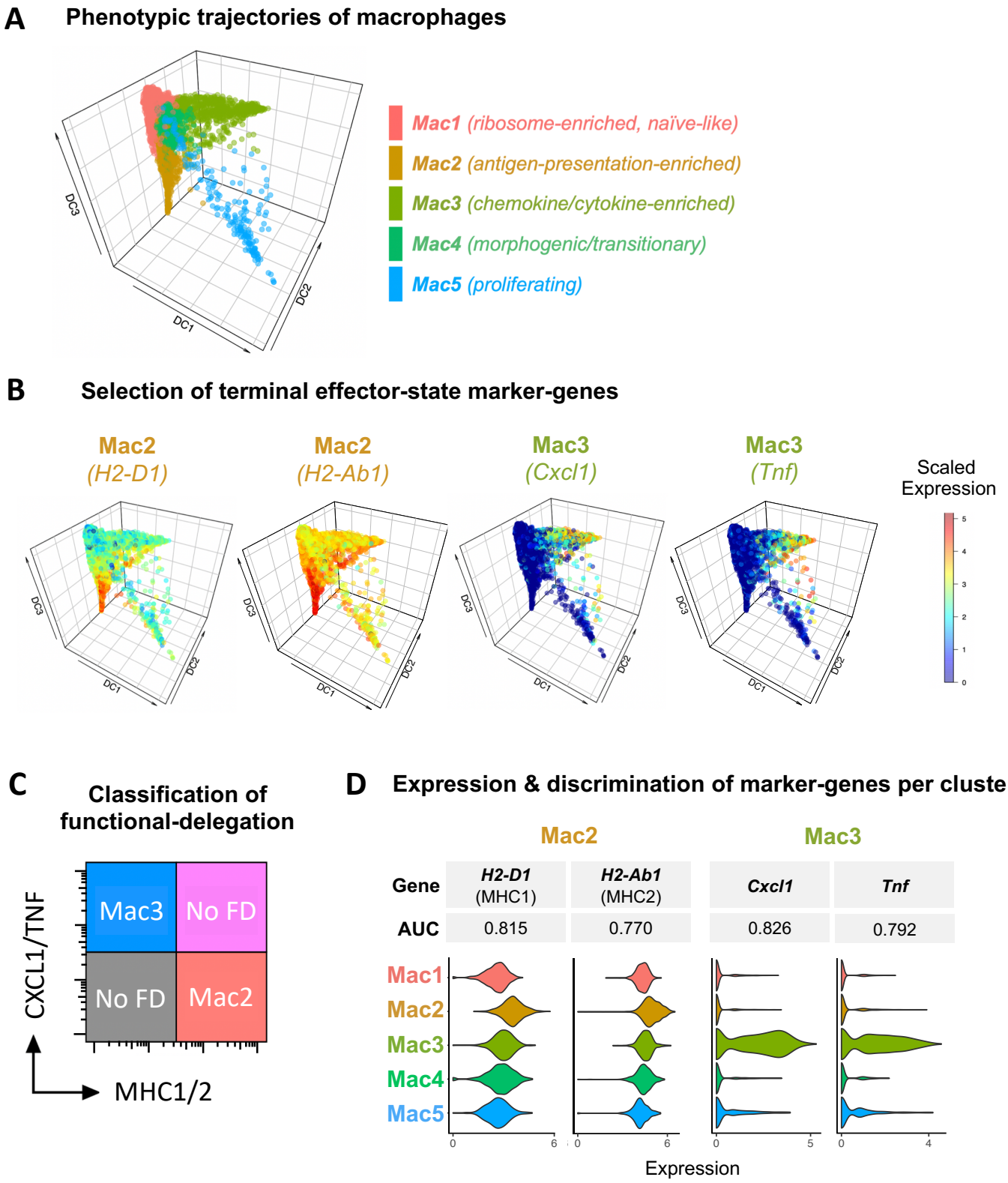


Figure 4.3 | Selection of marker genes to detect functionally-delegated macrophage effector states. Re-analysis of single-cell RNAseq data from **Chapter 3**. **(A)** Phenotypic trajectories of macrophages by diffusion mapping. Adapted from **Figure 3.9**. **(B)** Expression of selected marker genes of **Mac2** and **Mac3** representing the termini of macrophage effector states. **(C)** Classification of functional delegation on a schematic FACS plot. **(D)** Expression and discrimination of selected marker genes per cluster by violin plot, including Area Under the Curve (AUC) values following performance of Receiver Operating Characteristic (ROC) analysis on all macrophage clusters. Abbreviations: FACS = Fluorescence-Activated Cell Sorting, FD = Functional Delegation

Given that antibodies targeting the **Mac2** markers, MHC1 (*H2-D1*) and MHC2 (*H2-Ab1*), had previously been validated on tumour macrophages in this model (264), and demonstrated a TNC-dependent phenotype, both markers were selected to represent **Mac2**. For the **Mac3** markers, CXCL1 and TNF, prior validation of antibodies for flow cytometry had not been performed. In order to validate **Mac3** flow cytometry antibodies, a positive control cell-type with high expression of both markers was required. Bone-marrow derived macrophages (BMDMs) were stimulated with LPS for 24 hours and assessed for CXCL1 and TNF protein in the supernatant by ELISA (**Figure 4.4A**). Dose-dependent induction of both proteins was observed, demonstrating that BMDMs could serve as a positive control for assessment of both anti-CXCL1 and anti-TNF flow cytometry antibodies.

However, fortuitously, the availability of surplus tumour tissue from a NT193 tumour engraftment (1 TNC-high and 1 TNC-low tumour, pooled) enabled the performance of a preliminary optimisation experiment assessing these flow cytometry antibodies in tumour macrophages, rather than in BMDMs, as per the gating strategy ($CD45^+CD11b^+F4/80^+L/D^-$) illustrated in **Figure 4.4B**. Intracellular staining of CXCL1 and TNF was performed using the high end of the titration limit recommended by the manufacturer. Both CXCL1 and TNF achieved a positive signal compared to isotype or Fluorescence Minus One (FMO) controls (**Figure 4.4C-D**), meaning BMDM validation by flow cytometry was no longer required, as CXCL1 and TNF expression had already been validated in NT193-tumour-derived macrophages. The FMO control was used for TNF in this initial optimisation experiment as the isotype control was unavailable. However, the staining in tumour macrophages was able to more distinctly resolve the CXCL1-positive population from the CXCL1-negative population (**Figure 4.4C**), whereas the TNF-positive population was more poorly separated from the TNF-negative population by comparison, albeit with a higher total proportion of TNF-positive cells detected (**Figure 4.4D)**.

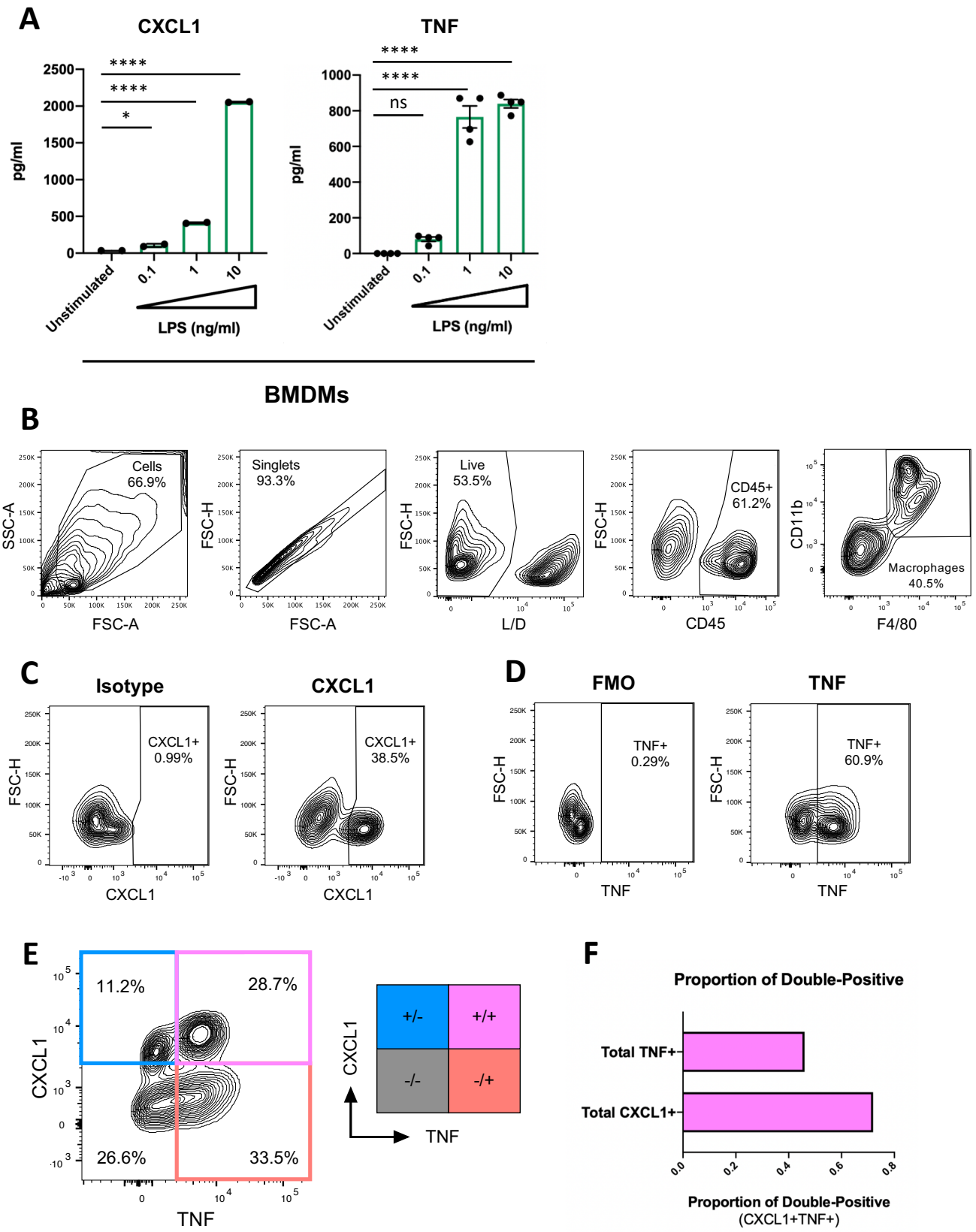


Figure 4.4 | Macrophage delegation-marker optimization (A) Expression of soluble CXCL1 or TNF in the supernatant of mouse BMDMs stimulated with a titration of LPS for 24h. For CXCL1, supernatants from 4 wildtype FVB mice were pooled for ELISA and analysed as technical duplicates. For TNF, supernatants from each mouse were run independently as four biological replicates. Ordinary one-way ANOVA was performed with Dunnett's multiple testing correction, comparing the mean of each condition to the unstimulated control. (*continued on next page*).

Figure 4.4 (continued). * P < 0.05; ** P < 0.01; *** P < 0.001; **** P < 0.0001, ns = not significant. Error bars represent the standard deviation. **(B-F)** Syngeneic NT193 tumour cells transduced with lentiviral shRNA targeting either *Tnc* (TNC-low, n=1) or a non-targeting control (TNC-high, n=1) were orthotopically engrafted into the mammary gland of wildtype female FVB mice for 18 days before tumours were harvested, pooled, and analysed by flow cytometry. **(B)** Gating strategy for flow cytometric analysis of live intra-tumoural macrophages ($CD45^+CD11b^+F4/80^+L/D^-$). **(C)** Validation of intracellular CXCL1 expression in macrophages, by comparison to an isotype control **(D)** Validation of intracellular TNF expression in macrophages by comparison to a FMO control (isotype unavailable at the time). **(E)** Co-expression of CXCL1 and TNF in tumour macrophages, with the pink gate representing the overlapping double-positive population, broadly representative of **Mac3**. **(F)** Enumeration of **Mac3** marker specificity, showing the proportion of CXCL1-positive cells that were co-positive for TNF, and the proportion of TNF-positive cells that were co-positive for CXCL1. Acknowledgment: Differentiation of BMDMs used in panel A was performed by Dr Anja Schwenzer at the Kennedy Institute of Rheumatology (University of Oxford). Abbreviations: BMDM = Bone Marrow-Derived Macrophage, ELISA = Enzyme-Linked Immunosorbent Assay, FMO = Fluorescence Minus One.

To determine whether CXCL1 or TNF varied in their quality as a **Mac3** marker, co-expression of both CXCL1 and TNF was assessed. The proportion of double-positive ($CXCL1^+TNF^+$) macrophages resembling **Mac3** were a similar proportion to that observed in the scRNAseq analysis (**Figure 3.11C, Appendix Figure 3.4A**), indicating reasonable consistency in **Mac3** detection across technologies and providing confidence that this subset was successfully detected (**Figure 4.4E**). Given that macrophage expression of *Cxcl1* and *Tnf* was exclusive to **Mac3** at the transcriptional level (**Figure 3.10**), the intersecting population of double-positive ($CXCL1^+TNF^+$) cells at the protein level was expected to be high, with minimal single-positivity for either marker. However, substantial differences between the markers was observed. The total TNF-positive compartment was composed of a relatively large proportion of CXCL1-negative cells, whereas for the total CXCL1-positive compartment, a higher proportion were co-positive for TNF (**Figure 4.4F**), indicating that CXCL1 staining marked a narrower population of cells and more accurately represented **Mac3**. On this basis, CXCL1 was selected as the **Mac3** marker for subsequent experiments.

After selecting subset-specific macrophage markers representing two key effector functions, and validating the tools required for flow cytometry-based detection, the next challenge was to assess whether these two seemingly opposing phenotypes are indeed segregated. To validate that **Mac2** and **Mac3** effector functions are delegated to discrete non-overlapping populations, co-staining of markers for both subsets was performed in tumour macrophages assessed by flow cytometry at 18 days post-engraftment (n=6 each of TNC-high and TNC-low). This revealed a segregation of **Mac2** (MHC1/2^{hi}CXCL1^{-}) and **Mac3** (MHC1/2^{lo}CXCL1^{+}) phenotypes (**Figure 4.5A**). Comparison of the double-positive overlapping (MHC1/2^{hi}CXCL1^{+}) population versus the single-positive functionally-delegated populations (**Mac2** or **Mac3**) demonstrated highly significant differences, indicating that these states co-exist relatively infrequently (**Figure 4.5B**).

To compare these 'delegated' macrophage phenotypes with classical macrophage phenotypic markers, two further surface markers were assessed: the co-stimulatory molecule, CD86, and the immune-suppressive macrophage scavenger receptor, CD206. A similar trend of phenotypic segregation occurred for the co-stimulatory molecule, CD86, as was observed for MHC1 and MHC2, featuring virtually no double-positive populations (CD86^{+}CXCL1^{+}) but high enrichment of single-positive populations (CD86^{+}CXCL1^{-} or CD86^{-}CXCL1^{+}), indicating that the **Mac2** phenotype is not MHC-specific, but rather representative of antigen-presentation capacity more broadly (**Figure 4.5B**). Interestingly, despite minimal detection at the transcriptional level (**Figure 3.10**), the macrophage scavenger receptor, CD206, followed a similar trend of functional segregation at the protein level (high single-positive but low double-positive), albeit with an additionally large double-negative 'undelegated' population, which was not observed for any other marker (**Figure 4.5B**). This potentially indicates that the dynamics of functional delegation may change depending on the effector function – in this case, CD206-relevant functions such as mannose-scavenging or phagocytic/endocytic-like processes.

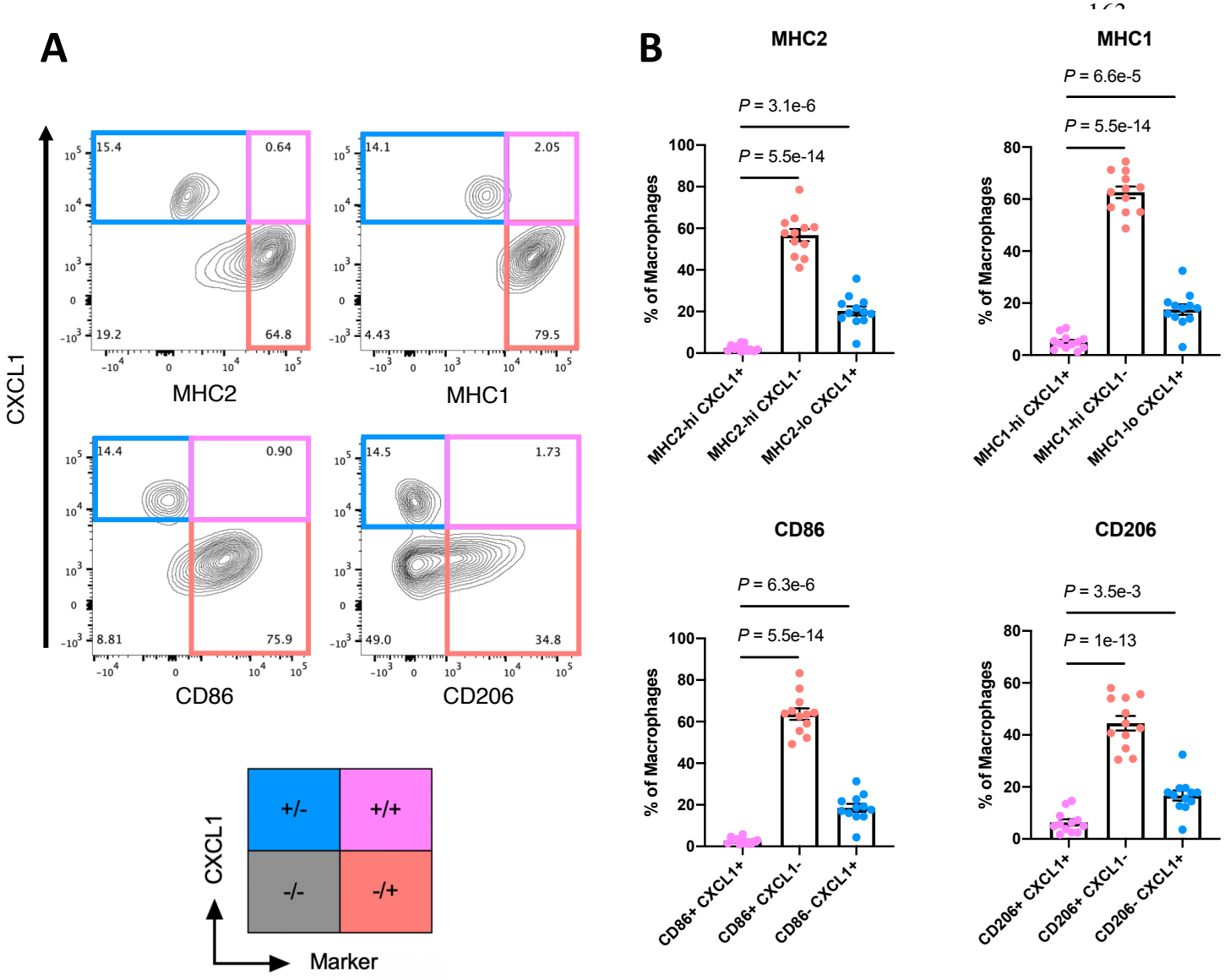


Figure 4.5 | Functional delegation of chemokine-high (Mac3-like) and antigen-presentation-enriched (Mac2-like) macrophages. Syngeneic NT193 tumour cells transduced with lentiviral shRNA targeting either *Tnc* (TNC-low, n=6) or a non-targeting control (TNC-high, n=6) were orthotopically engrafted into the mammary gland of wildtype female FVB mice for 18 days before tumours were harvested. **(A)** Intra-tumoural macrophages were assessed by flow cytometry (gating: $CD45^{+}CD11b^{+}F4/80^{+}$) for expression of surface and intracellular antigens resembling functionally-delegated states: **Mac2** (MHC1, MHC2) or **Mac3** (CXCL1), alongside the classical phenotypic markers, CD86 and CD206. A single representative tumour is shown for all markers. All data was collected at 18 days post-engraftment. **(B)** Comparison of functionally-delegated (single-positive) versus non-delegated (double-positive) macrophages. All samples are shown irrespective of TNC status to provide an average of macrophage states across the NT193 model of breast cancer. Singular function-independent (double-negative) cells are excluded from this analysis. One-way ANOVA of all-vs-all subsets was performed without assuming equivalent standard deviations between groups (using the Brown-Forsythe & Welch test), followed by adjustment for multiple comparisons using the Dunnett T3 correction. Error bars represent the standard error of the mean. Abbreviations: ANOVA = Analysis of Variance.

In summary, these data confirm that specific macrophage effector functions of antigen-presentation (**Mac2**) and chemokine production (**Mac3**) are indeed delegated to functionally-specialised subsets, as initially indicated by single-cell RNAseq.

4.1.4 | Isolation of Mac3 and Mac2 by FACS-sorting

To characterise functionally delegated macrophage subsets in greater depth, the ability to selectively isolate or deplete macrophage populations of interest would be advantageous. To facilitate this, surface-membrane-localised marker genes are required to simplify isolation of cell subsets by FACS-sorting or depletion by antibody-mediated cellular cytotoxicity (ADCC). Suitable surface markers for **Mac2** (MHC1/MHC2) had serendipitously been validated in our previous investigation of the NT193 model of breast cancer (264), however this was not the case for **Mac3**, and consequently, surface-marker triaging was required for this subset. The benefit of validating surface-surrogate markers for **Mac3** was not just to aid isolation and depletion strategies, but also to simplify downstream flow cytometry and tissue immunofluorescence imaging protocols, as intracellular stains for CXCL1 could be avoided.

To triage surface markers for **Mac3**, scRNAseq data from **Chapter 3** was re-visited. Of the top differentially expressed genes discriminating **Mac3** from all other macrophage subsets, only a small proportion were surface-localized, accounting for just 1 gene in the top 25, and just 7 of the top 38, when ranked by their statistical significance in differential expression (**Figure 4.6A**). Four of these genes were selected for further analysis. Triaging of surface markers was performed on the basis of differential expression rank (average log2 fold-change), cluster expression level and exclusivity, visualised by violin plot distributions (**Figure 4.6B**), as well as the availability of validated flow cytometry antibodies from reputable sources. This strategy led to the selection of CD83, CCRL2, PLAUR and TLR2 for further investigation as **Mac3** surface marker candidates. The remaining three markers were discounted for the following reasons: *Sdc4* expression was excessively high in one other cluster (**Mac4**), *Cd14* was highly differentially expressed but had high expression frequency in all other clusters (ie lower cluster-exclusivity), and *Plek* was expressed marginally lower than the next best alternative, *Plaur*, as well as having comparatively higher expression in one cluster (**Mac5**).

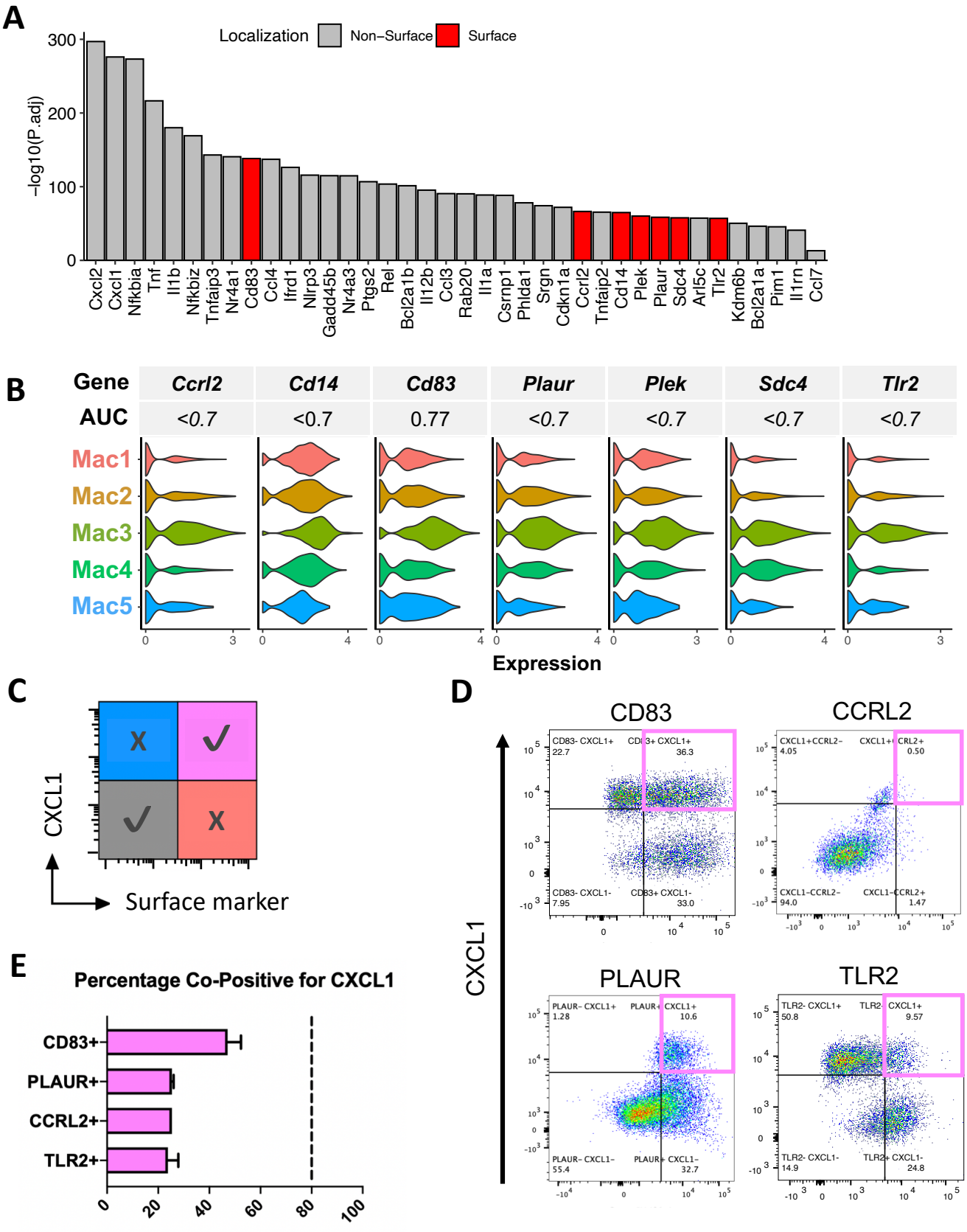


Figure 4.6 | Triaging of surface-localised candidate marker genes for isolation or depletion of Mac3. **(A-B)** Re-analysis of single-cell RNAseq data from **Chapter 3**. **(A)** Top differentially expressed marker genes discriminating **Mac3** from all other macrophage clusters. A non-parametric Wilcoxon Rank sum test was performed to calculate statistical significance, and P values were adjusted for multiple comparisons with the Bonferroni correction (P.adj). (*continued on next page*)

Figure 4.6 (continued). Only genes with an average log2 fold-change of >0.6 compared to all other clusters are shown. Proteins that are surface-localized to the plasma membrane, and are thus potential surface candidate markers, are indicated. **(B)** Violin plots of **Mac3** candidate surface marker gene expression per macrophage cluster, including Area Under the Curve (AUC) values following performance of Receiver Operating Characteristic (ROC) analysis on all macrophage clusters. **(C)** Schematic of desired Mac3 surface marker expression profile, demonstrating minimal single-positive populations and strong overlapping double-positive (**Mac3**), or double-negative (non-**Mac3**) populations. **(D-E)** Syngeneic NT193 tumour cells transduced with lentiviral shRNA targeting either *Tnc* (TNC-low, n=2) or a non-targeting control (TNC-high, n=2) were orthotopically engrafted into the mammary gland of wildtype female FVB mice for 18 days, before tumours were harvested and pooled irrespective of TNC condition. **(D)** Flow cytometric analysis of surface candidate markers of **Mac3** in combination with intracellular CXCL1, gated on live macrophages only ($CD45^+CD11b^+F4/80^+L/D^-$ cells). CXCL1 gating was determined by comparison to an isotype control, as per **Figure 4.4C**. **(E)** Enumeration of **Mac3-**surface marker candidates that are co-positive for intracellular CXCL1. Dotted line represents the threshold for co-positivity required to be considered a valid **Mac3**-surface marker. Samples were run in technical duplicates and error bars represent the standard deviation.

The ideal expression profile of a **Mac3** surface marker is one that strongly overlaps with the intracellular **Mac3** marker, CXCL1, and therefore has high double-positivity and low single-positivity, as illustrated by the schematic in **Figure 4.6C**. Expression of each surface candidate marker was therefore assessed for its co-positivity with intracellular CXCL1 in NT193 tumour-derived macrophages by flow cytometry. These data revealed high variability in the proportion of CXCL1 co-positivity per surface marker (**Figure 4.6D**), and demonstrated that no single candidate gene exhibited sufficiently high co-positivity with CXCL1 to be considered a valid surface marker of **Mac3** (**Figure 4.6E**).

To determine whether **Mac3** could be enriched by an alternative gating strategy, flow cytometry data from **Figure 4.5** was re-analysed to determine whether CXCL1 was asymmetrically distributed in macrophage subpopulations (**Figure 4.7A**). Given that MHC1 and MHC2 previously exhibited a segregated expression profile compared with CXCL1 (**Figure 4.5**), macrophages expressing lower MHC1/2 were assessed for their CXCL1 content, and were observed to harbour a subpopulation of strongly CXCL1-expressing cells.

Assessment of CXCL1 distributions within this gate revealed a proclivity for $CD11b^{hi}$ cells to express high CXCL1, as well as cells with a high side-scatter-area (SSC-A) profile, indicative of increased cellular granularity. Together, these observations formed the basis of an alternative gating strategy ($CD45^{+}CD11b^{hi}MHC1/2^{lo}SSC^{hi}$) hypothetically capable of enriching **Mac3** $CXCL1^{+}$ macrophages up to ~97% purity, according to previous staining data (**Figure 4.7A**).

To test whether a high purity of **Mac3** enrichment could be achieved in practise, FACS-sorting of NT193 tumour suspensions was performed according to this gating strategy, isolating $MHC1/2^{lo}CD11b^{hi}SSC^{hi}$ macrophages (**Mac3**) alongside a control population of $MHC1/2^{hi}$ cells, resembling a **Mac2**-like state (**Figure 4.7B**). The gate-purity of this isolation was high for **Mac2**-like cells, at >95%, however the gate-purity was substantially lower for **Mac3** due to the nature of the separation of a normal (SSC) rather than bimodal (protein marker) distribution, as explained further in the Technical Discussion (**Section 4.2**). The efficacy of **Mac3**-sorting with the novel gating strategy was then assessed by subsequent intracellular staining of FACS-sorted populations for CXCL1. This demonstrated that gating of $MHC1/2^{lo}$ $CD11b^{hi}$ SSC^{hi} macrophages successfully enriched the $CXCL1^{+}$ subpopulation (**Mac3**) to a purity of ~91%, compared to a baseline of ~11% in unsorted tumour macrophages, thus representing an ~8-fold enrichment (**Figure 4.8A**). The control population of $MHC1/2^{hi}$ sorted macrophages contained just ~4% $CXCL1^{+}$ cells, indicating relative depletion of this population compared to baseline.

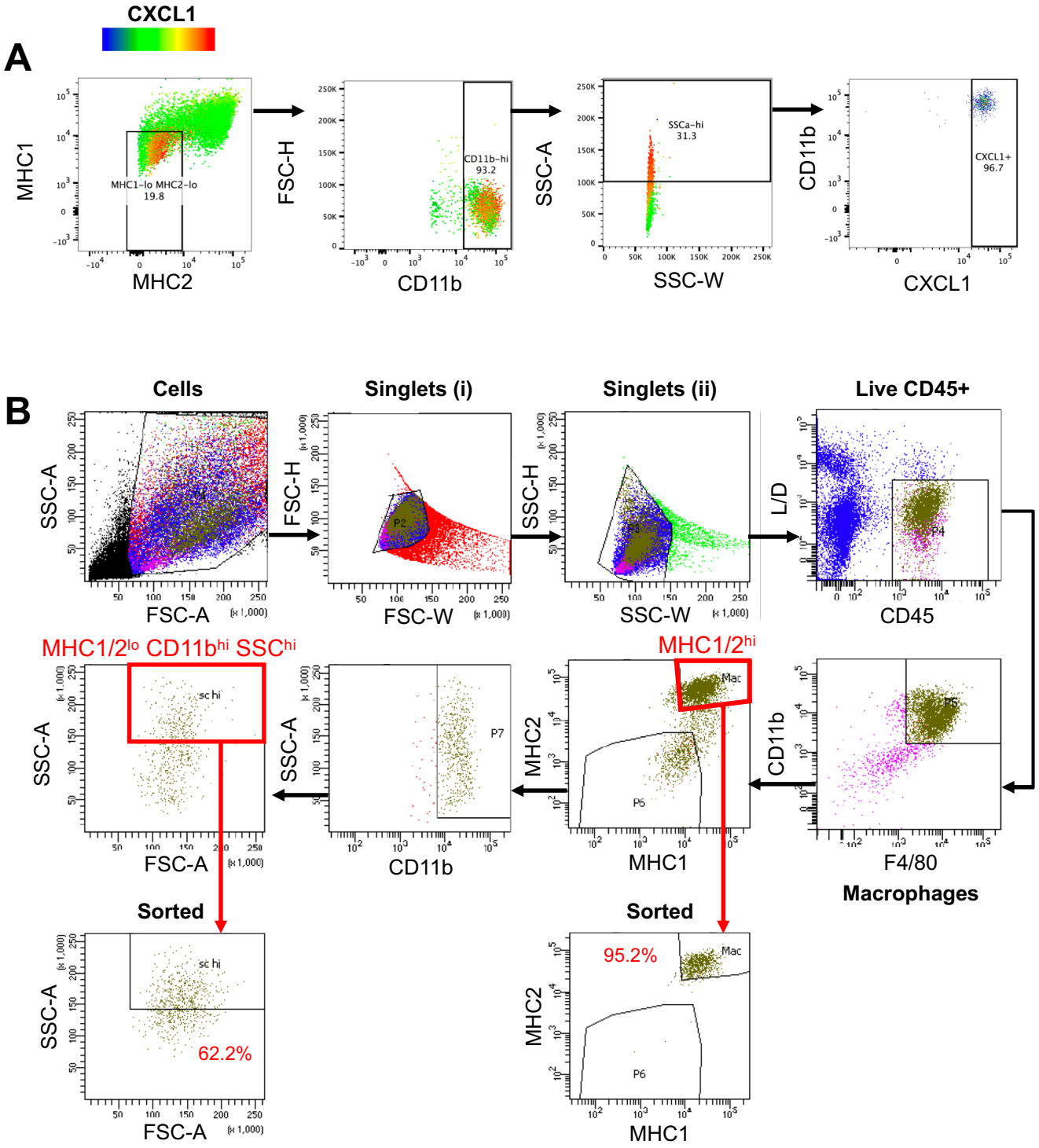


Figure 4.7 | Alternative gating strategy for the isolation of Mac3 by FACS-sorting (A) Re-analysis of previous flow cytometry data from **Figure 4.5**, showing only live macrophages ($CD45^+CD11b^+F4/80^+L/D^-$ cells) from harvested NT193 breast tumours. This novel gating strategy was devised for the isolation of **Mac3**, showing the disproportionate enrichment of $CXCL1^+$ macrophages within a $MHC1/2^{lo}CD11b^{hi}$ SSC^{hi} subset of macrophages. One representative tumour is shown. **(B)** Syngeneic NT193 tumour cells transduced with lentiviral shRNA targeting either *Tnc* (TNC-low, n=5) or a non-targeting control (TNC-high, n=4) were orthotopically engrafted into the mammary gland of wildtype female FVB mice for 18 days, before tumours were harvested and pooled irrespective of TNC condition. FACS-sorting of tumours using the novel gating strategy in **(A)** was used to discriminate **Mac2**-like ($MHC1/2^{hi}$) versus **Mac3**-like ($MHC1/2^{lo}CD11b^{hi}SSC^{hi}$) populations. The singular pooled sample is shown. The two bottom plots represent a purity check of the sorted **Mac2**-like and **Mac3**-like samples. Percentages represent the amount of cells relative to the parent gate. Abbreviations: FACS = Fluorescence-Activated Cell Sorting, SSC = Side Scatter.

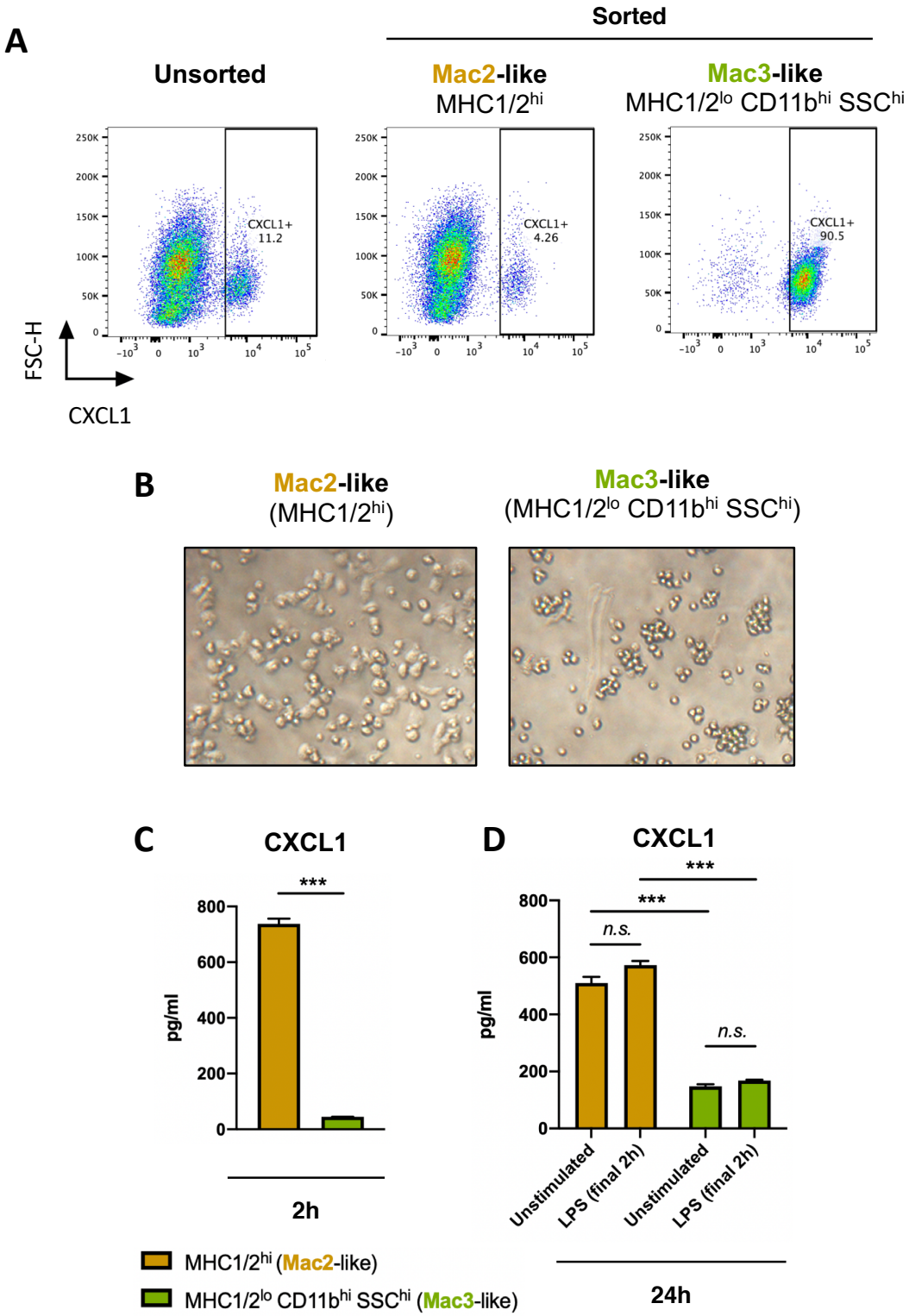


Figure 4.8 | Validation of post-sorting Mac3 phenotypes. Syngeneic NT193 tumour cells transduced with lentiviral shRNA targeting either *Tnc* (TNC-low, n=3) or a non-targeting control (TNC-high, n=4) were orthotopically engrafted into the mammary gland of wildtype female FVB mice for 18 days, before tumours were harvested and pooled irrespective of TNC condition. FACS-sorting of tumours to enrich **Mac3**-like (MHC1/2^{lo}CD11b^{hi}SSC^{hi}) cells was performed, alongside a control population of **Mac2**-like (MHC1/2^{hi}) cells for comparison. **(A)** Sorted macrophages populations and the unsorted tumour suspension were assessed for the presence of macrophages expressing CXCL1 by intracellular flow cytometry, validating the enrichment of CXCL1^{+} cells in the **Mac3**-like sorted population and the depletion of CXCL1^{+} cells in the **Mac2**-like sorted population, relative to the unsorted suspension. Plots are gated on live macrophages (CD45^{+}CD11b^{+}F4/80^{+}L/D^{-}). (*continued on next page*)

Figure 4.8 (continued). Percentages represent the amount of cells in the CXCL1-positive gate relative to the total population of macrophages. **(B)** Light microscopy imaging of FACS-sorted **Mac2**-like and **Mac3**-like populations cultured for 2 hours in 96-well flat-bottom plates at 37°C. Images were acquired at 20X magnification. **(C)** ELISA for soluble CXCL1 protein in the supernatants of FACS-sorted population cultures incubated for 2 hours at 37°C. Individual culture supernatants were run as technical duplicates, and statistical significance was assessed with an unpaired T test using Welch's correction. **(D)** ELISA for soluble CXCL1 protein in the culture supernatants of FACS-sorted population after 24 hours in culture, with or without LPS stimulation for the final 2 hours. Individual culture supernatants were run as technical duplicates and statistical significance was assessed with multiple unpaired T tests corrected for multiple comparisons using the Holm-Sidak method. All error bars represent the standard deviation. *** $P<0.001$, n.s. = not significant. Acknowledgment: FACS-sorting was performed by Dr. Jonathan Webber at the Kennedy Institute of Rheumatology (University of Oxford). Abbreviations: ELISA = Enzyme-Linked Immunosorbent Assay, FACS = Fluorescence-Activated Cell Sorting, LPS = Lipopolysaccharide.

To confirm that FACS-sorted **Mac3** cells secrete CXCL1, rather than merely sequestering it within intracellular granules, FACS-sorted cells were cultured *in vitro* for 2 hours before images of the cells were acquired, and supernatants were harvested for CXCL1 ELISA. Light microscopy imaging revealed differences in the size, morphology and clustering characteristics of FACS-sorted macrophage populations, with CXCL1-enriched **Mac3** cells being smaller, rounder, and predominantly forming large clusters (**Figure 4.8B**). Unexpectedly, ELISA analysis of culture supernatants revealed that soluble (secreted) CXCL1 protein levels were ~16-fold lower in the **Mac3** supernatant compared to **Mac2** (**Figure 4.8C**), despite intracellular flow cytometry data indicating the inverse relationship.

Given the high granularity of **Mac3**, evidenced by its high side-scatter profile, I reasoned that cellular activation might be required to release any potential CXCL1-containing granules, and that the extent of CXCL1 secretion may be more apparent after increased time in culture. A fresh batch of FACS-sorted **Mac3** and **Mac2** populations were therefore cultured for 24 hours, with or without LPS stimulation for the final 2 hours. A similarly inverse relationship was observed in both the presence and absence of LPS stimulation, with CXCL1 levels significantly higher in **Mac2** compared with **Mac3** supernatants, though this effect was

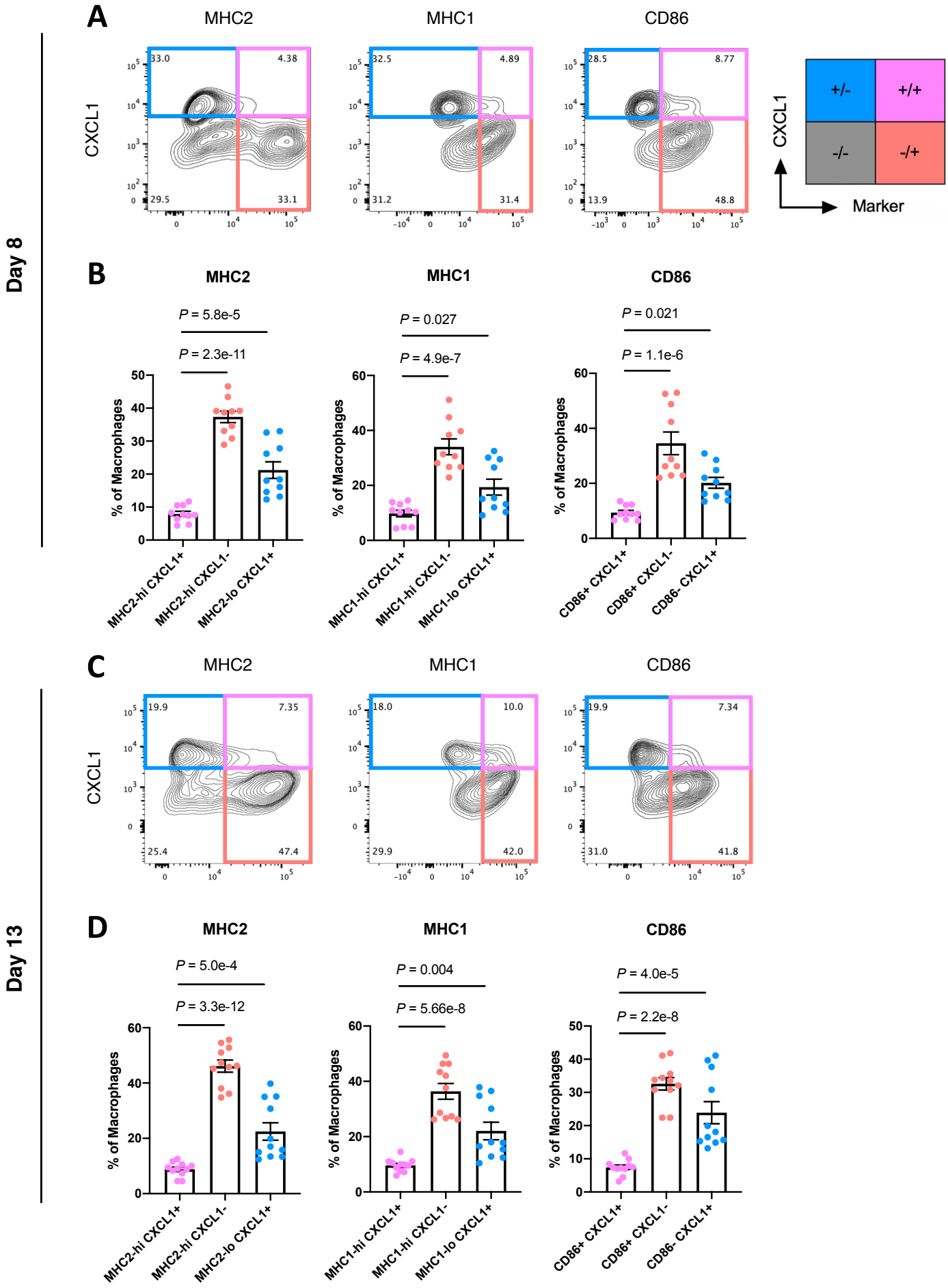
considerably weaker at just ~3.5-fold elevation. This indicates that activating stimuli are not required for **Mac3** to release intracellular CXCL1 granules, instead suggesting that *in vitro* culturing drastically alters macrophage phenotypes on short timescales of less than 2 hours (see **Section 4.2** Technical Discussion for further commentary).

Given that FACS-sorted macrophage phenotypes could not be maintained long enough *ex vivo* to perform functional assays and characterisation, macrophage isolation was deprioritised in favour of direct *in vivo* characterisation and modulation approaches.

4.1.5 | Temporal dynamics of functionally-delegated macrophage states

To understand the dynamics of macrophage effector-function acquisition over time, TNC-high and TNC-low tumours were harvested at Days 8, 13 and 18 for flow cytometric analysis. The earliest timepoint of Day 8 was selected because, before this time, tumours were palpable but not yet solid and fibrous, meaning they are likely to disintegrate upon harvesting.

To assess the general behaviour of functionally delegated macrophages over time, samples were analysed irrespective of TNC status. At both Day 8 (**Figure 4.9A**) and Day 13 (**Figure 4.9C**), **Mac2** and **Mac3** markers were strongly segregated into single-positive populations with limited overlap, similar to previous observations at Day 18 (**Figure 4.5A**), indicating that establishment of functional-delegation occurs early in tumour development, at least preceding Day 8. Assessment of the kinetics of functional delegation revealed that the relative abundance of **Mac2** consistently and significantly increased over time, compared to Day 8, whereas the 'non-delegated' double-negative macrophages decreased (**Figure 4.10**). This is consistent with the notion of a naive-like subset (**Mac1**) acquiring an antigen-presentation-enriched effector phenotype (**Mac2**). Supporting this, the proportion of double-negative cells (expected to be predominantly naive-like **Mac1**) consistently decreased over time.



(*figure legend on next page*)

Figure 4.9 | Macrophage functional-delegation exists at earlier stages of tumour progression. Syngeneic NT193 tumour cells transduced with lentiviral shRNA targeting either *Tnc* (TNC-low) or a non-targeting control (TNC-high) were orthotopically engrafted into the mammary gland of wildtype female FVB mice before tumours were harvested at either 8 days **(A,B)** or 13 days **(C,D)** post-engraftment. At Day 8, a total of 5 TNC-low and 5 TNC-high tumours were assessed. At Day 13, a total of 5 TNC-low and 6 TNC-high tumours were assessed. Intra-tumoural macrophages were assessed by flow cytometry (gating: $CD45^{+}CD11b^{+}F4/80^{+}$) for expression of surface and intracellular antigens resembling functionally-delegated states: **Mac2** (MHC1, MHC2) or **Mac3** (CXCL1), alongside the classical phenotypic markers, CD86 and CD206. A single representative tumour is shown for all markers. Comparison of functionally-delegated (single-positive) versus non-delegated (double-positive) macrophages. All samples are shown irrespective of TNC status to provide an average of macrophage states across the NT193 model of breast cancer. Singular function-independent (double-negative) cells are excluded from this analysis. One-way ANOVA of all-vs-all subsets was performed without assuming equivalent standard deviations between groups (using the Brown-Forsythe & Welch test), followed by adjustment for multiple comparisons using the Dunnett T3 correction. Error bars represent the standard error of the mean. Abbreviations: ANOVA = Analysis of Variance.

To determine the TNC-dependent effects on functionally delegated macrophages over time, TNC-high and TNC-low conditions were compared at each timepoint for every macrophage subset. Interestingly, the proportion of double-positive macrophages co-expressing **Mac2** and **Mac3** markers diverged in a time- and TNC-dependent manner consistently across all **Mac2** markers (**Figure 4.10**). This progressive weakening of the **Mac2**/**Mac3** segregation over time, specifically in the presence of TNC, suggests that TNC may be driving a cumulatively dysfunctional delegation of macrophage effector functions over time. Furthermore, the proportion of double-negative macrophages ($MHC2^{lo}CXCL1^{-}$), which are expected to be predominantly **Mac1**, were significantly higher in TNC-high compared to TNC-low tumours, corroborating the accrual of **Mac1** previously observed in TNC-high tumours by scRNAseq (**Figure 3.11C**).

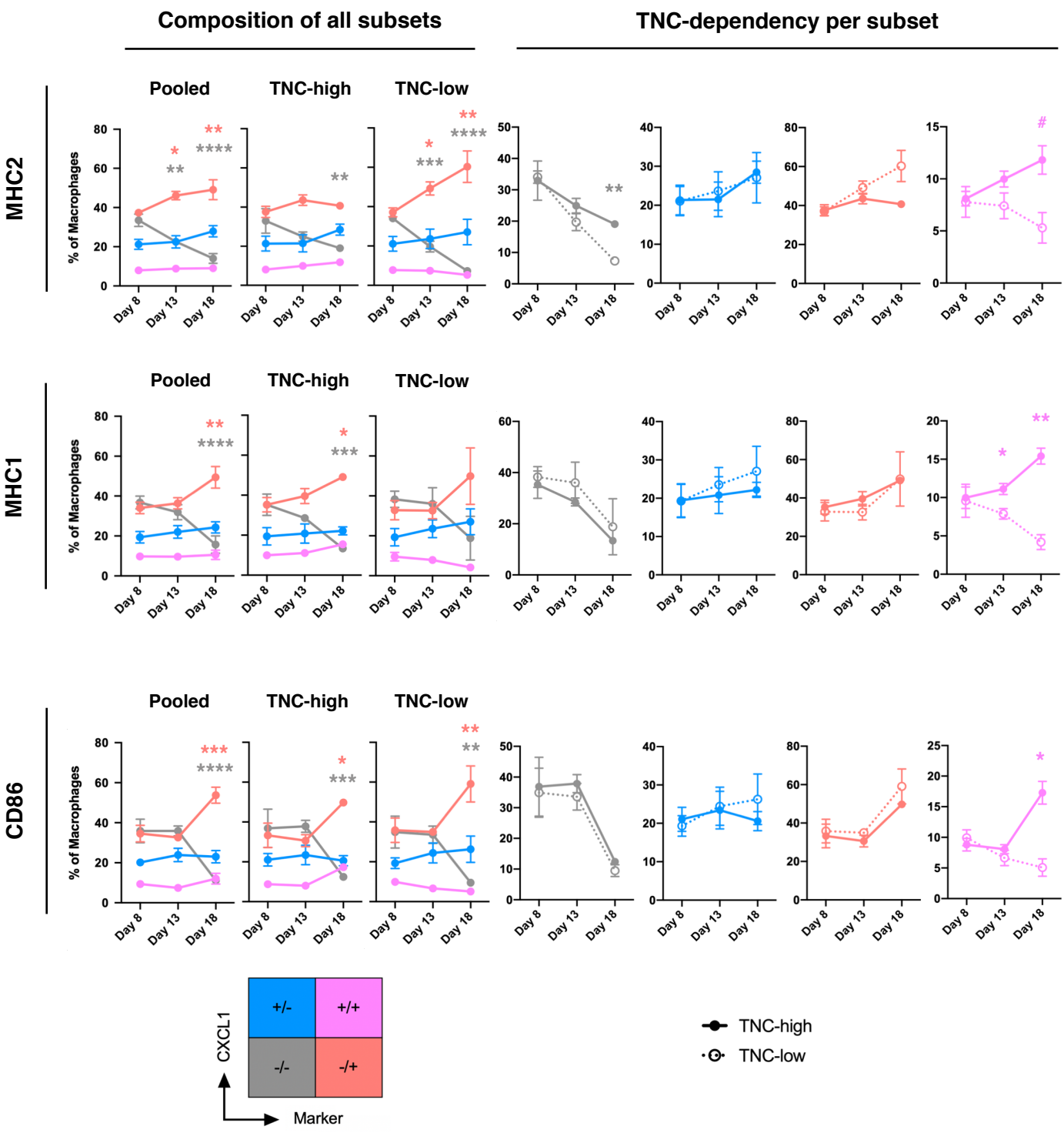


Figure 4.10 | Time-course analysis of functionally-delegated macrophages. Syngeneic NT193 tumour cells transduced with lentiviral shRNA targeting either *Tnc* (TNC-low, n=3) or a non-targeting control (TNC-high, n=4) were orthotopically engrafted into the mammary gland of wildtype female FVB mice and tumours were harvested at 18 days post-engraftment. Flow cytometry was performed to identify functionally delegated subsets of macrophages producing CXCL1 (**Mac3**-like) and/or elevated antigen presentation machinery (**Mac2**-like): MHC1, MHC2, alongside the co-stimulatory molecule CD86. Data were collated with **Figure 4.9** to perform a time-course analysis of tumours at days 8, 13 and 18 post-engraftment. The number of independent tumours analysed at Days 8, 13, and 18 was n=5, n=6, n=4 for TNC-high tumours, respectively, and n=5, n=5, n=3 for TNC-low tumours, respectively. (*continued on next page*)

Figure 4.10 (continued). All samples are plotted per macrophage subset, irrespective of TNC status to provide an average of macrophage states across the NT193 model of breast cancer. To assess changes in the relative abundance of functionally-delegated macrophage subsets over time, a two-way ANOVA was performed for each subset, compared to its value at Day 8, and then adjusted for multiple comparisons using the Dunnett correction. Asterisks are coloured according to macrophage subset. To assess the TNC-dependency of each cell subset, an unpaired T test was performed for each subset per timepoint and corrected for multiple comparisons using the Holm-Sidak method. All asterisks represent the following levels of statistical significance: * P ≤ 0.05, ** P ≤ 0.01, *** P ≤ 0.001, **** P ≤ 0.0001, # P = 0.069. All error bars represent the standard error of the mean. Abbreviations: ANOVA = Analysis of Variance.

4.1.6 | Temporal dynamics of monocyte-to-macrophage transition

Although scRNAseq did identify a very small number of monocytic (*Ly6c2*$^{+}$) cells, these were strongly and distinctly localised within the dendritic cell cluster in UMAP space (**Figure 3.7**), indicating these seeded the monocyte-derived DC compartment. An explicit monocytic cluster seeding the TAM compartment was not identified, and thus the presence of a proliferating macrophage cluster (**Mac5**) suggested that the ontogeny of TAMs could possibly be explained by self-renewal of tissue-resident macrophages rather than a new wave of monocytic replenishment.

To determine whether monocytes were present in NT193 tumours and potentially seeding/replenishing the TAM compartment, and whether the levels of monocytes changed over time, or in the presence of tumour-derived TNC, TNC-high and TNC-low tumours were harvested at Days 8, 13 and 18 post-engraftment for a time-course flow cytometric analysis. The monocyte-to-macrophage "waterfall" transition was assessed using Ly6C and MHC2 markers. Given the known presence of a Ly6G^{+}Ly6C^{+} (**MDSC**-like) population that might confound the analysis, Ly6G^{+} cells were excluded. Analysis of monocyte-macrophage states over time, compared to Day 8, not only indicated that Ly6C^{+} monocytes were present within tumours, but that the proportion of Ly6C^{hi} MHC2^{lo} monocytes (blue gate) significantly decreased ~2-fold over time (**Figure 4.11**, blue gate), suggesting that monocytes are an early

seeding population giving rise to at least part of the tumour-associated macrophage compartment. By comparison, the proportion of Ly6C^{hi} MHC2^{hi} 'transitionary' cells (pink gate) modestly increased over time, suggesting there was an accumulation of non-fully-differentiated cells. Unexpectedly, there was no sustained increase in fully differentiated macrophages (Ly6C^{lo} MHC2^{hi}, red gate) over time. Indeed, their abundance oscillated with time, perhaps alluding to the existence of temporally-regulated waves of differentiation from the transitionary state towards *bona fide* fully-differentiated macrophages. Taken together, these data indicate that monocytes are present within NT193 breast tumours and at least partially contribute to replenishment of the TAM population.

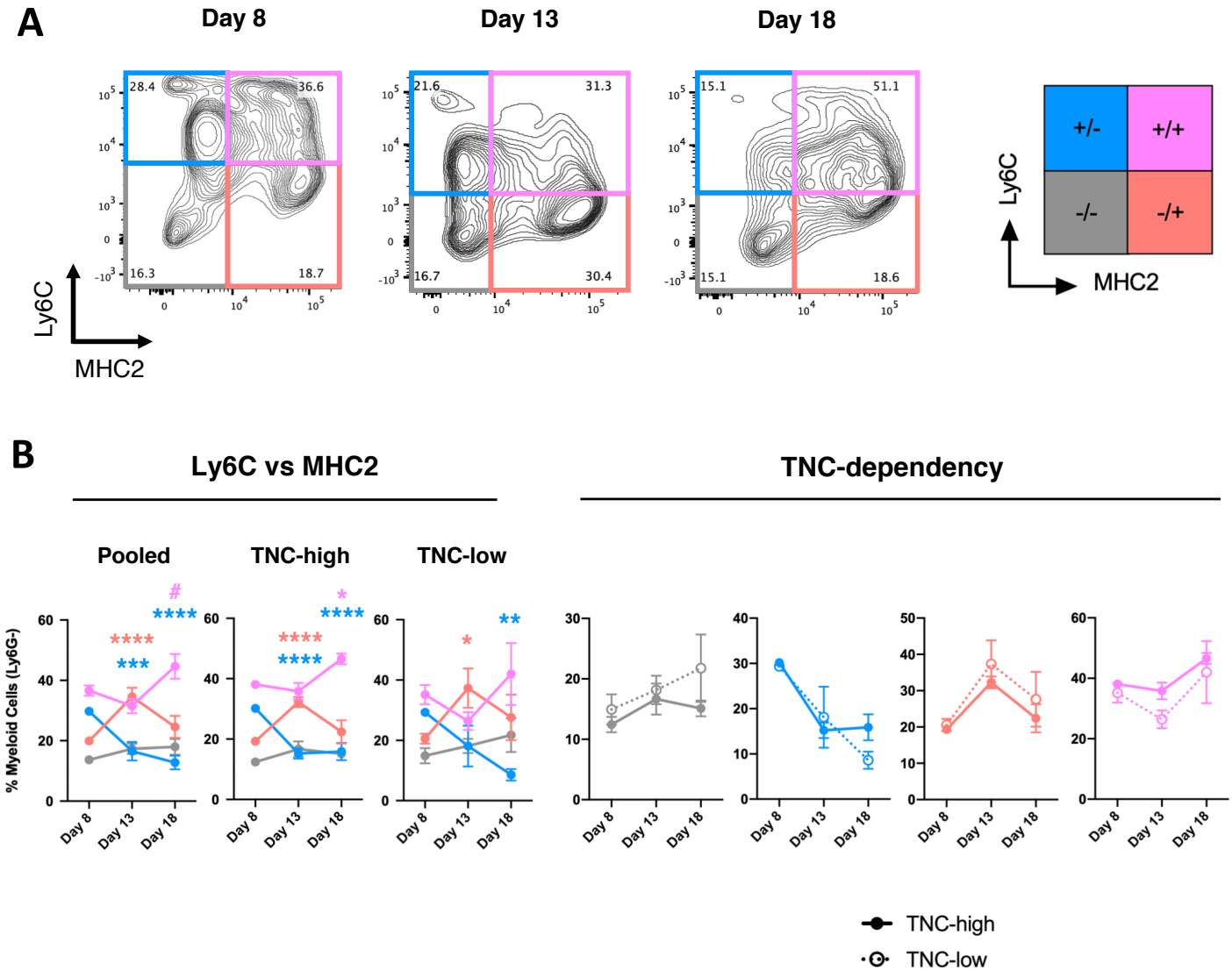


Figure 4.11 | Time-course analysis of monocyte-macrophage 'waterfall' transition. Syngeneic NT193 tumour cells transduced with lentiviral shRNA targeting either *Tnc* (TNC-low) or a non-targeting control (TNC-high) were orthotopically engrafted into the mammary gland of wildtype female FVB mice and tumours were harvested at 8, 13 or 18 days post-engraftment. The number of independent tumours analysed at Days 8, 13, and 18 was n=5, n=6, n=4 for TNC-high tumours, respectively, and n=5, n=5, n=3 for TNC-low tumours, respectively. Flow cytometry was performed on tumour samples to identify monocyte-to-macrophage transition. **(A)** Flow cytometric analysis of $CD11b^{+}Ly6G^{-}$ cells (all myeloid cells except neutrophils or MDSC) over time in a monocyte-macrophage transition 'waterfall' using Ly6C and MHC2 markers. A single representative tumour is shown for each timepoint. **(B)** To assess changes in the relative abundance of monocyte-macrophage transition states over time, a two-way ANOVA was performed for each subset, compared to its value at Day 8, and then adjusted for multiple comparisons using the Dunnett correction. Asterisks are coloured according to each subset. To assess the TNC-dependency of each cell subset, an unpaired T test was performed per timepoint and corrected for multiple comparisons using the Holm-Sidak method. All asterisks represent the following levels of statistical significance: * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$, # $P = 0.053$. Abbreviations: ANOVA = Analysis of Variance, MDSC = Myeloid-derived Suppressor Cell.

4.1.7 | Spatial distribution of functionally delegated macrophage states

To understand whether functionally-delegated TAM effector states are localized to distinct regions of the tumour microenvironment, confocal immunofluorescence imaging was performed for **Mac2** and **Mac3** populations across whole tumour sections, using MHC2 and CXCL1 as markers, respectively. For optimal detection of macrophages, co-localization of F4/80 expression with either MHC2 or CXCL1 was required for assignment of **Mac2** ($F4/80^{+}MHC2^{+}$) or **Mac3** ($F4/80^{+}CXCL1^{+}$). Tiling of whole tumours with this strategy revealed distinct spatial distributions of TAM subsets, with **Mac2** being distributed relatively evenly throughout the tumour, while **Mac3** was largely restricted to the periphery and excluded from the tumour core (**Figure 4.12**). Consistent with published literature, TNC was organized into distinctive matrix tracks (365, 366), and these tracks were enriched at the tumour periphery compared to the tumour core (**Figure 4.12**). Interestingly, the reduced presence of TNC in the tumour core coincided with a high proportion of **Mac2** and low **Mac3**, aligning with previous observations showing that $MHC2^{hi}$ macrophages (**Mac2**) are expanded in low TNC conditions (264) (**Figure 3.11C**). Taken together, these data suggested that at least two TAM effector states, **Mac2** and **Mac3**, exhibit distinctive localization across the tumour microenvironment, and may potentially be TNC-regulated.

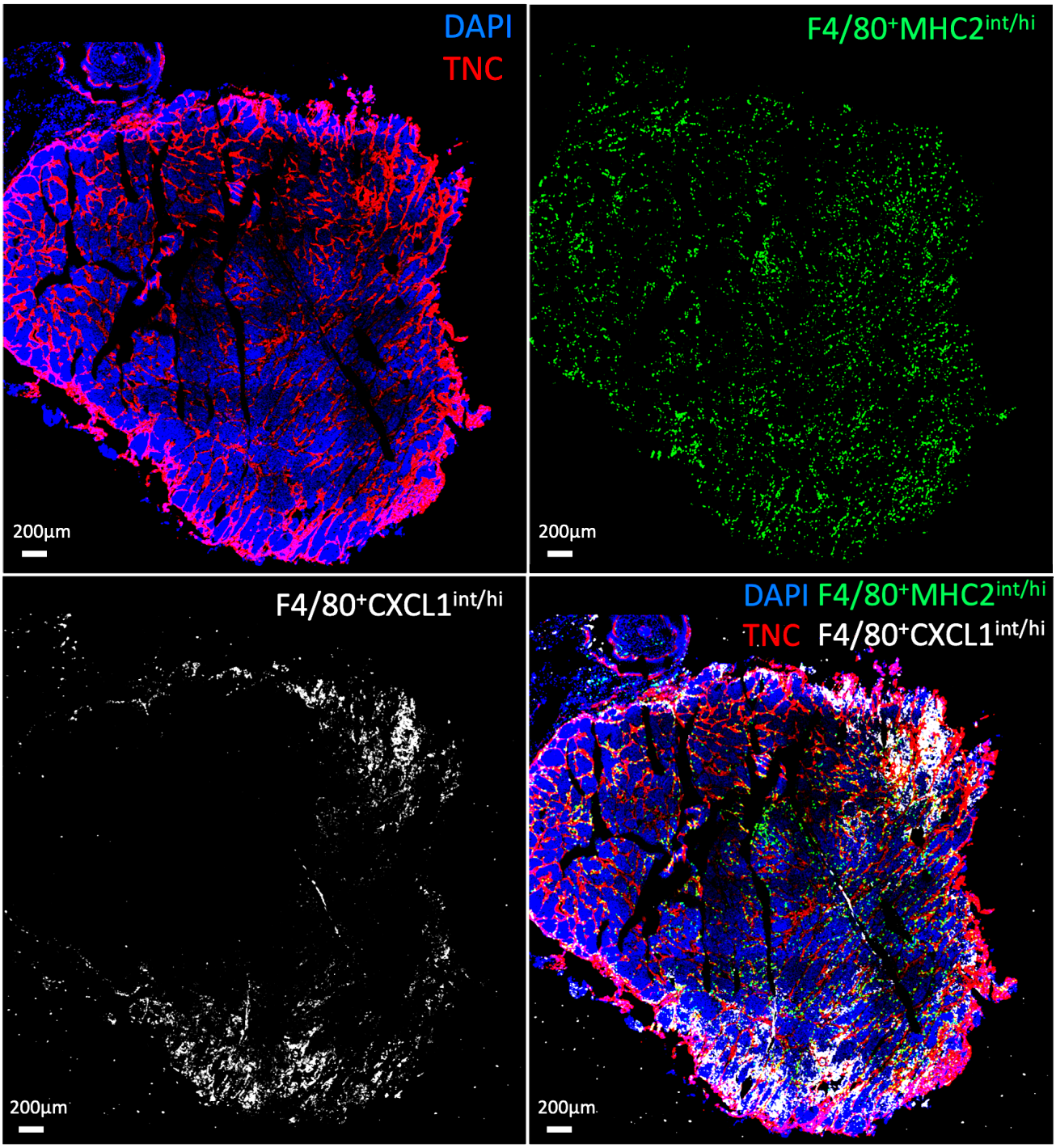


Figure 4.12 | Spatial distribution of delegated macrophage subsets by confocal immunofluorescence imaging. Syngeneic NT193 tumour cells transduced with lentiviral shRNA targeting either *Tnc* (TNC-low, n=2) or a non-targeting control (TNC-high, n=2) were orthotopically engrafted into the mammary gland of wildtype female FVB mice, and tumours were harvested, fixed and frozen in OCT at 18 days post-engraftment. Tumour sections were stained for DAPI (to detect all nucleated cells), F4/80 (to detect macrophages), CXCL1 (to detect **Mac3** cells in combination with F4/80), MHC2 (to detect **Mac2** cells in combination with F4/80), as well as TNC. Tiling of a whole tumour at 10X magnification is shown, illustrating the distinct localization of macrophages ($F4/80^+$ cells) expressing CXCL1 (**Mac3**) versus those expressing MHC2 (**Mac2**). Co-localization thresholds of these markers were determined using Imaris software. Acknowledgment: Tumour processing, immuno-staining and imaging was performed by Julie Mazet at the Kennedy Institute of Rheumatology (University of Oxford). Abbreviations: OCT = Optimum Cutting Temperature compound, DAPI = 4′,6-diamidino-2-phenylindole

4.1.8 | *In vivo* alteration of functional-delegation by KDM6B: a Mac3 regulator

As *ex vivo* manipulation of macrophages with functional assays was not possible due to prohibitive phenotypic changes upon FACS-sorting (**Figure 4.8C-D**), strategies to disrupt functional-delegation of macrophages directly *in vivo* were sought. In **Chapter 3**, diffusion pseudotime analysis of macrophage trajectories suggested that naive-like **Mac1** transitioning to chemokine-enriched **Mac3** strongly upregulated the epigenetic histone demethylase, *Kdm6b*, and the kinase, *Pim1*, immediately preceding the induction of inflammatory cytokines/chemokines such as *Cxcl1*, *Cxcl2*, *Tnf* and *Il1b* (**Figure 3.14B**). This observation led to the hypothesis that *Kdm6b* and *Pim1* might together represent a kinase-inhibitor-targetable epigenetic program regulating the acquisition of **Mac3** effector function (**Table 3.1**). I reasoned that inhibiting this putative epigenetic-regulatory axis, using potent and selective commercially-available small molecule inhibitors of KDM6B and PIM1, would disrupt the functional transition between **Mac1** and **Mac3** (**Figure 4.13A**). Owing to time constraints, KDM6B inhibition was prioritised ahead of PIM1 inhibition, as induction of *Kdm6b* expression in pseudotime was modestly stronger than *Pim1* (**Figure 3.14B**).

KDM6B possesses a Jumonji-C domain, which is responsible for lysine demethylase activity, and is a characteristic of this subfamily of histone demethylases (367). The KDM6B inhibitor, GSKJ4, was chosen for investigation based on its high selectivity for KDM6B over other Jumonji-domain family demethylases (368), thus limiting off-target effects. GSKJ4 has previously been administered in pre-clinical models of cancer (369-371), owing to its cytotoxic effects in cancer cells at high doses. However, the dosage in these studies was very high (50-100mg/kg), and the readouts of these studies focussed on tumour growth rather than on immune modulation. As the aforementioned dose is likely to be cytotoxic towards immune cells as well as tumour cells, the most relevant comparator studies were those assessing GSKJ4 in murine models of inflammatory colitis and sepsis (372, 373).

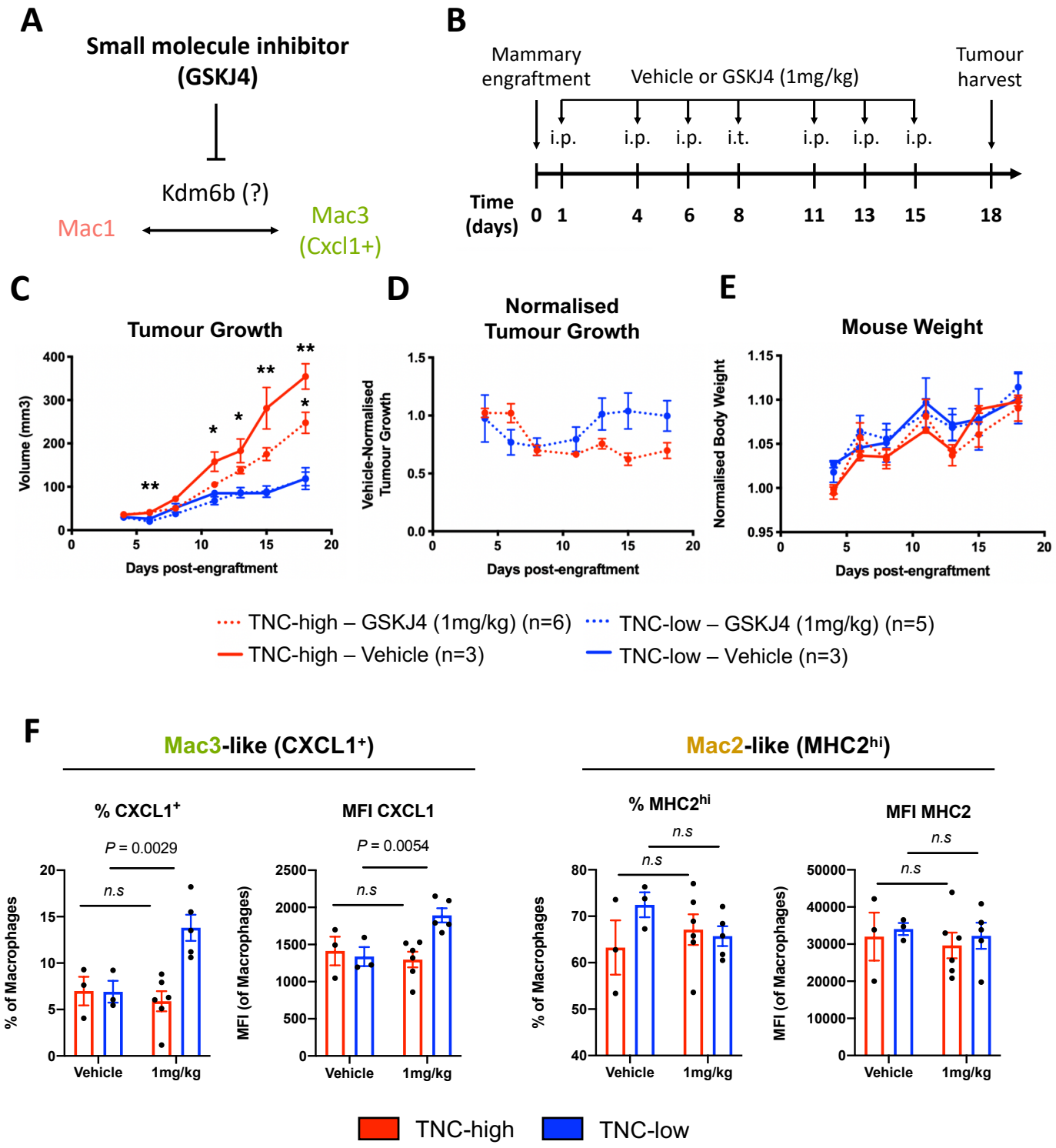


Figure 4.13 | *In vivo* alteration of macrophage functional-delegation by inhibition of KDM6B – an epigenetic regulator of macrophage effector function. Syngeneic NT193 tumour cells transduced with lentiviral shRNA targeting either *Tnc* (TNC-low) or a non-targeting control (TNC-high) were orthotopically engrafted into the mammary gland of wildtype female FVB mice. **(A)** Schematic diagram summarizing hypothesized role of KDM6B regulating Mac1<>Mac3 transition, and the use of the KDM6B inhibitor, GSKJ4, to disrupt this transition**. (B)** Timeline of experimental procedures for dosing of mice in the NT193 orthotopic mammary engraftment model. The following number of tumours were engrafted per condition: TNC-high with GSKJ4 (n=6), TNC-high with vehicle (n=3), TNC-low with GSKJ4 (n=5), TNC-low with vehicle (n=3). I.p. = Intraperitoneal, I.t. = intratumoural. (*continued on next page*)

Figure 4.13 (continued). (C) Tumour growth curves per treatment and per TNC-status. The TNC-dependent effects of each treatment were compared by two-way ANOVA, and adjusted for multiple comparisons with the Tukey correction. All asterisks represent the following levels of statistical significance: * $P \leq 0.05$, ** $P \leq 0.01$. **(D)** GSKJ4-treated tumour growth curves were normalized to the average vehicle-treated tumour volume per timepoint. Treatment-effects were assessed for each TNC-condition by two-way ANOVA, and adjusted for multiple comparisons with the Sidak correction. No comparisons were statistically significant. **(E)** Body weight of mice over time, normalized to the value on Day 1 for each mouse. Treatment-effects were assessed by two-way ANOVA and adjusted for multiple comparisons with the Tukey correction. No comparisons were statistically significant. **(F)** Flow cytometric analysis of intra-tumoural macrophages ($CD45^+CD11b^+F4/80^+$) expressing either **Mac3** or **Mac2** markers. Multiple T tests were performed and adjusted for multiple comparisons using the Holm-Sidak method. Abbreviations: ANOVA = Analysis of Variance.

These studies demonstrated potent immune modulation at a dose of 0.5-1mg/kg with daily intraperitoneal injections, and a dose of 1mg/kg was therefore replicated in the NT193 model of breast cancer, according to the timeline illustrated in **Figure 4.13B**, and in accordance with the maximum number of dosings permitted under the Home Office Project License. A total of six intraperitoneal and one intra-tumoural injections were performed per tumour-bearing mouse.

This analysis revealed that tumour growth in vehicle-treated mice was TNC-dependently regulated, with TNC-high tumours being significantly larger than TNC-low tumours at most timepoints between Day 6 and Day 18 post-engraftment (**Figure 4.13C**), which was consistent with all previous model validation data here (**Figure 4.2A**) and in published literature (264). However, in GSKJ4-treated mice, the TNC-dependent tumour growth was more modest, and only reached statistical significance at Day 18 rather than at earlier timepoints. Normalisation of GSKJ4-treated tumour volumes to the average volume of vehicle-treated tumours indicated a trend towards slowed tumour growth in GSKJ4-treated TNC-high tumours, but this did not reach statistical significance ($p=0.11$) (**Figure 4.13D**). Across all groups, the weight of mice was unaffected by GSKJ4 treatment or TNC-status, suggesting a tolerable toxicity profile for

GSKJ4, although in all groups, body weight did increase over time (**Figure 4.13E**), possibly due to the volume of i.p. injections and increase in tumour size.

Mice were culled at Day 18 and assessment of tumour macrophage effector states by flow cytometry revealed that GSKJ4 treatment led to a significant accrual of $CXCL1^{+}$ macrophages (**Mac3**) at the percentage and MFI level, compared to vehicle-treated mice (**Figure 4.13F**). By contrast, $MHC2^{hi}$ macrophages (**Mac2**) were unaffected at the percentage and MFI level (**Figure 4.13F**), suggesting that the acquisition of **Mac3** effector function, but not **Mac2**, was specifically modulated by KDM6B, thus supporting the initial hypothesis (**Figure 4.13A, Table 3.1**). Curiously, this KDM6B-dependent phenotype was observed only in TNC-low tumours, perhaps indicating that the absence of tumour-derived TNC creates a physically permissive environment for this type of macrophage transition to occur, or that the presence of tumour-derived TNC induces dysfunction in macrophages which supersedes the KDM6B-dependence of this process. Taken together, these data suggest that pharmacological inhibition of KDM6B specifically regulates the acquisition of **Mac3** effector function, and in a TNC-dependent manner.

4.2 | Technical Discussion

The aforementioned experiments and analyses have strengths and limitations that are worth noting. Firstly, and perhaps most importantly, key observations from the single-cell RNAseq analysis were successfully recapitulated at the protein level with sufficient independent biological replicates to make statistically robust conclusions. This recapitulation of key data, such as the functional delegation of TAM effector functions, and the identification of KDM6B as a regulator of **Mac3**, together serve to validate the scRNAseq findings and provide confidence that other insights emerging from this dataset may also be valid.

Another strength of these data in comparison to recent single-cell atlases of the human breast cancer immune infiltrate (102) or entire microenvironment (101) is the increased read depth per sequenced cell, with an average of 65k reads per cell, compared to only 21k in one study (102), and 30-40k in the most recent study (101) (*private author communication, 29th September 2021*). This difference in read depth may, for instance, partly explain the divergent findings regarding TAM phenotypes in these studies – it may be the case that the continuous activation trajectory of macrophages described by Azizi *et al.* (102) would not have been observed with increased depth and resolution, perhaps instead arguing for a more multifaceted model of TAM activation. Moreover, in contrast to single-cell atlases which capture a descriptive snapshot of the transcriptional landscape in breast cancer, the present work has the benefit of actively perturbing the microenvironment with knockdown of tumour-derived TNC, thus enabling the observation of phenotypic continua that might not otherwise be apparent, and the assignment of mechanistic causality to such phenomena. However, despite the strengths of these data, it is important to note that the analyses presented here have thus far been restricted to a single murine model, and consequently the conservation of these phenotypes in human tissue or clinical genomic datasets is required.

Discussion of individual experiments in this chapter will now be considered chronologically:

In the validation experiments of the mammary tumour engraftment model, RT-qPCR of FACS-sorted macrophages indicated elevated expression of *Tnf* mRNA in macrophages from TNC-low tumours (**Figure 4.2D**). However, this did not align with previous observations in the scRNAseq data, which indicated that neither *Tnf* production at the transcriptional level (**Figure 3.12A**) nor cellular level within **Mac3** (**Figure 3.11C**), was elevated in TNC-low tumours. While this might be explained by differences in the transcriptomic technologies employed, it might also be a consequence of differences in FACS sorting strategies. For the RT-qPCR, the gating strategy enriched $CD11b^{+}F4/80^{+}$ cells (macrophages) (**Figure 4.2C**), whereas for the scRNAseq of FACS-sorted myeloid cells, F4/80 expression was not considered, and a CD3 dump channel was included to ensure exclusion of T cells, as per published protocols (331), resulting in enrichment of a $CD11b^{+}CD3^{-}$ population. However, given emerging insights from other pathological settings regarding the existence of $CD3^{+}TCR^{+}$ myeloid cells that produce and are regulated by TNF (374), it remains possible that our scRNAseq analysis did not capture this population, perhaps explaining the difference in TNF levels between RT-qPCR and scRNAseq experiments. Arguing against this, though, is the fact that $CD3^{+}TCR^{+}$ macrophages have not been observed in the largest single-cell transcriptomic atlas of the breast tumour microenvironment (102). Thus, the source of additional TNF in TNC-low tumours is uncertain and requires further investigation.

Regarding the isolation of **Mac3** by FACS-sorting, a number of technical issues were encountered. First, it appeared that the sorting-purity of the SSC^{hi} (**Mac3**-like) population was low at only ~62% (**Figure 4.7B**), due to cells universally exhibiting lower SSC-A after resting on ice. However, it is important to note that, when attempting to separate a population such as SSC^{hi} which lies at the edge of a normal distribution (rather than a bimodal distribution with clear positive and negative populations), FACS-sorting often brings part of that normal distribution along with it, even if it did not strictly fall within the original gate. The key goal of this isolation was to segregate populations, and, importantly, analysis of the sorted population confirmed a total absence of SSC^{lo} cells, alongside enrichment of a homogemous population

of SSC^{hi} cells, indicating successful isolation. Indeed, subsequent intracellular staining for CXCL1 confirmed a purity of >90% $CXCL1^{+}$ cells (**Mac3**) from the SSC^{hi} population, meaning any negative impact of this SSC-based sorting strategy could not have been drastic, if present at all.

Another key issue was the drastic phenotypic changes observed in **Mac3** following FACS sorting, whereby supernatants of **Mac3** cultures contained greatly reduced CXCL1 protein compared to control cultures of **Mac2**-like cells (**Figure 4.8C-D**). The challenges of FACS-sorting macrophages are well documented (331, 375, 376), and while it should not be surprising that macrophages are capable of displaying remarkable phenotypic plasticity, the magnitude and timescale in which this occurred (after just 2 hours in culture) makes this unlikely to be explained by *de novo* plasticity alone. There are several potential explanations for why this inverse CXCL1 relationship may have occurred:

i) **Retention of CXCL1 within intracellular granules** in **Mac3**, even following strong activating stimuli such as LPS. To investigate this possibility, FACS-sorted populations could once again be subjected to ELISA of the supernatant, alongside cell lysates to confirm presence or absence of sequestered CXCL1.

ii) **Differential CXCL1 degradation in culture** perhaps caused by secretion of **Mac3**- or **Mac2**-specific proteases, MMPs, or binding proteins which alter CXCL1 half-life.

iii) **Rebalancing the disrupted equilibrium of TAM effector functions**, whereby an over-concentration of one macrophage population in culture such as **Mac2** (highly-antigen-presenting) leads to a compensatory transition of cells towards a **Mac3** phenotype, leading to enhanced CXCL1 production. Likewise, **Mac3** cells in culture

reciprocally transition towards a **Mac2** phenotype. This could be assessed by performing flow cytometry for MHC1/2 markers on **Mac3** cells post-culturing.

iv) **Tumour dissociation releases the breaks of suppressive TME factors**, leading activated **Mac2** to acquire a chemokine/cytokine effector function (**Mac3**). Alternatively, the inverse might occur, where **Mac3** is relieved of suppressive TME cues and therefore ceases production of tumour-promoting inflammatory molecules such as CXCL1,

One explanation that can be confidently dismissed is the notion that early secretion of CXCL1 during or before FACS-sorting is the reason for low CXCL1 levels in the supernatant by ELISA. As **Figure 4.8A** demonstrates, intracellular CXCL1 expression was still detected at 91% positivity at the same point in time which cells were placed in culture, meaning there was no significant loss of CXCL1 signal prior to culturing.

In terms of the functional delegation phenotype, there is a mild discrepancy between the amount of undelegated double-positive ($CXCL1^+$ $MHC2^+$) cells in the initial experiment at Day 18 (**Figure 4.5**), versus the time-course data (**Figure 4.9-4.10**). There might be several technical reasons for this, including **(i)** the usage of new batches of anti-CXCL1 antibody and isotype control in the time-course data, and **(ii)** the usage of different passages of NT193 cells in both experiments (P7 for time-course, and a combination of P8 & P16 for the Day 18 experiment). The NT193 cell morphology appeared very similar in all passages (data not shown), so this is not expected to be a major source of variation. However, it might be possible that older passages of NT193 cells induce a more extreme functionally-delegated phenotype in macrophages in unknown ways. In any case, while the exact source of this minor variation remains uncertain, it does not alter the overarching interpretation of the data, as the functional delegation phenotype remains pronounced in both contexts.

Assessment of monocyte-to-macrophage transition was an area in which scRNAseq and flow cytometry data did not closely align. For instance, scRNAseq did not predict any macrophages to be derived from Ly6C$^+$ monocytes due to almost universally absent *Ly6c2* transcript expression within macrophage clusters, yet Ly6C protein expression was detected on monocytes/macrophages at all timepoints by flow cytometry, though this decreased over time (**Figure 4.11**). Biologically, a simple explanation for this discrepancy may be the imperfect correlation between mRNA and protein (377), but from a technical perspective, this might be due to the low expression (and thus insufficient read depth for detection) of *Ly6c2* by scRNAseq. Complementing this, a very low rate of mRNA transcription for *Ly6c2* coupled with a long protein half-life at the plasma-membrane, perhaps facilitated by receptor recycling, would explain the high surface-protein signal by flow cytometry (as discussed in **Section 3.2**). The biological interpretation of Ly6C expression data and its implications for macrophage ontogeny will be discussed in the **Section 4.3**. However, one possibility might be that the discrepancy regarding monocyte detection may be partly due to scRNAseq being performed on intra-tumoural myeloid cells harvested 21 days post-engraftment, whereas subsequent flow cytometric analyses were performed on samples harvested at Days 8, 13 and 18. Thus, by Day 21, a greater proportion of monocytes may have already undergone full differentiation into macrophages, leading to low monocyte detection at Day 21. Finally, it is also worth noting a minor technical issue that occurred on Day 8 of the monocyte-macrophage-transition time-course, whereby a faulty cytometer meant an alternative machine had to be used with a different laser setup. The voltages and compensation were therefore different at Day 8 compared to other timepoints, as seen in the apparently increased Ly6C signal in **Figure 4.11A**. However, internal FMO and isotype controls were applied to ensure comparability of gating between experiments and instruments, so this technical issue did not negatively affect analysis or interpretation.

In terms of immunofluorescence imaging of tumour sections, no quantification of functionally-delegated macrophage subset localization is presented, necessitating that at least a semi-

quantification is performed across all replicates to confirm statistical significance. This analysis is in fact ongoing.

Finally, the lack of highly expressed cluster-discriminatory surface markers for TAM subsets impeded isolation strategies for subset-specific FACS-sorting, alongside antibody-mediated depletion strategies. One alternative considered was a knockout-mediated or Diptheria toxin receptor (DTR)-mediated ablation (378) of delegated macrophage subsets, as this would remove the requirement for surface protein expression. However, given that many of the strongest marker genes for macrophage subsets were immune-mediators required systemically throughout mammalian physiology, such as *Cxcl1*, *Cxcl2*, *Tnf*, *Il1b* for **Mac3** (**Figure 4.6A**), constitutive ablation of all cell populations expressing one of these genes would lead to extensive off-target effects outside of the myeloid compartment and across many tissues. Generation of a double-gene conditional cell ablation model – for instance, a myeloid-restricted ($CD11b^+$) ablation of CXCL1 – a would have been an ideal tool for investigation of functional delegation, however this does not currently exist, and would have been prohibitively time-consuming to develop. Thus, *in vivo* modulation approaches to disrupting macrophage effector functions were prioritised.

4.3 | Biological Discussion

4.3.1 | Introduction

This chapter sought to validate and explore the phenotypes of intra-tumoural myeloid populations first identified by scRNAseq. The central hypothesis that macrophages delegate key effector functions to discrete subsets was confirmed by flow cytometry (**Figure 4.4**), and deeper characterisation in spatial (**Figure 4.12**) and temporal (**Figure 4.10**) dimensions yielded important insights that reinforced this. Yet these insights also demonstrated an underlying biological complexity which raise further questions about the functionality, ontogeny, conservation, and context-dependence of macrophage subsets in the tumour microenvironment. This section will discuss the biological implications of these data, and will be divided into the following subsections:

4.3.2 Spatial insights into the pathological relevance of TAM effector functions

4.3.3 TAM ontogeny, maturity, and the role of TNC

4.3.4 Conserved and dichotomous TAM phenotypes

4.3.5 KDM6B as a selective regulator of **Mac3** effector function

4.3.6 Summary

4.3.2 | Spatial insights into the pathological relevance of TAM effector functions

Visualising the spatial dimension of two TAM effector states, **Mac2** and **Mac3**, revealed distinct distribution patterns of macrophages in the TME (**Figure 4.12**). The restriction of **Mac3** to the tumour periphery yet the widespread infiltration of **Mac2** in the tumour core, hints at their biological function and pathological relevance.

One possibility is that chemokine-enriched **Mac3** are a perivascular-associated subset, hence explaining their exclusion from the tumour core which is typically poorly vascularised in solid tumours. As previously discussed (**Section 3.3.3**), literature proposes a prominent role for TAM-derived CXCL1 in promoting metastasis in human and murine breast cancer (308). Other work revealed the central role of cancer-cell-derived CXCL1 and CXCL2 in promoting breast tumour growth and metastasis via the pathologic recruitment of suppressive $Gr1^+$ (ie $Ly6C^+$ and/or $Ly6G^+$) myeloid cells expressing the TLR4 ligand, calprotectin (S100A8/S100A9) – a subset with striking similarity to **MDSC** (307). Moreover, in the same study, chemotherapy treatment induced high TNFα production by stromal and endothelial cells, causing further upregulation of CXCL1 and CXCL2 by cancer cells, which ultimately led to chemotherapy-resistance. Thus, **Mac3** and its chemokines/cytokines (CXCL1, CXCL2 and TNF) may be the central coordinators of a pathological signalling loop in breast cancer, involving concerted participation by endothelial, myeloid, stromal and carcinoma cells. The localisation of putative perivascular **Mac3** to the periphery of tumours may therefore facilitate this pathological function, for instance by aiding **MDSC** recruitment from the blood, or by directly facilitating EMT or metastatic dissemination through proximity to the vasculature. Further detail of this myeloid-endothelial crosstalk was recently described, demonstrating that perivascular macrophages stimulated via the TNC-TLR4 axis were crucial for endothelial cell production of vascular-niche components that promoted lung metastasis (363), thus potentially proposing a novel pathogenic function of **Mac3** in breast cancer. However, to confirm this putative

association of **Mac3** with the vasculature, complementary immunofluorescence imaging with co-staining for the endothelial cell marker, CD31, should be performed.

By contrast, the antigen-presentation-enhanced **Mac2** are uniformly distributed throughout the tumour (**Figure 4.12**), and their expansion in slow-growing TNC-low tumours (**Figure 1A-B**) suggests they may be in an anti-tumoural state of activation, mediating immune-destruction and limiting tumour growth. The presence of **Mac2** in the tumour core suggests they are capable of performing their function while residing in a likely hostile environment with limited vascularisation, low pH, hypoxia, dense matrix and low nutrient availability. This raises the possibility that the mere existence of delegated TAM effector functions might itself be a metabolic necessity, imposed by the impermissibility of the local microenvironment for certain effector functions over others. For example, the environment of the tumour core might lend itself more towards antigen-presenting phenotypes (**Mac2**), whereas the high metabolic demand of a likely more energy-intensive function like chemokine/cytokine production (**Mac3**) might only be possible in more hospitable locales. Thus, functional delegation might be an energy-conservation mechanism used by macrophages to adapt to the hostile conditions of the tumour microenvironment.

4.3.3 | TAM ontogeny, maturity, and the role of TNC

Given the absence of a macrophage-seeding monocytic cluster in the scRNAseq data, yet the presence of Ly6C^{hi} monocytes by flow cytometry, the ontogeny of these breast TAMs is unclear. Confirmation of monocyte infiltration by flow cytometry excluded the possibility that TAMs are purely derived from self-renewing tissue-resident macrophages. However, the relative contribution of tissue-resident versus monocyte-derived TAMs in this model is not fully understood. Elegant investigation of macrophage ontogeny in normal mammary tissue versus the PyMT model of breast cancer demonstrated that Ly6C^{+} CCR2^{+} monocytes seed the population of normal mammary tissue macrophages as well as the TAM population, but that TAMs had a higher capacity for subsequent proliferation (188). The mammary gland is thus one of the few tissues where the replenishment of tissue macrophages is solely dependent on blood-borne monocytes, as opposed to (at least partial) local proliferation of tissue-resident macrophages seeded during foetal development. While this Chapter did not directly investigate naive-like **Mac1** or proliferating **Mac5**, the low expression of *Ccr2* in both these clusters (**Figure 3.10**) may therefore suggest they represent early-differentiated monocyte-derived TAMs (**Mac1**) which then undergo proliferation (**Mac5**), where they account for the majority of the proliferating macrophage population.

One complementary possibility might be that **Mac1** in fact resembles a late monocytic transitionary state, where Ly6C transcript expression has decreased to an undetectably low level for scRNAseq, but is still detectable at the protein level by flow cytometry. In support of this, time-course analysis of the monocyte-macrophage 'waterfall' indicated that the proportion of 'transitionary' Ly6C^{hi} MHC2^{hi} cells modestly increased over time (**Figure 4.11B**), while the Ly6C^{hi} MHC2^{lo} monocytes decreased, suggesting there was a reduction in monocyte infiltration over time and an accumulation of transitionary cells, perhaps reminiscent of naive-like **Mac1**. Indeed, the root-like topology of **Mac1** in diffusion map space (**Figure 3.13-3.14**) lends support to the idea of it potentially representing a late monocyte undergoing

differentiation, or an early macrophage state, given that all TAM effector-function trajectories emanated from it, while also expressing the terminal differentiation gene, *Mafb* (**Figures 3.6 & 3.10)**. Moreover, previous scRNAseq studies have shown that the ribosomal protein upregulation, characteristic of **Mac1**, is associated with developmental immaturity in lymphoid (leukaemia) cells (359) and HSCs (360), while pseudotime analysis of brain mononuclear phagocytes (MNPs) revealed that monocytes expressed elevated ribosomal genes which decreased upon differentiation to terminal macrophage states (379). Finally, as-yet-unpublished observations from collaborators studying human breast tumours by scRNAseq has identified ribosomal enrichment in late monocytes undergoing differentiation to macrophages via pseudotime analysis (*Chi et al., 2021, manuscript in progress*), thus confirming the translational relevance of this phenomenon in human TAMs. Together these insights are suggestive of **Mac1** representing a transitionary state of late monocytes or early macrophages, though this would ultimately require validation by subsequent *in vivo* studies involving lineage tracing approaches, CCR2 knockout, or DTR-mediated monocyte ablation.

Interestingly, the ribosomal protein upregulation that is characteristic of **Mac1** has also been observed in activated tissue-resident brain macrophages (microglia) following exposure to developing glioma tumours, compared to homeostatic non-activated microglia from non-tumour-bearing controls (380). This implies that ribosomal enrichment may not simply be a marker of developmental immaturity in immune cells, but rather, more precisely, a conserved state which precedes the acquisition of effector functions. Indeed, a similar phenomenon has recently been described in naive T cells, where stockpiling of translational machinery poises T cells for rapid immune activation (381), thus suggesting that this mechanism of translationally-poised naive-like immune cells is conserved across lymphoid and myeloid lineages.

Given that glioma is a cancer where TNC expression has high pathogenic and prognostic significance (239), it might be that TNC-associated TAMs in glioma are similarly arrested in a

naive-like state, unable to acquire anti-tumoural effector functions such as antigen-presentation, as our findings suggest for breast cancer. Importantly, rather than being exclusive to TNC, this induction of a naive-like ribosome-enriched state in macrophages may signify a broader conserved mechanism of pathogenic immune-suppression in the TME, and thus may hold merit as a therapeutic target.

Aside from a failure to delegate TAM effector functions, arresting TAMs in a state of naivety, TNC may also mediate its pathogenic role in another way: driving the dysfunctional delegation of effector functions, leading to the co-existence of phenotypes that would otherwise not exist (**Figure 4.10**). With the hypothesis that functional delegation might be an energy-conservation mechanism, as proposed earlier, it may be the case that accumulation of dysfunctional **Mac2^{+}Mac3^{+}** cells in TNC-high tumours are indicative of a shift in metabolic environment or energy utilisation, which forces consolidation of delegated phenotypes into a co-existing state that would otherwise be avoided.

4.3.4 | Conserved & dichotomous TAM phenotypes

It is worth noting that the tumour microenvironment is known to induce 'phenotypic expansion' in the myeloid compartment of breast cancer (102), whereby novel phenotypes are specifically observed in tumour tissue compared to normal tissue. Functionally-delegated effector states may therefore be tumour-specific phenomena, offering an opportunity to therapeutically target tumour-specific TAM states with minimal off-tumour effects on macrophage populations from other tissues. Indeed, a recent scRNAseq analysis of infiltrating myeloid cells across 210 patients in 15 human cancer types revealed that macrophage subsets from different cancers exhibited high complexity, and could not be effectively clustered together, as confirmed by an entropy-based metric for quantifying the mixability of data (199). Interestingly, the same study by Cheng *et al.* observed a functional dichotomy between $SPP1^+$ (pro-angiogenic) and $C1QC^+$ (pro-phagocytic) macrophage phenotypes in breast and colorectal cancers (199), thus demonstrating delegation of phagocytic effector function to a distinct subset, which is analogous at an abstract level to the delegation of macrophage effector functions described in **Chapter 3** and validated in this chapter (eg antigen-presentation-enrichment versus chemokine production). However, the supposedly dichotomous phenotypes in Cheng *et al.* were also shown to co-exist in a macrophage subpopulation of endometrial carcinoma, suggesting that the influence of the local tissue environment drives phenotypic diversity among TAMs, and means that functional delegation can be both subset- and tissue-specific, or even abolished altogether. While previous investigation of tumour-infiltrating myeloid cells by Cheng *et al.* and Azizi *et al.* consistently showed that M1 and M2 genes exhibit a positive correlation in macrophages (for breast cancer and other cancers), thus opposing the classic M1/M2 paradigm of macrophage activation, this does not preclude the possibility of dichotomous macrophage polarisation trajectories that are based on effector functions rather than M1/M2 genes. Indeed, the decentralisation of macrophage effector functions may be a more appropriate classifier of macrophage phenotypes in the TME, rather than the M1/M2 paradigm.

Taken together, these insights have important implications for understanding macrophage phenotypes in the TME, and for the development of TAM-targeted therapies, as such therapies may need to be tailored to individual TAM subpopulations depending on the tissue of origin, rather than being developed as pan-cancer TAM-modulatory agents.

4.3.5 | KDM6B as a selective regulator of **Mac3** effector function

The epigenetic eraser enzyme, KDM6B, is a histone demethylase that catalyses the removal of transcriptionally-repressive trimethylated lysine groups, specifically at histone 3 lysine 27 (H3K27), thus serving to activate gene expression (382-384). KDM6B has known roles in development, alongside the differentiation of various cell types including T cells (385), and its inhibition with the selective inhibitor, GSKJ4, was able to block the acquisition of an inflammatory state in LPS-stimulated primary human macrophages, reducing expression of cytokines including TNFα, IL1α and IL1β (368). However, aside from resolving ambiguity on the role of H3K27 demethylases in regulating inflammatory responses, GSKJ4 has not since been explored (in published literature) for its therapeutic relevance in modulating macrophages. To our knowledge, this inhibitor has not been used to either block tumour-promoting inflammation in immune cells, or to limit inflammatory macrophage effector functions *in vivo*. The original hypothesis from the scRNAseq pseudotime analysis in **Chapter 3** was that KDM6B induction immediately precedes the acquisition of **Mac3** effector function from a naive-like **Mac1** state, and that KDM6B is therefore responsible for regulating and enabling this acquisition (**Figure 3.14** & **Table 3.1**). Inhibition of KDM6B *in vivo* did indeed alter the acquisition of **Mac3** effector function, measured by the percentage of CXCL1$^+$ cells and the intensity of CXCL1 expression in macrophages. (**Figure 4.13F**). Surprisingly, however, KDM6B inhibition led to an accrual rather than a depletion of CXCL1$^+$ cells (**Mac3**), perhaps indicating that KDM6B mediates a different regulatory programme in (mouse) tumour

macrophages *in vivo* compared to (human) LPS-stimulated macrophages *in vitro* (368), and that KDM6B's role in regulating **Mac1**>**Mac3** transition is not unidirectional. Curiously, this KDM6B-dependent accrual of **Mac3** was observed only in the TNC-low environment, perhaps due to the TNC-high environment disproportionately arresting macrophages in a naive-like state (**Mac1**), which KDM6B inhibition was unable to overcome to promote **Mac3** accrual. Taken together, these data suggest that KDM6B restrains the acquisition of **Mac3** effector function, and that its inhibition promotes selective expansion of the **Mac3** population in a TNC-low microenvironment.

4.3.6 | Summary

In summary, this chapter confirms that macrophage effector functions, such as (anti-tumoural) antigen-presentation (**Mac2**) and (pro-metastatic) chemokine production (**Mac3**), are segregated into specialised subsets which are phenotypically, spatially, and temporally distinct in the NT193 mouse model of breast cancer. Notably, transitions between macrophage effector states can be specifically altered by pharmacological inhibition of key regulators *in vivo*, such as the epigenetic eraser enzyme, KDM6B, which modulates acquisition of the **Mac3** phenotype. While these data provide novel insights into the phenotype, regulation, ontogeny and pathological relevance of TAMs in the breast tumour microenvironment, as well as the role of tumour-derived TNC, these findings are limited to a single murine model of breast cancer. Therefore, the translational relevance of these insights will be assessed in human breast cancer in **Chapter 5**.

Chapter 5 | Conservation and translational relevance of macrophage gene-signatures in human breast cancer

5.1 | Results

5.1.1 | Introduction

The allure of the single-cell genomics revolution is the unprecedented detail it provides for understanding and disentangling the complexity of biological systems. Harnessing a variety of published single-cell and bulk transcriptomic datasets, this Chapter focusses on determining the conservation and translational relevance of mouse-derived myeloid subsets in human breast cancer.

First, to assess whether the functionally-delegated macrophage subsets identified in the NT193 mouse model are conserved in human disease, the two most comprehensive public scRNAseq datasets of the human breast cancer microenvironment (Azizi *et al.* (102) and Wu *et al.* (101)) were analysed. While the Azizi *et al.* dataset possessed some conserved myeloid subsets, the Wu *et al.* dataset demonstrated much stronger conservation, identifying 4 out of 5 macrophage subsets. Second, as the scaling of scRNAseq technologies across large patient cohorts has not yet become feasible, the prognostic association of mouse-derived myeloid subsets were assessed in two bulk clinical transcriptomic datasets (TCGA (66) and METABRIC (290)), using subset-specific gene signatures. This approach led to the identification of myeloid ecotype profiles, which were significantly associated with patient survival. Together, these insights confirm the validity of the previous observations in mice, and reinforce the relevance and importance of specific myeloid subsets in breast cancer, with implications for novel treatment strategies.

This Results section will be divided into the following subsections:

5.1.2 Conversion of mouse genes to human homologs

5.1.3 Conservation of mouse myeloid subsets in human single-cell RNAseq data

5.1.4 Quality & refinement of gene signatures in human bulk RNAseq data (TCGA)

5.1.5 Gene signature enrichment & hierarchical clustering of TCGA data

5.1.6 Survival analysis of patient tumour ecotypes in TCGA data

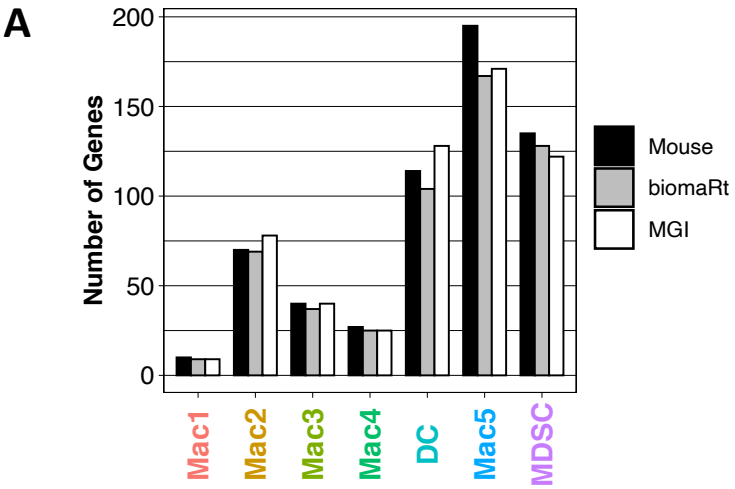
5.1.7 Quality of myeloid gene signatures in human bulk array data (METABRIC)

5.1.8 Gene signature enrichment & hierarchical clustering of METABRIC data

5.1.9 Survival analysis of patient tumour ecotypes in METABRIC data

5.1.2 | Conversion of mouse genes to human homologs

To assess the conservation and translational relevance of mouse myeloid subsets in human breast cancer datasets, the marker genes of each mouse subset identified in **Chapter 3** were first converted to their human homologs. Two methods for homolog conversion were compared. The R Package 'biomaRt' provides functions to perform inter-species gene homolog conversion, and the Mouse Genome Informatics (MGI) database provides a comprehensive list of mouse-human gene homology which was manually parsed to provide equivalent human homologs for each mouse gene. Both methods successfully converted the vast majority of genes, however the MGI method captured a modestly higher proportion of human homologs in most cases (**Figure 5.1A**). Upon inspection of individual homolog conversions, the biomaRt method led to false-negative conversions for the two most highly differentially expressed marker genes defining **Mac3** identity (*Cxcl1* and *Cxcl2*) whereas manually parsed data from the MGI database successfully identified the correct human homologs (*CXCL2* and *CXCL3*, respectively) (**Figure 5.1B**).



C

	MGI	Mac1	Mac2	Mac3	Mac4	DC	Mac5	MDSC
1	# of mouse genes total	10	70	40	27	114	195	135
2	# of mouse genes converted to a human homolog (including all duplicates)	9 (90%)	76* (109%)	40 (100%)	25 (93%)	128* (112%)	171 (88%)	122 (90%)
3	# of mouse genes with a 1-to-1 conversion	9 (90%)	61 (87%)	37 (93%)	25 (93%)	96 (84%)	158 (81%)	110 (81%)
4	# of mouse genes with no human homolog	1 (10%)	3 (4%)	0 (0%)	1 (4%)	10 (9%)	3 (2%)	7 (5%)
5	**# of unique converted human homologs (no duplicates)**	**9 (90%)**	**67 (96%)**	**38 (95%)**	**25 (93%)**	**100 (88%)**	**163 (95%)**	**114 (84%)**

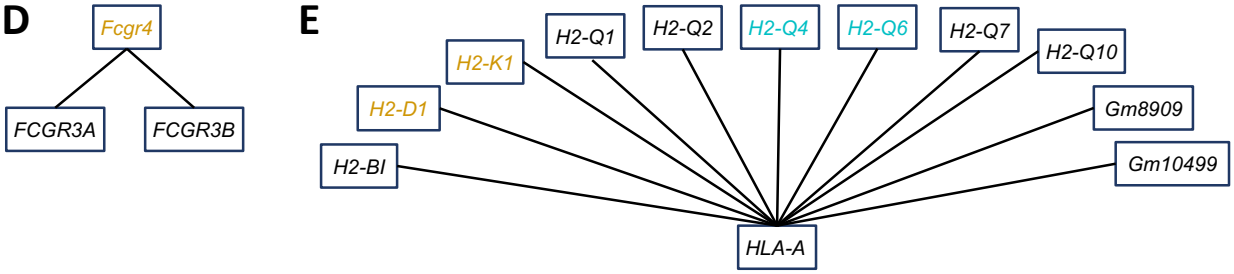


Figure 5.1 | Conversion of mouse genes to human homologs. (A) Mouse marker genes for each myeloid subset identified in **Chapter 3** were converted to their human homologs either using the R Package 'biomaRt', or manually parsed from the the Mouse Genome Informatics (MGI) database, and then compared by their number of genes. **(B)** Conversion of genes using the biomaRt package in R leads to false-negative homolog conversions for the most important (highly differentially expressed) marker genes in **Mac3**, *Cxcl1* and *Cxcl2*, whereas manually parsed data from the Mouse Genome Informatics (MGI) database successfully identifies the correct human homologs. **(C)** Table enumerating the positive marker genes for each mouse-derived myeloid subset, and their conversion to human homologs. The unique converted homologs used for downstream analysis are highlighted in bold. (*continued on next page*)

Figure 5.1 (continued). *Note: there are a greater number of converted mouse genes than there are total mouse genes due to the fact that a single mouse gene can give rise to multiple human homologs (and vice versa). All percentages are calculated relative to row 1 (the total number of mouse marker genes per subset). **(D-E)** Example of mouse-to-human gene expansion, wherein a single mouse gene (*Fcgr4*) gives rise to two human homologs **(D)** or homolog contraction, wherein multiple mouse genes converge on the same human homolog (*HLA-A*) **(E)**. Both these sets of genes were derived from **Mac2** and **DC** as an example. Colour coding represents the mouse myeloid subset for which the gene is a marker. All gene names in capital letters are human genes.

Thus, the MGI method was selected to preserve the fidelity of myeloid subset identities, especially for **Mac3**.

The top differentially expressed marker genes per myeloid subset, the top 20 of which are shown in **Figure 3.6**, were then assessed in more detail for their fidelity, drop-out rate and multi-homology (**Figure 5.1C**). The number of mouse marker genes per myeloid subset varied substantially to begin with, from a minimum of 10 genes in **Mac1** to a maximum of 195 genes in **Mac5**. In general, the number of mouse genes that mapped to a human homolog in a 1-to-1 conversion was relatively high for all myeloid subsets (>81% in all cases) (**Figure 5.1C**). However, some mouse genes possessed more than one human homolog (ie homolog expansion) (**Figure 5.1D**), while other mouse genes converged on the same human homolog (ie homolog contraction) (**Figure 5.1E**). This process of contraction was particularly pronounced in antigen presentation genes such as *HLA-A* (**Figure 5.1E**) which were disproportionately represented in **Mac2** and **DC** subsets. Furthermore, **DC** possessed a relatively high number of mouse genes with no human homolog, amounting to 9% of total mouse genes (**Figure 5.1C**). It is therefore important to note that while most mouse genes successfully converted to a human homolog in a 1-to1 mapping, this conversion is imperfect and may modestly affect the representativeness of each myeloid subset in downstream analyses, particularly **Mac2** and **DC**.

5.1.3 | Conservation of mouse myeloid subsets in human single-cell RNAseq data

To determine whether the myeloid subsets identified by scRNAseq in mouse breast tumours (**Chapter 3**) were conserved in human breast cancer patients, two large scRNAseq atlases of the human breast cancer landscape were assessed for their expression of homolog-converted mouse myeloid subset marker genes.

The first scRNAseq dataset, published by Azizi and colleagues in 2018 (102), describes the immune cell infiltrate of 8 human breast tumours, alongside matched normal breast tissue, peripheral blood, and lymph nodes, totalling ~45,000 ($CD45^+$) cells. The authors' BISCUIT algorithm assigns clusters with a Bayesian approach, harnessing previous transcriptomic datasets to arrive at a finalised cluster assignment. To assess conservation of macrophage phenotypes in this dataset, only the BISCUIT-inferred clusters assigned by the authors as "macrophage", "monocyte" or "monocyte precursor" were selected, and any clusters predominantly (>65%) derived from non-tumour samples were excluded (**Table 5.1**). This led to a total of 4124 cells (across 14 clusters) being taken forward for expression analysis.

The expression of homolog-converted marker genes from mouse myeloid subsets was assessed across all 14 human clusters (**Figure 5.2**). In general, converted human homologs were well expressed in the Azizi *et al.* dataset, however the extent of their conservation within discrete subsets was highly variable. The only subsets exhibiting clear conservation between mouse and human were the 'Cluster 54' monocyte (analogous to **Mac1**), and the 'Cluster 68' and 'Cluster 94' monocytes (analogous to **Mac3**). Most remaining human myeloid clusters exhibited some degree of phenotypic overlap, enriching marker genes from multiple mouse subsets. Indeed, this overlap was particularly pronounced in all three macrophage clusters, which resembled a mixture of **Mac2**, **Mac3**, **Mac4** and **Mac5** phenotypes, indicating a poor discrimination of human macrophage subsets with mouse-derived markers.

Cluster	Cell Type	# of Cells	Top Marker Gene	% from Tumour Tissue	% from Normal Tissue	% from Blood	% from Lymph Node
23	Macrophage	800	C1QC	99.4	0.5	0.1	0
25	Macrophage	677	SPP1	99.6	0.3	0.15	0
28	Macrophage	627	CD109	100	0	0	0
36	Monocyte	308	CDH23	8.4	3.9	87.7	0
40	Monocyte	187	RBP7	86.6	9.6	2.7	1.1
54	Monocyte	79	HACD4	100	0	0	0
67	Monocyte	50	SLC16A10	100	0	0	0
68	Monocyte	50	GZF1	100	0	0	0
86	Monocyte	37	FLCN	100	0	0	0
91	Monocyte	35	IQSEC2	100	0	0	0
94	Monocyte	33	CCL2	100	0	0	0
24	Monocyte Precursor	700	MRC1	69	28.9	2.1	0
27	Monocyte Precursor	658	FCGBP	75.1	24.2	0.5	0.3
42	Monocyte Precursor	165	TNFAIP8L2	2.4	0	97.6	0
49	Monocyte Precursor	97	LYVE1	100	0	0	0
64	Monocyte Precursor	57	ADGRE1	93	3.5	3.5	0

Table 5.1 | Cluster metadata for Azizi *et al.* scRNAseq atlas of human breast cancer. A total of 16 clusters were assigned by Azizi *et al.* (102) as monocytes, macrophages or monocyte precursors. Metadata for all these clusters is presented. Percentages refer to the amount of each cluster originating from each tissue. Clusters 36 and 42 are both majority-derived (>65%) from non-tumour samples. The top marker gene for each subset was defined as the gene with the highest Z score following a differential 'immune' gene expression analysis across all clusters, as per the authors' description. A total of 8 tumour samples were reported in this study.

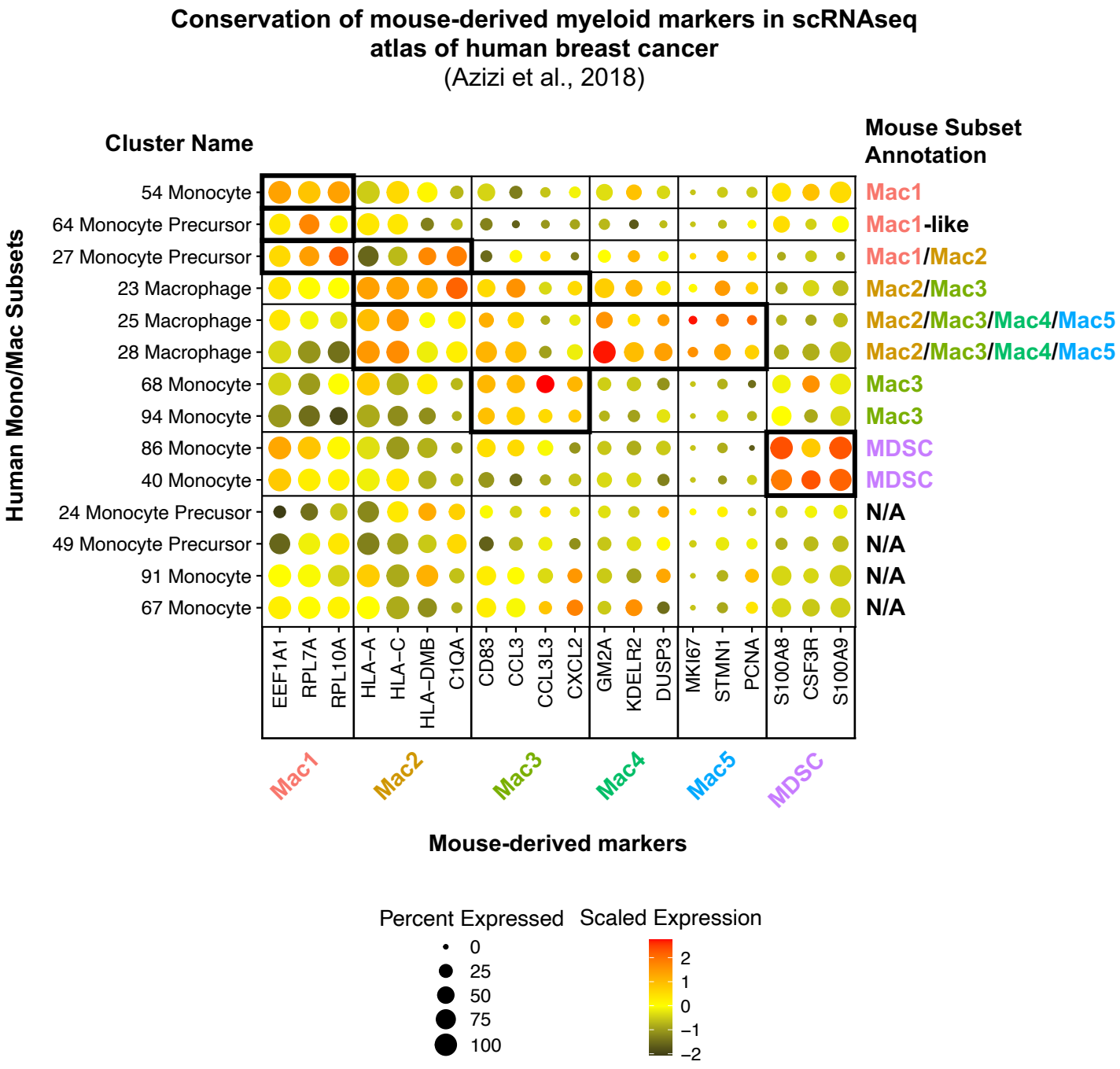


Figure 5.2 | Conservation of mouse myeloid subsets in Azizi *et al.* human breast cancer single-cell RNAseq atlas. Expression bubble plot of top mouse-derived myeloid marker genes, converted to human homologs, and assessed in a human breast cancer immune scRNAseq atlas published by Azizi *et al.* 2018 (102). This analysis was restricted only to clusters assigned by the authors as 'macrophage', 'monocyte' or 'monocyte precursor'. A total of 4124 cells across 14 clusters are shown, from n=8 patient tumours, plus matched normal tissue, lymph node and blood samples. Only clusters that were majority (>65%) derived from tumour samples are considered, and two clusters of monocyte precursors were therefore removed due to their majority presence in the blood (clusters 36 and 42). The genes most closely resembling the marker genes of each mouse subset are highlighted with a thick black border. The mouse subset most closely resembling each human subset is annotated on the right where possible, or defined as N/A for subsets that could not be assigned. (*continued on next page*)

Figure 5.2 (continued). Expression values are scaled per gene (column) to be comparable between genes. Acknowledgement: the Azizi *et al.* dataset was retrieved, cleaned and clustered using the authors' BISCUIT algorithm to recapitulate the original publication, which was entirely performed by Yuechan Chi from the Department of Oncology (University of Oxford). All further analysis, which is presented above, was performed alone with no assistance.

Interestingly, the strong expression of **MDSC** genes within the 'Cluster 86' and 'Cluster 40' monocytes suggest that these clusters may potentially have been mis-assigned by the authors, especially 'Cluster 40 Monocyte' which expresses very high levels of *CSF3R* (the G-CSF receptor), a gene specifically associated with granulocytic cells. Several clusters possessed no clear similarity to any mouse-derived markers, likely indicating several human-specific phenotypes. In the case of the monocyte precursor 'Cluster 24', this may in fact be a normal-mammary tissue specific phenotype, given that this cluster is present in 28% of matched normal breast tissue. The other cluster with relatively high representation in matched normal breast tissue is the monocyte precursor 'Cluster 27', which mimics both **Mac1** and **Mac2** phenotypes in mice (**Figure 5.2**).

While the number of tumour-derived myeloid cells captured by Azizi *et al.* was greater than the number captured by the mouse scRNAseq presented in **Chapter 3** (6074 versus 4024 myeloid cells), the sequencing depth per cell was at least 3-fold lower in the Azizi *et al.* study at 21k reads per cell across tumour samples (102). Given that Azizi *et al.* were able to resolve only three macrophage subsets, which were poorly discriminated using mouse-derived markers, it is reasonable to hypothesise that the reduced sequencing depth in this dataset may have prevented the identification of deeper phenotypic features, such as those defining the functionally-delegated macrophage subsets in mouse.

To address this possibility, the most recent scRNAseq atlas of the human breast tumour microenvironment was assessed once it became available. This dataset was published by Wu and colleagues in 2021 (101), and includes ~100k cells from the entire microenvironment of

26 patient tumours, of which 9675 were assigned as myeloid cells, including a total of 8720 cells labelled as “macrophage” or “monocyte” (**Table 5.2**), which were taken forward for expression analysis (**Figure 5.3**). Using a selection of the same mouse-derived marker genes as were used for the Azizi *et al.* analysis (**Figure 5.2**), the Wu *et al.* dataset possessed strong conservation of 4 out of 5 functionally-delegated macrophage states (**Mac1**, **Mac2**, **Mac3** & **Mac5**, but not **Mac4**). Interestingly, the ribosome-enriched **Mac1** state appears to be conserved in *FCGR3A* Monocytes, suggesting that **Mac1** shares features of monocytic cells. The *APOE* Lipid-associated macrophages (LAMs) showed strong hallmarks of highly antigen-presenting **Mac2**, while *EGR1* Macrophages expressed a chemokine signature resembling **Mac3**. Notably, the *CXCL10* Macrophage subset identified by Wu *et al.* possessed strong expression of both **Mac2** and **Mac3** characteristics simultaneously, indicating the existence of an intermediate state between these two phenotypes. This, in fact, closely aligns with flow cytometric analyses of mouse breast tumours in **Chapter 4**, where a minor population of non-delegated $CXCL1^{+}MHC2^{+}$ cells representing both **Mac2** and **Mac3** states was observed (**Figures 4.5, 4.9, 4.10**). The presence of ‘Cycling Myeloid’ cells was strongly reminiscent of proliferating **Mac5**, and while these were not explicitly assigned as macrophages, it is unclear what other myeloid subset these could possibly represent, given that DCs, monocytes and granulocytes are not known to proliferate. Intriguingly, in both Azizi *et al.* (**Figure 5.2**) and Wu *et al.* (**Figure 5.3**) datasets, the **MDSC** markers appeared to be conserved in Monocytic subsets, despite the expression of the granulocytic gene, *CSF3R*, in these clusters, which would indicate a partly granulocytic origin. The only mouse subset to possess no clear human conservation was the ‘transitionary’ **Mac4** subset, which was poorly conserved across both human datasets. Similar to the Azizi *et al.* dataset, there were at least two human subsets for which no clear mouse conservation existed in the Wu *et al.* dataset, indicating the presence of human-specific myeloid phenotypes. These included the *SIGLEC1* Macrophages which expressed virtually no markers associated with any mouse myeloid subset, and the *FABP5* LAMs which expressed modest amounts of **Mac2** and **Mac3** markers (**Figure 5.3**). Finally, the *IL1B* Monocytes also exhibited some resemblance to **Mac1** through their enrichment of

ribosomal proteins, although they did also express modest amounts of **Mac3** markers, unlike the **Mac1**-equivalent *FCGR3A* Monocytes which did not, and these *IL1B* Monocytes were therefore designated '**Mac1**-like'.

Myeloid Cluster	Cell Name	# of Cells	Marker Gene
N/A	Cycling_Myeloid	463	SPP1
c0	DC	131	LAMP3
c1	LAM1	2122	FABP5
c2	LAM2	1126	APOE
c3	cDC1	156	CLEC9A
c4	pDC	331	IRF7
c5	Macrophage_3	26	SIGLEC1
c7	Monocyte_3	54	FCGR3A
c8	Monocyte_2	935	S100A9
c9	Macrophage_2	505	CXCL10
c10	Macrophage_1	2150	EGR1
c11	cDC2	337	CD1C
c12	Monocyte_1	1339	IL1B

Table 5.2 | Cluster metadata for Wu *et al.* scRNAseq atlas of human breast cancer. A total of 13 clusters were assigned as myeloid cells by Wu *et al.* (101), comprising monocytes, macrophages, lipid-associated macrophages (LAM), cycling cells, plasmacytoid and conventional dendritic cells (DCs). A total of 26 tumour samples were reported in this study. The marker gene for each cluster was assigned by the authors in all cases except the 'Cycling_myeloid' cluster, which was selected on the basis of its highest average log2 fold-change in a differential gene expression analysis of all clusters, performed by the authors. Abbreviations: LAM = Lipid-Associated Macrophage, cDC = Conventional Dendritic Cell, pDC = Plasmacytoid Dendritic Cell.

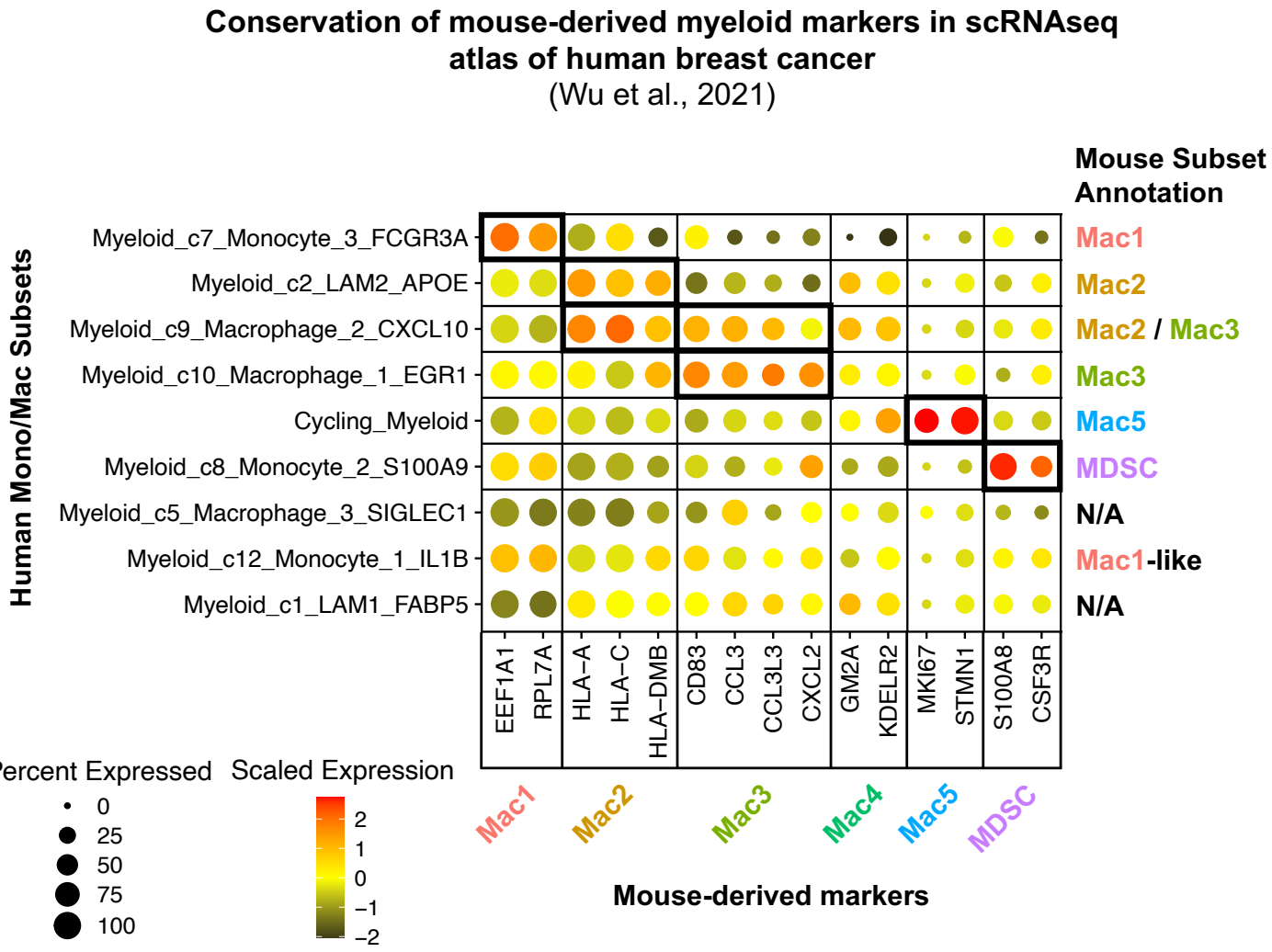


Figure 5.3 | Conservation of mouse myeloid subsets in Wu *et al.* human breast cancer single-cell RNAseq atlas. Expression bubble plot of top mouse-derived marker genes, converted to human homologs, and assessed in a human breast cancer scRNAseq atlas published by Wu *et al.* 2021 (101). This analysis was restricted only to clusters assigned by the authors as 'macrophage' or 'monocyte'. A total of 8720 cells across 9 clusters are shown, from n=26 patient tumours. The genes most closely resembling the marker genes of each mouse subset are highlighted with a thick black border. The mouse subset most closely resembling each human subset is annotated on the right where possible, or defined as N/A for subsets that could not be assigned. Expression values are scaled per gene (column) to be comparable between genes. Abbreviations: LAM = Lipid-associated Macrophage.

Expression of the **Mac3** epigenetic regulator, *KDM6B*, which was identified in **Chapter 3** and validated in **Chapter 4**, was also assessed across the whole breast tumour microenvironment using the Wu *et al.* dataset (**Figure 5.4**). *KDM6B* was expressed almost exclusively within the myeloid compartment (aside from low expression in luminal cells), and specifically within the macrophage compartment, *KDM6B* was highly expressed only in *EGR1* Macrophages (orthologous to **Mac3**) and *SIGLEC1* Macrophages (for which no orthologous mouse subset existed). This data suggests *KDM6B* may be a conserved regulator of **Mac3** in human breast cancer. In contrast, the putative **Mac3** regulator, *PIM1*, was relatively lowly expressed within myeloid cells but much more highly expressed in the T cell compartment, indicating poor specificity of *PIM1* for the myeloid compartment.

Though the conservation of dendritic cells was not the focus of this chapter, the heterogeneous **DC** mouse subset was also conserved in the Wu *et al.* dataset, where its marker genes segregated into 3 out of the 4 human DC clusters (**Appendix Figure 5.1**).

Finally, to address previous concerns in **Chapter 3** that the FACS sorting strategy used for scRNAseq may have excluded an unusual population of $CD3^+$ macrophages, the Wu *et al.* dataset was assessed for expression of CD3 genes (**Appendix Figure 5.2**). This analysis verified an absence of *CD3D*, *CD3E* and *CD3G* expression in all myeloid subsets, suggesting that this unusual $CD3^+$ macrophage population is not conserved in humans, and may merely be an artefact or mouse-specific phenomenon.

Taken together, human scRNAseq atlases of breast cancer both suggest conservation of myeloid phenotypes between mouse and human breast tumours, though the more recent Wu *et al.* dataset provides much greater clarity and conservation, likely owing to its greater sequencing depth, cell numbers, cohort size and subtype representation.

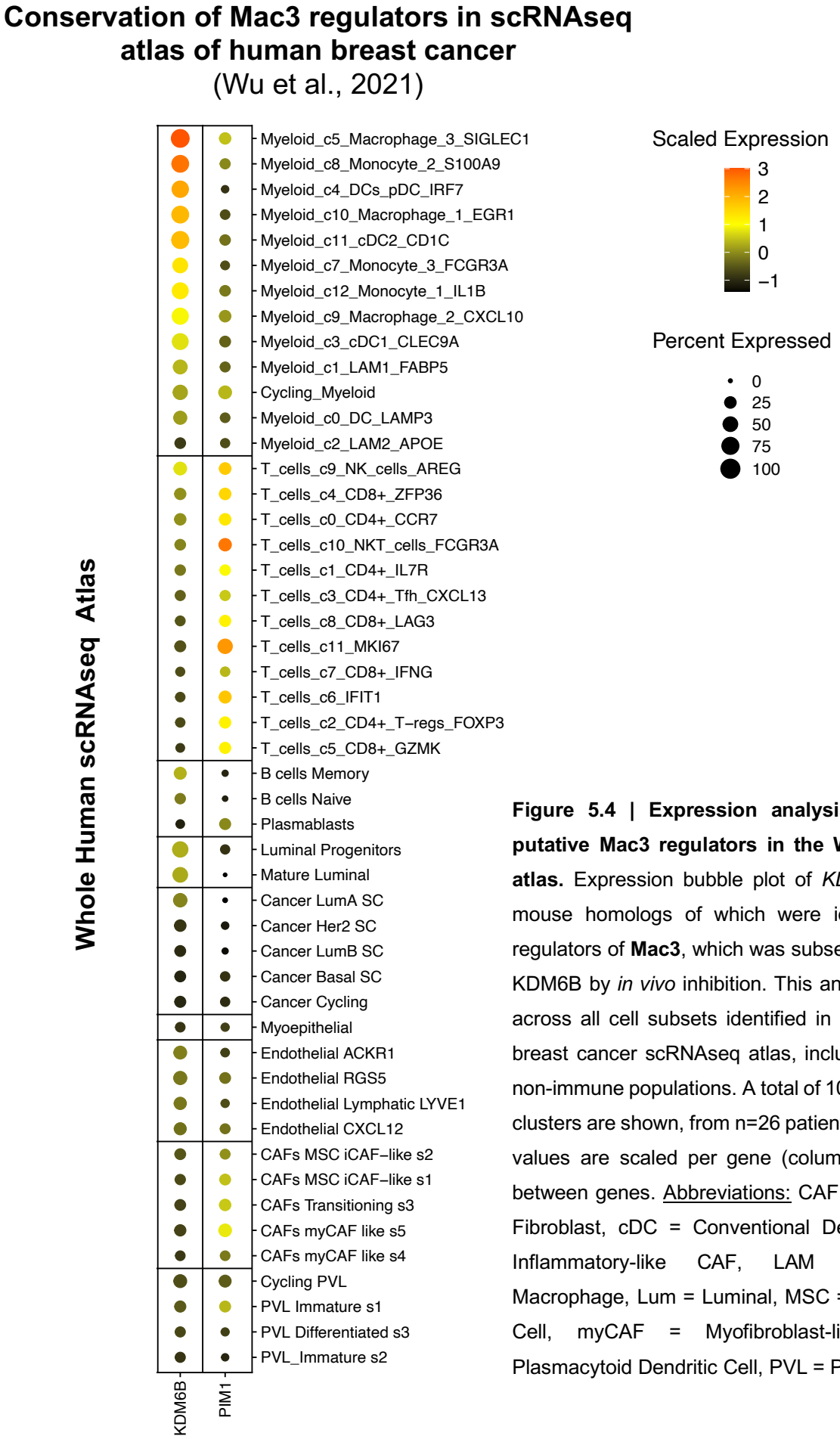


Figure 5.4 | Expression analysis of validated and putative Mac3 regulators in the Wu et al. scRNAseq atlas. Expression bubble plot of *KDM6B* and *PIM1,* the mouse homologs of which were identified as possible regulators of **Mac3**, which was subsequently confirmed for KDM6B by *in vivo* inhibition. This analysis was performed across all cell subsets identified in the Wu et al. human breast cancer scRNAseq atlas, including all immune and non-immune populations. A total of 100,064 cells across 49 clusters are shown, from n=26 patient tumours. Expression values are scaled per gene (column) to be comparable between genes. Abbreviations: CAF = Cancer-associated Fibroblast, cDC = Conventional Dendritic Cell, iCAF = Inflammatory-like CAF, LAM = Lipid-associated Macrophage, Lum = Luminal, MSC = Mesenchymal Stem Cell, myCAF = Myofibroblast-like CAF, pDC = Plasmacytoid Dendritic Cell, PVL = Perivascular-Like.

5.1.4 | Quality & refinement of gene signatures in human RNAseq data (TCGA)

To assess the prognostic association of mouse-derived myeloid subsets in breast cancer patients, bulk transcriptomic studies with large clinical cohorts were sought. The Cancer Genome Atlas (TCGA) contains bulk RNAseq data for 1045 breast tumour samples (BRCA), and this dataset (65, 66) was initially selected for assessing mouse-derived myeloid gene signatures.

Each gene signature was defined as the unique converted human homologs for a given mouse myeloid subset, as shown in **Figure 5.1B**. To assess the behaviour and quality of each gene signature, the R package SigQC (293) was utilised. Signatures were assessed for their expression, variability, compactness, structure and scoring metrics across all samples of the TCGA BRCA dataset. The ideal signature is one that is compact (highly autocorrelated), with reasonably high expression and high variability across samples, as well as fulfilling a variety of statistical characteristics. To summarise this multitude of properties in a single visualisation, a radar plot was generated for each signature and compared (**Figure 5.5A**). A global value was assigned per signature which calculates the proportion of area in the radar plot occupied by each summary polygon, with higher scores being preferable. For reference, the authors of SigQC report a summary value of 0.39 for a high quality 'Metastasis Signature' in the very same TCGA BRCA RNAseq dataset, compared to a randomly generated signature which achieved 0.16 (293). Comparing the radar plot in **Figure 5.5A**, the majority of myeloid subset signatures (5 out of 7: **Mac2**, **Mac3**, **Mac4**, **Mac5** & **DC**) bear hallmarks of high-quality signatures (global value of 0.33-0.37), however **Mac1** and **MDSC** had relatively low scores of 0.22 and 0.25, respectively.

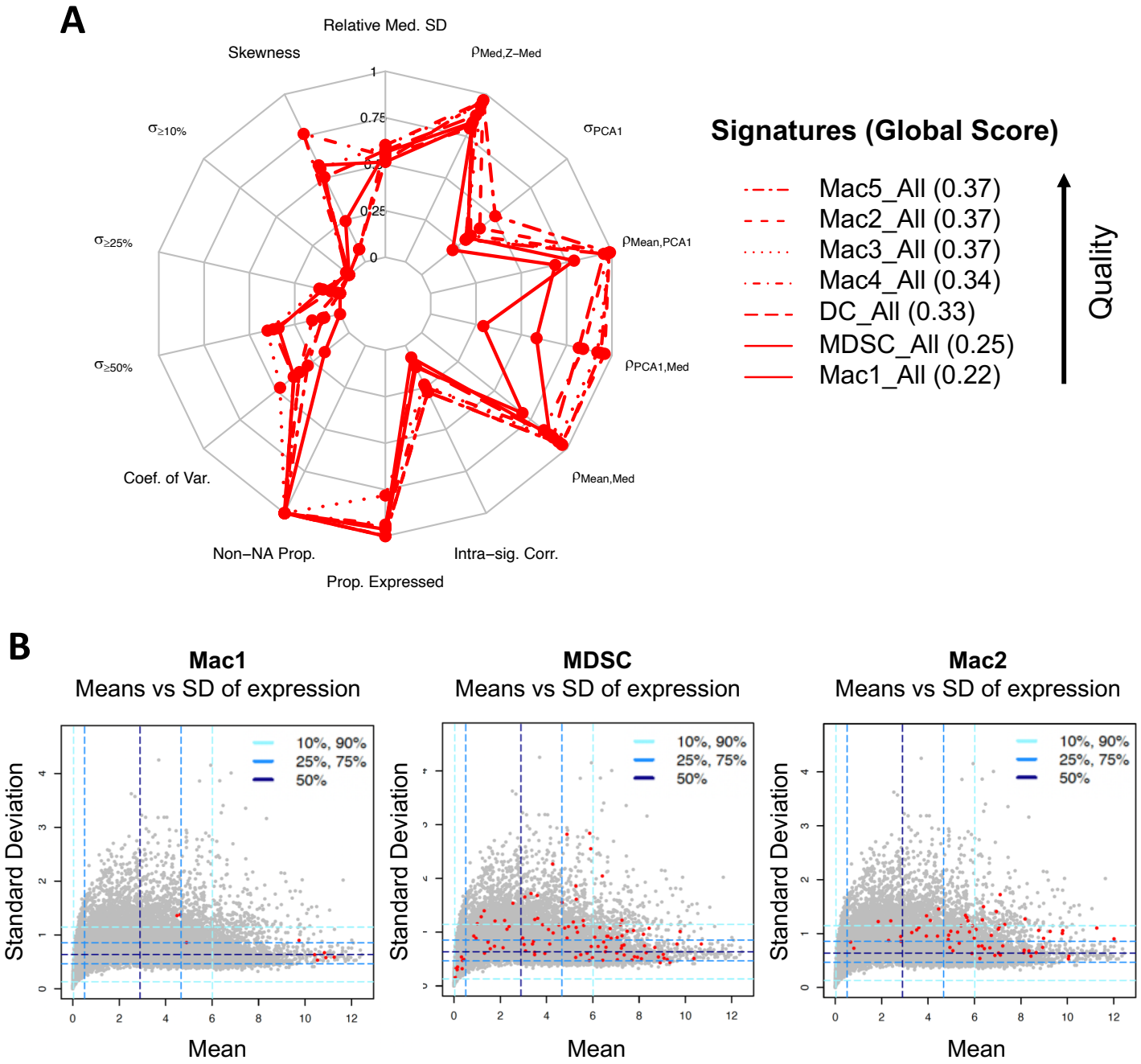


Figure 5.5 | Quality summary of mouse-derived myeloid gene signatures in TCGA bulk RNAseq data. A total of 1021 bulk RNAseq human breast tumour samples from TCGA (65, 66) were analysed. Expression values in this dataset represent TPM and were log2(x+1) transformed prior to analysis with the R Package, SigQC (293). **(A)** Radar plot summarising key properties of each gene signature, generated by SigQC. The value in parentheses represents the proportion of area occupied by each polygon, thus approximating global quality of the signature. **(B)** Evaluation of expression and variability of individual genes in each signature (red) compared to all other detected genes in the dataset (grey). Coloured dotted lines represent the percentiles of the distribution of all detected genes. Abbreviations: TCGA = The Cancer Genome Atlas, TPM = Transcripts Per Million.

Both **Mac1** and **MDSC** signatures performed consistently poorly in terms of the amount of variance explained by their first principal component (PC) in a principal component analysis (PCA), represented by their low standard deviation in PC1 (σ PCA1). Performance of a Spearman's rank correlation (ρ) between the median expression of the signature genes in each sample against the first PC of that sample (ρ PCA1,Med) informs whether the expression of signature genes is changing in a coherent and compact manner, and that these changes are well represented by the midpoint of the distribution. In all cases, **Mac1** and **MDSC** score substantially lower in this metric than other myeloid gene signatures. For **Mac1**, the reason for this issue likely stems from the fact the signature is small and mostly comprised of ribosomal protein genes which are constitutively expressed across all samples, meaning there is a high mean expression per sample but low inter-sample variability (standard deviation), as illustrated in **Figure 5.5B**. While the **MDSC** signature has much higher variance than **Mac1**, it also contains a large proportion of genes with low levels of expression, which may partly explain its suboptimal performance. By contrast, the exemplar higher quality signature, **Mac2**, possesses genes with a relatively high mean expression level as well as a relatively high variance (**Figure 5.5B**).

The issue of constitutive gene expression hindering signature variability in fact highlights a wider strategic problem with this analysis. The gene signatures of each mouse myeloid subject were derived by differential expression analysis between individual FACS-sorted myeloid cells, meaning the background on which these genes are differential is relative to the myeloid compartment, and is therefore not representative of the background of the bulk tumour samples, which contain a range of other cell types.

A

	Mac1	Mac2	Mac3	Mac4	DC	Mac5	MDSC
Original Gene Signature (# of unique converted human homologs)	**9**	**67**	**38**	**25**	**100**	**163**	**114**
# of genes deducted by gene signature refinement	**5**	**0**	**0**	**1**	**8**	**55**	**2**
Refined Gene Signature (after subtraction of pan-cancer / breast-cancer fitness genes)	**4** (44%)	**67** (100%)	**38** (100%)	**24** (96%)	**92** (92%)	**108** (66%)	**112** (98%)

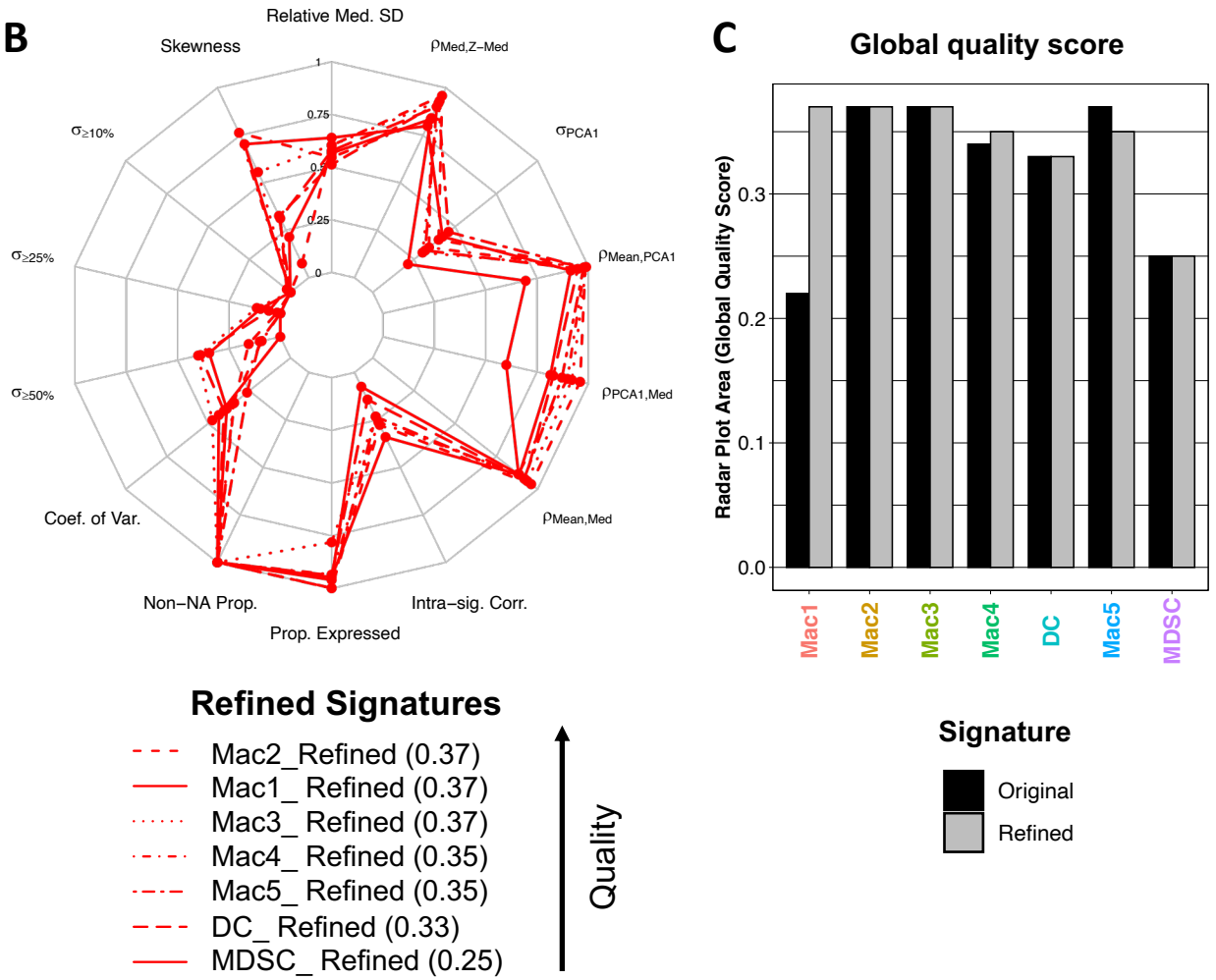


Figure 5.6 | Quality assessment of myeloid signatures refined via fitness-gene deduction. (A) Mouse myeloid gene signatures which had been converted to human homologs, were refined according to a list of 732 'fitness genes' that are required by all tumour cells and breast-specific tumour cells to survive, as determined by a large-scale genome-wide synthetic lethality screen across 324 cancer cell lines (292).This refinement excludes constitutively expressed essential genes, such as ribosomal proteins, that lower inter-sample variability, and are not specific to the myeloid compartment, meaning they would likely not be considered "differential genes" if they had been analysed against the whole tumour background (rather than only the myeloid background). Percentages are shown relative to the original gene signature**. (B)** Radar plot summarising key quality metrics for each refined signature, outputted by the R package 'SigQC' (293). **(C)** Comparison of global signature quality scores as determined by the area occupied by each radar plot polygon. Abbreviations: PCA = Principal Component Analysis, SD = Standard Deviation.

To mitigate the issue of tumour cells – which represent the majority of each bulk clinical sample – from biasing the gene signature due to their shared expression of constitutive/essential genes (such as the ribosomal proteins in the **Mac1** signature), and to improve the inter-sample variability of each signature, a new method was devised to refine the gene signatures. Harnessing recent advances in genome-wide synthetic-lethality screens, involving CRISPR knockouts of all genes across 324 human cancer cell lines (292), a list of 'core fitness' genes was obtained that are essential for the survival of all human cancer cell lines. Supplementation of this list with any fitness genes that were universally essential in breast cancer cell lines (n=25 lines) yielded a total list of 732 fitness genes, 87 of which were ribosomal protein genes. This final list, representing a heuristic for all common-essential genes in human breast cancer cells, was then deducted from the original myeloid gene signatures to provide a list of refined signatures which will not overlap with the background transcriptional landscape of tumour cells (**Figure 5.6A**). The proportion of genes deducted from the original signature varied greatly between subsets, with the largest deduction being within **Mac5** by absolute number, and within **Mac1** by relative proportion. This inadvertent reduction in the **Mac5** gene signature occurred due to proliferation-related genes that were shared between proliferating macrophages and proliferating cancer cells, and thus the deduction of 55 fitness genes should reduce the biological noise associated with cancer-cells expressing this proliferation signature, serving to improve interpretability of the bulk sample data. Finally, quality assessment of the refined signatures by radar plot (**Figure 5.6B**) revealed a substantial increase in the quality score for the refined **Mac1** signature compared to its original score (**Figure 5.6C**), while other signatures were relatively unaffected, highlighting the benefits of the fitness-gene-deduction approach in this context.

Finally, to confirm the quality and significance of each gene signature, random resampling was performed to assess the underlying null distribution of each QC metric, serving as a negative control against which genuine signatures could be compared (**Appendix Figure 5.3** & **Appendix Figure 5.4**). This analysis demonstrated that genuine signature genes generally

lay outside of the 'control' null distribution for each QC metric, indicating reasonable quality, however this was not the case for all metrics within the **Mac1**, **Mac3** and **MDSC** signatures. Taken together, these data indicate that signature quality is variable but generally of a reasonable quality, and is therefore sufficient to justify downstream survival analysis for all myeloid subsets.

5.1.5 | Gene signature enrichment & hierarchical clustering of TCGA data

Numerous methods exist to assess the presence of gene signatures in bulk transcriptomic samples. Rather than calculating the net expression of all genes in a given signature, gene set enrichment (GSE) methods focus on the enriched pathways of a given signature, serving to reduce dimensionality and noise while aiding biological interpretability. The Gene Set Variation Analysis (GSVA) algorithm is a non-parametric unsupervised GSE method which estimates the variation of pathway activity across samples, and provides increased power to detect subtle changes in pathway activity compared to other methods, including in survival analyses (294). GSVA enrichment scores for the TCGA BRCA RNAseq dataset were outputted by the R package, SigQC, and these were used for subsequent analyses.

Previous single-cell and spatial transcriptomic analyses of the human breast tumour microenvironment by Wu and colleagues demonstrated the importance of defining tumour ecotypes – niches of co-enriched cell types – in predicting patient survival from bulk clinical genomic data (101). By using single-cell RNAseq signatures to deconvolute the abundance of cancer, immune and stromal cell types in bulk transcriptomic samples, followed by hierarchical clustering, the authors showed that tumour ecotypes not only prognostically stratified patients, but also showed no association with the 'integrative genomic clusters' of breast cancer (386), suggesting that ecotypes are not merely a surrogate for, or confounder of, molecular and genomic subtypes. Thus, rather than considering each cell subset in

isolation, a more rich and representative understanding of tumour biology can be gleaned with ecotype-based stratification.

Emulating this approach, GSVA scores of all refined myeloid signatures were hierarchically clustered across all TCGA BRCA tumour samples (n=1021) to yield groups of patients with similar myeloid ecotype profiles, hereafter known as ecotypes. The hierarchical cluster dendrogram was partitioned by k-means clustering with a high number of iterations (n=1000) to ensure the robustness and stability of each ecotype (**Figure 5.7**). Partitioning into 10 ecotypes (k=10) produced a well-demarcated heatmap showing distinct strata of differentially enriched myeloid cell profiles. These included ecotypes: **(i)** strongly enriching only 1-2 myeloid signatures with low enrichment of others (E6, E8, E9, E10), **(ii)** those strongly enriching only 1-2 myeloid signatures with modest enrichment of others (E7), **(iii)** those strongly enriching ≥5 myeloid signatures (E1, E3, E5), and **(iv)** those moderately-to-strongly enriching ≥5 myeloid signatures (E2, E4, E7).

The ecotypes of E1-5 all universally enriched the four signatures **Mac2**, **Mac3**, **DC** and **MDSC**, while ecotypes E6-10 mostly did not, with the exception of E6 which modestly enriched some signatures. This suggests that these four myeloid signatures are positively co-varying. While this stark bifurcation of the ecotypes E1-5 and E6-10 can partly be explained by these four signatures, it is important to note that further substructure exists in the data, leading to the segregation of discrete populations that are ‘single-positive’ or ‘double-positive’ for distinct myeloid signatures. The ecotypes E10, E8 and E9 strongly enriched signatures for **Mac1** alone, **Mac5** alone**,** and both combined, respectively (**Figure 5.7**). Furthermore, the enrichment of **Mac1** and **Mac5** was mutually exclusive in all ecotypes except E3 and E9 where both were present, and in E6 where neither were present, indicating that the presence of these subsets is highly context-dependent. Notably, aside from **Mac1** and **Mac5**, the only other myeloid signature to create a ‘single-positive’ ecotype (an ecotype which enriched only one myeloid signature) was E6, which modestly enriched **Mac4**.

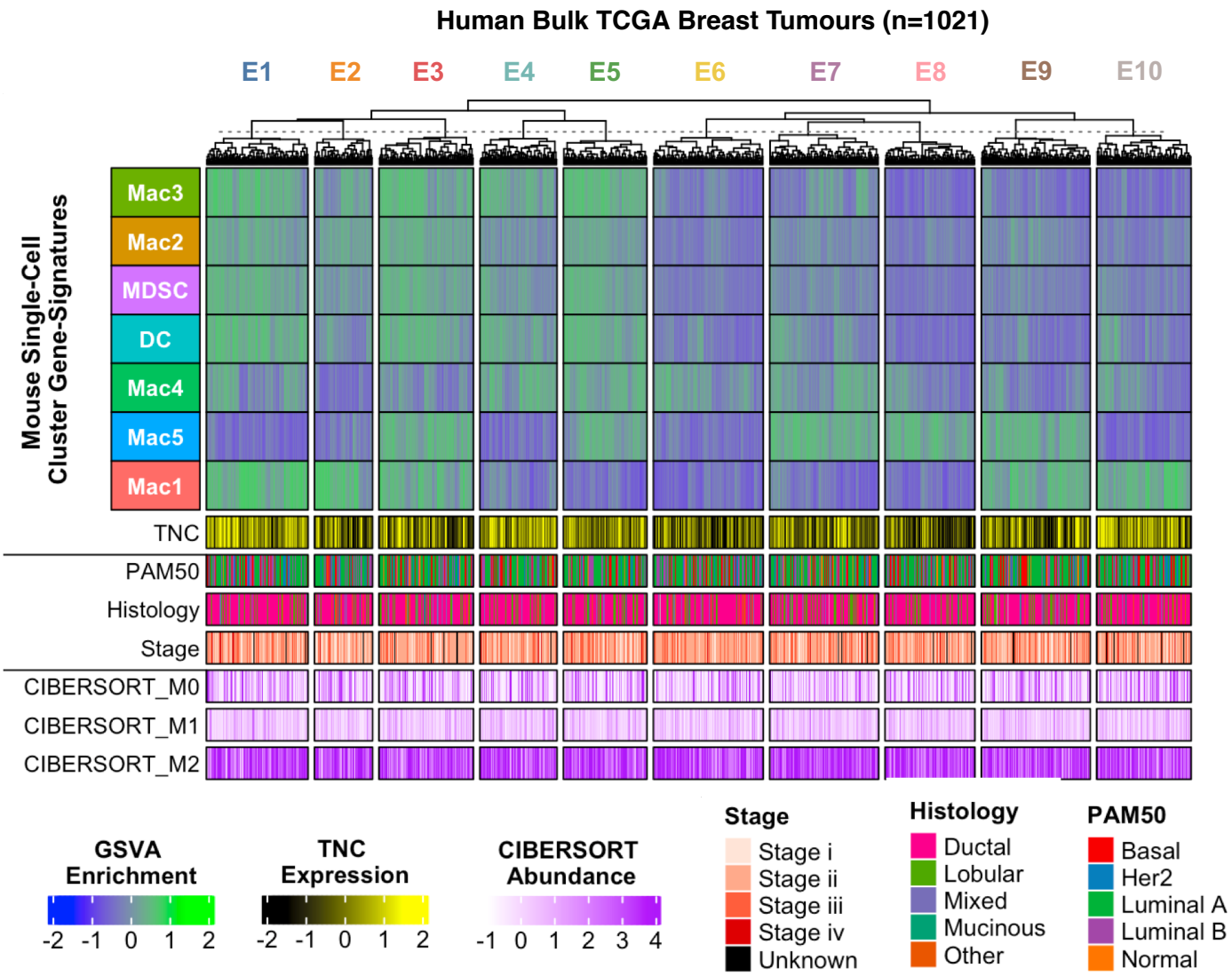


Figure 5.7 | Hierarchical clustering of GSVA enrichment scores to derive myeloid ecotypes in TCGA data. A total of 1021 bulk RNAseq human breast tumour samples from TCGA were analysed. Expression values in this dataset represent TPM and were log2(x+1) transformed prior to analysis with the R Package, SigQC (293). GSVA enrichment scores estimate the variation of pathway activity across samples, and GSVA scores for all myeloid signatures (rows) were used to hierarchically cluster TCGA bulk RNAseq samples (columns) in a heatmap. The dendrogram was partitioned by k-means clustering with 1000 iterations to ensure the robustness and stability of each cluster, representing an ecotype of myeloid signatures. A total of 10 ecotypes (E1-10) were defined. Annotation of clinical metadata and TNC expression are indicated for each tumour sample (columns), including the PAM50 molecular subtype, the histological subtype, and disease stage. CIBERSORT deconvolution analysis (387) estimates the abundance of infiltrating immune cells in bulk data, and served as a positive control to confirm the existence and distribution of intra-tumoural macrophages across ecotypes. CIBERSORT macrophage signatures are denoted as "pro-inflammatory" (M1), "anti-inflammatory" (M2) or "unpolarised" (M0) as per the LM22 gene signature matrix (387). Abbreviations: CIBERSORT = Cell Type Identification By Estimating Relative Subsets of RNA Transcripts, GSVA = Gene Set Variation Analysis, LM22 = Leukocyte Signature Matrix 22, PAM50 = Prediction Analysis of Microarray 50, TCGA = The Cancer Genome Atlas, TPM = Transcripts Per Million.

Annotation of each ecotype with clinical metadata and TNC expression revealed that, while TNC levels varied substantially between ecotypes, the clinical features of histology and disease stage were mostly unaffected. The only exception to this appears to be E1 which includes a modestly elevated proportion of 'basal' tumours, which is a subtype known to consist mostly of triple-negative tumours (63, 64, 67). Finally, as a positive control to confirm the presence of macrophages in this TCGA dataset, CIBERSORT deconvolution analysis with the LM22 gene signature matrix (387) was performed to estimate the abundance of infiltrating immune cells within tumour samples, and this confirmed a high infiltration of suppressive "M2-like" macrophages, and a low infiltration of pro-inflammatory "M1-like" and unpolarised "M0" macrophages (**Figure 5.7**). This analysis not only validates that all ecotypes do indeed enrich macrophages, but that the conventional M0, M1 and M2 signatures ascribed by CIBERSORT cannot explain the true diversity of macrophage phenotypes identified by scRNAseq.

5.1.6 | Survival analysis of patient tumour ecotypes in TCGA data

To assess the prognostic implications of patients with diverse myeloid ecotypes, Kaplan-Meier survival analysis was performed across all 10 ecotypes that were previously identified by hierarchical clustering (**Figure 5.8A**). While differential trends in survival were observed for various ecotypes, performance of a log-rank test demonstrated that ecotypes were not significantly different from each other at the whole group level (P=0.26).

The ecotypes E10, E8 and E9 – which strongly enriched signatures for **Mac1** alone, **Mac5** alone**,** and both combined, respectively (**Figure 5.7**) – all bear a similarly poor prognosis (**Figure 5.8**), indicating that naive-like macrophages (**Mac1**) and proliferating macrophages (**Mac5**) are equally pathogenic constituents of the breast tumour microenvironment. Interestingly, ecotypes E4 and E10 had relatively high amounts of TNC expression and were both associated with among the least favourable prognoses, indicating that TNC may be able

to stratify pathogenic myeloid ecotypes. However, ecotypes with low levels of TNC expression in fact harboured a wide range of favourable and unfavourable prognoses (E2, E6, E8), suggesting that any association between TNC and prognosis was likely context-dependent, perhaps depending on specific myeloid profiles.

Ecotypes with trends towards highly differential prognoses, either favourable (E2, E5) or unfavourable (E4, E10), were selected for further assessment by pairwise Kaplan-Meier analysis (**Figure 5.8B**). This revealed a nominal level of significance (P=0.046) by a log-rank test indicating a difference between E2 versus E10 patient survival, and all other comparisons were only marginally above this level of significance at $P \leq 0.06$ (E2 vs E4, E4 vs E5, and E5 vs E10). Interestingly, ecotypes with a more favourable prognosis (E2 and E5) shared in common an enrichment of **Mac2**, **Mac3**, **DC** and **MDSC**, while ecotypes with a less favourable prognosis (E4 and E10) shared in common an enrichment of **Mac5**.

Given these promising trends of differential prognoses between specific ecotypes, yet modest levels of statistical confidence, a new dataset was sought with a larger sample size, increased duration of patient follow-up, and thus improved statistical power. The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) dataset contains bulk microarray data for a large cohort of n=1991 breast cancer patients (290). Comparison of patient survival data from METABRIC and TCGA revealed a greatly improved duration of patient follow-ups in the METABRIC dataset, leading to a broader distribution of survival times (**Figure 5.9**). The METABRIC dataset was therefore selected for subsequent analysis.

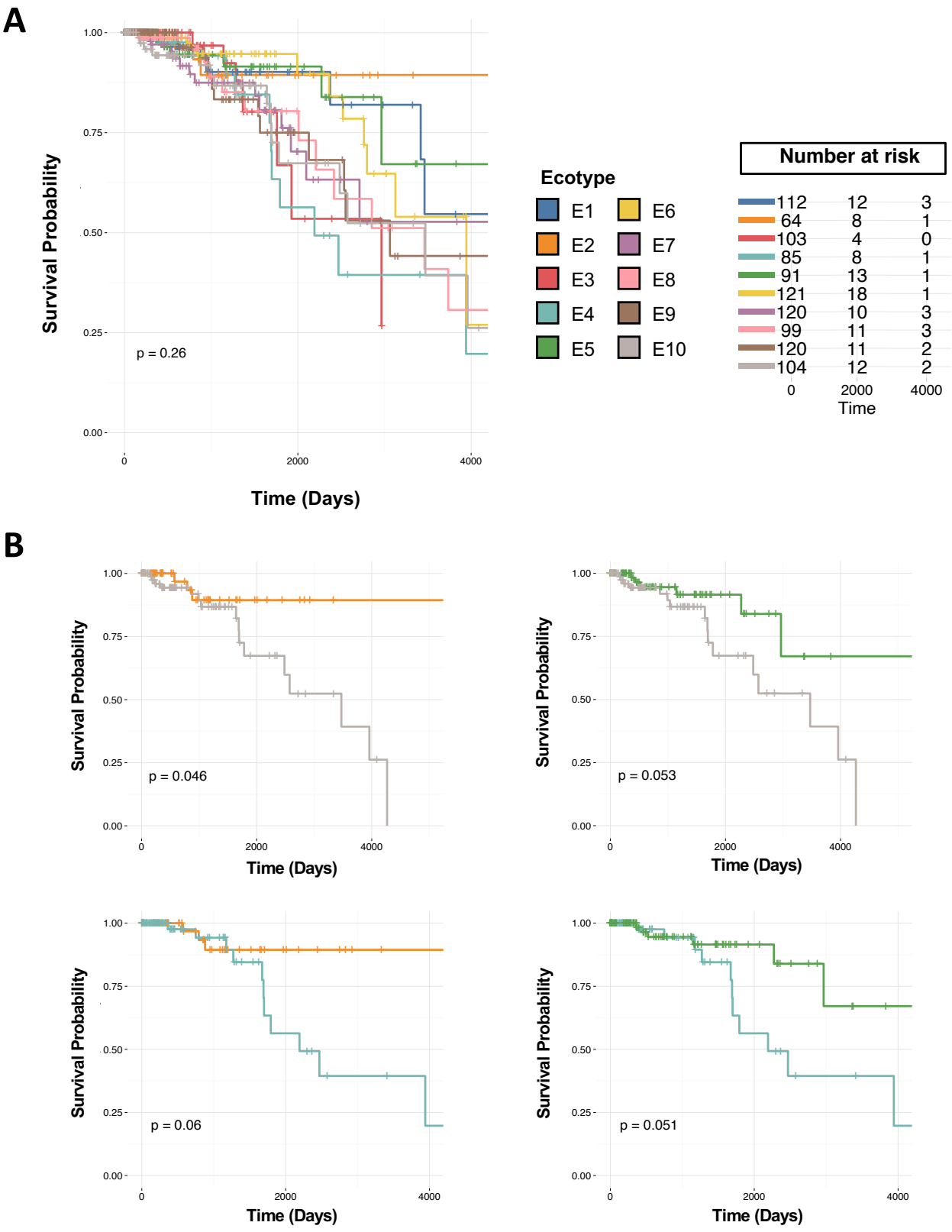


Figure 5.8 | Survival analysis of myeloid ecotypes in the TCGA BRCA dataset. A total of 1021 bulk RNAseq human breast tumour samples from TCGA were analysed. Expression values in this dataset represent TPM and were log2(x+1) transformed prior to analysis with the R Package, SigQC (293). Myeloid ecotypes were defined by consensus hierarchical clustering of GSVA enrichment scores for each mouse-derived myeloid signature. **(A)** Kaplan-Meier survival curves of all patient groups (n=10) harbouring distinct myeloid ecotype profiles (left). (*continued on next page*)

Figure 5.8 (continued). The number of patients at risk at each major timepoint is indicated (right), measured in days. **(B)** Kaplan-Meier survival curves of selected tumour ecotype profiles showing highly differential prognoses for patients between selected ecotypes. For all plots, P values were calculated by performance of a Log-rank test, and values <0.05 were considered statistically significant. Abbreviations: GSVA = Gene Set Variation Analysis, TCGA = The Cancer Genome Atlas, TPM = Transcripts Per Million.

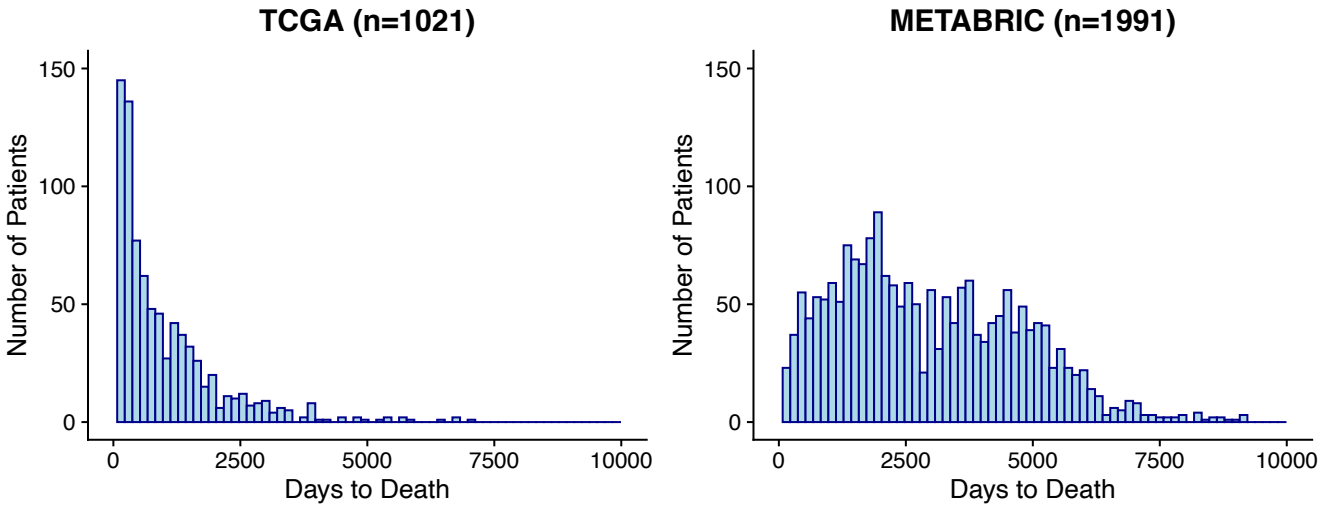


Figure 5.9 | Comparison of survival follow-up data between TCGA and METABRIC datasets. Histograms comparing the duration of patient survival 'follow-up' between two bulk clinical transcriptomic datasets of breast cancer, indicating a greatly increased duration and range of follow-up in the bulk microarray METABRIC dataset (n=1991) (290) compared with the bulk RNAseq TCGA dataset (n=1021) (65, 66). Acknowledgment: For METABRIC, clinical metadata was cleaned and amalgamated by Yuechan Chi, and the dataset was provided by the group of Professor Francesca Buffa, both from the Department of Oncology (University of Oxford). All analysis presented above was performed alone with no assistance. Abbreviations: METABRIC = Molecular Taxonomy of Breast Cancer International Consortium, TCGA = The Cancer Genome Atlas.

5.1.7 | Quality of myeloid gene signatures in human bulk array data (METABRIC)

To successfully harness the METABRIC array dataset (290), refined mouse-derived myeloid signatures were first assessed for their quality characteristics with SigQC (293), in the same manner as was performed for the TCGA dataset. To summarise these characteristics in a single visualisation, a radar plot was generated to compare each signature (**Figure 5.10A**). While most signature scores were of a high quality at ≥0.37 (**Mac1**, **Mac2**, **Mac5**, **DC**), the **Mac4** signature performed poorly with a score of 0.20. This was mostly due to a very low Spearman correlation between the median expression of the signature genes in each sample against the first PC of that sample (ρPCA1,Med), indicating that expression of signature genes does not change in a coherent and compact manner. Indeed, the expression level and inter-sample variability of most **Mac4** signature genes was relatively low, with only 7 out of 24 genes falling above the 75th percentile for both expression and variance (**Figure 5.10B**). By contrast, the exemplar higher quality signature, **Mac2**, possesses a greater proportion of genes with a high mean expression level and variance (**Figure 5.10B**). Comparison of signature quality scores between METABRIC and those previously observed in TCGA revealed that quality was generally higher in the METABRIC dataset, substantially outperforming TCGA quality scores for most signatures, with the exception of **Mac4** which saw a large decline, and **Mac3** which saw a modest decline (**Figure 5.10C**).

To further assess the quality and significance of each gene signature, random resampling was performed to assess the underlying null distribution of each QC metric, serving as a negative control against which genuine signatures could be compared (**Appendix Figure 5.5** & **Appendix Figure 5.6**). This analysis demonstrated that genuine signature genes generally lay outside of the 'control' null distribution for each QC metric, indicating high quality, however this was not the case for **Mac4** which only rarely lay outside the null distribution. Similar to observations in the TCGA dataset (**Appendix Figure 5.3** & **Appendix Figure 5.4**), QC

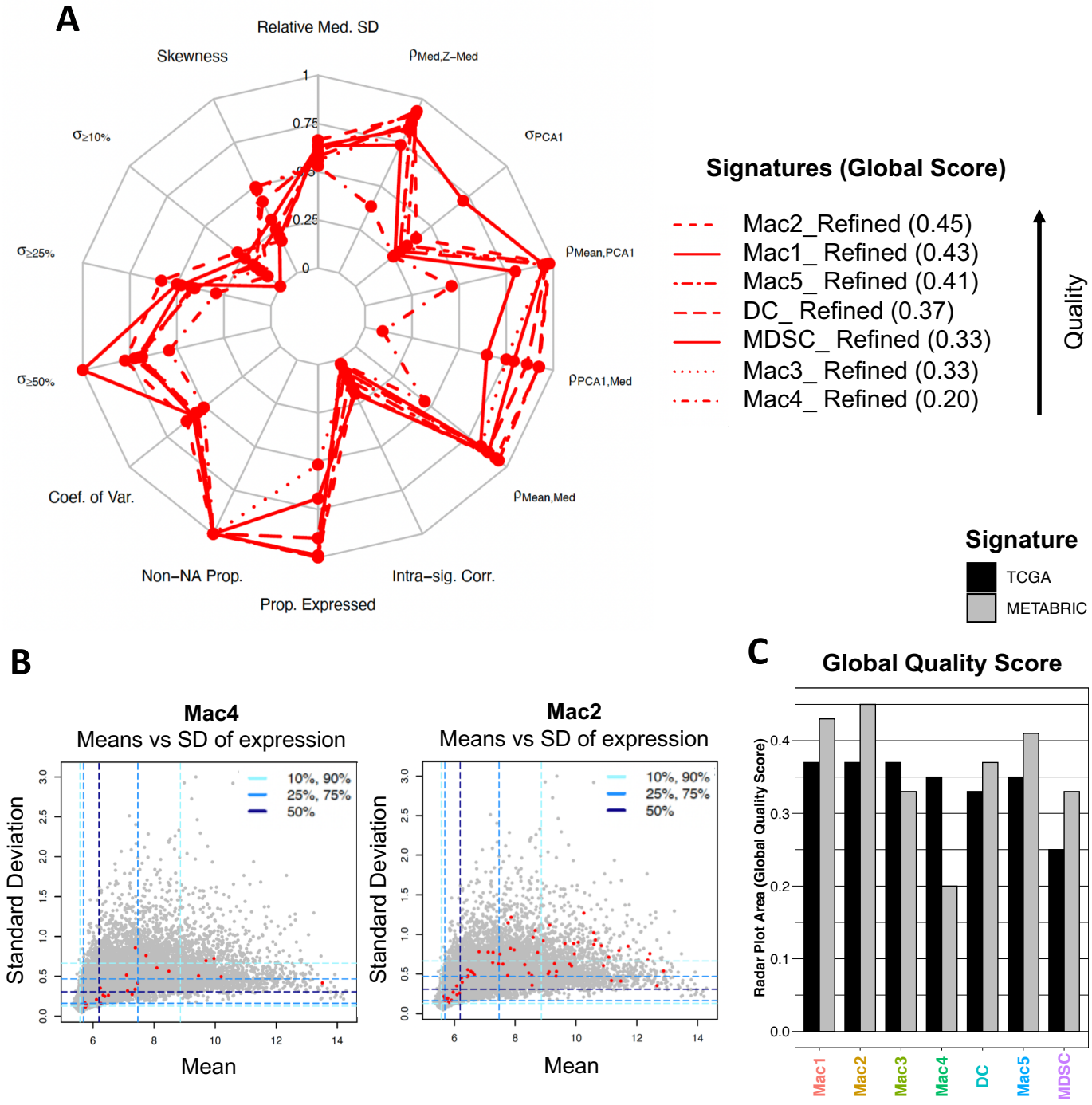


Figure 5.10 | Quality summary of mouse-derived myeloid gene signatures in METABRIC data. A total of 1991 bulk microarray human breast tumour samples from METABRIC (290) were analysed with the R Package, SigQC (293). **(A)** Radar plot summarising key properties of each gene signature. The value in parentheses represents the proportion of area occupied by each polygon, thus approximating global quality of the signature. **(B)** Evaluation of expression and variability of individual genes in each signature (red) compared to all other detected genes in the dataset (grey). Coloured dotted lines represent the percentiles of the distribution of all detected genes. **(C)** Comparison of global signature quality scores between TCGA and METABRIC breast cancer datasets, as determined by the area occupied by each radar plot polygon. (*continued on next page*)

Figure 5.10 (continued). Acknowledgment: Clinical metadata was cleaned and amalgamated by Yuechan Chi, and the dataset was provided by the group of Professor Francesca Buffa, both from the Department of Oncology (University of Oxford). All analysis presented above was performed alone with no assistance. Abbreviations: METABRIC = Molecular Taxonomy of Breast Cancer International Consortium, TCGA = The Cancer Genome Atlas.

metrics for **Mac1** and **Mac3** lay mostly outside or towards the end of the null distribution, but not in all cases. For instance, both **Mac1** and **Mac3** had a low 'skewness' score, indicating the skew of their expression distribution over all samples was not significantly different (skewed) compared to the overall expression distribution of all genes. Taken together, these QC data indicate that myeloid signatures are generally of a high quality in the METABRIC dataset, and are sufficient to justify downstream survival analysis for all signatures, although **Mac4** data must be interpreted with particular caution.

5.1.8 | Gene signature enrichment & hierarchical clustering of METABRIC data

Recapitulating the approach taken for TCGA bulk RNAseq tumour samples in **Section 5.1.5**, GSVA scores of all refined myeloid signatures were hierarchically clustered across all METABRIC bulk microarray tumour samples (n=1911) to define groups of patients with similar myeloid ecotype profiles. The hierarchical cluster dendrogram was partitioned by k-means clustering with a high number of iterations (n=1000) to ensure the robustness and stability of each ecotype (**Figure 5.11**). Partitioning into 10 ecotypes (k=10) produced a well-demarcated heatmap showing distinct strata of differentially enriched myeloid cell profiles. These included ecotypes: **(i)** strongly enriching 1-2 myeloid signatures (E4, E10), **(ii)** those mildly enriching 1-2 myeloid signatures (E3, E9), **(iii)** those strongly enriching 1-2 myeloid signatures plus mildly enriching ≥5 myeloid subsets (E6, E8), **(iv)** those strongly enriching >5 myeloid signatures (E1, E2, E5), and, **(v)** those mildly enriching ≥5 myeloid subsets (E7).

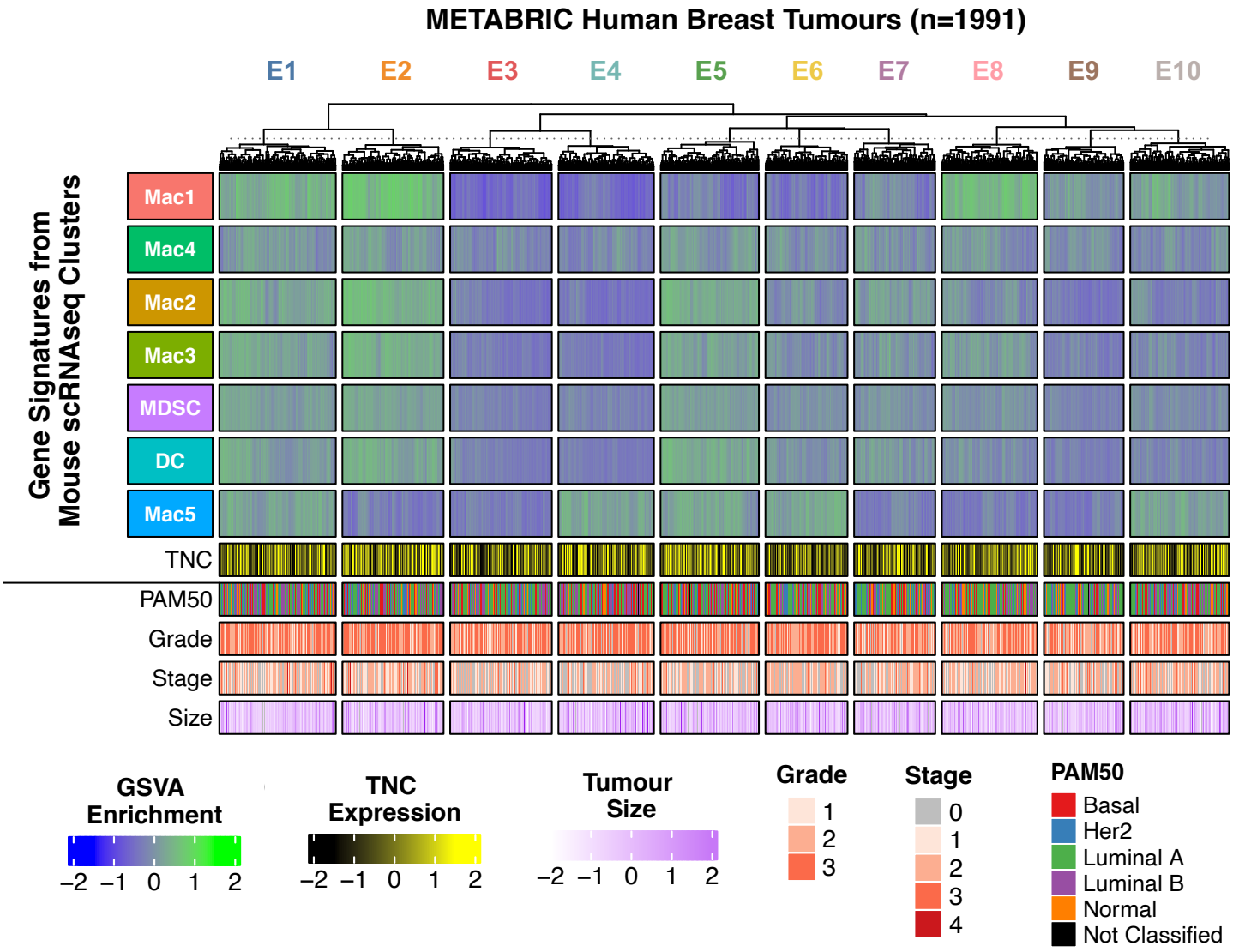


Figure 5.11 | Hierarchical clustering of GSVA enrichment scores to derive myeloid ecotypes in METABRIC data. A total of 1991 bulk microarray human breast tumour samples from METABRIC (290) were analysed with the R Package, SigQC (293). GSVA enrichment scores estimate the variation of pathway activity across samples, and GSVA scores for all myeloid signatures (rows) were used to hierarchically cluster bulk microarray breast tumour samples (columns) from METABRIC in a heatmap. The dendrogram was partitioned by k-means clustering with 1000 iterations to ensure the robustness and stability of each cluster, representing an ecotype of myeloid signatures. A total of 10 ecotypes (E1-10) were defined. Annotation of clinical metadata and TNC expression are indicated for each tumour sample (columns), including the PAM50 molecular subtype, tumour size, tumour grade and disease stage. Acknowledgment: Clinical metadata was cleaned and amalgamated by Yuechan Chi, and the dataset was provided by the group of Professor Francesca Buffa, both from the Department of Oncology (University of Oxford). All analysis presented above was performed alone with no assistance. Abbreviations: GSVA = Gene Set Variation Analysis, METABRIC = Molecular Taxonomy of Breast Cancer International Consortium, PAM50 = Prediction. Analysis of Microarray 50, TNC= Tenascin-C.

The ecotypes of E1, E2 and E5 all universally co-enriched the same four core myeloid signatures (**Mac2**, **Mac3**, **DC**, **MDSC**), while several other ecotypes (E3, E4, E9, E10) downregulated these same signatures (**Figure 5.11**). Interestingly, a similar segregation of ecotypes according to these same 4 myeloid signatures was also observed in the TCGA dataset for five ecotypes (E1-5) (**Figure 5.7**). This similarity between datasets, which both differentially enrich the same core signatures, serves to cross-validate the findings between these studies.

Further reinforcing these similarities, the ecotypes E4, E8 and E10 enriched signatures for **Mac1** alone, **Mac5** alone**,** and both combined, respectively (**Figure 5.11**), mimicking the same findings as observed in TCGA (**Figure 5.7**). Notably, in the METABRIC dataset, **Mac1** and **Mac5** were the only signatures that created a 'single-positive' ecotype (an ecotype which enriched only one myeloid signature), whereas the TCGA dataset additionally included a **Mac4** single-positive ecotype. However, the absence of a **Mac4**-specific ecotype in the METABRIC dataset can be explained by the substantially lowered quality of the **Mac4** signature in this dataset, which showed a sporadic enrichment across all clusters and did not segregate into any particular ecotype (**Figure 5.11**).

Expression of TNC was relatively evenly distributed across all ecotypes, with modest upregulation observed in E3 and downregulation observed in E4 and E5 (**Figure 5.12A**). The distribution of clinical features was also assessed across each ecotype, such as for the PAM50 molecular subtype (**Figure 5.12B**), tumour grade (**Figure 5.12C**) and disease stage (**Figure 5.12D**). PAM50 subtypes were highly variably distributed across ecotypes, though each PAM50 subtype was represented within each ecotype. The ecotype E6 displayed a particularly high percentage of Basal and HER2-enriched tumours, together comprising 68% of this ecotype, while the ecotypes E1 and E7 had similarly high percentages of 45% and 47%, respectively (**Figure 5.12B**). These same three ecotypes (E1, E6, E7) also contained the largest proportion of high-grade tumours (grade 3) (**Figure 5.12C**), indicating a potential

association of these ecotypes with tumour progression. In contrast, ecotypes E9 and E10 were predominantly composed of Luminal A and normal-like subtypes, which corresponded to the lowest proportion of high-grade tumours (**Figure 5.12B-C**). However, such ecotype associations are likely to confounded by differences in PAM50 distribution, and must therefore be interpreted with caution. Finally, the disease stage of patients was relatively evenly distributed across all ecotypes, with only minor differences observed, such as a moderately lower proportion of Stage 4 patients in E4 and a higher proportion in E6 (**Figure 5.12D**).

Taken together, these data demonstrate a successful segregation of myeloid ecotypes within the METABRIC dataset, and indicate a high degree of concordance with TCGA ecotypes, mutually reinforcing the findings from both analyses and providing confidence in their validity and reproducibility.

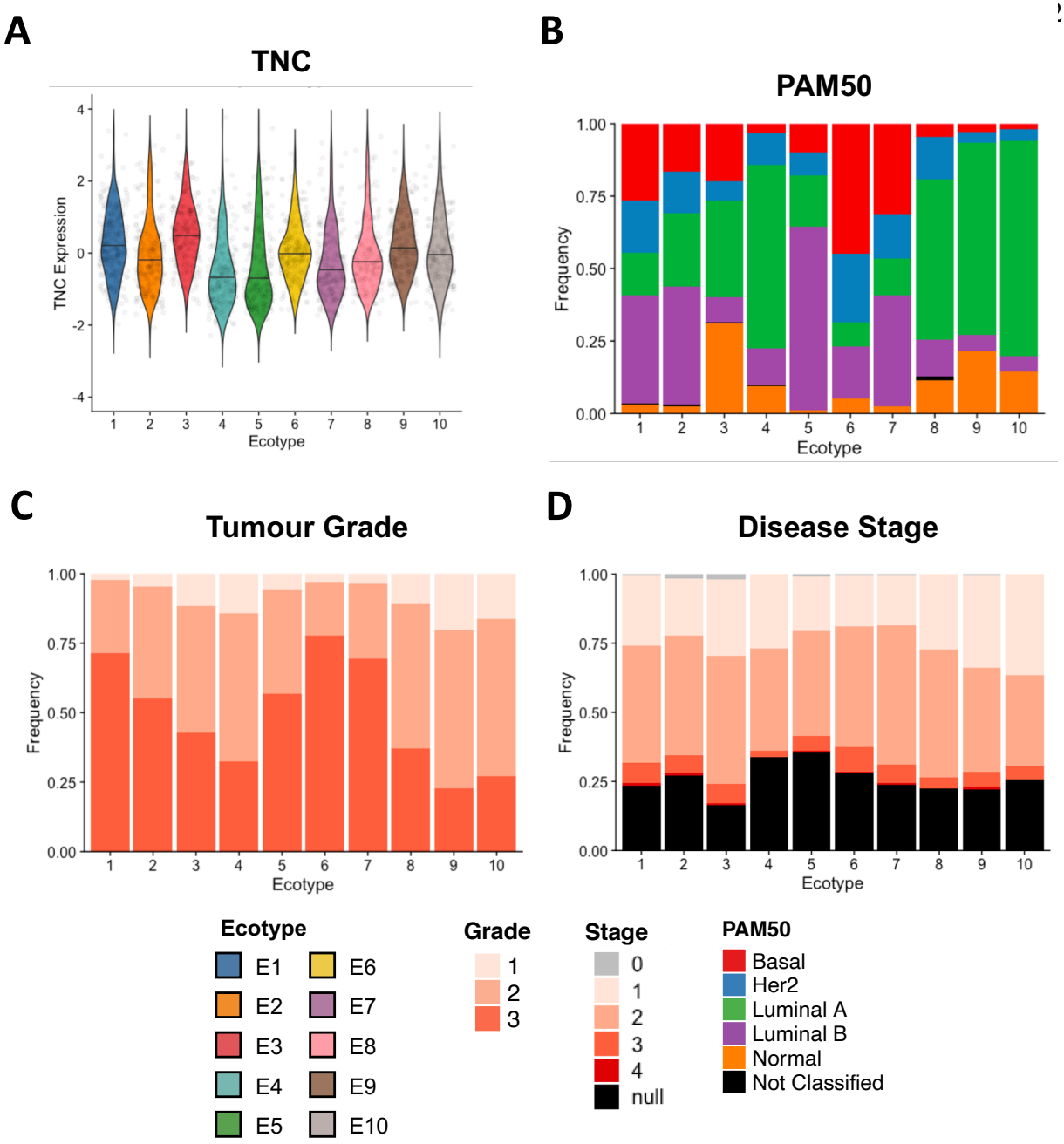


Figure 5.12 | Distribution of TNC expression and clinical metadata per ecotype in the METABRIC dataset. A total of 1991 bulk microarray human breast tumour samples from METABRIC (290) were analysed with the R Package, SigQC (293). Myeloid ecotypes were defined by hierarchical clustering of GSVA enrichment scores for each mouse-derived myeloid signature. **(A)** Violin plots showing the distribution of TNC expression per ecotype. Expression values were scaled, and the median expression value per violin is represented by a horizontal line. Individual tumours are represented by translucent grey dots. **(B-D)** Stacked barplots showing the proportional composition of each ecotype in terms of its clinical metadata, such as the PAM50 molecular subtype **(B)**, tumour grade **(C)**, and disease stage **(D)**. Acknowledgment: Clinical metadata was cleaned and amalgamated by Yuechan Chi, and the dataset was provided by the group of Professor Francesca Buffa, both from the Department of Oncology (University of Oxford). All analysis presented above was performed alone with no assistance. Abbreviations: GSVA = Gene Set Variation Analysis, METABRIC = Molecular Taxonomy of Breast Cancer International Consortium, PAM50 = Prediction. Analysis of Microarray 50, TNC= Tenascin-C.

5.1.9 | Survival analysis of patient tumour ecotypes in METABRIC data

The METABRIC dataset offers a greatly increased sample size and duration of patient follow-up which together improve the statistical power of downstream survival analyses. Using the 10 myeloid ecotypes previously defined by hierarchical clustering (**Figure 5.11**), the survival effect of each ecotype was assessed by Kaplan-Meier analysis (**Figure 5.13A**). Performance of a log-rank test revealed statistically significant differences in survival across all ecotypes (P=0.016), indicating that myeloid cell infiltration can stratify patients prognostically. The ecotypes E4, E8 and E10 enriched **Mac1** alone, **Mac5** alone, and both combined, respectively. Each of these ecotypes harboured a very similar prognosis, falling in the middle of the survival distribution (**Figure 5.13A**), thus indicating that the sole presence of these myeloid subsets in isolation or combination confers an 'average' prognosis.

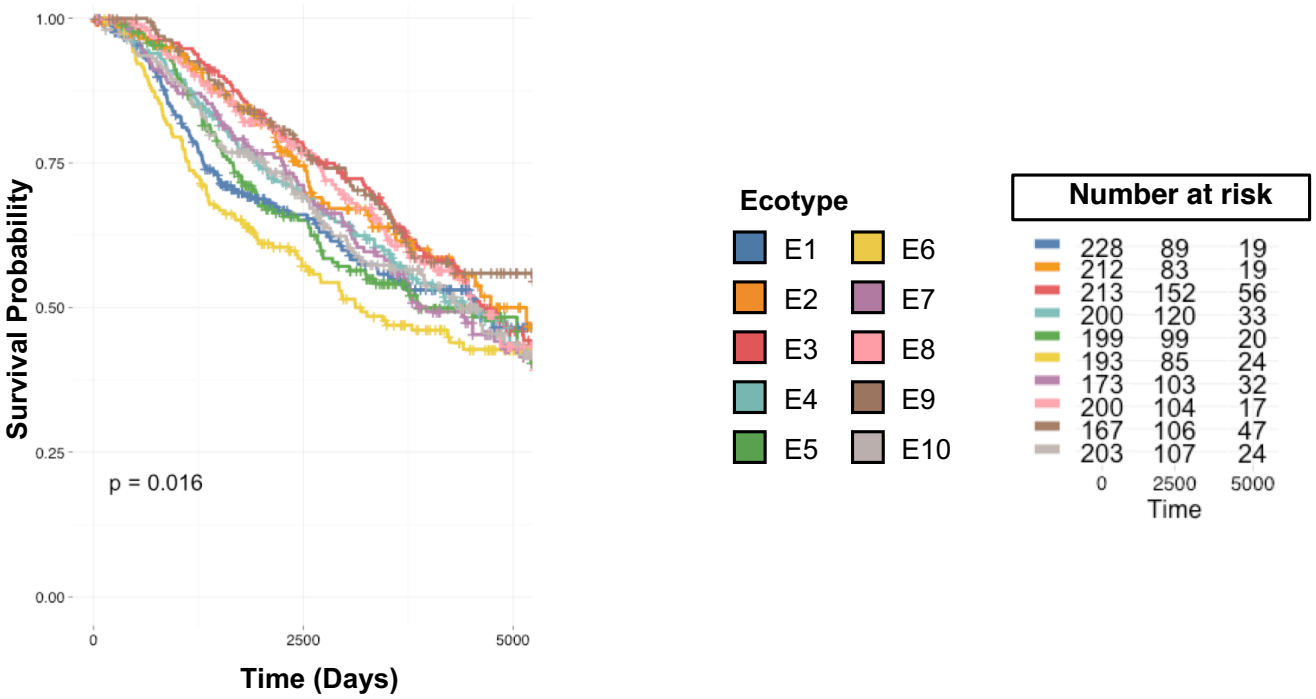
Ecotypes with seemingly differential prognoses, either favourable (E3, E9) or unfavourable (E1, E6), were selected for further assessment by pairwise Kaplan-Meier analysis (**Figure 5.13B**). This revealed highly significant differences in survival between E6 versus E9 (P=0.0019) and between E3 versus E6 (P=0.0031), while more modest differences were observed for E1 versus E9 (P=0.043) and E1 versus E3 (P=0.088). Interestingly, the two most prognostically favourable ecotypes (E3, E9) had in common a low enrichment of **Mac5** and the set of 4 signatures: **Mac2**, **Mac3**, **DC**, and **MDSC**, implying a pathogenic role for **Mac5** and these co-enriched subsets. In contrast, the two most prognostically unfavourable ecotypes (E1, E6) displayed the exact inverse – a high enrichment of **Mac5** as well as the same group of 4 co-enriched signatures alongside (**Figure 5.11**). However, these 4 signatures alone were not sufficient to confer poor prognosis, as the ecotype E2, which enriches all of them, still held a reasonably favourable prognosis (**Figure 5.13A**).

Taken together, these data indicate that survival of breast cancer patients is influenced by context-dependent myeloid interactions in the tumour microenvironment, and demonstrates

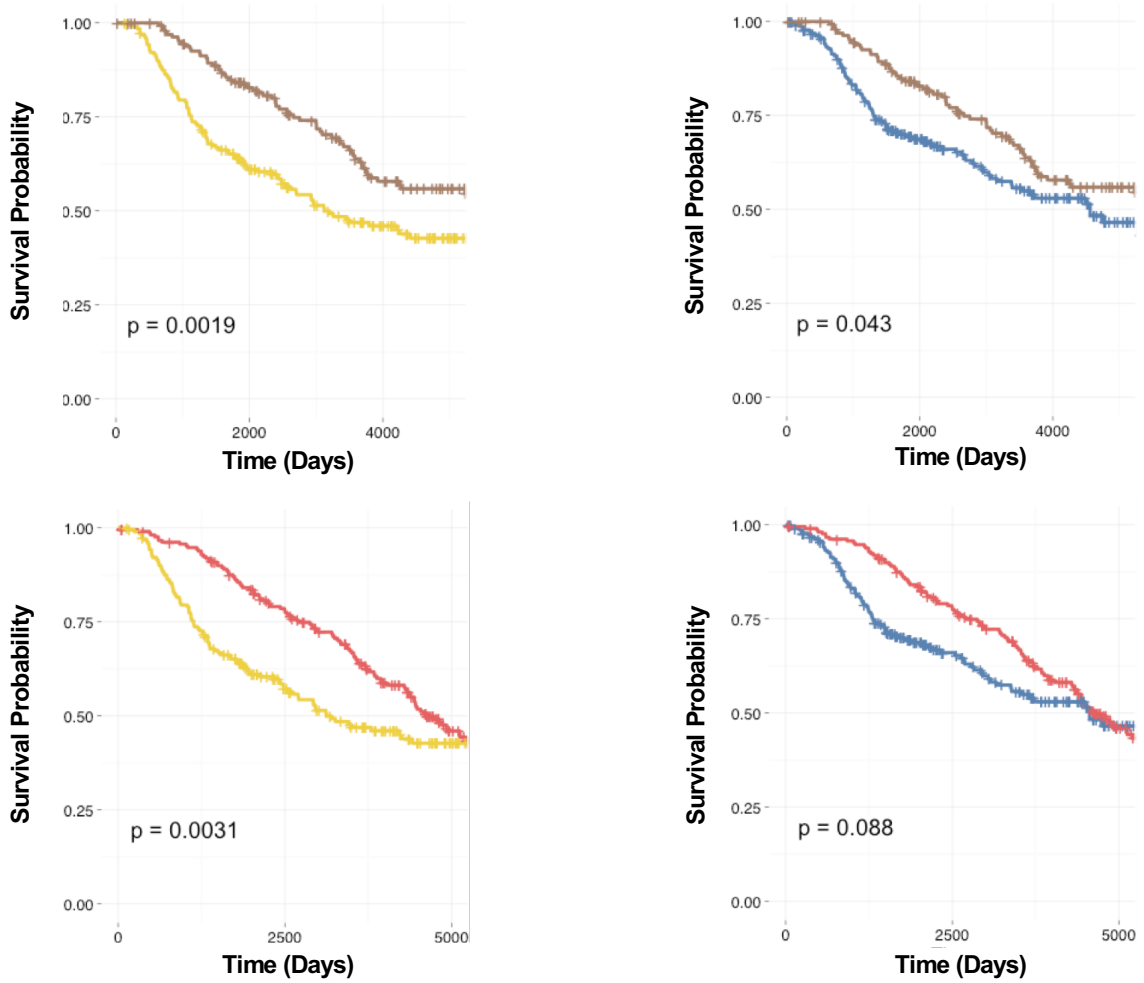
through ecotype analysis that pan-myeloid depletion is positively prognostic, while additionally proposing a pathogenic role for proliferating **Mac5** in tumour progression.

Figure 5.13 (next page) | Survival analysis of myeloid ecotypes in the METABRIC dataset. A total of 1991 bulk microarray human breast tumour samples from METABRIC (290) were analysed with the R Package, SigQC (293). Myeloid ecotypes were defined by consensus hierarchical clustering of GSVA enrichment scores for each mouse-derived myeloid signature. **(A)** Kaplan-Meier survival curves of all patient groups (n=10) harbouring distinct myeloid ecotype profiles (left), as defined by hierarchical clustering. The number of patients at risk at each major timepoint is indicated (right), measured in days. **(B)** Kaplan-Meier survival curves of selected tumour ecotype profiles showing highly differential prognoses. For all plots, P values were calculated by performance of a Log-rank test, and values <0.05 were considered statistically significant. Acknowledgment: Clinical metadata was cleaned and amalgamated by Yuechan Chi, and the dataset was provided by the group of Professor Francesca Buffa, both from the Department of Oncology (University of Oxford). All analysis presented above was performed alone with no assistance. Abbreviations: GSVA = Gene Set Variation Analysis.

A


Survival Probability
1.00
0.75
0.50
0.25
0.00
p = 0.016
0
2500
5000
Time (Days)
Ecotype
E1
E2
E3
E4
E5
E6
E7
E8
E9
E10
Number at risk
228 89 19
212 83 19
213 152 56
200 120 33
199 99 20
193 85 24
173 103 32
200 104 17
167 106 47
203 107 24
0
2500
5000
Time

B


Survival Probability
1.00
0.75
0.50
0.25
0.00
p = 0.0019
0
2000
4000
Time (Days)
p = 0.043
p = 0.0031
2500
5000
p = 0.088

5.2 | Technical Discussion

5.2.1 | Introduction

This chapter harnesses a wide range of published single-cell and bulk transcriptomic atlases of the breast tumour microenvironment, all of which have limitations and advantages. Likewise, the analytical procedures used to derive insights from these datasets have a number of caveats and assumptions that are worth noting. These issues will be considered chronologically, and this section will be divided into the following subsections:

5.2.2 Conversion of mouse genes to human homologs

5.2.3 Conservation of mouse myeloid subsets in human single-cell RNAseq data

5.2.4 Gene signature approaches in bulk transcriptomic datasets

5.2.5 Analysis of TNC in bulk transcriptomic datasets

5.2.6 Interpretation of survival data

5.2.2 | Conversion of mouse genes to human homologs

The process of converting mouse genes to human homologs involves a number of caveats. Firstly, not all mouse genes possess 1-to-1 human homologs, meaning that 'homolog expansion' occurs, whereby conversion of a given mouse gene leads to multiple homologs, or vice versa, whereby multiple mouse genes converge on the same human homolog, causing 'homolog contraction'. This can not only cause the significant loss of homologs from a given signature, as it did for **Mac2** and **DC** (**Figure 5.1B**), which risks reducing the specificity of the signature, but can also cause the appearance of new genes which may not represent the underlying biology. One potential way to mitigate this issue would be to devise a method that enables weighting of signature genes to account for multi-homology. For instance, the

additional MHC Class I genes in **Mac2** that corresponded to only 1 human homolog (*HLA-A*) could hypothetically be weighted higher to account for this, or perhaps be expression-adjusted to increase accuracy.

Ultimately, of course, the greatest caveat of any animal model is its limited homology to humans. While it remains possible that false-negative errors in the MGI homology database may be responsible for some apparent lack of homology, there do appear to be genes which have no direct human homolog at the sequence level, but instead achieve the same functionality through a different gene of potentially distinct evolutionary ancestry (ie an example of convergent evolution). For example, the co-stimulatory gene involved in antigen presentation and T cell activation, *Tnfsf9* (*4-1BBL*), does not have a human homolog according to the MGI database, despite humans possessing a functionally similar molecule named *TNFSF9* (*4-1BBL*). This potentially highlights how sequence homology may not necessarily translate to functional homology at the systems level. Supporting this non-equivalence of homologs at a functional level is the example of the discriminatory surface molecule, F4/80 (*Adgre1*), which specifically marks murine but not human macrophages and monocytes, despite *ADGRE1* being a direct human homolog. A similar non-equivalence occurs for human monocyte markers CD14 (*CD14*) and CD16 (*FCGR3A*/*FCGR3B*) which do not mark the same subsets in mice. Thus, it is worth noting that functional differences may occur, even for closely homologous genes, due to systems-level differences between species.

In summary, gene signatures must be interpreted with the caveat that a small minority of genes may be influenced by homolog expansion and contraction, and with the assumption of equivalent functionality of member genes between species, which is mostly but not always accurate.

5.2.3 | Conservation of mouse myeloid subsets in human single-cell RNAseq data

While some of the most prominent myeloid marker genes in mouse possessed clear conservation in human myeloid subsets from published scRNAseq datasets, this analysis was not unsupervised and therefore introduces biases. To perform a systematic unbiased conservation analysis across human scRNAseq studies of breast cancer, an unsupervised algorithmic integration of datasets could be performed, with a method such as that pioneered by Satija and colleagues, which leverages both canonical correlation analysis (CCA) and mutual nearest neighbour (MNN) approaches to robustly integrate diverse 'omic single-cell data via anchor-based transfer learning (269). This would provide robust orthogonal validation to support the analyses already performed, however was not possible within the given time constraints.

As previously mentioned, substantial differences exist in the size and quality of the Azizi *et al.* and Wu *et al.* scRNAseq atlases of human breast cancer. Consequently, the more recent Wu *et al.* dataset should be preferentially weighted in importance, owing to its larger myeloid cell population, increased sequencing depth, expanded patient numbers and improved representation of disease subtypes such as the HER2-enriched subtype, which was lacking in the Azizi *et al.* dataset.

In general, myeloid subsets from mouse scRNAseq data were highly concordant with the Wu *et al.* dataset (**Figure 5.3**), demonstrating that the myeloid landscape in murine breast cancer is reasonably representative of human disease. While the absence of certain human myeloid subsets in the mouse data indicated the presence of human-specific subsets, *FABP5* Lipid-associated macrophages were in fact mostly observed within ER^+ and triple-negative tumours, but much less so in $HER2^+$ tumours (101), suggesting that the absence of a conserved *FABP5* subset in the mouse data was likely due to the HER2-positive status of the mouse model. Perhaps the largest caveat in the present analysis of the Wu *et al.* dataset is that the **Mac1**-

equivalent cells (*FCGR3A* Monocytes) have a very low relative abundance compared to other myeloid subsets (**Table 5.2**) (101), which does not correspond to **Mac1** in mouse data, where it is the most abundant macrophage subset, comprising ~26% of the myeloid compartment (**Ch3 Appendix Figure 4**). However, the *IL1B* Monocyte subset identified by Wu *et al.*, which was initially designated as '**Mac1**-like' due to its upregulation of ribosomal proteins, may in fact represent the **Mac1** subset due to its much higher relative abundance (**Table 5.2**), though it also modestly expressed **Mac3** markers, either indicating that this mapping is imperfect, or that a spectrum of **Mac1**/**Mac3** phenotype has been identified, which may represent an intermediate state (**Figure 5.3**).

One final discrepancy worth noting was with regard to the **MDSC** subset, which was assigned in mice on the basis of its mixed expression of monocytic (*Ly6c2*) and granulocytic (*Ly6g*, *Csf3r*) genes (**Figure 3.8**) while expressing extremely high quantities of calprotectin (*S100a8*, *S100a9*) (**Figure 3.7A**) which is common to both neutrophils and monocytes, though more highly expressed in neutrophils. However, Wu *et al.* defined their *S100A9*-expressing cluster as a monocyte (101), despite its high expression of *CSF3R* (**Figure 5.3**), indicating that this cell type might be a mis-assigned granulocyte. Indeed, Wu *et al.* did not identify any granulocytic subsets in their data, which was postulated to be due to the sensitivity of neutrophils to tissue dissociation protocols (101), though it remains possible that the **MDSC**-like *S100A8* Monocytes may in fact represent the granulocytes that were presumed to be absent. Confusingly, however, a recent pan-cancer analysis of tumour-infiltrating myeloid cells reported that blood-derived monocytes indeed show higher expression of neutrophil associated genes (*S100A8*, *S100A9*, *CSF3R*) (199), begging the question as to what truly defines each subset, and showing that strong phenotypic overlap between these subsets is conserved across all solid tumours.

5.2.4 | Gene signature approaches in bulk transcriptomic datasets

The enrichment of gene signatures in bulk transcriptomic datasets has several important limitations which must be acknowledged:

The signal-averaging inherent to bulk transcriptomics is the largest and most obvious caveat of this analysis, as the relative contribution of gene expression by each cell subset in the sample is unknown, as is the relative infiltration of each cell subset. Consequently, it is not possible to confirm that signatures are truly derived from the distinct cellular subtypes or states that the single-cell data proposed. Thus, signature enrichments represent only an estimation of the presence of a given cell subset, rather than a robust enumeration of it. To mitigate this issue, recent deconvolution algorithms such as CIBERSORTx could be used, which estimate cell-type-specific expression profiles in bulk samples from single-cell data, using a machine learning method (*v*-support vector regression) (295).

Another caveat of the signature enrichment approach is that the differential expression background of purified myeloid cells is not the same as the background for entire tumours (as represented in bulk data), and therefore the marker genes defining these subsets may not in fact be truly differentially expressed when considering the whole microenvironment. Thus, to improve this analysis, an unsupervised conservation analysis could first be performed between mouse and human datasets to verify that these genes are indeed exclusively expressed within a given subset, before performing CIBERSORTx analysis as stated above. This would additionally resolve the caveat that was addressed by the 'fitness gene deduction' approach, whereby constitutively expressed essential genes in tumour cells may confound myeloid signatures that share these same genes (eg ribosomal protein genes in **Mac1** are also highly expressed by tumour cells).

While the 'fitness gene deduction' approach did successfully mitigate this issue, and improved signature quality (**Figure 5.6**), it may not completely resolve the issue. For instance, it remains possible that the survival disadvantage associated with myeloid ecotypes that enrich **Mac5** (**Figures 5.11** & **5.13**) might in fact be due to the residual presence of proliferation-associated genes in the **Mac5** signature overlapping with highly proliferative tumour cells expressing the same genes (which are likely to bear poor prognosis). Thus, the **Mac5** signature might be partly enriching bulk samples with highly proliferative tumour cells, as well as enriching **Mac5**. Arguing against this, however, is the fact that sole enrichment of **Mac5** in ecotype E4 (**Figure 5.11**) led to a relatively 'average' prognosis compared to the rest of the cohort (**Figure 5.13**), which does not align with the notion of a highly proliferative tumour, and therefore suggests that the signature refinement strategy successfully mitigated this issue.

While it is plausible that the agreement of ecotype profiles between the TCGA and METABRIC datasets (**Figures 5.7** & **5.11**) is due to the identification of orthologous myeloid signatures in each study, this may equally be driven by strong inter-signature correlations, due to an overlap in the pathways each signature enriches by GSVA. For instance, **Mac2** and **DC** are both highly antigen-presenting subsets, and their enrichment of shared biological pathways by GSVA will lead to tumour samples co-enriching both **Mac2** and **DC** even if only one of these subsets is present. This co-enrichment due to shared biological pathways would lead to a strong pairwise correlation in **Mac2** and **DC** enrichment scores, which would be present in both datasets. Thus, an inherent limitation of the GSVA method is that it prevents different gene signatures that enrich similar biological functions from being effectively segregated. Reassuringly, however, the comprehensive ecotype analysis performed by Wu *et al.* using CIBERSORTx demonstrated that both conserved **Mac2** and **DC** subsets (*APOE* LAMs and *LAMP3* DCs, respectively) were clustered within ecotype 4 (101), corroborating the co-enrichment and correlation of the **Mac2** and **DC** signatures observed in analyses of both TCGA (**Figure 5.7**) and METABRIC data (**Figure 5.11**). This either suggests that **Mac2** and **DC** are genuinely co-

enriched, or that the more superior CIBERSORTx method is susceptible to the same biases as the GSVA signature enrichment method.

Finally, it is possible that the poor quality of the **Mac4** signature specifically in the METABRIC dataset (**Figure 5.10**) may represent genuine biology, especially given the lack of apparent conservation of **Mac4** marker genes in human scRNAseq datasets (**Figures 5.2** & **5.3**). However, this may also signify a platform-specific issue with microarray probesets in the METABRIC dataset, which may not be representative of underlying expression patterns of the **Mac4** signature. This does seem unlikely, however, given that this issue did not manifest for any other signatures except **Mac4**.

5.2.5 | Analysis of TNC in bulk transcriptomic datasets

There are several caveats associated with studying TNC in bulk transcriptomic datasets. Our previous investigation of the breast tumour microenvironment in mice demonstrated that the cellular source of TNC is important to disease pathogenesis, with tumour-cell-derived TNC inducing tumour-supportive phenotypes in myeloid cells, in contrast to host-stroma-derived TNC which mediated a tumour-eliminating phenotype (264). Given the importance of the cellular source of TNC, the unknown cellular composition of bulk transcriptomic samples complicates interpretation. This is especially problematic in breast cancer, which has the 6th highest level of non-cancer cell infiltration out of 21 cancer types in the TCGA according to consensus estimations (388), meaning there is an increased probability that TNC is being expressed by the high proportion of non-cancer cells present. Indeed, a variety of cell types produce *TNC*, with CAFs likely to be a major source (241), alongside tumour cells (235), and immune cells (240). Moreover, a large number of alternatively spliced variants of *TNC* exist with markedly different functional roles (242), and only the sum of these is represented by *TNC* transcript quantification in TCGA and METABRIC, rather than quantification of individual

splice variants. Furthermore, for bulk microarray studies such as METABRIC, it remains possible that probe-sets for the *TNC* gene might not have sufficient coverage to detect all potential splice variants. Finally, TNC undergoes substantial post-translational modifications that modulate its function (233, 242), and this is not captured in transcriptomic data, and may be partly responsible for the biological differences previously observed in host- versus stroma-derived TNC (264).

Thus, in summary, *TNC* expression data has numerous caveats which complicate interpretation, and caution is advised when comparing experimental TNC knockdown data (from a controlled mouse model using single-cell technologies such as scRNAseq, flow cytometry and IF), with generalised TNC expression from bulk clinical samples.

5.2.6 | Interpretation of survival data

The ultimate purpose of performing gene signature enrichment in clinical transcriptomic datasets was to assess the survival of patient groups with defined myeloid phenotypes. However, as previously stated, large differences in the sample size and duration of patient 'follow-up' between studies limited the statistical power of survival analyses in the TCGA dataset. As a consequence, the prognostic insights gleaned from the METABRIC dataset should be weighted with higher importance than TCGA.

However, it is important to consider the biases and differences of each dataset. For instance, patient samples were collected by different means in different studies: the use of fresh-frozen breast tumour samples enabled longer patient follow-up in METABRIC data, whereas TCGA used only fresh tumour resections. Furthermore, in the METABIC dataset, no HER2-positive patients received trastuzumab therapy, and nearly all ER-positive and/or LN-negative patients did not receive chemotherapy, whereas ER-negative and LN-positive patients did (290). By

contrast, clinical treatment information was not available for TCGA cohort patients without specific granted access, and thus biases have not been determined for this dataset.

Finally, it is also worth noting that depending on the stage and subtype of breast cancer, adjuvant (post-surgery) treatment options vary substantially for patients. Individual survival times are therefore a function of a patient's response to future treatment, as well as the probability of tumour recurrence and metastasis. What this prognostic information cannot determine is whether a given ecotype is pathologically important in the preceding stages of pathogenesis before surgery, such as tumour initiation and progression.

In any case, one improvement that would aid interpretation of survival data would be to include a Cox Proportional Hazard model (389) to determine the relative hazard of each ecotype, alongside other clinical metadata, such as PAM50 subtypes and tumour grade, which are likely to be confounders of each ecotype. The hazard model could therefore be adjusted to account for these co-variates.

5.3 | Biological Discussion

5.3.1 | Introduction

This chapter demonstrates the conservation of almost all mouse myeloid subsets in human breast cancer, and assesses their translational relevance via gene signature enrichment and survival analysis in clinical bulk transcriptomic datasets. The biological implications of these findings will be discussed in the following subsections:

5.3.2 Conservation of mouse myeloid subsets in human single-cell RNAseq data

5.3.3 Phenotypic expansion of myeloid subsets as a hallmark of pathogenesis

5.3.4 Assessment of **Mac3** regulators: *KDM6B* and *PIM1*

5.3.5 The prognostic impact of myeloid ecotypes

5.3.6 Elucidating a role for TNC

5.3.2 | Conservation of mouse myeloid subsets in human single-cell RNAseq data

The existence of two large scRNAseq atlases of human breast cancer offers an excellent opportunity to assess the conservation and translational relevance of mouse-derived myeloid signatures. While these datasets were limited in their agreement, the most comprehensive and high-quality dataset published by Wu and colleagues (101) demonstrated considerable conservation of virtually all mouse myeloid subsets (**Figure 5.3**). The only exception to this was **Mac4**, which could not be identified in either Azizi *et al.* or Wu *et al.* datasets, reinforcing the idea of this subset as a transitionary, unstable and non-discrete state, as proposed earlier in **Chapter 3**.

The successful conservation of the remaining mouse myeloid subsets confirmed the presence of functionally-delegated macrophage states in human breast cancer, demonstrating the translational relevance of these findings. However, this conservation analysis also offered a range of novel insights regarding the mouse data itself:

Firstly, the two human subsets most closely resembling **Mac1** were both assigned as monocytes (either *FCGR3A* or *IL1B* Monocytes) (101), which corroborates the notion of **Mac1** as a naive-like state forming a 'root' like topology in diffusion map space in **Chapter 3** (**Figures 3.13A-B** & **3.14**). It also reinforces the idea of ribosomal protein upregulation as a conserved state of developmental immaturity within immune cells, and perhaps mechanistically implies that modulation of ribosomal translational is key for achieving the required level of developmental control.

Secondly, the human subset most closely resembling **Mac2** was the *APOE* Lipid-associated macrophages (LAMs), which spatial transcriptomic analysis demonstrated was present at invasive cancer regions and in areas with high levels of lymphocytes, stromal and adipose cells (101). *APOE* LAMs were also correlated with the infiltration of $CD4^+$ and $CD8^+$ T cells across multiple tumours (representing all three subtypes: ER^+, TNBC and $HER2^+$), and these regions were additionally enriched for expression of PD-1 (*PDCD1*), PD-L1 (*CD274*) and PD-L2 (*PDCD1LG2*) (101), suggesting a role for *APOE* LAMs in regulating anti-tumour immunity and perhaps the therapeutic response to immune checkpoint blockade.

Another subset that positively correlated with CD8 T cell infiltration across many tumours and was enriched in invasive regions was the *CXCL10* Macrophage, which bears hallmarks of both **Mac2** and **Mac3** markers (**Figure 5.3**), and perhaps represent an intermediate state between these phenotypes.

In summary, this conservation analysis successfully identified functionally-delegated macrophage states in human breast tumours, and additionally suggested that **Mac1** are monocytic, and that **Mac2** are LAMs situated at the invasive front which correlate with exhausted/inhibited T cells.

5.3.3 | Phenotypic expansion of myeloid subsets as a hallmark of pathogenesis

One of the unique advantages of the Azizi *et al.* dataset is its analysis of matched blood, lymph node and normal mammary tissue per patient (102), providing an important comparator that enables the distinction of tumour-specific phenotypes. Indeed, Azizi *et al.* described a 'phenotypic expansion' of myeloid cells within tumours compared to normal tissue. Tumour-infiltrating monocytes displayed the highest 'phenotypic volume' by orders of magnitude, demonstrating a unique susceptibility of monocytes to the impact of their surrounding microenvironment (102). Further analysis of the Azizi *et al.* dataset in **Figure 5.2** showed that the 'monocyte precursor', cluster 27, which exhibited features of both **Mac1** and **Mac2**, was highly abundant within matched normal breast tissue as well as tumour tissue. This raises the possibility that a subset resembling naive-like **Mac1** and antigen-presentation enriched **Mac2** are present in steady-state healthy mammary tissue, but that the impact of the tumour microenvironment induced tumour-specific functionally-delegated phenotypes (though the *et al.* Azizi dataset does not have the depth or coverage to adequately detect them).

Supporting this possibility, a recent pan-cancer scRNAseq analysis of myeloid cells indicated the conservation of a tumour-infiltrating monocyte subset, expressing *CXCL2* and *CCL4* (both **Mac3**-enriched chemokines), which was specific to tumours but absent in normal tissue (199). While this suggests **Mac3** may be monocytic, the fact that these chemokines are specifically expressed in such a wide range of solid tumours suggests that the **Mac3** state is indeed a conserved tumour-induced state, perhaps therefore offering an opportunity to therapeutically

target a unique tumour-specific phenotype while sparing healthy tissue macrophages. Finally, the fact that pan-myeloid signature enrichment in ecotype E1 of the METABRIC dataset (**Figure 5.11**) is negatively prognostic (**Figure 5.13B**) might possibly suggest that a high degree of phenotypic expansion is responsible for this worsened prognosis.

5.3.4 | Assessment of Mac3 regulators: *KDM6B* and *PIM1*

Expression of the validated regulator of **Mac3** effector function, *KDM6B*, and the putative regulator, *PIM1*, were differentially distributed across the tumour microenvironment in human breast cancer. *KDM6B* expression was almost exclusively restricted to the myeloid compartment, whereas *PIM1* was expressed lowly in myeloid cells but highly and consistently across the T cell compartment (**Figure 5.4**). This not only suggests that the myeloid-modulatory effects of *PIM1* may not be conserved in humans, but also indicates that modulation of the PIM1 axis may encounter large off-target effects in T cells, though it remains unclear whether *Pim1* expression is also upregulated in murine T cells.

Within the macrophage compartment specifically, *KDM6B* was enriched in *EGR1* Macrophages (equivalent to **Mac3**) and *SIGLEC1* Macrophages (which had no equivalent mouse subset). While these data indicate conservation of *KDM6B* as a regulator of **Mac3** effector function in humans, its expression in *SIGLEC1* Macrophages alongside other myeloid subsets suggests that therapeutic targeting of KDM6B would encounter moderate off-target effects in other myeloid cells, though much less so than PIM1. Given the broadly pathogenic role of myeloid cells in breast cancer and most solid tumours (120, 155, 157, 158), minor disruption to other myeloid cells is a likely acceptable off-target effect. This data therefore supports *KDM6B* as a potential target for therapeutic modulation of macrophages in breast cancer. However, to confirm the same dynamics of *KDM6B* expression in regulating the acquisition of **Mac3** effector function in human *EGR1* Macrophages, trajectory inference and

pseudotime analyses should be performed in the Wu *et al.* dataset, aiming to recapitulate the finding that *KDM6B* is expressed immediately preceding the induction of metastasis-promoting chemokines, as observed in mice (**Figure 3.14**).

5.3.5 | The prognostic impact of myeloid ecotypes

Emulating the holistic approach employed by Wu *et al.* to define cellular context in the TME (101), myeloid ecotypes were defined by hierarchical clustering and assessed for their impact on survival. This approach led to the identification of ecotypes with significantly favourable (E3, E9) and unfavourable (E1, E6) prognoses (**Figure 5.13B**), while the composition of these ecotypes differed principally in their co-enrichment of **Mac5**, **Mac2**, **Mac3**, **DC**, and **MDSC**, (**Figure 5.11**), implying a pathogenic role for these subsets. These data also reinforce the notion that this core set of 4 signatures (**Mac2**, **Mac3**, **DC**, **MDSC**) co-exists within breast tumours in a binary manner, suggesting potentially distinct pathogenic mechanisms of pathogenesis for patients with or without co-enrichment of these four signatures. Indeed, it seems that, with the exception of **Mac5** and **Mac1**, the structure and distribution of myeloid signatures is compact and binary across ecotypes, with tumours either enriching or not enriching these 4 signatures in concert. The binary nature of this trend perhaps also suggests that a 'phenotypic expansion' of myeloid cells is in itself pathogenic, as ecotypes enriching multiple myeloid signatures (including functionally-delegated macrophages) confer a poorer prognosis. The finding that pan-myeloid signature enrichment is correlated with worsened survival aligns with the literature (157, 349), and serves to validate this analytical approach. However, the virtual absence of all myeloid cell enrichment in ecotype E3 would be expected to garner the most favourable prognosis, yet E9 in fact held a superior survival advantage, which modestly enriched **Mac1** in comparison to E3 (**Figures 5.11 & 5.13**). This suggests that complete ablation of the myeloid compartment is not necessarily optimal for improved survival, and that a small proportion of naive-like macrophages (**Mac1**) in isolation might be beneficial,

perhaps due to their ability to transition towards an anti-tumoural state. This implies that, while a generally low myeloid cell infiltration is always favourable, **Mac1** may not be intrinsically pathogenic as previously hypothesised in **Chapters 3 & 4**, but that the extent of its pathogenicity is dependent on the surrounding myeloid context. While these data do not align with the hypothesis that **Mac1** is an exclusively pathogenic naive-like macrophage state which fails to acquire effector functions, they do not exclude the possibility that **Mac1**, in the presence of other functionally-delegated states, may still be pathogenic. For instance, in a situation where acquisition of a certain effector function might aid anti-tumour immunity, such as enhanced antigen-presentation (**Mac2**), a failure to acquire such an effector function may be pathogenic; whereas, in contrast, a failure to acquire the **Mac3** effector function of metastatic-chemokine production may in fact be beneficial. Thus, there may be an added layer of contextual complexity to consider regarding **Mac1** pathogenicity, which depends on the systems-level requirements of its niche.

It is important to note that PAM50 subtypes were unevenly distributed across ecotypes, and that every PAM50 subtype was represented within each ecotype (**Figure 5.12B**). This suggests that ecotype analysis stratifies patient groups in a way that is influenced by, but can only be partly explained by, PAM50 subtyping, and that these ecotypes are also potentially relevant biomarkers with applicability across all PAM50 breast cancer subtypes. Furthermore, the asymmetric distribution of PAM50 subtypes across ecotypes demonstrates that certain subtypes promote the formation or maintenance of specific cellular niches, with important implications for prognosis. However, the confounding effect of PAM50 proportion must be accounted for, and this could be resolved either by performing ecotype analyses for each individual PAM50 subtype, or by performing Cox regression models (389) and adjusting for PAM50 as a co-variate.

To gain a more complete understanding of cellular diversity in the tumour microenvironment, and its impact on patient survival, assessment of broader ecotypes containing additional

immune and stromal subsets is required. Indeed, Wu *et al.* performed a comprehensive ecotype analysis using CIBERSORTx that included all detected subsets from the whole microenvironment, and performed subsequent survival analysis in the METABRIC dataset (101). Interestingly, this analysis showed that the 'Wu ecotype' with the most favourable prognosis (ecotype 2) contained few monocyte or macrophage subsets, enriching only *IL1B* Monocytes (analogous to **Mac1**) and *SIGLEC1* macrophages (which were not conserved in mouse). The other **Mac1**-analogus subset, *FCGR3A* Monocytes, was associated with the second-most favourable prognosis (Wu ecotype 9) (101). Thus, the presence of **Mac1** yet again within human ecotypes that harbour the most favourable prognoses reinforces the notion that this subset is not intrinsically pathogenic, despite previous mouse data proposing it was, due to the expansion of **Mac1** in TNC-high tumours (**Figure 3.11C**). Furthermore, the presence of *CLEC9A* DCs in this Wu ecotype suggests that this subset may contribute to the successful activation of the T cell compartment and the induction of anti-tumoural immunity. In contrast, the ecotype with the most unfavourable prognosis was Wu ecotype 7, which was primarily defined by strong enrichment of $HER2^+$ tumour cells, alongside enrichment of *S100A9* Monocytes (which closely resemble **MDSC** and are likely mis-assigned as monocytes). This confirms the pathogenic role of **MDSC** alongside its most highly expressed genes – the TLR4 ligands, *S100a8* and *S100a9* – in worsening patient survival. Finally, the depletion of the *FABP5* LAM subset specifically within $HER2^+$ tumours, and the fact that high enrichment of *FABP5* LAMs was positively prognostic (101) suggests that the HER2-positive tumour microenvironment somehow regulates LAM fate. Indeed, the generally worsened prognosis of (untreated) $HER2^+$ patients may in part be due to the absence of this subset, reinforcing the pathological importance of macrophage fate determination in influencing tumour progression and prognosis.

5.3.6 | Elucidating a role for TNC

The technical issues associated with quantifying and interpreting TNC expression in bulk transcriptomic data (discussed in **Section 5.2.5**) limit what can be directly gleaned from these human datasets. However, the known effects of exposure to tumour-derived TNC on mouse myeloid subsets allows inferences to be made about their conserved counterparts in human breast cancer. The four mouse myeloid subsets found to be TNC-dependently regulated at the level of cellular abundance were: **Mac1** (*FCGR3A*:Monocytes, *IL1B*:Monocytes), **Mac2** (*APOE*:LAMs), **DC** (*LAMP3*:DC, *CLEC9A*:DC, *IRF7*:pDC), and **MDSC** (*S100A9*:Monocytes).

Firstly, the hypothesis put forward based on mouse data that **Mac1** represents a pathogenic, naive-like arrested state induced by TNC, is likely an incomplete picture. Analysis of METABRIC data by Wu *et al.* indicated that **Mac1**-analogous cells (*IL1B* Monocytes and *FCGR3A* Monocytes) were associated with ecotypes bearing the most favourable prognoses (Wu ecotypes 2 and 9) (101), suggesting **Mac1** are not intrinsically pathogenic, and may in fact be positively prognostic depending on the context. Thus, the role of TNC in **Mac1**-analogous cells, if any, is unclear from this.

Secondly, myeloid cells from TNC-low (but not TNC-high) mouse tumours upregulated several genes including *Apoe*, *B2m* and MHC Class I (**Figure 3.12**), all of which are **Mac2** markers. Given the high expression of these genes in the **Mac2**-equivalent human subset, *APOE* LAMs, it is possible that variable levels of TNC in patients may modulate these *APOE* LAMs cells at the direct transcriptional level, as well as at the level of cellular abundance.

Thirdly, in terms of **MDSC**, the ecotype with the most unfavourable prognosis was Wu ecotype 7, which was primarily defined by strong enrichment of $HER2^+$ tumour cells, alongside enrichment of *S100A9* Monocytes (101) (which closely resemble **MDSC**). This confirms the pathogenic role of **MDSC** alongside its most highly expressed genes – the fellow TLR4

ligands, *S100a8* and *S100a9* – in worsening patient survival. Thus, it is possible that Wu ecotype 7 represents a TNC-dependently regulated niche of cells, and that their poor prognosis may be due to the additive stimulation with alternative TLR4 ligands.

Fourthly, DCs are more challenging to interpret, as the singular mouse **DC** subset is divided across three distinct human DC subsets. However, *CLEC9A* DCs may be a strong candidate for being TNC-dependent, given that they are present in a group with favourable prognosis (Wu ecotype 2) and relatively downregulated in those with poor prognosis (Wu ecotype 7) (101), which aligns with mouse data showing ~2-fold reduction of DC in the presence of TNC (**Figure 3.11C**).

Finally, the Wu *et al.* dataset (101) offers the opportunity to identify the exact source(s) of stromal TNC in human breast tumours. This analysis confirmed that *TNC* is highly expressed in four main subsets: CAFs (subset 5, myofibroblast-like), Cycling perivascular-like cells (PVLs), Immature PVLs (subset 1), and Myoepithelial cells (which are usually only found in glandular tissues). Moderate *TNC* expression was also observed in all remaining CAF subsets (subset 1-4) (101). To gain mechanistic insights into the role of TNC in human breast tumours, further bioinformatic analyses could be performed on these subsets. For instance, a ligand-receptor analysis could help delineate the potential interactions between these *TNC*-producing subsets and key TNC-dependently regulated myeloid subsets outlined above. Complementing this, *TNC* expression could be visualised in the spatial transcriptomic dataset to assess its proximity to myeloid subsets and other cellular niches, aiding understanding of its role in human disease.

Chapter 6 | General Discussion

6.1 | Introduction

This investigation into the role of myeloid cells and TNC in the breast tumour microenvironment provides novel insights into the features and mechanisms of pathogenesis, while corroborating key literature in the field. In this chapter, these findings will be placed into context with the proposal of a unified model integrating the heterogeneity of myeloid subsets with the multi-faceted impact of TNC.

This chapter will be divided into the following subsections:

6.2 The myeloid landscape in breast cancer

6.3 The role of TNC

6.4 Prognostic and therapeutic implications

6.5 Future directions

6.2 | The myeloid landscape in breast cancer

One of the principal findings of this work is that macrophages delegate key effector functions to discrete subsets rather than performing both functions simultaneously. This delegation has important implications for breast cancer pathogenesis, as distinct subsets ultimately drive either anti-tumoural (**Mac2**) or pro-metastatic (**Mac3**) functions. Considering only the myeloid subsets which were conserved in the most comprehensive scRNAseq atlas of human breast cancer (**Figure 5.3**) (101), **Figure 6.1** summarises the key characteristics of functionally-delegated macrophage subsets, and the mechanism by which TNC influences myeloid cells to promote tumour progression and metastasis.

To summarise this working model, **Mac1** is a naive-like subset that upregulates the ribosomal translation apparatus and is poised to acquire effector functions. Human scRNAseq data assigned **Mac1**-analogous cells as monocytic, which aligns with their immature phenotype, though whether they represent a 'late monocyte' or 'early macrophage' is unclear. In any case, **Mac1** is capable of transitioning to states of enhanced antigen-presentation (**Mac2**), high chemokine production (**Mac3**), and proliferation (**Mac5**). **Mac2** is the only macrophage subset with anti-tumoural properties, owing to its upregulation of MHC genes and antigen processing machinery, which enable the presentation of tumour antigens to $CD8^+$ T cells *in situ*, with the potential to stimulate a cytotoxic response that eliminates tumour cells. By contrast, **Mac3** is responsible for the majority of chemokine and cytokine production among macrophages, including the production of chemokines CXCL1, CXCL2, CCL3 and CCL4, that are not only known to directly promote EMT and metastasis of breast cancer cells (251, 307-311), but also to recruit **MDSC**, which have widespread tumour-enabling roles (390-392). For example, **MDSC** express high quantities of the TLR4 ligand, calprotectin (S100A8/S100A9) which contributes to breast tumour progression, metastasis and chemotherapy resistance through a

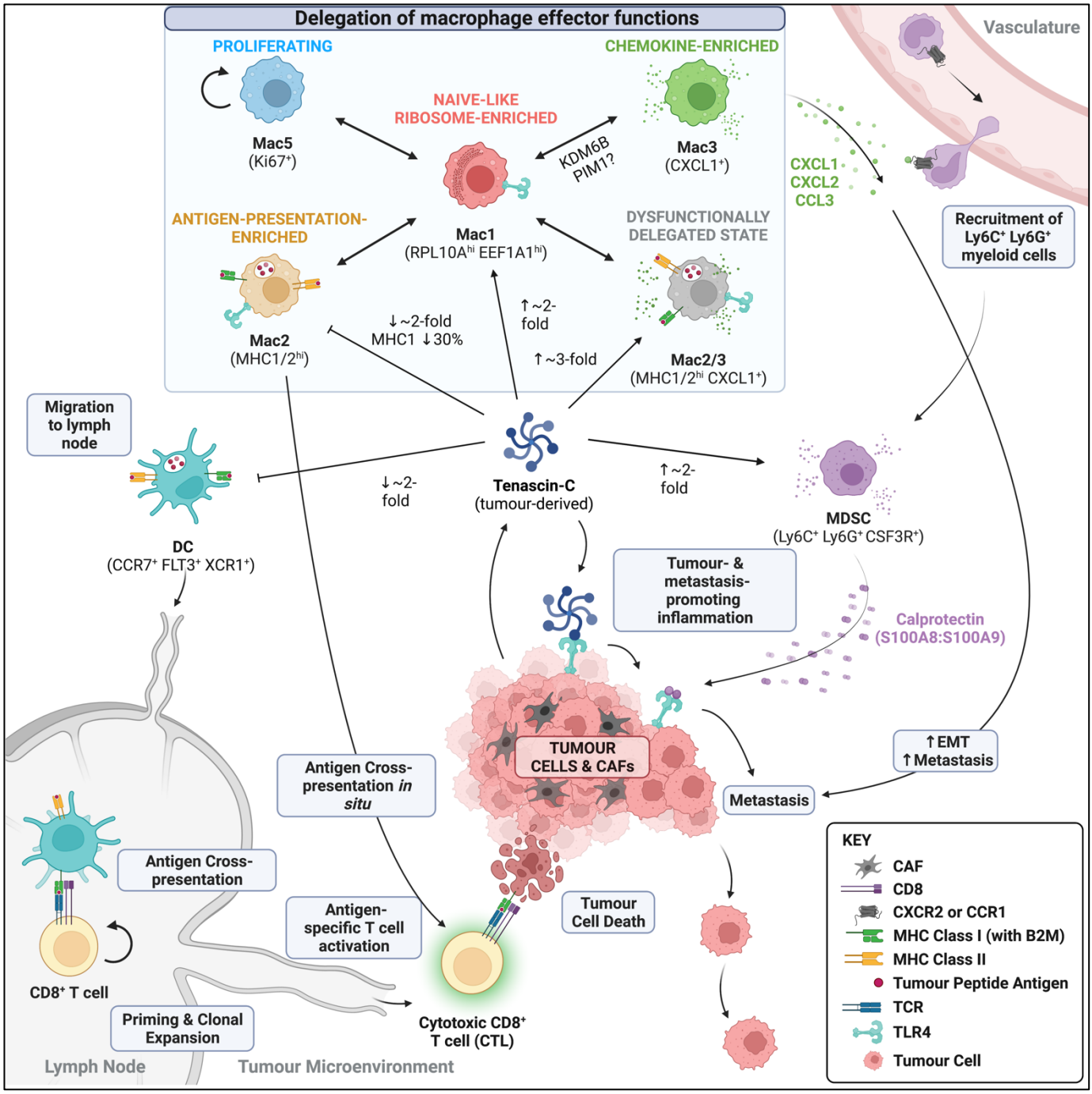


Figure 6.1 | Proposed mechanism of myeloid-TNC interactions in breast cancer pathogenesis. Background: Single-cell RNAseq of intra-tumoural myeloid cells in mouse NT193 breast tumours expressing either high or low levels of tumour-derived TNC led to the identification of five macrophage subsets (**Mac1-5**) with delegated effector functions, one **DC** subset and one **MDSC** subset. All myeloid subsets that were conserved in the most comprehensive scRNAseq atlas of human breast cancer (Wu ref) are displayed, thus excluding **Mac4** – a transitionary, minor and non-conserved state. Macrophages are numbered in order of size, with **Mac1** being the most abundant subset. Explanation: Breast cancer cells and CAFs secrete large quantities of the extracellular matrix protein and TLR4 ligand, TNC, which exerts its impact on the myeloid compartment in numerous ways. Firstly, tumour-derived TNC mediates a concerted ~2-fold depletion of both highly-antigen-presenting myeloid subsets, **Mac2** and **DC**, while additionally reducing the transcription of MHC1 in **Mac2** by ~30%. (*continued on next page*).

Figure 6.1 (continued). Together, this reduces the ability of **Mac2** and **DC** to cross-present tumour peptide-antigens (acquired by phagocytosis) to $CD8^+$ T cells, thereby limiting the ability of the myeloid compartment to efficiently activate adaptive anti-tumoural immunity via cytotoxic $CD8^+$ T cells. The TNC-dependent depletion of **Mac2** is caused by a failure of the naive-like **Mac1** subset to acquire **Mac2** effector function, leading to a reciprocal ~2-fold expansion of **Mac1** in TNC-high tumours. **Figure 6.1 (continued).** The two final macrophage subsets shown are proliferating **Mac5**, and chemokine-enriched **Mac3**, neither of which are directly regulated by TNC. **Mac5** replenishes the macrophage compartment, while **Mac3** produces large quantities of chemokines with known roles in direct promotion of breast cancer cell metastasis, as well as recruitment of $Ly6C^+$ $Ly6G^+$ myeloid cells from peripheral blood, which resemble **MDSC** in the tissue. This **MDSC** subset exhibits phenotypic features of both granulocytic (CSF3R, Ly6G) and monocytic (Ly6C) subtypes alongside extremely high expression of the TLR4 ligand, calprotectin (S100A8:S100A9), which is expressed by both monocytes and granulocytes (though higher in granulocytes). **MDSC** is expanded ~2-fold in TNC-high tumours, though the precise mechanism is unclear, and this resultant increase in calprotectin leads to enhanced TLR4 stimulation, aiding cancer cell survival and promoting metastasis. This TLR4 stimulation likely synergises with the paracrine/autocrine TLR4 stimulation of cancer cells mediated by tumour-derived TNC, resulting in enhanced tumour progression and metastasis. Acquisition and retention of the **Mac3** effector function appears to be regulated by the epigenetic eraser enzyme, KDM6B, though the mechanism in tumours appears to be the inverse to *in vitro* macrophage models. Finally, **Mac2/3** is a low abundance population identified by flow cytometry alone, and resembles a dysfunctionally-delegated state (featuring both **Mac2** and **Mac3** properties). **Mac2/3** is not considered a *bona fide* subset (identified by transcriptome-wide scRNAseq clustering) but is included here due to its ~3-fold expansion in TNC-high tumours, suggesting that TNC may induce this dysfunctional state. Caveats: A number of omissions have been made for the purposes of clarity. Co-stimulatory molecules such as CD28 and CD80/CD86 (B7) which are required for successful antigen-specific T cell activation have been excluded for simplicity, alongside cytotoxic molecules such as perforin and granzyme B). In the conservation analysis of human scRNAseq atlases, **MDSC** was not defined by Ly6C or Ly6G (as these are murine myeloid markers), but instead by calprotectin (S100A8/S100A9) and CSF3R. In addition, the mouse chemokines indicated as CXCL1 and CXCL2, are homologous to the human genes CXCL2 and CXCL3, respectively. All cellular abundance estimates relate to fully developed tumours at Day 18-21 post-engraftment. Finally, the effects of TNC on myeloid cells in this model represents the 'end-state' mediated by TNC, whereas the actual role of TNC likely lies in altering the transitions between these myeloid states. For instance, the ~2-fold depletion of **Mac2** is likely not a direct effect on **Mac2**, but on the transition between **Mac1** and **Mac2**. Citation: This illustration was generated using Biorender. Abbreviations: B2M = β-2-Microglobulin, CAF = Cancer-associated Fibroblast, CD8 = Cluster of Differentiation 8, CCR1 = C-C Motif Chemokine Receptor 1, CXCR2 = C-X-C Motif Chemokine Receptor 2, DC = Dendritic Cell, EMT = Epithelial-to-Mesenchymal Transition, Mac = Macrophage, MDSC = Myeloid-derived Suppressor Cell, TCR = T Cell Receptor, TNC = Tenascin-C, TLR4 = Toll-like Receptor 4.

paracrine signalling loop in which **MDSC**-derived calprotectin stimulates tumour cells to express CXCL1 and CXCL2, which in turn, recruit further **MDSC** (307, 393). Thus, **Mac3** is a key co-ordinator of breast tumour pathogenesis through expression of a pro-metastatic and pro-tumoural milieu of chemokines and cytokines. **DC** is the only remaining myeloid subset with an anti-tumoural phenotype, showing enrichment of antigen-presentation machinery and expression of CCR7, enabling its migration to the lymph node where it can prime and clonally expand tumour-specific $CD8^+$ T cells. Finally, proliferating **Mac5** is responsible for replenishing the intra-tumoural macrophage population, and strong enrichment of the **Mac5** gene signature was a common feature among breast cancer patients with myeloid ecotypes bearing the worst prognosis in the METABRIC dataset (E1 & E6) but not the best prognosis (E3 & E9) (**Figure 5.11 & 5.13**), suggesting its pathological importance when combined with the enrichment of other myeloid subsets.

The nature of functionally-delegated macrophage subsets is explored further in **Figure 6.2**, illustrating the unique spatial and temporal dynamics of the key effector functions: antigen presentation (**Mac2**) and chemokine/cytokine production (**Mac3**). Temporally, the abundance of **Mac3** remains unchanged over time, whereas **Mac2** increases in abundance. Finally, macrophages with neither **Mac2** nor **Mac3** characteristics (ie signifying all remaining macrophage subsets) decrease in abundance over time to the same extent that **Mac2** is increased, suggesting a reciprocal transition away from non-delegated subsets towards **Mac2** effector function acquisition. Spatially, **Mac2** are uniformly distributed throughout mouse breast tumours, whereas **Mac3** are excluded from the tumour core and restricted to the periphery. While this could possibly signify a perivascular association of **Mac3**, the lower availability of oxygen, nutrients and physiological pH in the tumour core compared to the periphery might also signify that macrophage effector functions have a differential tolerance for their surrounding physiological conditions.

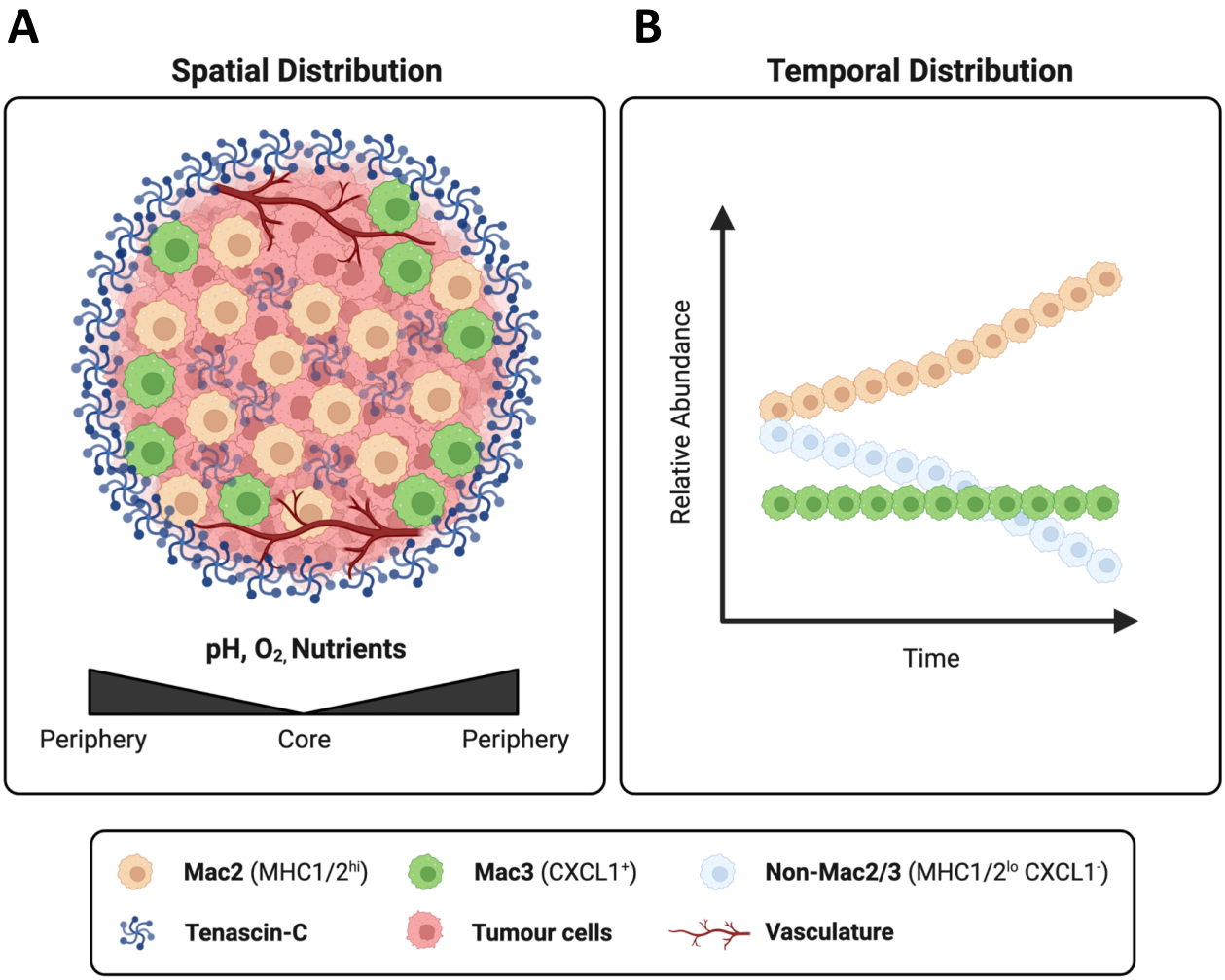


Figure 6.2 | Spatio-temporal differences in the delegation of macrophage effector functions. Intra-tumoural macrophage subsets with specialised effector functions for enhanced antigen presentation (**Mac2**) and chemokine production (**Mac3**) exhibit distinct spatial and temporal dynamics, and this occurs independently of TNC status. **(A)** Spatially, **Mac2** are uniformly distributed throughout tumours, whereas **Mac3** are largely excluded from the tumour core and restricted to the periphery, possibly signifying a perivascular-association. As the availability of oxygen, nutrients and physiological pH are likely to be much lower in the tumour core compared to the periphery, this invites speculation that these delegated macrophage states may localise distinctly in order to optimally perform their function, or due to a differential tolerance for physiological conditions. **(B)** Temporally, the abundance of **Mac3** remains unchanged over time, whereas **Mac2** increases in abundance. Finally, macrophages with neither **Mac2** nor **Mac3** characteristics (ie signifying all remaining macrophage subsets) decrease in abundance over time to the same extent as **Mac2** is increased, suggesting a reciprocal transition away from non-delegated subset towards **Mac2** effector function acquisition.

Indeed, while this thesis confirms the differential properties of functionally-delegated macrophage subsets, the underlying biological purpose of this delegation remains speculative. Okabe and Medzhitov propose that the progressive specialisation of cell types during metazoan evolution represents a “division of labour” between primary and supportive functions, and that macrophages may have inherently evolved to serve as “accessory” cells that complement the functional repertoire of their clients (including cancer cells) to support the performance of primary client functions (394). Delegated macrophage subsets are perhaps therefore communally working towards a common meta-goal to optimally support the needs of a malignant microenvironment. Indeed, macrophages may delegate their effector functions as a direct response to the hostile conditions imposed by the tumour microenvironment, thus necessitating a streamlining of key effector functions to optimise energetic conservation and performance in these conditions. Interestingly, this enforced delegation would also explain the massive ‘phenotypic expansion’ of myeloid cells in human breast tumours compared to normal mammary tissue, observed by Azizi *et al.* (102), as well as the finding that a **Mac3**-like chemokine-expressing subset is specific to many solid tumours and not found in normal tissue (199).

6.3 | The role of TNC

The impact of TNC on the myeloid compartment is extensive and multi-faceted, as illustrated by the working hypothetical model in **Figure 6.1**, which is based on the literature as well as the experimental evidence described herein.

First, breast cancer cells and cancer-associated fibroblasts (CAFs) produce large quantities of TNC, and previous work indicated that this tumour-derived TNC induces an immune-suppressive state in myeloid cells while host-derived TNC does not (264). Once deposited, tumour-derived TNC exerts its pathogenic effects in several ways. Firstly, TNC mediates a ~2-fold depletion of the highly-antigen-presenting myeloid subsets, **Mac2** and **DC**, while additionally reducing the expression of MHC1 in **Mac2** by ~30%. Together, this impairs the ability of **Mac2** and **DC** to cross-present tumour peptide-antigens to $CD8^+$ T cells, thereby limiting the ability of the myeloid compartment to efficiently activate adaptive anti-tumoural immunity. The TNC-dependent depletion of **Mac2** is caused by a failure of the naive-like **Mac1** subset to acquire **Mac2** effector function, leading to a reciprocal ~2-fold expansion of **Mac1** in TNC-high tumours. The mechanism by which TNC depletes **DC** is unclear, but may possibly be via preventing their differentiation from monocytic precursors, the remnants of which can be observed in this subset by *Ly6c2* expression. **MDSC** are expanded ~2-fold in TNC-high tumours, and literature describes a wide array of pro-tumoural immune-suppressive functions of these cells (390-392), including an increase in calprotectin-mediated TLR4 stimulation which aids cancer cell survival and promotes metastasis (307, 393, 395, 396). As TLR4 is expressed in human breast cancer cells and TLR4 stimulation is associated with a pro-invasive phenotype (397), the TLR4 stimulation mediated by **MDSC** likely synergises with the paracrine/autocrine TLR4 stimulation of cancer cells mediated by tumour-derived TNC, resulting in enhanced tumour progression and metastasis in TNC-high tumours. However, the precise mechanism for how **MDSC** become expanded by tumour-derived TNC is unclear and requires further investigation. It is also important to note that a wide range of other cell types

in the TME also express TLR4 (for example, CAFs) meaning any TLR4-dependent mechanisms are likely more complex, implicating a multitude of cell types. The remaining macrophage subsets outlined in **Figure 6.1** are proliferating **Mac5**, and chemokine-enriched **Mac3**, neither of which are directly regulated by TNC.

However, a low-abundance population was identified by flow cytometry, resembling a dysfunctionally-delegated state featuring both **Mac2** and **Mac3** markers. While **Mac2/3** was not identified as a *bona fide* subset by transcriptome-wide scRNAseq clustering, and was only observed as a minor population of $CXCL1^{+}$ $MHC1/2^{hi}$ macrophages by flow cytometry, **Mac2/Mac3** are notable for their increasing expansion in TNC-high tumours over time, reaching ~3-fold by day 18 post engraftment, suggesting that cumulative exposure to tumour-derived TNC may induce a dysfunctionally-delegated state in macrophages. Importantly, the recent Wu *et al.* human scRNAseq atlas of breast cancer (101) which sequenced a greater number of myeloid cells, did in fact identify a **Mac2/Mac3**-like subset, perhaps suggesting that this is a relevant conserved macrophage subset in breast cancer, but that the present analysis of murine myeloid cells did not have the statistical power to resolve it.

One further area of ambiguity is the mechanism by which high levels of tumour-derived TNC mediate increased infiltration of macrophages. As this cannot be explained by changes in proliferating **Mac5**, it may be the case that TNC is able to trap monocytic cells in the TME, perhaps also explaining why naive-like **Mac1** (which may represent a monocytic or early macrophage state) is expanded in TNC-high tumours. Indeed, there is a precedent for such a 'monocyte trapping' mechanism, given that different domains of TNC are known to promote or inhibit migration of human monocyte-derived DCs (398). Finally, it is important to note that the pathogenic functions of TNC are unlikely to be exclusively mediated via its interaction with TLR4. However the TLR4 interaction is thought to be a major contributor to this phenotype, as previous work showed that antibody-mediated blockade of the TNC:TLR4 interaction in a related NeuNT model of breast cancer significantly reduced metastasis and increased MHC2

expression in macrophages (ie **Mac2**-like) (264), aligning with the effects of TNC proposed in **Figure 6.1**. Thus, this working model hypothesises that myeloid subsets whose abundance is regulated by tumour-derived TNC (**Mac1**, **Mac2**, **DC**, **MDSC**) all express surface-localized TLR4, whereas the remaining subsets do not (**Mac3**, **Mac4**, **Mac5**), although it is possible that differences in co-receptor expression (eg MD-2) may also be relevant. This model additionally proposes that therapeutic blockade of the TNC:TLR4 interaction phenocopies a low tumour-derived TNC environment (TNC-low), which is evidenced by the expansion of a **Mac2**-like subset upon this treatment in previous work (264).

6.4 | Prognostic and therapeutic implications

The delegation of macrophage effector functions to distinct anti-tumoural and pro-metastatic subsets provides an opportunity to specifically target pathogenic cell populations while sparing beneficial ones. Given the central role that **Mac3** appears to play in co-ordinating breast cancer metastasis, and the possibility that this is a tumour-specific subset not found in normal tissue, treatment strategies that deplete or reprogram **Mac3** are likely to prevent metastasis with limited off-target effects, thus representing an attractive therapeutic target. *In vivo* inhibition of the epigenetic **Mac3** regulator, KDM6B, increased the abundance of **Mac3** and would therefore not be appropriate therapeutically. However, this implicates the trimethylation status of H3K27 in maintaining **Mac3** abundance, potentially meaning that either therapeutic induction of KDM6B, or inhibition of a H3K27-specific methyltransferase (such as EZH2) may reduce **Mac3** abundance.

Ecotype analysis in a large clinical transcriptomic dataset (METABRIC) verified that low or absent myeloid enrichment was positively prognostic, somewhat reinforcing the value of pan-myeloid-depletion strategies in breast cancer patients as a post-surgical (adjuvant) intervention. However, the knowledge that mild enrichment of naive-like **Mac1** subsets is associated with ecotypes bearing the most positive patient prognosis both in this analysis and in Wu *et al.* (101), suggests that treatment strategies may need to leave certain elements of the myeloid compartment intact enough to inform subsequent adaptive immune responses and anti-tumour immunity. It is important to note that the anti-tumoural subsets, **Mac2** and **DC,** were enriched alongside all other myeloid subsets in patients that harboured poor prognosis, suggesting that their mere presence is not sufficient to adequately induce anti-tumour immunity. However, the absence of any ecotypes solely (or mostly) enriching **Mac2** and/or **DC** might suggest that this state is so highly anti-tumoural that cancer could not form or progress in this scenario. Thus, the value of **Mac2** and **DC** should not be discounted due to the pathogenic effects of the remaining myeloid compartment. One final therapeutic possibility

would be to reduce the myeloid burden in breast tumours by depleting or inhibiting the pathogenic proliferation of **Mac5**, which replenishes the TAM population. In fact, perhaps the optimal solution would be to devise a dual therapeutic approach that reduces myeloid burden (ie by selective inhibition or depletion of **Mac5**) while also reprogramming macrophages towards non-pathogenic phenotypes (ie shifting pro-metastatic **Mac3** to anti-tumoural **Mac2**).

Finally, and perhaps most importantly, this work proposes a central role for TNC in pathogenically manipulating the myeloid compartment, and the conservation of these myeloid subsets in human breast cancer suggests that anti-TNC therapy may indeed translate into clinically effective immune activation in patients. In combination with immune checkpoint blockade, anti-TNC therapy holds particular promise, as demonstrated by its synergy with anti-PD-L1 treatment in the NeuNT model of breast cancer (264).

6.5 | Future directions

In conclusion, this investigation has answered important questions about the myeloid landscape in breast cancer, and the impact of TNC in shaping myeloid phenotypes. However, it has also raised a range of other important questions which require further assessment. In the short term, several follow-up analyses should be performed to complement the current findings and address working hypotheses. First, to confirm that the chemokine-enriched subset, **Mac3**, is indeed associated with a perivascular niche, further immunofluorescence imaging of tumour tissue sections could be performed with co-staining for the endothelial cell marker, CD31. Second, to confirm the exact ontogeny of TAMs in this mouse model lineage tracing experiments or anti-CCR2 monocyte depletion could be performed to establish the contribution of monocytes to functional delegation. Third, to confirm whether tumour-derived TNC traps monocytes, thus explaining the accumulation of macrophages in TNC-high tumours, immunofluorescence imaging of monocyte markers such as Ly6C and CCR2 in combination with TNC could be performed on TNC-high and TNC-low tumour sections. Fourth, to confirm the precise role of TLR4 in mediating TNC-dependent myeloid phenotypes, scRNAseq of mice undergoing therapeutic blockade of the TNC:TLR4 interaction could be performed, verifying that this treatment phenocopies the state of myeloid cells in a TNC-low microenvironment. Fifth, a range of bioinformatic analyses could be performed that harness the recent Wu *et al.* scRNAseq atlas of human breast cancer (101). For instance, a systematic unsupervised conservation analysis could be performed to confirm previous findings, while a pseudotime analysis (which was not performed by the authors) could be used to verify trajectories of macrophage activation and effector function acquisition. This would not only provide insights into the conservation of macrophage effector trajectories, and any potential regulators that could be targeted for therapeutic reprogramming, but it would also address whether the model of ‘continuous activation’ for breast cancer macrophages (proposed by Azizi *et al.* (102)) still holds true when using greater cell numbers, sequencing depth, cohort size and disease subtype representation. The inclusion of the whole tumour microenvironment

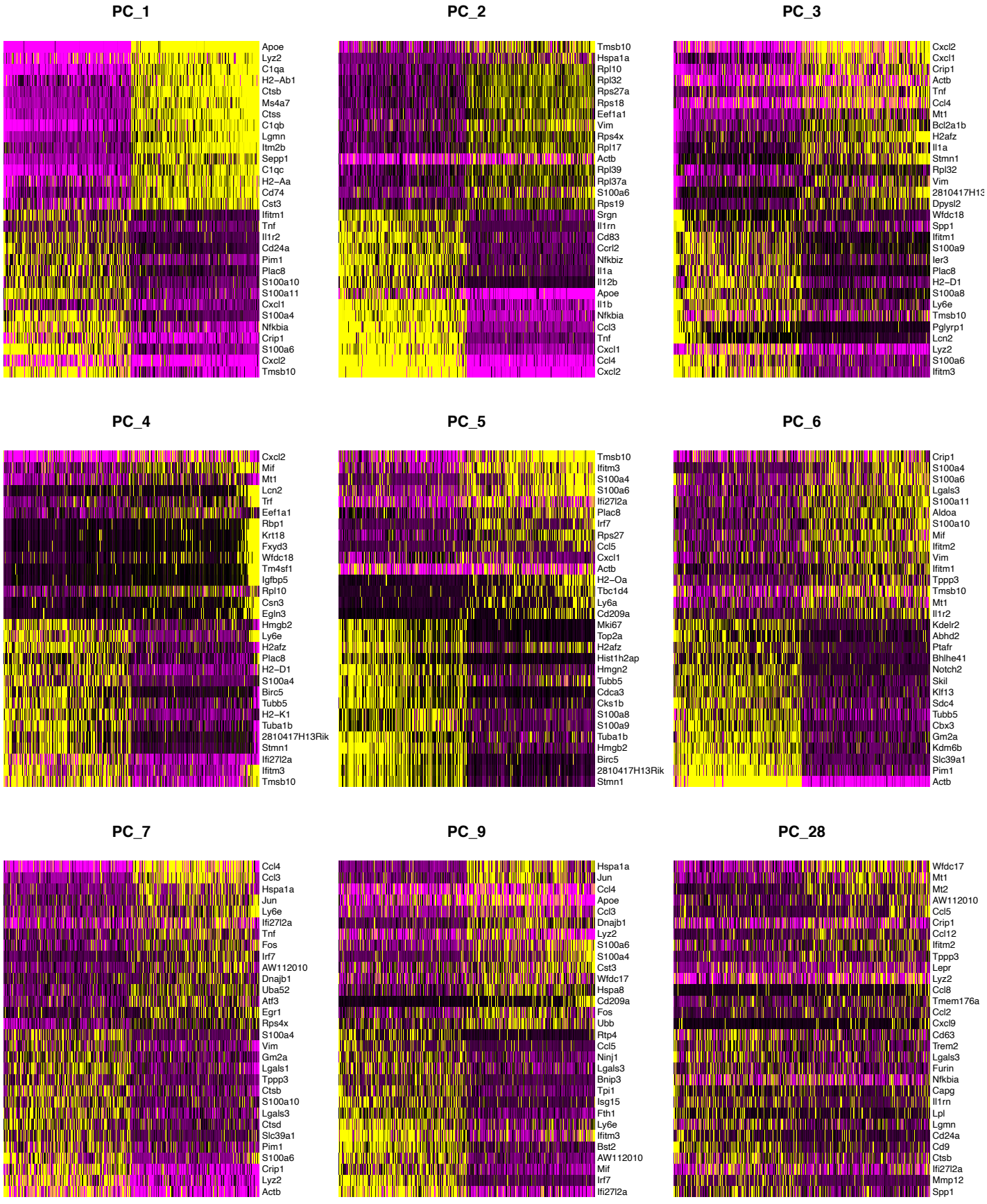
in the. Wu *et al.* dataset also offers a valuable opportunity to perform microenvironment-wide ligand-receptor interaction analysis, which may uncover novel interactions between conserved myeloid subsets and other cell types. Finally, the availability of spatial transcriptomic data enables analysis of the localisation of distinct macrophage states - for example, whether **Mac2** and **Mac3** are indeed differentially localized in human tumours, whether **Mac3** is perivascular-associated, and whether the presence of *TNC* correlates with localisation of specific myeloid subsets.

In the long term, several important avenues should be explored. Firstly, delineating the properties of host-stroma-derived versus tumour-derived TNC remains a critical question, as the former is able to stimulate beneficial anti-tumoural immunity while the latter worsens tumour progression and metastasis. Deepened understanding of the molecular and immunologic mechanism underpinning this stark difference may offer new therapeutic approaches for selectively inducing the beneficial activity of TNC while limiting pathogenic functions. Second, devising further targeting strategies for key pathogenic subsets such as the pro-metastatic subset, **Mac3**, and the proliferating subset, **Mac5**, may hold great therapeutic potential, and strategies that reprogram TAMs towards the anti-tumoural **Mac2** state should also be prioritised. Third, a deeper analysis of myeloid ecotypes within different disease subtypes (such as HER2-enriched and basal), alongside the efficacy of therapeutic regimens across different ecotypes (such as chemotherapy, radiotherapy and others), may lead to effective ecotype-based patient stratification and personalised medicine. Finally, the analysis of ever-larger cohorts of breast cancer patients at single-cell resolution will enable direct associations to be established between cell subsets and patient survival, without making inferences through bulk datasets. Achieving the statistical power necessary for single-cell survival analysis will likely require the use of historical frozen samples with long-term patient follow-up data, at least in the near term. Importantly, the routine collection of matched normal tissue, blood and lymph node samples would also greatly aid our understanding of tumour-specific phenotypes and the development of myeloid-targeted therapies.

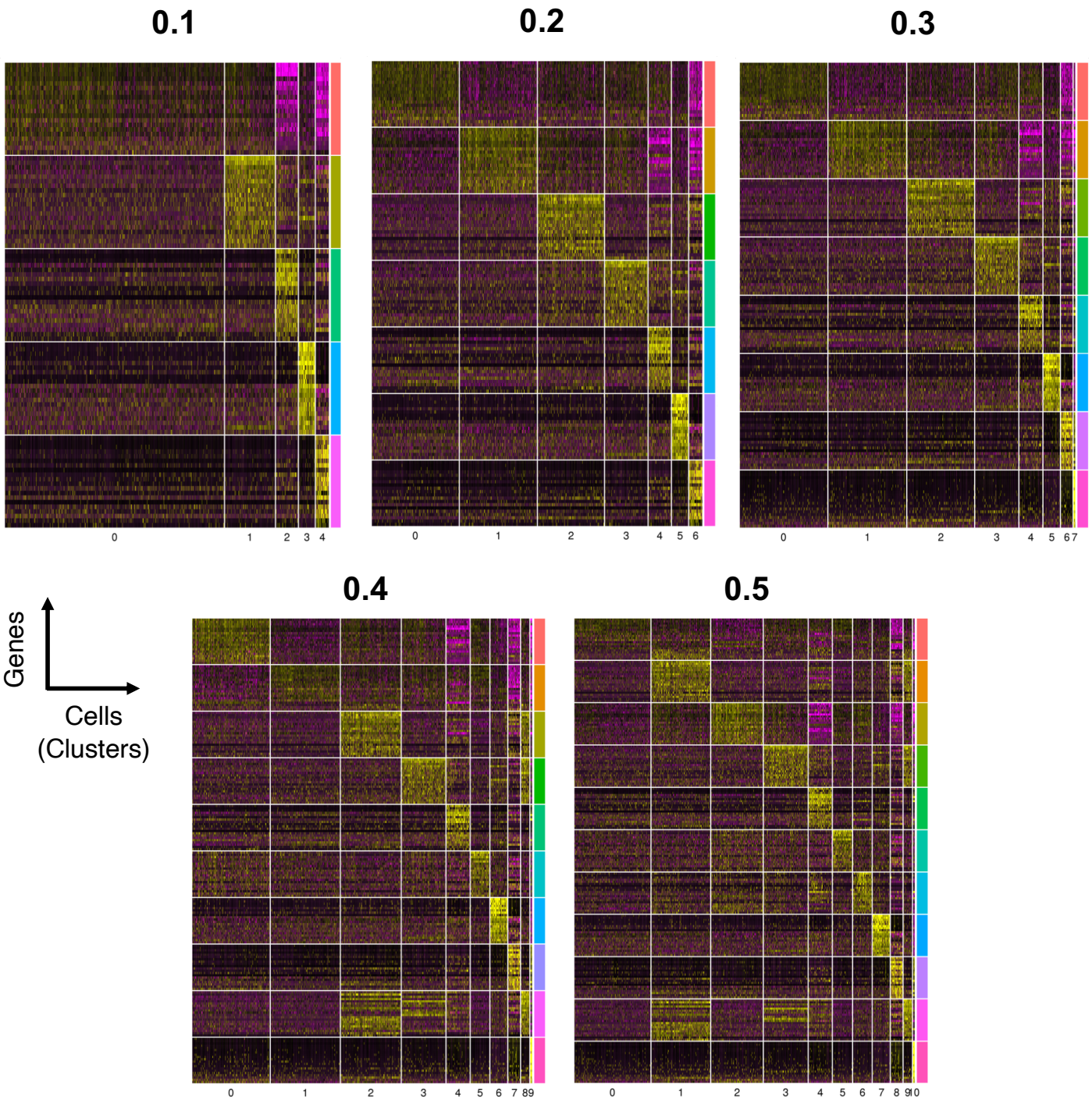
Appendix

Metric	TNC-high (shCTL)	TNC-low (shTNC)
Estimated Number of Cells	2,711	2,105
Mean Reads per Cell	60,686	71,178
Median Genes per Cell	1,546	1,532
Number of Reads	164,520,161	149,830,278
Valid Barcodes	98.40%	98.40%
Sequencing Saturation	83.80%	86.60%
Q30 Bases in Barcode	96.60%	96.60%
Q30 Bases in RNA Read	70.20%	71.20%
Q30 Bases in Sample Index	92.30%	92.90%
Q30 Bases in UMI	96.90%	96.90%
Reads Mapped to Genome	91.50%	91.40%
Reads Mapped Confidently to Genome	87.00%	87.50%
Reads Mapped Confidently to Intergenic Regions	3.00%	2.80%
Reads Mapped Confidently to Intronic Regions	11.60%	11.50%
Reads Mapped Confidently to Exonic Regions	72.50%	73.20%
Reads Mapped Confidently to Transcriptome	69.50%	70.00%
Reads Mapped Antisense to Gene	1.50%	1.50%
Fraction Reads in Cells	84.00%	86.20%
Total Genes Detected	15,033	15,090
Median UMI Counts per Cell	5,447	5,340

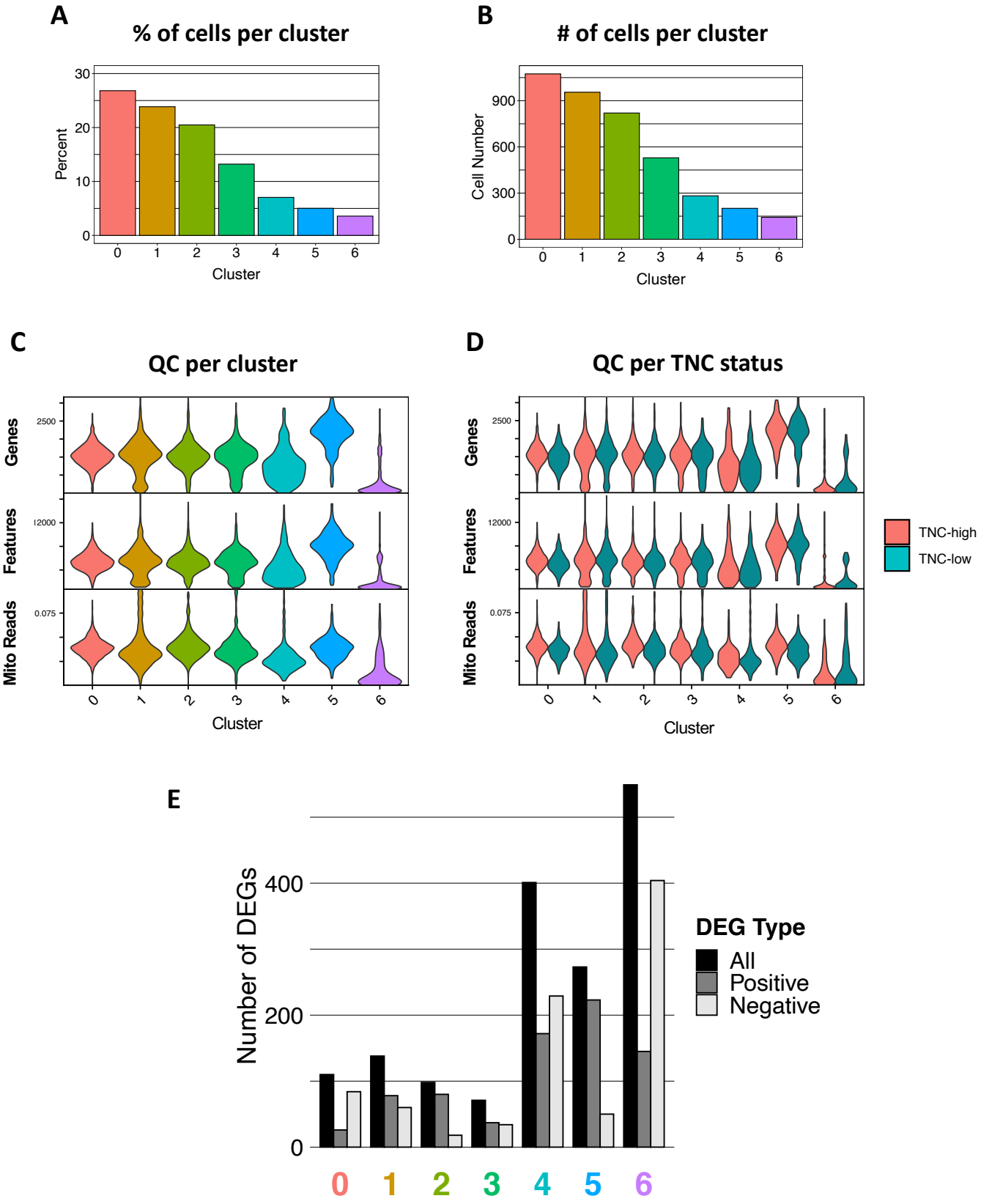
Appendix Figure 3.1 | Summary metrics of 10X single-cell RNAseq run. Quality control metrics of 10X Chromium single-cell RNAseq run for both TNC-high and TNC-low conditions, generated by CellRanger.



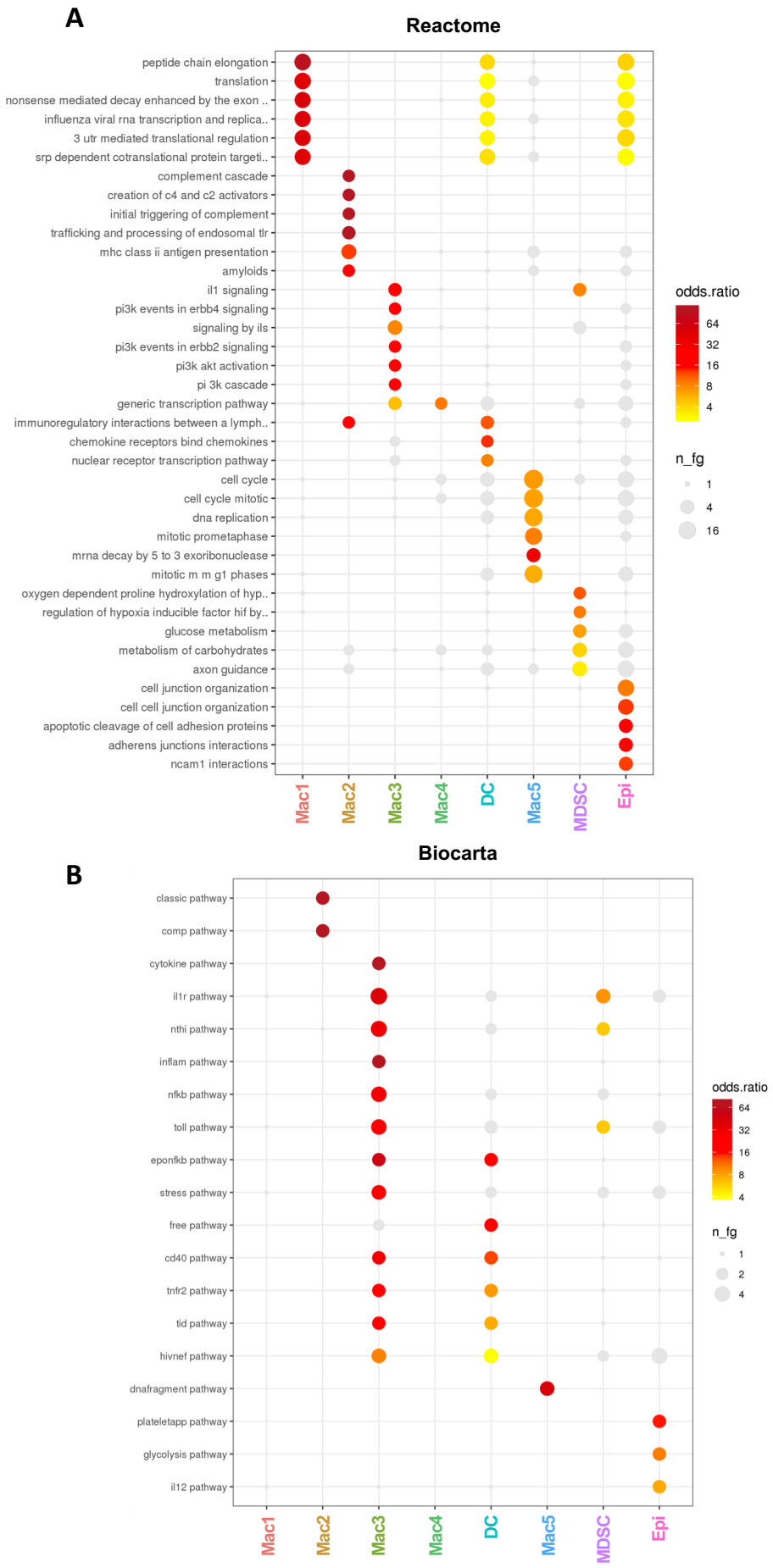
Appendix Figure 3.2 | Extended linear Dimensionality Reduction. Heatmaps visualizing the top 15 highly and top 15 lowly expressed genes for all statistically significant principal components (PCs), as determined by the Jackstraw procedure (272, 273). Only the top 500 cells are shown per PC. Note: the first three PCs are shown in **Figure 3.3C**.



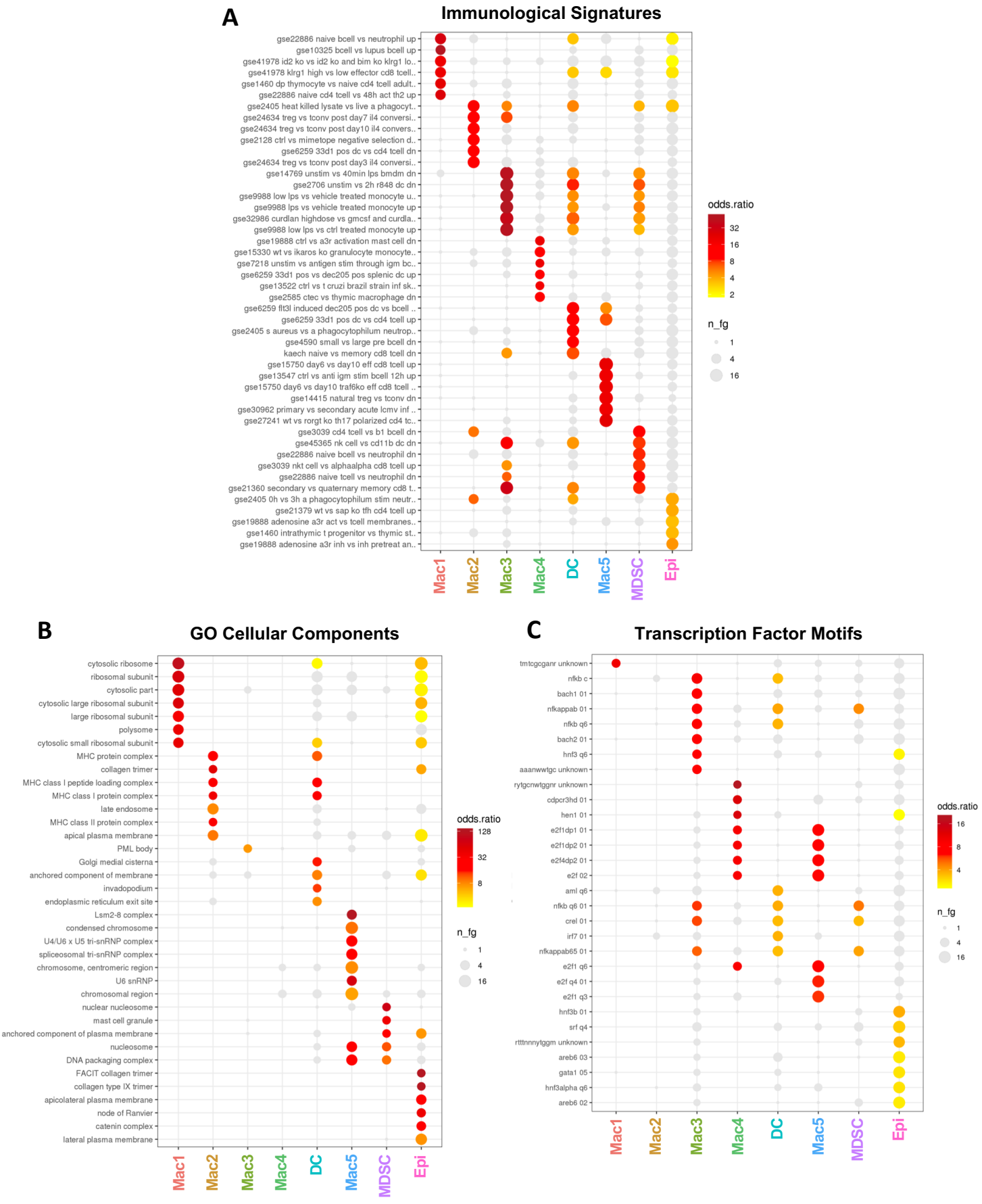
Appendix Figure 3.3 | Comparison of marker gene heatmaps across several cluster resolutions. Differential expression analysis between clusters was performed on aggregated data (from both TNC conditions) using the Wilcox test, at five different cluster resolutions (0.1-0.5). Testing was limited to genes: with a log fold-change of >0.25, detected in at least 10% of cells, and conserved between TNC-high and TNC-low conditions. The top 20 differentially expressed marker genes per cluster (ie rows), defined by their adjusted P value (Benjamini-Hochberg correction) are displayed by heatmap. A total of 4024 cells (ie columns) selected after QC and downsampling were included (2012 from each condition). Gene names were removed for clarity.



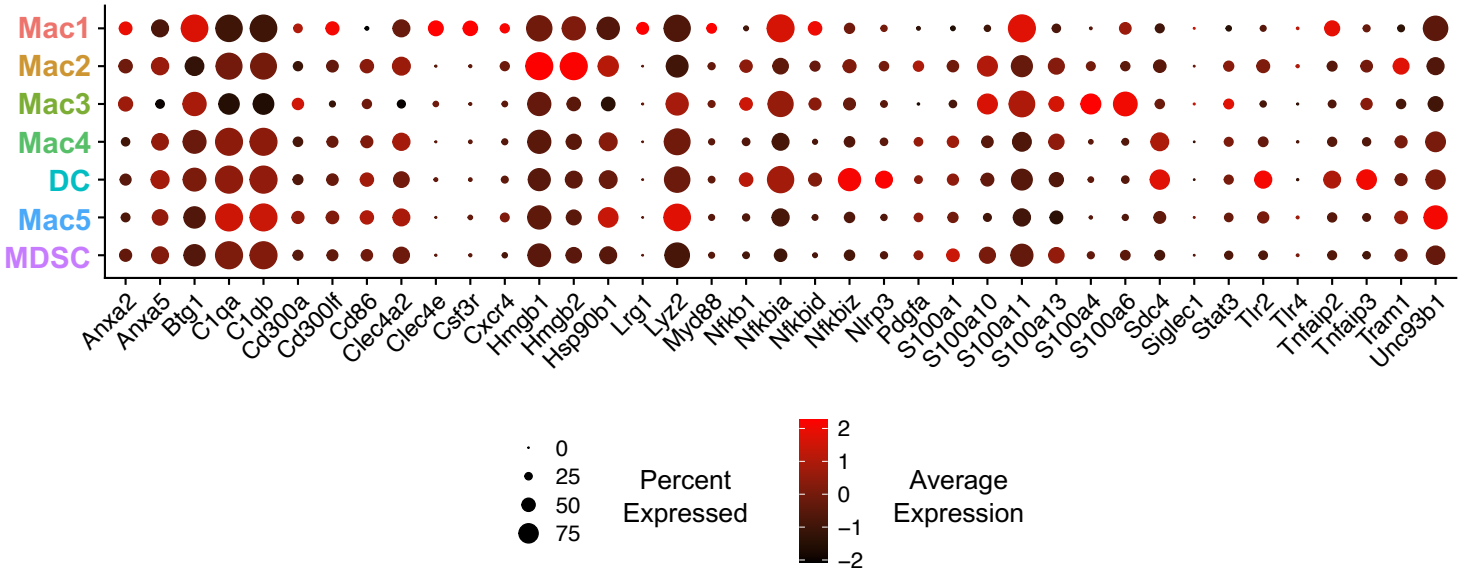
Appendix Figure 3.4 | Cluster metadata & quality control metrics of myeloid cells. **(A)** Barplot of the percentage of cells in each cluster. **(B)** Barplot of the absolute number of cells in each cluster. **(C)** Violin plots of QC metrics representing the number of genes, UMIs and the proportion of mitochondrial reads, per cluster. **(D)** Violin plots of QC metric according to TNC status, per cluster. **(E)** Barplot enumerating the differentially expressed genes per cluster. Abbreviations: DEG = Differentially Expressed Gene, QC = Quality Control, UMI = Unique Molecular Identifier.



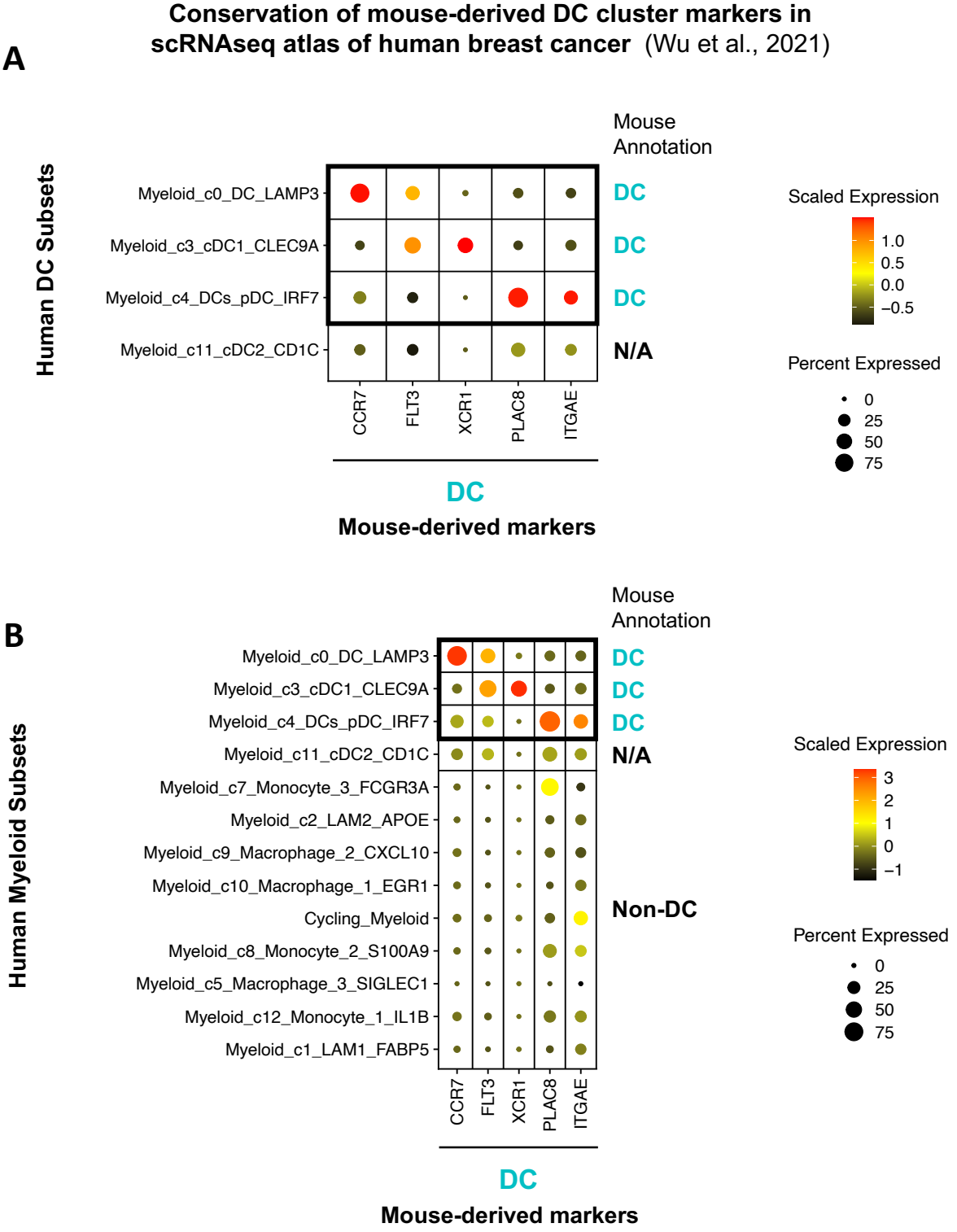
Appendix Figure 3.5 | Extended geneset over-representation analysis using pathway-related annotations. All genes significantly differentially expressed between clusters (BH–adjusted *P* value < 0.05) were assessed for their over-representation of the following functionally annotated genesets: **(A)** Reactome **(B)** Biocarta. The minimum odds ratio for inclusion of a geneset was 1.5, and the minimum number of over-represented foreground genes ('n_fg') was 3. Abbreviations: BH = Benjamini-Hochberg correction



Appendix Figure 3.6 | Extended geneset over-representation analysis using cellular- and molecular-signature-related annotations. All genes significantly differentially expressed between clusters (BH–adjusted *P* value < 0.05) were assessed for their over-representation of the following functionally annotated genesets: **(A)** Immunological Signatures **(B)** GO Cellular Components **(C)** Transcription Factor motifs. The minimum odds ratio for inclusion of a geneset was 1.5, and the minimum number of over-represented foreground genes ('n_fg') was 3. Abbreviations: BH = Benjamini-Hochberg correction; GO = Gene Ontology.

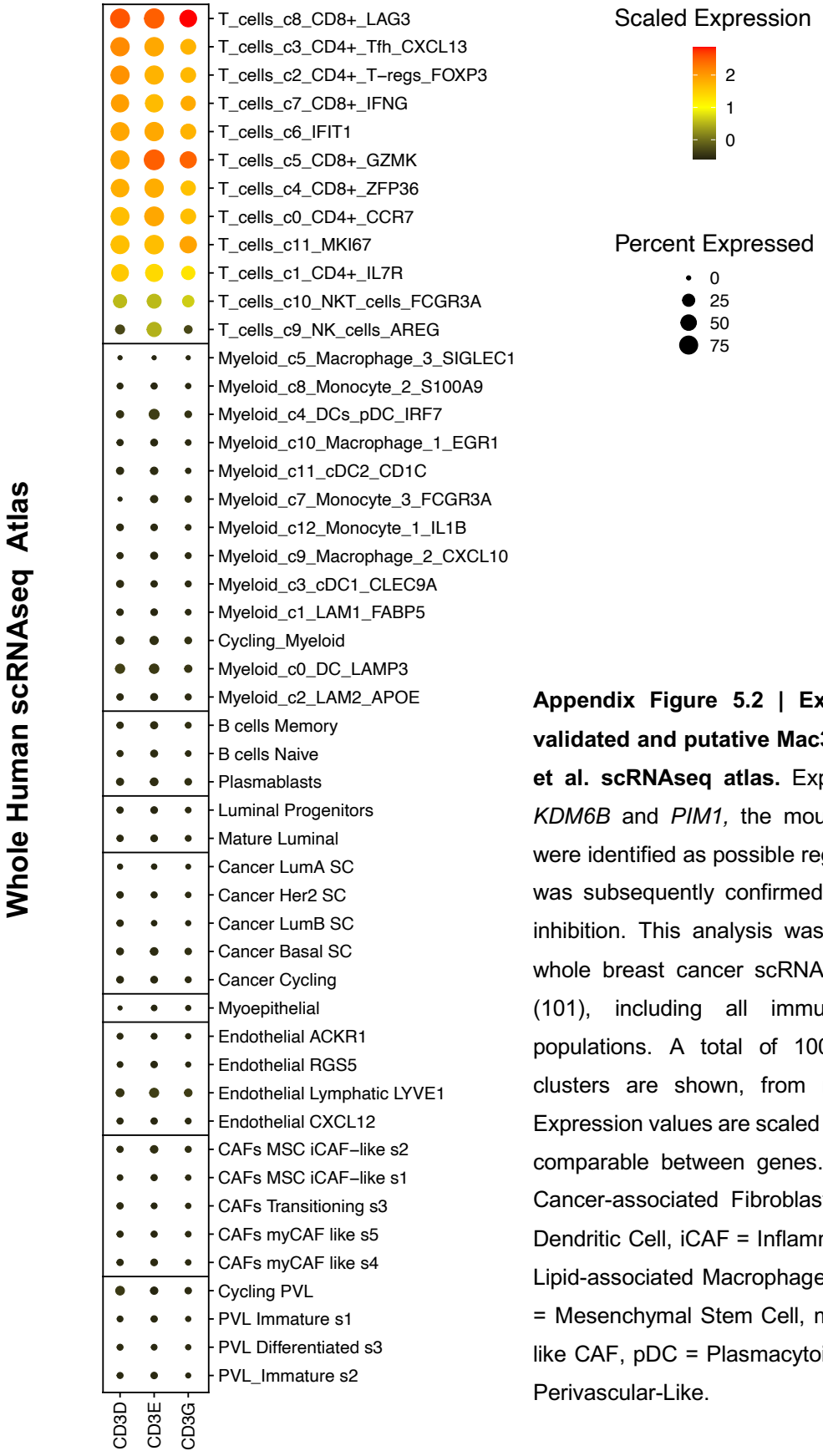


Appendix Figure 3.7 | Extended phenotypic characterization of myeloid subsets by scRNAseq expression analysis of known immune genes. Dot plots presenting expression of well-described genes involving myeloid cell functions such as pattern recognition signaling, complement production, and angiogenesis, alongside other lineage-defining and phenotype-defining markers of the myeloid compartment. The size of dots represents the percentage of cells expressing a given gene within a cluster, while the color represents the mean expression level of a given gene within a cluster.

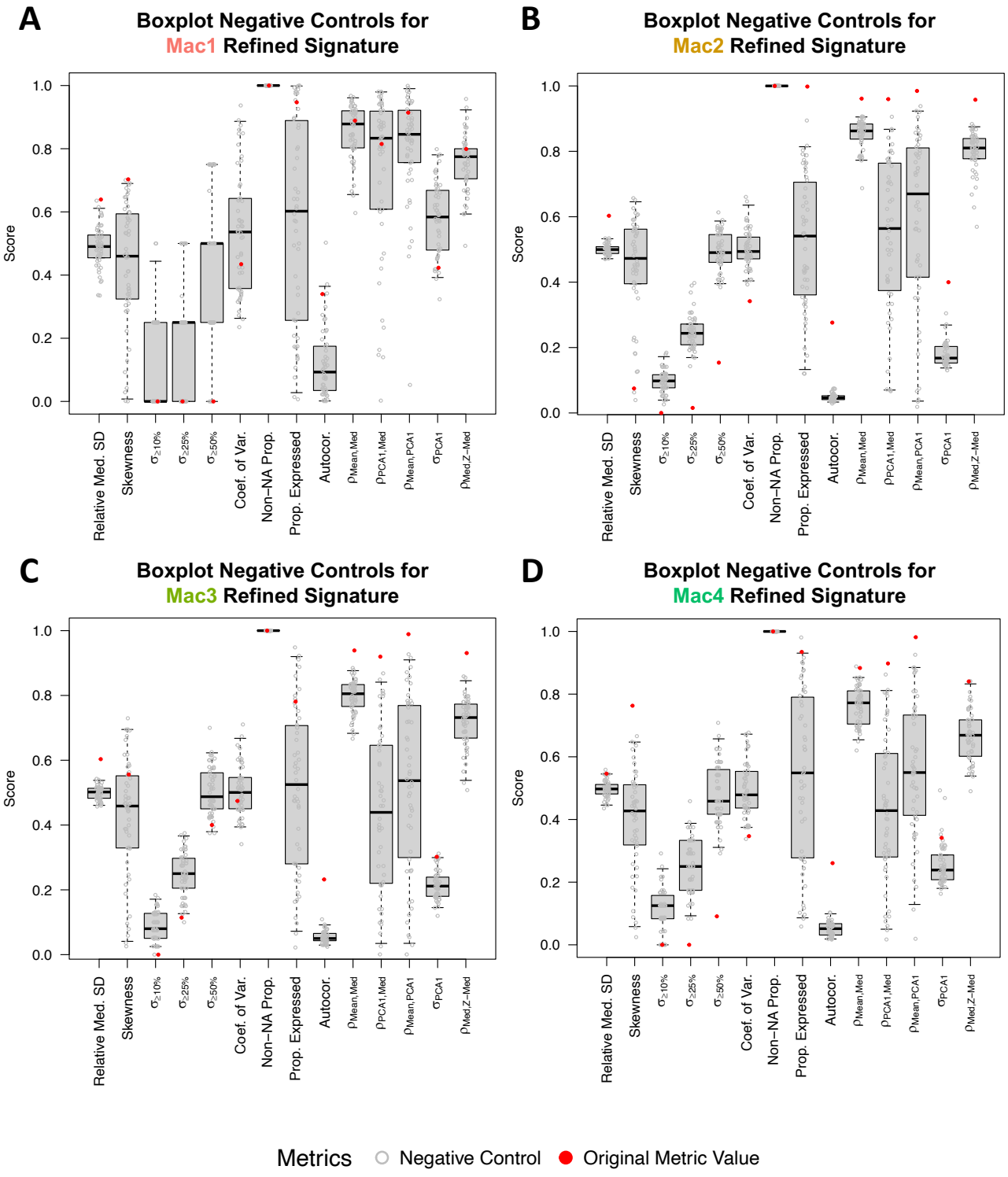


Appendix Figure 5.1 | Conservation of the heterogeneous mouse DC cluster in the Wu *et al.*, human breast cancer single-cell RNAseq atlas. Expression bubble plot of mouse-derived **DC** marker genes, converted to human homologs, and assessed in the Wu *et al.* human breast cancer scRNAseq atlas (101). This analysis was either restricted to **(A)** the four subsets assigned as ”DC” by the authors (618 cells) or **(B)** the entire myeloid compartment (9675 cells) comprised of 13 subsets. This dataset was derived from n=26 patient breast tumours. Mouse-derived marker genes which could successfully identify a human subset were annotated where possible, or defined as N/A for subsets that could not be assigned. Expression values are scaled per gene (column) to be comparable between genes. Abbreviations: cDC = Conventional Dendritic Cell, LAM = Lipid-associated Macrophage, pDC = Plasmacytoid Dendritic Cell.

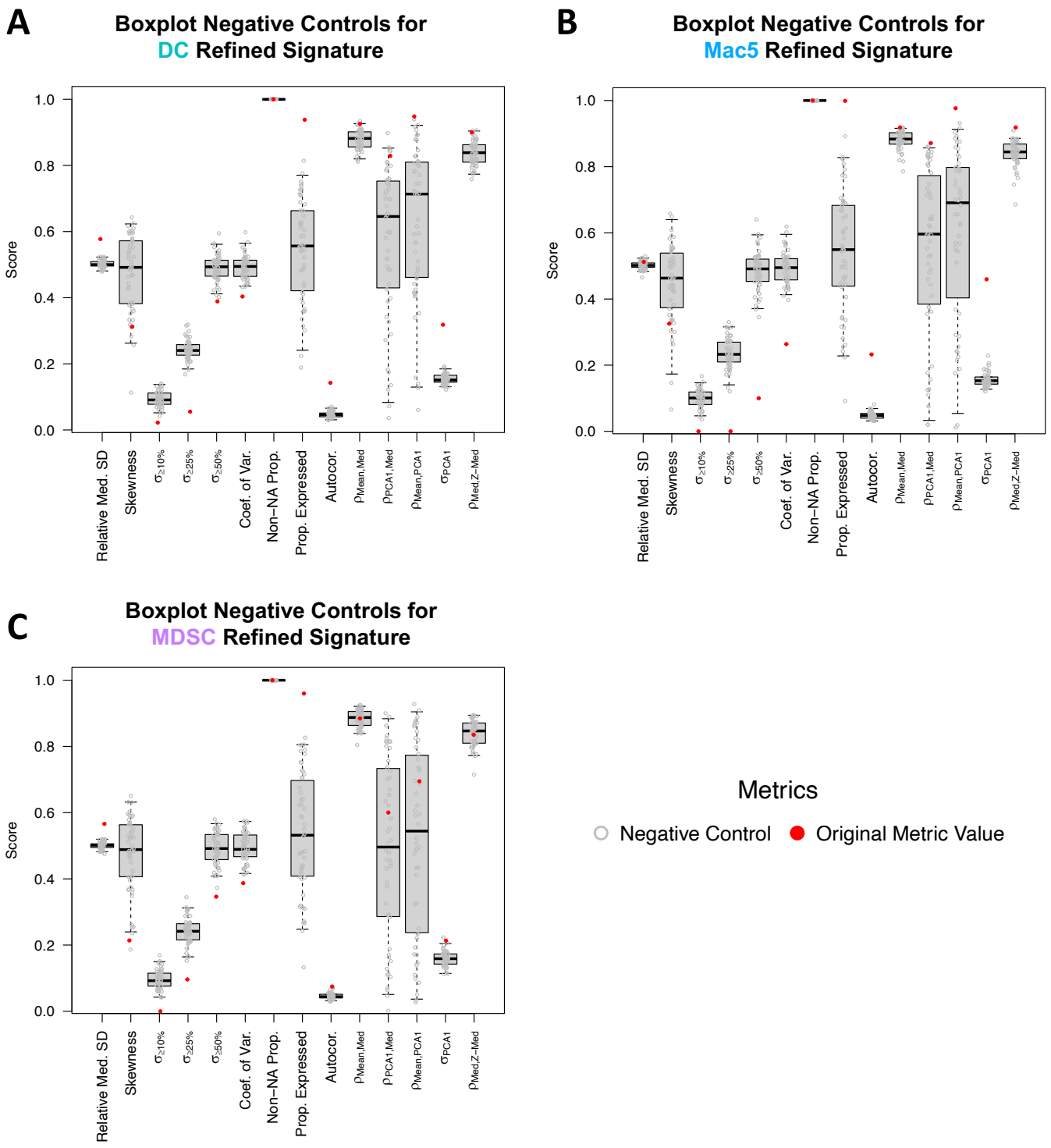
Conservation of Mac3 regulators in scRNAseq atlas of human breast cancer (Wu et al., 2021)



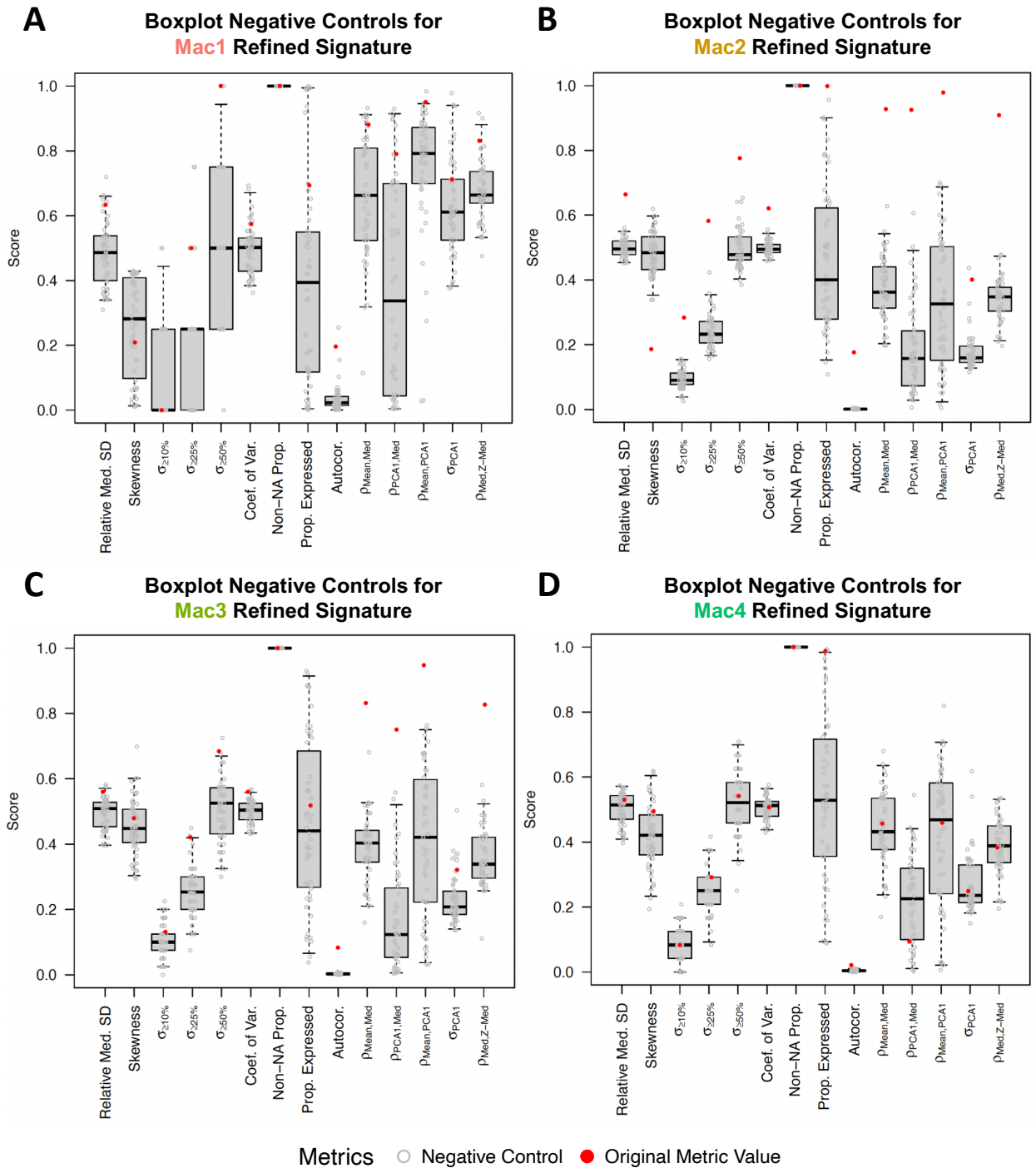
Appendix Figure 5.2 | Expression analysis of validated and putative Mac3 regulators in the Wu et al. scRNAseq atlas. Expression bubble plot of *KDM6B* and *PIM1,* the mouse homologs of which were identified as possible regulators of **Mac3**, which was subsequently confirmed for KDM6B by *in vivo* inhibition. This analysis was performed across the whole breast cancer scRNAseq atlas of Wu et al. (101), including all immune and non-immune populations. A total of 100,064 cells across 49 clusters are shown, from n=26 patient tumours. Expression values are scaled per gene (column) to be comparable between genes. Abbreviations: CAF = Cancer-associated Fibroblast, cDC = Conventional Dendritic Cell, iCAF = Inflammatory-like CAF, LAM = Lipid-associated Macrophage, Lum = Luminal, MSC = Mesenchymal Stem Cell, myCAF = Myofibroblast-like CAF, pDC = Plasmacytoid Dendritic Cell, PVL = Perivascular-Like.



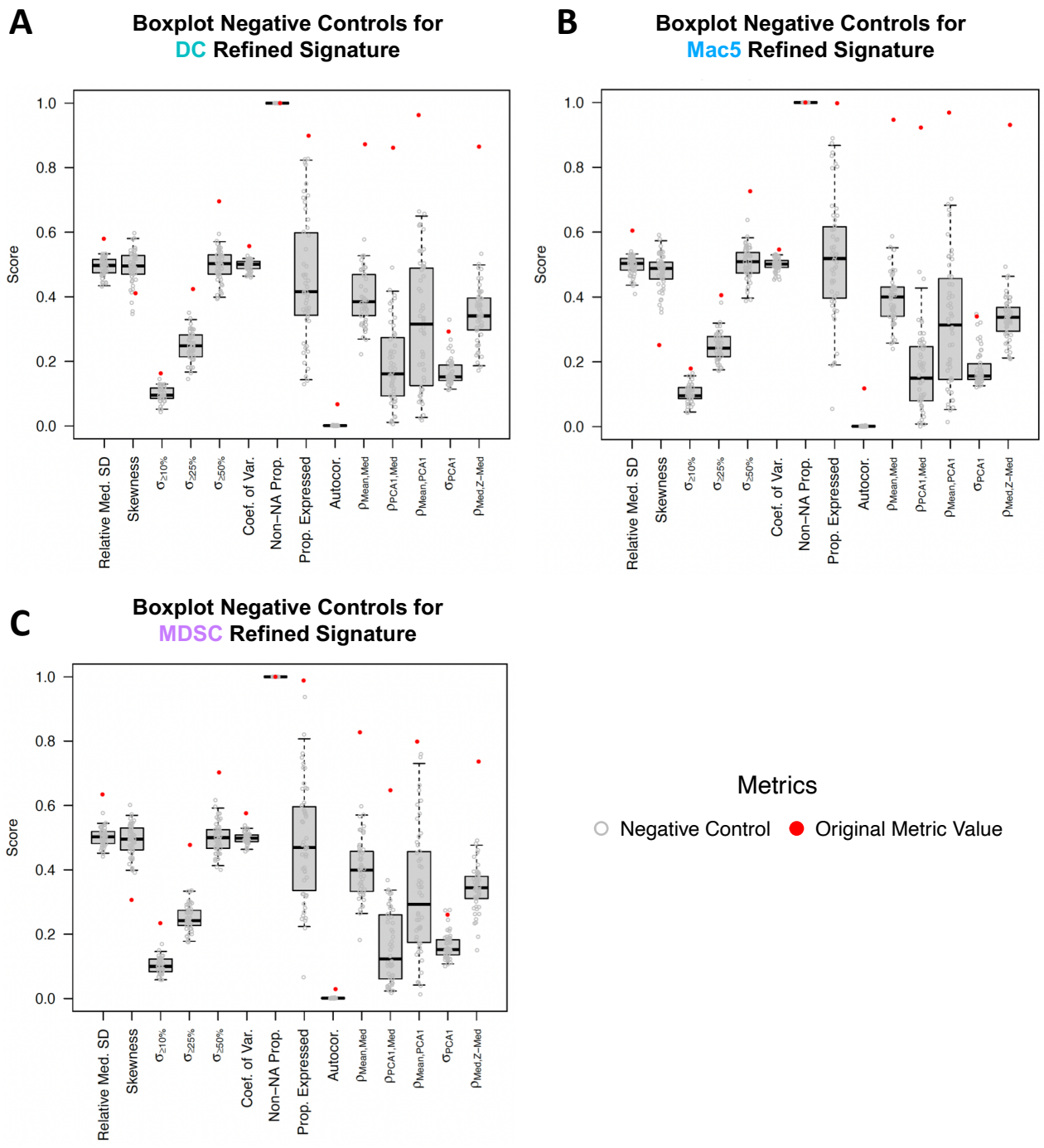
Appendix Figure 5.3 | Random resampling to generate negative control signatures in TCGA BRCA data (Part 1). A total of 1021 bulk RNAseq human breast tumour samples from TCGA were analysed. Expression values in this dataset represent TPM and were log2(x+1) transformed prior to analysis with the R Package, SigQC (293). Boxplots visualising the null distribution (negative control) of each quality control metric for myeloid signatures in the TCGA BRCA dataset were derived by resampling of random gene signatures (grey) of an equal length to the genuine signature genes (red) comprising: **(A)** Mac1, **(B)** Mac2, **(C)** Mac3 **(D)** Mac4**.** Only the the first 4 out of 7 signatures are shown, with remaining plots presented in **Appendix Figure 5.4**. The centre line denotes the median value, while edges indicate quartiles, and outliers are shown as individual replicates. Abbreviations: TCGA = The Cancer Genome Atlas, TPM = Transcripts Per Million.



Appendix Figure 5.4 | Random resampling to generate negative control signatures in TCGA BRCA data (Part 2). A total of 1021 bulk RNAseq human breast tumour samples from TCGA were analysed. Expression values in this dataset represent TPM and were log2(x+1) transformed prior to analysis with the R Package, SigQC (293). Boxplots visualising the null distribution (negative control) of each quality control metric for myeloid signatures in the TCGA BRCA dataset were derived by resampling of random gene signatures (grey) of an equal length to the genuine signature genes (red) comprising: **(A)** DC, **(B)** Mac5, and **(C)** MDSC**.** Only the final 3 out of 7 signatures are shown, with remaining plots presented in **Appendix Figure 5.3**. The centre line denotes the median value, while edges indicate quartiles, and outliers are shown as individual replicates. Abbreviations: TCGA = The Cancer Genome Atlas, TPM = Transcripts Per Million.



Appendix Figure 5.5 | Random resampling to generate negative control signatures in METABRIC data (Part 1). A total of 1991 bulk microarray human breast tumour samples from METABRIC were analysed with the R Package, SigQC (293). Boxplots visualising the null distribution (negative control) of each quality control metric for myeloid signatures in the METABRIC dataset were derived by resampling of random gene signatures (grey) of an equal length to the genuine signature genes (red) comprising: **(A)** Mac1, **(B)** Mac2, **(C)** Mac3 **(D)** Mac4**.** Only the the first 4 out of 7 signatures are shown, with remaining plots presented in **Appendix Figure 5.6**. The centre line denotes the median value, while edges indicate quartiles, and outliers are shown as individual replicates. Acknowledgment: Clinical metadata was cleaned and amalgamated by Yuechan Chi, and the dataset was provided by the group of Professor Francesca Buffa, both from the Department of Oncology (University of Oxford). All analysis presented above was performed alone with no assistance. Abbreviations: METABRIC = Molecular Taxonomy of Breast Cancer International Consortium.



Appendix Figure 5.6 | Random resampling to generate negative control signatures in METABRIC data (Part 2). A total of 1991 bulk microarray human breast tumour samples from METABRIC were analysed with the R Package, SigQC (293). Boxplots visualising the null distribution (negative control) of each quality control metric for myeloid signatures in the METABRIC dataset were derived by resampling of random gene signatures (grey) of an equal length to the genuine signature genes (red) comprising: **(A)** DC, **(B)** Mac5, and **(C)** MDSC**.** Only the final 3 out of 7 signatures are shown, with remaining plots presented in **Appendix Figure 5.5**. The centre line denotes the median value, while edges indicate quartiles, and outliers are shown as individual replicates. Acknowledgment: Clinical metadata was cleaned and amalgamated by Yuechan Chi, and the dataset was provided by the group of Professor Francesca Buffa, both from the Department of Oncology (University of Oxford). All analysis presented above was performed alone with no assistance. Abbreviations: METABRIC = Molecular Taxonomy of Breast Cancer International Consortium.

Bibliography

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Ca-a Cancer Journal for Clinicians. 2021;71(3):209-49.
2. Abbafati C, Abbas KM, Abbasi M, Abbasifard M, Abbasi-Kangevari M, Abbastabar H, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet. 2020;396(10258):1204-22.
3. Fitzmaurice C, Abate D, Abbasi N, Abbastabar H, Abd-Allah F, Abdel-Rahman O, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017 A Systematic Analysis for the Global Burden of Disease Study. Jama Oncology. 2019;5(12):1749-68.
4. Cairns J. MUTATION SELECTION AND NATURAL-HISTORY OF CANCER. Nature. 1975;255(5505):197-200.
5. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Ng AWT, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578(7793):94-+.
6. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. Science. 2015;349(6255):1483-9.
7. Gerstung M, Jolly C, Leshchiner I, Dentro SC, Gonzalez S, Rosebrock D, et al. The evolutionary history of 2,658 cancers. Nature. 2020;578(7793):122-+.
8. Albuquerque TAF, do Val LD, Doherty A, de Magalhaes JP. From humans to hydra: patterns of cancer across the tree of life. Biological Reviews. 2018;93(3):1715-34.
9. Aktipis CA, Boddy AM, Jansen G, Hibner U, Hochberg ME, Maley CC, et al. Cancer across the tree of life: cooperation and cheating in multicellularity. Philosophical Transactions of the Royal Society B-Biological Sciences. 2015;370(1673):21.
10. Nedelcu AM. The evolution of multicellularity and cancer: views and paradigms. Biochemical Society Transactions. 2020;48(4):1505-18.
11. Davies PCW, Lineweaver CH. Cancer tumors as Metazoa 1.0: tapping genes of ancient ancestors. Physical Biology. 2011;8(1):7.
12. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. Cell. 2011;144(5):646-74.
13. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57-70.
14. Marusyk A, Janiszewska M, Polyak K. Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. Cancer Cell. 2020;37(4):471-84.
15. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. Nature Reviews Clinical Oncology. 2018;15(2):81-94.
16. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. Cell. 2017;168(4):613-28.
17. Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. Nature. 2008;454(7203):436-44.
18. Greten FR, Grivennikov SI. Inflammation and Cancer: Triggers, Mechanisms, and Consequences. Immunity. 2019;51(1):27-41.
19. Baghban R, Roshangar L, Jahanban-Esfahlan R, Seidi K, Ebrahimi-Kalan A, Jaymand M, et al. Tumor microenvironment complexity and therapeutic implications at a glance. Cell Communication and Signaling. 2020;18(1):19.
20. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. Nature Medicine. 2013;19(11):1423-37.
21. Hanahan D, Coussens LM. Accessories to the Crime: Functions of Cells Recruited to the Tumor Microenvironment. Cancer Cell. 2012;21(3):309-22.
22. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nature Medicine. 2018;24(5):541-50.

23. Pua KH, Chew CL, Lane DP, Tergaonkar V. Inflammation-associated genomic instability in cancer Genome Instability & Disease. 2020(1):1-9.
24. Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO, et al. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. Journal of Biological Chemistry. 1999;274(26):18470-6.
25. Martinez-Lostao L, Anel A, Pardo J. How Do Cytotoxic Lymphocytes Kill Cancer Cells? Clinical Cancer Research. 2015;21(22):5047-56.
26. Fu CM, Jiang AM. Dendritic Cells and CD8 T Cell Immunity in Tumor Microenvironment. Frontiers in Immunology. 2018;9:11.
27. Muntjewerff EM, Meesters LD, van den Bogaart G. Antigen Cross-Presentation by Macrophages. Frontiers in Immunology. 2020;11:12.
28. Burnet FM. CONCEPT OF IMMUNOLOGICAL SURVEILLANCE. Progress in Experimental Tumor Research. 1970;13:1-+.
29. Shankaran V, Ikeda H, Bruce AT, White JM, Swanson PE, Old LJ, et al. IFN gamma and lymphocytes prevent primary tumour development and shape tumour immunogenicity. Nature. 2001;410(6832):1107-11.
30. Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoediting: from immunosurveillance to tumor escape. Nature Immunology. 2002;3(11):991-8.
31. Dunn GP, Old LJ, Schreiber RD. The immunobiology of cancer immunosurveillance and immunoediting. Immunity. 2004;21(2):137-48.
32. Schreiber RD, Old LJ, Smyth MJ. Cancer Immunoediting: Integrating Immunity's Roles in Cancer Suppression and Promotion. Science. 2011;331(6024):1565-70.
33. Vinay DS, Ryan EP, Pawelec G, Talib WH, Stagg J, Elkord E, et al. Immune evasion in cancer: Mechanistic basis and therapeutic strategies. Seminars in Cancer Biology. 2015;35:S185-S98.
34. Groth A, Kloess S, von Strandmann EP, Koehl U, Koch J. Mechanisms of Tumor and Viral Immune Escape from Natural Killer Cell-Mediated Surveillance. Journal of Innate Immunity. 2011;3(4):344-54.
35. Lee JB, Ha SJ, Kim HR. Clinical Insights Into Novel Immune Checkpoint Inhibitors. Frontiers in Pharmacology. 2021;12:19.
36. Robert C. A decade of immune-checkpoint inhibitors in cancer therapy. Nature Communications. 2020;11(1):3.
37. Hamid O, Robert C, Daud A, Hodi FS, Hwu WJ, Kefford R, et al. Safety and Tumor Responses with Lambrolizumab (Anti-PD-1) in Melanoma. New England Journal of Medicine. 2013;369(2):134-44.
38. Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, et al. Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. New England Journal of Medicine. 2010;363(8):711-23.
39. Wolchok JD, Kluger H, Callahan MK, Postow MA, Rizvi NA, Lesokhin AM, et al. Nivolumab plus Ipilimumab in Advanced Melanoma. New England Journal of Medicine. 2013;369(2):122-33.
40. Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. New England Journal of Medicine. 2015;373(17):1627-39.
41. Brahmer J, Reckamp KL, Baas P, Crino L, Eberhardt WEE, Poddubskaya E, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. New England Journal of Medicine. 2015;373(2):123-35.
42. Yarchoan M, Johnson BA, Lutz ER, Laheru DA, Jaffee EM. Targeting neoantigens to augment antitumour immunity. Nature Reviews Cancer. 2017;17(4):209-22.
43. Matsuda T, Leisegang M, Park J-H, Kato T, Ikeda Y, Harada M, et al. Effective screening of neoantigen-specific cytotoxic t cells. Journal of Clinical Oncology. 2018;36(15).
44. Chen DS, Mellman I. Oncology Meets Immunology: The Cancer-Immunity Cycle. Immunity. 2013;39(1):1-10.

45. Galluzzi L, Zitvogel L, Kroemer G. Immunological Mechanisms Underneath the Efficacy of Cancer Therapy. Cancer Immunology Research. 2016;4(11):895-902.
46. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. New England Journal of Medicine. 2015;372(26):2509-20.
47. Lei YN, Li XY, Huang Q, Zheng XF, Liu M. Progress and Challenges of Predictive Biomarkers for Immune Checkpoint Blockade. Frontiers in Oncology. 2021;11:14.
48. Martins F, Sofiya L, Sykiotis GP, Lamine F, Maillard M, Fraga M, et al. Adverse effects of immune-checkpoint inhibitors: epidemiology, management and surveillance. Nature Reviews Clinical Oncology. 2019;16(9):563-80.
49. Vareki SM. High and low mutational burden tumors versus immunologically hot and cold tumors and response to immune checkpoint inhibitors. Journal for Immunotherapy of Cancer. 2018;6.
50. Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. Science. 2018;359(6382):1350-+.
51. Neelapu SS, Locke FL, Bartlett NL, Lekakis LJ, Miklos DB, Jacobson CA, et al. Axicabtagene Ciloleucel CAR T-Cell Therapy in Refractory Large B-Cell Lymphoma. New England Journal of Medicine. 2017;377(26):2531-44.
52. Maude SL, Laetsch TW, Buechner J, Rives S, Boyer M, Bittencourt H, et al. Tisagenlecleucel in Children and Young Adults with B-Cell Lymphoblastic Leukemia. New England Journal of Medicine. 2018;378(5):439-48.
53. Schuster SJ, Svoboda J, Chong EA, Nasta SD, Mato AR, Anak O, et al. Chimeric Antigen Receptor T Cells in Refractory B-Cell Lymphomas. New England Journal of Medicine. 2017;377(26):2545-54.
54. Rafiq S, Hackett CS, Brentjens RJ. Engineering strategies to overcome the current roadblocks in CAR T cell therapy. Nature Reviews Clinical Oncology. 2020;17(3):147-67.
55. Sadelain M, Brentjens R, Riviere I. The Basic Principles of Chimeric Antigen Receptor Design. Cancer Discovery. 2013;3(4):388-98.
56. Sterner RC, Sterner RM. CAR-T cell therapy: current limitations and potential strategies. Blood Cancer Journal. 2021;11(4):11.
57. Majzner RG, Mackall CL. Tumor Antigen Escape from CAR T-cell Therapy. Cancer Discovery. 2018;8(10):1219-26.
58. Loibl S, Poortmans P, Morrow M, Denkert C, Curigliano G. Breast cancer. Lancet. 2021;397(10286):1750-69.
59. Johnson KS, Conant EF, Soo MS. Molecular Subtypes of Breast Cancer: A Review for Breast Radiologists. Journal of Breast Imaging. 2021;3(1):12-24.
60. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thurlimann B, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Annals of Oncology. 2013;24(9):2206-23.
61. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. Jama-Journal of the American Medical Association. 2006;295(21):2492-502.
62. Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, et al. Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. Journal of Clinical Oncology. 2009;27(8):1160-7.
63. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747-52.
64. Prat A, Pineda E, Adamo B, Galvan P, Fernandez A, Gaba L, et al. Clinical implications of the intrinsic molecular subtypes of breast cancer. Breast. 2015;24:S26-S35.
65. Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, et al. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490(7418):61-70.

66. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. Cell. 2015;163(2):506-19.
67. Garrido-Castro AC, Lin NU, Polyak K. Insights into Molecular Classifications of Triple-Negative Breast Cancer: Improving Patient Selection for Treatment. Cancer Discovery. 2019;9(2):176-98.
68. Yersal O, Barutca S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. World Journal of Clinical Oncology. 2014;5(3):412-24.
69. Campeau PM, Foulkes WD, Tischkowitz MD. Hereditary breast cancer: new genetic developments, new therapeutic avenues. Human Genetics. 2008;124(1):31-42.
70. Walsh T, Casadei S, Lee MK, Pennil CC, Nord AS, Thornton AM, et al. Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(44):18032-7.
71. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips KA, Mooij TM, Roos-Blom MJ, et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. Jama-Journal of the American Medical Association. 2017;317(23):2402-16.
72. Chen SN, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. Journal of Clinical Oncology. 2007;25(11):1329-33.
73. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell. 2018;173(2):371-+.
74. Angus L, Smid M, Wilting SM, van Riet J, Van Hoeck A, Nguyen L, et al. The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. Nature Genetics. 2019;51(10):1450-+.
75. Howlader N, Cronin KA, Kurian AW, Andridge R. Differences in Breast Cancer Survival by Molecular Subtypes in the United States. Cancer Epidemiology Biomarkers & Prevention. 2018;27(6):619-26.
76. Johansson ALV, Trewin CB, Hjerkind KV, Ellingjord-Dale M, Johannesen TB, Ursin G. Breast cancer-specific survival by clinical subtype after 7 years follow-up of young and elderly women in a nationwide cohort. International Journal of Cancer. 2019;144(6):1251-61.
77. Stover DG, Winer EP. Tailoring adjuvant chemotherapy regimens for patients with triple negative breast cancer. Breast. 2015;24:S132-S5.
78. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, et al. Triple-negative breast cancer: Clinical features and patterns of recurrence. Clinical Cancer Research. 2007;13(15):4429-34.
79. Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, et al. Breast cancer. Nature Reviews Disease Primers. 2019;5:31.
80. Chaffer CL, Weinberg RA. A Perspective on Cancer Cell Metastasis. Science. 2011;331(6024):1559-64.
81. Dillekas H, Rogers MS, Straume O. Are 90% of deaths from cancer caused by metastases? Cancer Medicine. 2019;8(12):5574-6.
82. Chen WJ, Hoffmann AD, Liu HP, Liu X. Organotropism: new insights into molecular mechanisms of breast cancer metastasis. Npj Precision Oncology. 2018;2:12.
83. Morgan AJ, Giannoudis A, Palmieri C. The genomic landscape of breast cancer brain metastases: a systematic review. Lancet Oncology. 2021;22(1):E7-E17.
84. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. New England Journal of Medicine. 2001;344(11):783-92.
85. Clarke M, Collins R, Davies C, Godwin J, Gray R, Peto R, et al. Tamoxifen for early breast cancer: An overview of the randomised trials. Lancet. 1998;351(9114):1451-67.
86. Andre F, Ciruelos E, Rubovszky G, Campone M, Loibl S, Rugo HS, et al. Alpelisib for PIK3CA-Mutated, Hormone Receptor-Positive Advanced Breast Cancer. New England Journal of Medicine. 2019;380(20):1929-40.

87. Robson M, Im SA, Senkus E, Xu BH, Domchek SM, Masuda N, et al. Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation. New England Journal of Medicine. 2017;377(6):523-33.
88. Barroso-Sousa R, Jain E, Cohen O, Kim D, Buendia-Buendia J, Winer E, et al. Prevalence and mutational determinants of high tumor mutation burden in breast cancer. Annals of Oncology. 2020;31(3):387-94.
89. Chalmers ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome Medicine. 2017;9:14.
90. Mei P, Freitag CE, Wei L, Zhang YX, Parwani AV, Li ZB. High tumor mutation burden is associated with DNA damage repair gene mutation in breast carcinomas. Diagnostic Pathology. 2020;15(1):7.
91. Ravaioli S, Limarzi F, Tumedei MM, Palleschi M, Maltoni R, Bravaccini S. Are we ready to use TMB in breast cancer clinical practice? Cancer Immunology Immunotherapy. 2020;69(10):1943-5.
92. Haricharan S, Bainbridge MN, Scheet P, Brown PH. Somatic mutation load of estrogen receptor-positive breast tumors predicts overall survival: an analysis of genome sequence data. Breast Cancer Research and Treatment. 2014;146(1):211-20.
93. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao YJ, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature. 2012;486(7403):395-9.
94. Chumsri S, Sokol ES, Soyano-Muller AE, Parrondo RD, Reynolds GA, Nassar A, et al. Durable Complete Response With Immune Checkpoint Inhibitor in Breast Cancer With High Tumor Mutational Burden and APOBEC Signature. *Journal of the National Comprehensive Cancer Network*. 2020(5):517-21.
95. Stanton SE, Adams S, Disis ML. Variation in the Incidence and Magnitude of Tumor-Infiltrating Lymphocytes in Breast Cancer Subtypes A Systematic Review. Jama Oncology. 2016;2(10):1354-60.
96. Denkert C, von Minckwitz G, Darb-Esfahani S, Lederer B, Heppner BI, Weber KE, et al. Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. Lancet Oncology. 2018;19(1):40-50.
97. Savas P, Salgado R, Denkert C, Sotiriou C, Darcy PK, Smyth MJ, et al. Clinical relevance of host immunity in breast cancer: from TILs to the clinic. Nat Rev Clin Oncol. 2016;13(4):228-41.
98. Dieci MV, Griguolo G, Miglietta F, Guarneri V. The immune system and hormone-receptor positive breast cancer: Is it really a dead end? Cancer Treat Rev. 2016;46:9-19.
99. Loi S, Drubay D, Adams S, Pruneri G, Francis PA, Lacroix-Triki M, et al. Tumor-Infiltrating Lymphocytes and Prognosis: A Pooled Individual Patient Analysis of Early-Stage Triple-Negative Breast Cancers. Journal of Clinical Oncology. 2019;37(7):559-+.
100. DeNardo DG, Coussens LM. Inflammation and breast cancer - Balancing immune response: crosstalk between adaptive and innate immune cells during breast cancer progression. Breast Cancer Research. 2007;9(4):10.
101. Wu SZ, Al-Eryani G, Roden DL, Junankar S, Harvey K, Andersson A, et al. A single-cell and spatially resolved atlas of human breast cancers. Nature Genetics. 2021;53(9):1334-+.
102. Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. Cell. 2018;174(5):1293-+.
103. Wagner J, Rapsomaniki MA, Chevrier S, Anzeneder T, Langwieder C, Dykgers A, et al. A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast Cancer. Cell. 2019;177(5):1330-+.
104. Liu FF, Lang RG, Zhao J, Zhang XM, Pringle GA, Fan Y, et al. CD8(+) cytotoxic T cell and FOXP3(+) regulatory T cell infiltration in relation to breast cancer survival and molecular subtypes. Breast Cancer Research and Treatment. 2011;130(2):645-55.

105. Vihervuori H, Autere TA, Repo H, Kurki S, Kallio L, Lintunen MM, et al. Tumor-infiltrating lymphocytes and CD8(+) T cells predict survival of triple-negative breast cancer. Journal of Cancer Research and Clinical Oncology. 2019;145(12):3105-14.
106. Oshi M, Asaoka M, Tokumaru Y, Yan L, Matsuyama R, Ishikawa T, et al. CD8 T Cell Score as a Prognostic Biomarker for Triple Negative Breast Cancer. International Journal of Molecular Sciences. 2020;21(18):16.
107. Savas P, Virassamy B, Ye CZ, Salim A, Mintoff CP, Caramia F, et al. Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. Nature Medicine. 2018;24(7):986-93.
108. Lin EY, Nguyen AV, Russell RG, Pollard JW. Colony-stimulating factor 1 promotes progression of mammary tumors to malignancy. Journal of Experimental Medicine. 2001;193(6):727-39.
109. Lin EY, Li JF, Gnatovskiy L, Deng Y, Zhu L, Grzesik DA, et al. Macrophages regulate the angiogenic switch in a mouse model of breast cancer. Cancer Research. 2006;66(23):11238-46.
110. Leek RD, Lewis CE, Whitehouse R, Greenall M, Clarke J, Harris AL. Association of macrophage infiltration with angiogenesis and prognosis in invasive breast carcinoma. Cancer Research. 1996;56(20):4625-9.
111. Medrek C, Ponten F, Jirstrom K, Leandersson K. The presence of tumor associated macrophages in tumor stroma as a prognostic marker for breast cancer patients. Bmc Cancer. 2012;12:9.
112. Gwak JM, Jang MH, Kim DI, Seo AN, Park SY. Prognostic Value of Tumor-Associated Macrophages According to Histologic Locations and Hormone Receptor Status in Breast Cancer. Plos One. 2015;10(4):14.
113. Riaz N, Burugu S, Cheng AS, Leung SCY, Gao DX, Nielsen TO. Prognostic Significance of CSF-1R Expression in Early Invasive Breast Cancer. Cancers. 2021;13(22):19.
114. Cassetta L, Fragkogianni S, Sims AH, Swierczak A, Forrester LM, Zhang H, et al. Human Tumor-Associated Macrophage and Monocyte Transcriptional Landscapes Reveal Cancer-Specific Reprogramming, Biomarkers, and Therapeutic Targets. Cancer Cell. 2019;35(4):588-+.
115. Mahmoud SMA, Lee AHS, Paish EC, Macmillan RD, Ellis IO, Green AR. Tumour-infiltrating macrophages and clinical outcome in breast cancer. Journal of Clinical Pathology. 2012;65(2):159-63.
116. Zhao XX, Qu JK, Sun YC, Wang JZ, Liu X, Wang FD, et al. Prognostic significance of tumor-associated macrophages in breast cancer: a meta-analysis of the literature. Oncotarget. 2017;8(18):30576-86.
117. Zhu L, Narloch JL, Onkar S, Joy M, Broadwater G, Luedke C, et al. Metastatic breast cancers have reduced immune cell recruitment but harbor increased macrophages relative to their matched primary tumors. Journal for Immunotherapy of Cancer. 2019;7(1):10.
118. Xiang XN, Wang JG, Lu D, Xu X. Targeting tumor-associated macrophages to synergize tumor immunotherapy. Signal Transduction and Targeted Therapy. 2021;6(1):12.
119. Poh AR, Ernst M. Targeting Macrophages in Cancer: From Bench to Bedside. Frontiers in Oncology. 2018;8:16.
120. Mantovani A, Marchesi F, Malesci A, Laghi L, Allavena P. Tumour-associated macrophages as treatment targets in oncology. Nature Reviews Clinical Oncology. 2017;14(7):399-416.
121. DeNardo DG, Ruffell B. Macrophages as regulators of tumour immunity and immunotherapy. Nature Reviews Immunology. 2019;19(6):369-82.
122. Cassetta L, Pollard JW. Targeting macrophages: therapeutic approaches in cancer. Nature Reviews Drug Discovery. 2018;17(12):887-904.
123. Chaudhuri S, Thomas S, Munster P. Immunotherapy in breast cancer: A clinician's perspective. Journal of the National Cancer Center. 2021;1(2):47-57.

124. Zhang SY, Song XY, Li Y, Ye LL, Zhou Q, Yang WB. Tumor-associated macrophages: A promising target for a cancer immunotherapeutic strategy. Pharmacological Research. 2020;161:8.
125. Duan ZJ, Luo YP. Targeting macrophages in cancer immunotherapy. Signal Transduction and Targeted Therapy. 2021;6(1):21.
126. Cassetta L, Kitamura T. Targeting Tumor-Associated Macrophages as a Potential Strategy to Enhance the Response to Immune Checkpoint Inhibitors. Frontiers in Cell and Developmental Biology. 2018;6:6.
127. Qian BZ, Li JF, Zhang H, Kitamura T, Zhang JH, Campion LR, et al. CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. Nature. 2011;475(7355):222-U129.
128. Strachan DC, Ruffell B, Oei Y, Bissell MJ, Coussens LM, Pryer N, et al. CSF1R inhibition delays cervical and mammary tumor growth in murine models by attenuating the turnover of tumor-associated macrophages and enhancing infiltration by CD8(+) T cells. Oncoimmunology. 2013;2(12):12.
129. Autio KA, Klebanoff CA, Schaer D, Kauh JSW, Slovin SF, Adamow M, et al. Immunomodulatory Activity of a Colony-stimulating Factor-1 Receptor Inhibitor in Patients with Advanced Refractory Breast or Prostate Cancer: A Phase I Study. Clinical Cancer Research. 2020;26(21):5609-20.
130. Powles T, Paterson A, McCloskey E, Schein P, Scheffler B, Tidy A, et al. Reduction in bone relapse and improved survival with oral clodronate for adjuvant treatment of operable breast cancer ISRCTN83688026. Breast Cancer Research. 2006;8(2):7.
131. Junankar S, Shay G, Jurczyluk J, Ali N, Down J, Pocock N, et al. Real-Time Intravital Imaging Establishes Tumor-Associated Macrophages as the Extraskeletal Target of Bisphosphonate Action in Cancer. Cancer Discovery. 2015;5(1):35-42.
132. Coscia M, Quaglino E, Iezzi M, Curcio C, Pantaleoni F, Riganti C, et al. Zoledronic acid repolarizes tumour-associated macrophages and inhibits mammary carcinogenesis by targeting the mevalonate pathway. Journal of Cellular and Molecular Medicine. 2010;14(12):2803-15.
133. Powles T, Paterson S, Kanis JA, McCloskey E, Ashley S, Tidy A, et al. Randomized, placebo-controlled trial of clodronate in patients with primary operable breast cancer. Journal of Clinical Oncology. 2002;20(15):3219-24.
134. Diel IJ, Solomayer EF, Costa SD, Gollan C, Goerner R, Wallwiener D, et al. Reduction in new metastases in breast cancer with adjuvant clodronate treatment. New England Journal of Medicine. 1998;339(6):357-63.
135. Paterson AHG, Anderson SJ, Lembersky BC, Fehrenbacher L, Falkson CI, King KM, et al. Oral clodronate for adjuvant treatment of operable breast cancer (National Surgical Adjuvant Breast and Bowel Project protocol B-34): a multicentre, placebo-controlled, randomised trial. Lancet Oncology. 2012;13(7):734-42.
136. De Henau O, Rausch M, Winkler D, Campesato LF, Liu CL, Hirschhorn-Cymerman D, et al. Overcoming resistance to checkpoint blockade therapy by targeting PI3K gamma in myeloid cells. Nature. 2016;539(7629):443-7.
137. Kaneda MM, Messer KS, Ralainirina N, Li HY, Leem CJ, Gorjestani S, et al. PI3K gamma is a molecular switch that controls immune suppression. Nature. 2016;539(7629):437-42.
138. Guerriero JL, Sotayo A, Ponichtera HE, Castrillon JA, Pourzia AL, Schad S, et al. Class IIa HDAC inhibition reduces breast tumours and metastases through anti-tumour macrophages. Nature. 2017;543(7645):428-+.
139. Willingham SB, Volkmer JP, Gentles AJ, Sahoo D, Dalerba P, Mitra SS, et al. The CD47-signal regulatory protein alpha (SIRPa) interaction is a therapeutic target for human solid tumors. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(17):6662-7.
140. Upton R, Banuelos A, Feng DD, Biswas T, Kao K, McKenna K, et al. Combining CD47 blockade with trastuzumab eliminates HER2-positive breast cancer cells and

overcomes trastuzumab tolerance. Proceedings of the National Academy of Sciences of the United States of America. 2021;118(29):8.
141. Harney AS, Karagiannis GS, Pignatelli J, Smith BD, Kadioglu E, Wise SC, et al. The Selective Tie2 Inhibitor Rebastinib Blocks Recruitment and Function of Tie2(Hi) Macrophages in Breast Cancer and Pancreatic Neuroendocrine Tumors. Molecular Cancer Therapeutics. 2017;16(11):2486-501.
142. Karagiannis GS, Pastoriza JM, Wang YR, Harney AS, Entenberg D, Pignatelli J, et al. Neoadjuvant chemotherapy induces breast cancer metastasis through a TMEM-mediated mechanism. Science Translational Medicine. 2017;9(397):15.
143. Georgoudaki AM, Prokopec KE, Boura VF, Hellqvist E, Sohn S, Ostling J, et al. Reprogramming Tumor-Associated Macrophages by Antibody Targeting Inhibits Cancer Progression and Metastasis. Cell Reports. 2016;15(9):2000-11.
144. Gao J, Deng FS, Jia WD. Inhibition of Indoleamine 2,3-Dioxygenase Enhances the Therapeutic Efficacy of Immunogenic Chemotherapeutics in Breast Cancer. Journal of Breast Cancer. 2019;22(2):196-209.
145. Schmid P, Adams S, Rugo HS, Schneeweiss A, Barrios CH, Iwata H, et al. Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. New England Journal of Medicine. 2018;379(22):2108-21.
146. Schmid P, Rugo HS, Adams S, Schneeweiss A, Barrios CH, Iwata H, et al. Atezolizumab plus nab-paclitaxel as first-line treatment for unresectable, locally advanced or metastatic triple-negative breast cancer (IMpassion130): updated efficacy results from a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Oncology. 2020;21(1):44-59.
147. Schmid P, Adams S, Rugo HS, Schneeweiss A, Barrios CH, Iwata H, et al. IMpassion130: updated overall survival (OS) from a global, randomized, double-blind, placebo-controlled, Phase III study of atezolizumab (atezo) plus nab-paclitaxel (nP) in previously untreated locally advanced or metastatic triple-negative breast cancer (mTNBC). Journal of Clinical Oncology. 2019;37(15):2.
148. Miles D, Gligorov J, Andre F, Cameron D, Schneeweiss A, Barrios C, et al. Primary results from IMpassion131, a double-blind, placebo-controlled, randomised phase III trial of first-line paclitaxel with or without atezolizumab for unresectable locally advanced/metastatic triple-negative breast cancer. Annals of Oncology. 2021;32(8):994-1004.
149. Bonapace L, Coissieux MM, Wyckoff J, Mertz KD, Varga Z, Junt T, et al. Cessation of CCL2 inhibition accelerates breast cancer metastasis by promoting angiogenesis. Nature. 2014;515(7525):17.
150. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science. 2017;357(6349):409-13.
151. Lemery S, Keegan P, Pazdur R. First FDA Approval Agnostic of Cancer Site - When a Biomarker Defines the Indication. New England Journal of Medicine. 2017;377(15):1409-12.
152. Mills AM, Dill EA, Moskaluk CA, Dziegielewski J, Bullock TN, Dillon PM. The Relationship Between Mismatch Repair Deficiency and PD-L1 Expression in Breast Carcinoma. American Journal of Surgical Pathology. 2018;42(2):183-91.
153. Cortes J, Cescon DW, Rugo HS, Nowecki Z, Im SA, Yusof MM, et al. Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for previously untreated locally recurrent inoperable or metastatic triple-negative breast cancer (KEYNOTE-355): a randomised, placebo-controlled, double-blind, phase 3 clinical trial. Lancet. 2020;396(10265):1817-28.
154. Schmid P, Cortes J, Pusztai L, McArthur H, Kummel S, Bergh J, et al. Pembrolizumab for Early Triple-Negative Breast Cancer. New England Journal of Medicine. 2020;382(9):810-21.
155. Gabrilovich DI, Ostrand-Rosenberg S, Bronte V. Coordinated regulation of myeloid cells by tumours. Nat Rev Immunol. 2012;12(4):253-68.

156. Noy R, Pollard JW. Tumor-Associated Macrophages: From Mechanisms to Therapy. Immunity. 2014;41(1):49-61.
157. Engblom C, Pfirschke C, Pittet MJ. The role of myeloid cells in cancer therapies. Nature Reviews Cancer. 2016;16(7):447-62.
158. Chaib M, Chauhan SC, Makowski L. Friend or Foe? Recent Strategies to Target Myeloid Cells in Cancer. Frontiers in Cell and Developmental Biology. 2020;8:18.
159. Qiu SQ, Waaijer SJH, Zwager MC, de Vries EGE, van der Vegt B, Schroder CP. Tumor-associated macrophages in breast cancer: Innocent bystander or important player? Cancer Treatment Reviews. 2018;70:178-89.
160. Wynn TA, Chawla A, Pollard JW. Macrophage biology in development, homeostasis and disease. Nature. 2013;496(7446):445-55.
161. Pollard JW. Trophic macrophages in development and disease. Nature Reviews Immunology. 2009;9(4):259-70.
162. Varol C, Mildner A, Jung S. Macrophages: Development and Tissue Specialization. In: Littman DR, Yokoyama WM, editors. Annual Review of Immunology Vol 33. Annual Review of Immunology. 33. Palo Alto: Annual Reviews; 2015. p. 643-75.
163. Sica A, Mantovani A. Macrophage plasticity and polarization: in vivo veritas. Journal of Clinical Investigation. 2012;122(3):787-95.
164. Gordon S, Pluddemann A. Tissue macrophages: heterogeneity and functions. Bmc Biology. 2017;15:18.
165. Glass CK, Natoli G. Molecular control of activation and priming in macrophages. Nature Immunology. 2016;17(1):26-33.
166. Zindel J, Kubes P. DAMPs, PAMPs, and LAMPs in Immunity and Sterile Inflammation. In: Abbas AK, Aster JC, Feany MB, editors. Annual Review of Pathology: Mechanisms of Disease, Vol 15, 2020. Annual Review of Pathology-Mechanisms of Disease. 15. Palo Alto: Annual Reviews; 2020. p. 493-518.
167. Tang DL, Kang R, Coyne CB, Zeh HJ, Lotze MT. PAMPs and DAMPs: signal 0s that spur autophagy and immunity. Immunological Reviews. 2012;249:158-75.
168. Roh JS, Sohn DH. Damage-Associated Molecular Patterns in Inflammatory Diseases. Immune Network. 2018;18(4):14.
169. Patten DA, Shetty S. More Than Just a Removal Service: Scavenger Receptors in Leukocyte Trafficking. Frontiers in Immunology. 2018;9:16.
170. Canton J, Neculai D, Grinstein S. Scavenger receptors in homeostasis and immunity. Nature Reviews Immunology. 2013;13(9):621-34.
171. Chen YB, Song YC, Du W, Gong LL, Chang HC, Zou ZZ. Tumor-associated macrophages: an accomplice in solid tumor progression. Journal of Biomedical Science. 2019;26(1):13.
172. Mantovani A, Sica A, Sozzani S, Allavena P, Vecchi A, Locati M. The chemokine system in diverse forms of macrophage activation and polarization. Trends in Immunology. 2004;25(12):677-86.
173. Amit I, Winter DR, Jung S. The role of the local environment and epigenetics in shaping macrophage identity and their effect on tissue homeostasis. Nature Immunology. 2016;17(1):18-25.
174. Lavin Y, Winter D, Blecher-Gonen R, David E, Keren-Shaul H, Merad M, et al. Tissue-Resident Macrophage Enhancer Landscapes Are Shaped by the Local Microenvironment. Cell. 2014;159(6):1312-26.
175. Pollard JW. Tumour-educated macrophages promote tumour progression and metastasis. Nature Reviews Cancer. 2004;4(1):71-8.
176. Biswas SK, Mantovani A. Macrophage plasticity and interaction with lymphocyte subsets: cancer as a paradigm. Nature Immunology. 2010;11(10):889-96.
177. Gordon S. Alternative activation of macrophages. Nature Reviews Immunology. 2003;3(1):23-35.
178. Italiani P, Boraschi D. From monocytes to M1/M2 macrophages: phenotypical vs. functional differentiation. Frontiers in Immunology. 2014;5:22.

179. Mantovani A, Sozzani S, Locati M, Allavena P, Sica A. Macrophage polarization: tumor-associated macrophages as a paradigm for polarized M2 mononuclear phagocytes. Trends in Immunology. 2002;23(11):549-55.
180. Mills CD, Kincaid K, Alt JM, Heilman MJ, Hill AM. M-1/M-2 macrophages and the Th1/Th2 paradigm. Journal of Immunology. 2000;164(12):6166-73.
181. Martinez FO, Gordon S. The M1 and M2 paradigm of macrophage activation: time for reassessment. F1000Prime Reports. 2014;6(13).
182. Sica A, Bronte V. Altered macrophage differentiation and immune dysfunction in tumor development. Journal of Clinical Investigation. 2007;117(5):1155-66.
183. Mantovani A, Bottazzi B, Colotta F, Sozzani S, Ruco L. THE ORIGIN AND FUNCTION OF TUMOR-ASSOCIATED MACROPHAGES. Immunology Today. 1992;13(7):265-70.
184. Liu Y, Cao XT. The origin and function of tumor-associated macrophages. Cellular & Molecular Immunology. 2015;12(1):4.
185. Epelman S, Lavine KJ, Randolph GJ. Origin and Functions of Tissue Macrophages. Immunity. 2014;41(1):21-35.
186. Perdiguero EG, Klapproth K, Schulz C, Busch K, Azzoni E, Crozet L, et al. Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors. Nature. 2015;518(7540):547-51.
187. Hoeffel G, Chen JM, Lavin Y, Low D, Almeida FF, See P, et al. C-Myb(+) Erythro-Myeloid Progenitor-Derived Fetal Monocytes Give Rise to Adult Tissue-Resident Macrophages. Immunity. 2015;42(4):665-78.
188. Franklin RA, Liao W, Sarkar A, Kim MV, Bivona MR, Liu K, et al. The cellular and molecular origin of tumor-associated macrophages. Science. 2014;344(6186):921-5.
189. Hashimoto D, Chow A, Noizat C, Teo P, Beasley MB, Leboeuf M, et al. Tissue-Resident Macrophages Self-Maintain Locally throughout Adult Life with Minimal Contribution from Circulating Monocytes. Immunity. 2013;38(4):792-804.
190. Kowal J, Kornete M, Joyce JA. Re-education of macrophages as a therapeutic strategy in cancer. Immunotherapy. 2019;11(8):677-89.
191. Suva ML, Tirosh I. Single-Cell RNA Sequencing in Cancer: Lessons Learned and Emerging Challenges. Mol Cell. 2019;75(1):7-12.
192. Ojalvo LS, King W, Cox D, Pollard JW. High-Density Gene Expression Analysis of Tumor-Associated Macrophages from Mouse Mammary Tumors. American Journal of Pathology. 2009;174(3):1048-64.
193. Pucci F, Venneri MA, Biziato D, Nonis A, Moi D, Sica A, et al. A distinguishing gene signature shared by tumor-infiltrating Tie2-expressing monocytes, blood "resident" monocytes, and embryonic macrophages suggests common functions and developmental relationships. Blood. 2009;114(4):901-14.
194. Lin YX, Xu JX, Lan HY. Tumor-associated macrophages in tumor metastasis: biological roles and clinical therapeutic applications. Journal of Hematology & Oncology. 2019;12:16.
195. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. Nature Reviews Genetics. 2019;20(11):631-56.
196. Lahnemann D, Koster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, et al. Eleven grand challenges in single-cell data science. Genome Biology. 2020;21(1):35.
197. Chung W, Eum HH, Lee HO, Lee KM, Lee HB, Kim KT, et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. Nat Commun. 2017;8.
198. An YJ, Adams JR, Hollern DP, Zhao A, Chang SG, Gams MS, et al. Cdh1 and Pik3ca Mutations Cooperate to Induce Immune-Related Invasive Lobular Carcinoma of the Breast. Cell Reports. 2018;25(3):702-+.
199. Cheng SJ, Li ZY, Gao RR, Xing BC, Gao YN, Yang Y, et al. A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. Cell. 2021;184(3):792-+.

200. Janiszewska M, Tabassum DP, Castano Z, Cristea S, Yamamoto KN, Kingston NL, et al. Subclonal cooperation drives metastasis by modulating local and systemic immune microenvironments. Nat Cell Biol. 2019;21(7):879-+.
201. Keren L, Bosse M, Marquez D, Angoshtari R, Jain S, Varma S, et al. A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. Cell. 2018;174(6):1373-+.
202. Marques O, Porto G, Rema A, Faria F, Paula AC, Gomez-Lazaro M, et al. Local iron homeostasis in the breast ductal carcinoma microenvironment. BMC Cancer. 2016;16:14.
203. McGinnis CS, Patterson DM, Winkler J, Conrad DN, Hein MY, Srivastava V, et al. MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. Nat Methods. 2019;16(7):619-+.
204. Opzoomer JW, Anstee JE, Dean I, Hill EJ, Bouybayoune I, Caron J, et al. Macrophages orchestrate the expansion of a proangiogenic perivascular niche during cancer progression. Science Advances. 2021;7(45):15.
205. Qiu S, Hong R, Zhuang Z, Li Y, Zhu L, Lin X, et al. A Single-Cell Immune Atlas of Triple Negative Breast Cancer Reveals Novel Immune Cell Subsets. BioRxiv; 2019.
206. Schapiro D, Jackson HW, Raghuraman S, Fischer JR, Zanotelli VRT, Schulz D, et al. histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. Nat Methods. 2017;14(9):873-+.
207. Szczerba BM, Castro-Giner F, Vetter M, Krol I, Gkountela S, Landin J, et al. Neutrophils escort circulating tumour cells to enable cell cycle progression. Nature. 2019;566(7745):553-+.
208. Muller S, Kohanbash G, Liu SJ, Alvarado B, Carrera D, Bhaduri A, et al. Single-cell profiling of human gliomas reveals macrophage ontogeny as a basis for regional differences in macrophage activation in the tumor microenvironment. Genome Biol. 2017;18:14.
209. Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O, et al. Phenotype molding of stromal cells in the lung tumor microenvironment. Nature Medicine. 2018;24(8):1277-+.
210. Gatti-Mays ME, Balko JM, Gameiro SR, Bear HD, Prabhakaran S, Fukui J, et al. If we build it they will come: targeting the immune response to breast cancer. Npj Breast Cancer. 2019;5:13.
211. Frantz C, Stewart KM, Weaver VM. The extracellular matrix at a glance. Journal of Cell Science. 2010;123(24):4195-200.
212. Hynes RO, Naba A. Overview of the Matrisome-An Inventory of Extracellular Matrix Constituents and Functions. Cold Spring Harbor Perspectives in Biology. 2012;4(1):16.
213. Hynes RO. The evolution of metazoan extracellular matrix. Journal of Cell Biology. 2012;196(6):671-9.
214. Doonan JH, Sablowski R. Walls around tumours-why plants do not develop cancer. Nature Reviews Cancer. 2010;10(11):793-802.
215. Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer. Embo Reports. 2014;15(12):1243-53.
216. Provenzano PP, Inman DR, Eliceiri KW, Knittel JG, Yan L, Rueden CT, et al. Collagen density promotes mammary tumor initiation and progression. Bmc Medicine. 2008;6:15.
217. Levental KR, Yu HM, Kass L, Lakins JN, Egeblad M, Erler JT, et al. Matrix Crosslinking Forces Tumor Progression by Enhancing Integrin Signaling. Cell. 2009;139(5):891-906.
218. Boghaert E, Gleghorn JP, Lee K, Gjorevski N, Radisky DC, Nelson CM. Host epithelial geometry regulates breast cancer cell invasiveness. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(48):19632-7.
219. Boyd NF, Guo H, Martin LJ, Sun LM, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. New England Journal of Medicine. 2007;356(3):227-36.

220. Acerbi I, Cassereau L, Dean I, Shi Q, Au A, Park C, et al. Human breast cancer invasion and aggression correlates with ECM stiffening and immune cell infiltration. Integrative Biology. 2015;7(10):1120-34.
221. Chang JM, Park IA, Lee SH, Kim WH, Bae MS, Koo HR, et al. Stiffness of tumours measured by shear-wave elastography correlated with subtypes of breast cancer. European Radiology. 2013;23(9):2450-8.
222. Wiseman BS, Werb Z. Development - Stromal effects on mammary gland development and breast cancer. Science. 2002;296(5570):1046-9.
223. Robinson GW, Karpf ABC, Kratochwil K. Regulation of mammary gland development by tissue interaction. Journal of Mammary Gland Biology and Neoplasia. 1999;4(1):9-19.
224. Kaushik S, Pickup MW, Weaver VM. From transformation to metastasis: deconstructing the extracellular matrix in breast cancer. Cancer and Metastasis Reviews. 2016;35(4):655-67.
225. Stramer BM, Mori R, Martin P. The inflammation-fibrosis link? A Jekyll and Hyde role for blood cells during wound repair. Journal of Investigative Dermatology. 2007;127(5):1009-17.
226. Sorokin L. The impact of the extracellular matrix on inflammation. Nature Reviews Immunology. 2010;10(10):712-23.
227. Houghton AM, Quintero PA, Perkins DL, Kobayashi DK, Kelley DG, Marconcini LA, et al. Elastin fragments drive disease progression in a murine model of emphysema. Journal of Clinical Investigation. 2006;116(3):753-9.
228. Hynes RO. The Extracellular Matrix: Not Just Pretty Fibrils. Science. 2009;326(5957):1216-9.
229. Peng DH, Rodriguez BL, Diao LX, Chen LM, Wang J, Byers LA, et al. Collagen promotes anti-PD-1/PD-L1 resistance in cancer through LAIR1-dependent CD8(+) T cell exhaustion. Nature Communications. 2020;11(1):18.
230. Chiquet-Ehrismann R, Orend G, Chiquet M, Tucker RP, Midwood KS. Tenascins in stem cell niches. Matrix Biol. 2014;37:112-23.
231. Midwood KS, Orend G. The role of tenascin-C in tissue injury and tumorigenesis. J Cell Commun Signal. 2009;3(3-4):287-310.
232. Midwood KS, Chiquet M, Tucker RP, Orend G. Tenascin-C at a glance. Journal of Cell Science. 2016;129(23):4321-7.
233. Midwood KS, Hussenet T, Langlois B, Orend G. Advances in tenascin-C biology. Cellular and Molecular Life Sciences. 2011;68(19):3175-99.
234. Orend G. Potential oncogenic action of tenascin-C in tumorigenesis. International Journal of Biochemistry & Cell Biology. 2005;37(5):1066-83.
235. Orend G, Tucker RP. Did Tenascin-C Co-Evolve With the General Immune System of Vertebrates? Frontiers in Immunology. 2021;12:6.
236. Marzeda AM, Midwood KS. Internal Affairs: Tenascin-C as a Clinically Relevant, Endogenous Driver of Innate Immunity. Journal of Histochemistry & Cytochemistry. 2018;66(4):289-304.
237. Midwood K, Sacre S, Piccinini AM, Inglis J, Trebaul A, Chan E, et al. Tenascin-C is an endogenous activator of Toll-like receptor 4 that is essential for maintaining inflammation in arthritic joint disease. Nat Med. 2009;15(7):774-U11.
238. Piccinini AM, Zuliani-Alvarez L, Lim JMP, Midwood KS. Distinct microenvironmental cues stimulate divergent TLR4-mediated signaling pathways in macrophages. Science Signaling. 2016;9(443):16.
239. Orend G, Chiquet-Ehrismann R. Tenascin-C induced signaling in cancer. Cancer Letters. 2006;244(2):143-63.
240. Goh FG, Piccinini AM, Krausgruber T, Udalova IA, Midwood KS. Transcriptional Regulation of the Endogenous Danger Signal Tenascin-C: A Novel Autocrine Loop in Inflammation. Journal of Immunology. 2010;184(5):2655-62.
241. Primac I, Maquoi E, Blacher S, Heljasvaara R, Van Deun J, Smeland HYH, et al. Stromal integrin alpha 11 regulates PDGFR beta signaling and promotes breast cancer progression. Journal of Clinical Investigation. 2019;129(11):4509-28.

242. Giblin SP, Midwood KS. Tenascin-C: Form versus function. Cell Adhesion & Migration. 2015;9(1-2):48-82.
243. Mackie EJ, Chiquetehrismann R, Pearson CA, Inaguma Y, Taya K, Kawarada Y, et al. TENASCIN IS A STROMAL MARKER FOR EPITHELIAL MALIGNANCY IN THE MAMMARY-GLAND. Proceedings of the National Academy of Sciences of the United States of America. 1987;84(13):4621-5.
244. Chiquetehrismann R, Mackie EJ, Pearson CA, Sakakura T. TENASCIN - AN EXTRACELLULAR-MATRIX PROTEIN INVOLVED IN TISSUE INTERACTIONS DURING FETAL DEVELOPMENT AND ONCOGENESIS. Cell. 1986;47(1):131-9.
245. Howeedy AA, Virtanen I, Laitinen L, Gould NS, Koukoulis GK, Gould VE. DIFFERENTIAL DISTRIBUTION OF TENASCIN IN THE NORMAL, HYPERPLASTIC, AND NEOPLASTIC BREAST. Laboratory Investigation. 1990;63(6):798-806.
246. Borsi L, Carnemolla B, Nicolo G, Spina B, Tanara G, Zardi L. EXPRESSION OF DIFFERENT TENASCIN ISOFORMS IN NORMAL, HYPERPLASTIC AND NEOPLASTIC HUMAN BREAST TISSUES. International Journal of Cancer. 1992;52(5):688-92.
247. Guttery DS, Hancox RA, Mulligan KT, Hughes S, Lambe SM, Pringle JH, et al. Association of invasion-promoting tenascin-C additional domains with breast cancers in young women. Breast Cancer Research. 2010;12(4):14.
248. Hancox RA, Allen MD, Holliday DL, Edwards DR, Pennington CJ, Guttery DS, et al. Tumour-associated tenascin-C isoforms promote breast cancer cell invasion and growth by matrix metalloproteinase-dependent and independent mechanisms. Breast Cancer Research. 2009;11(2):13.
249. Adams M, Jones JL, Walker RA, Pringle JH, Bell SC. Changes in tenascin-C isoform expression in invasive and preinvasive breast disease. Cancer Research. 2002;62(11):3289-+.
250. Oskarsson T, Acharyya S, Zhang XHF, Vanharanta S, Tavazoie SF, Morris PG, et al. Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs. Nature Medicine. 2011;17(7):867-U256.
251. Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu WP, Giri DD, et al. Genes that mediate breast cancer metastasis to lung. Nature. 2005;436(7050):518-24.
252. Tavazoie SF, Alarcon C, Oskarsson T, Padua D, Wang QQ, Bos PD, et al. Endogenous human microRNAs that suppress breast cancer metastasis. Nature. 2008;451(7175):147-U3.
253. Gurbuz I, Ferralli J, Roloff T, Chiquet-Ehrismann R, Asparuhova MB. SAP domain-dependent Mkl1 signaling stimulates proliferation and cell migration by induction of a distinct gene set indicative of poor prognosis in breast cancer patients. Molecular Cancer. 2014;13:16.
254. Ioachim E, Charchanti A, Briasoulis E, Karavasilis V, Tsanou H, Arvanitis DL, et al. Immunohistochemical expression of extracellular matrix components tenascin, fibronectin, collagen type IV and laminin in breast cancer: their prognostic value and role in tumour invasion and progression. European Journal of Cancer. 2002;38(18):2362-70.
255. Dandachi N, Hauser-Kronberger C, More E, Wiesener B, Hacker GW, Dietze O, et al. Co-expression of tenascin-C and vimentin in human breast cancer cells indicates phenotypic transdifferentiation during tumour progression: correlation with histopathological parameters, hormone receptors, and oncoproteins. Journal of Pathology. 2001;193(2):181-9.
256. Jahkola T, Toivonen T, Virtanen I, von Smitten K, Nordling S, von Boguslawski K, et al. Tenascin-C expression in invasion border of early breast cancer: a predictor of local and distant recurrence. British Journal of Cancer. 1998;78(11):1507-13.
257. Ishihara A, Yoshida T, Tamaki H, Sakakura T. Tenascin expression in cancer cells and stroma of human breast cancer and its prognostic significance. Clinical Cancer Research. 1995;1(9):1035-41.
258. Sun Z, Velazquez-Ouesada I, Murdamoothoo D, Ahowesso C, Yilmaz A, Spenle C, et al. Tenascin-C increases lung metastasis by impacting blood vessel invasions. Matrix Biol. 2019;83:26-47.

259. Saupe F, Schwenzer A, Jia YD, Gasser I, Spenle C, Langlois B, et al. Tenascin-C Downregulates Wnt Inhibitor Dickkopf-1, Promoting Tumorigenesis in a Neuroendocrine Tumor Model. Cell Reports. 2013;5(2):482-92.
260. Lowy CM, Oskarsson T. Tenascin C in metastasis: A view from the invasive front. Cell Adhesion & Migration. 2015;9(1-2):112-24.
261. Pollard JW. Macrophages define the invasive microenvironment in breast cancer. Journal of Leukocyte Biology. 2008;84(3):623-30.
262. Kaplan G. INVITRO DIFFERENTIATION OF HUMAN-MONOCYTES - MONOCYTES CULTURED ON GLASS ARE CYTO-TOXIC TO TUMOR-CELLS BUT MONOCYTES CULTURED ON COLLAGEN ARE NOT. Journal of Experimental Medicine. 1983;157(6):2061-72.
263. Deligne C, Midwood KS. Macrophages and Extracellular Matrix in Breast Cancer: Partners in Crime or Protective Allies? Frontiers in Oncology. 2021;11:12.
264. Deligne C, Murdamoothoo D, Gammage AN, Gschwandtner M, Erne W, Loustau T, et al. Matrix-Targeting Immunotherapy Controls Tumor Growth and Spread by Switching Macrophage Phenotype. Cancer Immunology Research. 2020;8(3):368-82.
265. Arpel A, Sawma P, Spenle C, Fritz J, Meyer L, Garnier N, et al. Transmembrane Domain Targeting Peptide Antagonizing ErbB2/Neu Inhibits Breast Tumor Growth and Metastasis. Cell Reports. 2014;8(6):1714-21.
266. Murdamoothoo D, Sun Z, Yilmaz A, Riegel G, Abou-Faycal C, Deligne C, et al. Tenascin-C immobilizes infiltrating T lymphocytes through CXCL12 promoting breast cancer progression. Embo Molecular Medicine. 2021;13(6):21.
267. Fantozzi A, Christofori G. Mouse models of breast cancer metastasis. Breast Cancer Research. 2006;8(4):11.
268. Sims D, Ilott NE, Sansom SN, Sudbery IM, Johnson JS, Fawcett KA, et al. CGAT: computational genomics analysis toolkit. Bioinformatics. 2014;30(9):1290-1.
269. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019;177(7):1888-+.
270. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, et al. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nature Biotechnology. 2015;33(2):155-60.
271. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 2019;20(1):15.
272. Chung NC, Storey JD. Statistical significance of variables driving systematic variation in high-dimensional data. Bioinformatics. 2015;31(4):545-54.
273. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015;161(5):1202-14.
274. van der Maaten L. Accelerating t-SNE using Tree-Based Algorithms. Journal of Machine Learning Research. 2014;15:3221-45.
275. van der Maaten L, Hinton G. Visualizing Data using t-SNE. Journal of Machine Learning Research. 2008;9:2579-605.
276. McInnes L, Healy J, Saul N, Grossberger L. UMAP: Uniform Manifold Approximation and Projection. The Journal of Open Source Software; 2018. p. 861.
277. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2019;37(1):38-+.
278. Zappia L, Oshlack A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. Gigascience. 2018;7(7):9.
279. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(43):15545-50.
280. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. Cell Systems. 2015;1(6):417-25.

281. Haghverdi L, Buettner F, Theis FJ. Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics. 2015;31(18):2989-98.
282. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. Bmc Genomics. 2018;19:16.
283. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. Methods. 2001;25(4):402-8.
284. Tweedie S, Braschi B, Gray K, Jones TEM, Seal RL, Yates B, et al. Genenames.org: the HGNC and VGNC resources in 2021. Nucleic Acids Research. 2021;49(D1):D939-D46.
285. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nature Protocols. 2009;4(8):1184-91.
286. Agapite J, Albou LP, Aleksander S, Argasinska J, Arnaboldi V, Attrill H, et al. Alliance of Genome Resources Portal: unified model organism research platform. Nucleic Acids Research. 2020;48(D1):D650-D8.
287. Sean D, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. Bioinformatics. 2007;23(14):1846-7.
288. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Research. 2016;44(8):11.
289. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. Bmc Bioinformatics. 2011;12:16.
290. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012;486(7403):346-52.
291. Cerami E, Gao JJ, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery. 2012;2(5):401-4.
292. Behan FM, Iorio F, Picco G, Goncalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;568(7753):511-+.
293. Dhawan A, Barberis A, Cheng WC, Domingo E, West C, Maughan T, et al. Guidelines for using sigQC for systematic evaluation of gene signatures. Nature Protocols. 2019;14(5):1377-400.
294. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. Bmc Bioinformatics. 2013;14:15.
295. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nature Biotechnology. 2019;37(7):773-+.
296. Gu ZG, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32(18):2847-9.
297. Bellman R. Dynamic programming. Princeton, NJ: Princeton University Press; 1957.
298. Brennecke P, Anders S, Kim JK, Kolodziejczyk AA, Zhang XW, Proserpio V, et al. Accounting for technical noise in single-cell RNA-seq experiments. Nat Methods. 2013;10(11):1093-5.
299. Fix E, Hodges JL. Discriminatory Analysis - Nonparametric Discrimination - Consistency Properties. International Statistical Review. 1989;57(3):238-47.
300. Altman NS. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. American Statistician. 1992;46(3):175-85.
301. Jarvis RA, Patrick EA. Clustering Using A Similarity Measure Based On Shared Near Neighbors. Ieee Transactions on Computers. 1973;C-22(11):1025-34.
302. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech-Theory Exp. 2008:12.
303. Waltman L, van Eck NJ. A smart local moving algorithm for large-scale modularity-based community detection. Eur Phys J B. 2013;86(11):14.

304. Gabrilovich DI, Velders MP, Sotomayor EM, Kast WM. Mechanism of immune dysfunction in cancer mediated by immature Gr-1(+) myeloid cells. J Immunol. 2001;166(9):5398-406.
305. Kelly LM, Englmeier U, Lafon I, Sieweke MH, Graf T. MafB is an inducer of monocytic differentiation. Embo J. 2000;19(9):1987-97.
306. Cuevas VD, Anta L, Samaniego R, Orta-Zavalza E, de la Rosa JV, Baujat G, et al. MAFB Determines Human Macrophage Anti- Inflammatory Polarization: Relevance for the Pathogenic Mechanisms Operating in Multicentric Carpotarsal Osteolysis. J Immunol. 2017;198(5):2070-81.
307. Acharyya S, Oskarsson T, Vanharanta S, Malladi S, Kim J, Morris PG, et al. A CXCL1 Paracrine Network Links Cancer Chemoresistance and Metastasis. Cell. 2012;150(1):165-78.
308. Wang N, Liu WP, Zheng YF, Wang SQ, Yang BW, Li M, et al. CXCL1 derived from tumor-associated macrophages promotes breast cancer metastasis via activating NF-kappa B/SOX4 signaling. Cell Death & Disease. 2018;9:18.
309. Kim MY, Oskarsson T, Acharyya S, Nguyen DX, Zhang XHF, Norton L, et al. Tumor Self-Seeding by Circulating Cancer Cells. Cell. 2009;139(7):1315-26.
310. Kitamura T, Qian BZ, Soong D, Cassetta L, Noy R, Sugano G, et al. CCL2-induced chemokine cascade promotes breast cancer metastasis by enhancing retention of metastasis-associated macrophages. Journal of Experimental Medicine. 2015;212(7):1043-59.
311. Sasaki S, Baba T, Nishimura T, Hayakawa Y, Hashimoto S, Gotoh N, et al. Essential roles of the interaction between cancer cell-derived chemokine, CCL4, and intra-bone CCR5-expressing fibroblasts in breast cancer bone metastasis. Cancer Letters. 2016;378(1):23-32.
312. Yuk JM, Kim TS, Kim SY, Lee HM, Han J, Dufour CR, et al. Orphan Nuclear Receptor ERR alpha Controls Macrophage Metabolic Signaling and A20 Expression to Negatively Regulate TLR-Induced Inflammation. Immunity. 2015;43(1):80-91.
313. Horber S, Hildebrand DG, Lieb WS, Lorscheid S, Hailfinger S, Schulze-Osthoff K, et al. The Atypical Inhibitor of NF-kappa B, I kappa B zeta, Controls Macrophage Interleukin-10 Expression. Journal of Biological Chemistry. 2016;291(24):12851-61.
314. Kryczek I, Banerjee M, Cheng P, Vatan L, Szeliga W, Wei S, et al. Phenotype, distribution, generation, and functional and clinical relevance of Th17 cells in the human tumor environments. Blood. 2009;114(6):1141-9.
315. Zhao HL, Chen Q, Alam A, Cui J, Suen KC, Soo AP, et al. The role of osteopontin in the progression of solid organ tumour. Cell Death Dis. 2018;9:15.
316. Bresnick AR, Weber DJ, Zimmer DB. S100 proteins in cancer. Nat Rev Cancer. 2015;15(2):96-109.
317. Klingelhofer J, Grum-Schwensen B, Beck MK, Knudsen RSP, Grigorian M, Lukanidin E, et al. Anti-S100A4 Antibody Suppresses Metastasis Formation by Blocking Stroma Cell Invasion. Neoplasia. 2012;14(12):1260-+.
318. Martinet L, Filleron T, Le Guellec S, Rochaix P, Garrido I, Girard JP. High Endothelial Venule Blood Vessels for Tumor-Infiltrating Lymphocytes Are Associated with Lymphotoxin beta-Producing Dendritic Cells in Human Breast Cancer. J Immunol. 2013;191(4):2001-8.
319. Verjans E, Noetzel E, Bektas N, Schutz AK, Lue HQ, Lennartz B, et al. Dual role of macrophage migration inhibitory factor (MIF) in human breast cancer. BMC Cancer. 2009;9:18.
320. Pang YL, Gara SK, Achyut BR, Li ZY, Yan HH, Day CP, et al. TGF-beta Signaling in Myeloid Cells Is Required for Tumor Metastasis. Cancer Discov. 2013;3(8):936-51.
321. Chung B, Esmaeili AA, Gopalakrishna-Pillai S, Murad JP, Andersen ES, Reddy NK, et al. Human brain metastatic stroma attracts breast cancer cells via chemokines CXCL16 and CXCL12. npj Breast Cancer. 2017;3:9.
322. Rabe DC, Frankenberger C, Bainer R, Sankarasharma D, Chada K, Krausz T, et al. Metastasis suppressors regulate the tumor microenvironment by blocking recruitment of pro-metastatic tumor-associated macrophages. Cancer Res. 2016;76:3.

323. Baitsch D, Bock HH, Engel T, Telgmann R, Muller-Tidow C, Varga G, et al. Apolipoprotein E Induces Antiinflammatory Phenotype in Macrophages. Arterioscler Thromb Vasc Biol. 2011;31(5):1160-U577.
324. Rashidi M, Bandala-Sanchez E, Lawlor KE, Zhang YX, Neale AM, Vijayaraj SL, et al. CD52 inhibits Toll-like receptor activation of NF-kappa B and triggers apoptosis to suppress inflammation. Cell Death Differ. 2018;25(2):392-405.
325. Miyasato Y, Shiota T, Ohnishi K, Pan C, Yano H, Horlad H, et al. High density of CD204-positive macrophages predicts worse clinical prognosis in patients with breast cancer. Cancer Sci. 2017;108(8):1693-700.
326. Wang ZX, Yin N, Zhang ZX, Zhang Y, Zhang GB, Chen WC. Upregulation of T-cell Immunoglobulin and Mucin-Domain Containing-3 (Tim-3) in Monocytes/Macrophages Associates with Gastric Cancer Progression. Immunol Invest. 2017;46(2):134-48.
327. Gao JJ, Ward JF, Pettaway CA, Shi LZ, Subudhi SK, Vence LM, et al. VISTA is an inhibitory immune checkpoint that is increased after ipilimumab therapy in patients with prostate cancer. Nat Med. 2017;23(5):551-+.
328. Qi YY, Chang Y, Wang ZW, Chen LL, Kong YY, Zhang PP, et al. Tumor-associated macrophages expressing galectin-9 identify immunoevasive subtype muscle-invasive bladder cancer with poor prognosis but favorable adjuvant chemotherapeutic response. Cancer Immunol Immunother.14.
329. Tanay A, Regev A. Scaling single-cell genomics from phenomenology to mechanism. Nature. 2017;541(7637):331-8.
330. Tritschler S, Buttner M, Fischer DS, Lange M, Bergen V, Lickert H, et al. Concepts and limitations for learning developmental trajectories from single cell genomics. Development. 2019;146(12):12.
331. Cassetta L, Noy R, Swierczak A, Sugano G, Smith H, Wiechmann L, et al. Isolation of Mouse and Human Tumor-Associated Macrophages. In: Koumenis C, Coussens LM, Giaccia A, Hammond E, editors. Tumor Microenvironment: Study Protocols. Advances in Experimental Medicine and Biology. 899. Cham: Springer International Publishing Ag; 2016. p. 211-29.
332. Chavez-Galan L, Olleros ML, Vesin D, Garcia I. Much more than M1 and M2 macrophages, there are also CD169(+) and TCR+ macrophages. Front Immunol. 2015;6:15.
333. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018;36(5):411-+.
334. 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(12):1289-+.
335. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8(1):118-27.
336. 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(5):421-+.
337. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. Nat Methods. 2018;15(12):1053-+.
338. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet. 2012;13(4):227-32.
339. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol. 2019;15(6):23.
340. Catalan V, Gomez-Ambrosi J, Rodriguez A, Ramirez B, Rotellar F, Valenti V, et al. Increased Tenascin C And Toll-Like Receptor 4 Levels in Visceral Adipose Tissue as a Link between Inflammation and Extracellular Matrix Remodeling in Obesity. J Clin Endocrinol Metab. 2012;97(10):E1880-E9.
341. Dong XR, Hao Y, Wang X, Tian WD. LEGO: a novel method for gene set over-representation analysis by incorporating network-based gene weights. Sci Rep. 2016;6:17.

342. Yoon S, Kim J, Kim SK, Baik B, Chi SM, Kim SY, et al. GScluster: network-weighted gene-set clustering analysis. BMC Genomics. 2019;20:14.
343. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. Nat Methods. 2020;17(2):159-62.
344. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat Protoc. 2020;15(4):1484-506.
345. Tanaka A, To J, O'Brien B, Donnelly S, Lund M. Selection of reliable reference genes for the normalisation of gene expression levels following time course LPS stimulation of murine bone marrow derived macrophages. BMC Immunol. 2017;18:12.
346. Saelens W, Cannoodt R, Todorov H, Saeys Y. A comparison of single-cell trajectory inference methods. Nat Biotechnol. 2019;37(5):547-54.
347. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Gottgens B, et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol. 2019;20:9.
348. Griffiths JA, Scialdone A, Marioni JC. Using single-cell genomics to understand developmental processes and cell fate decisions. Mol Syst Biol. 2018;14(4):12.
349. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng WG, Kim D, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nature Medicine. 2015;21(8):938-45.
350. Qiu S, Hong R, Zhuang Z, Li Y, Zhu L, Lin X, et al. A Single-Cell Immune Atlas of Triple Negative Breast Cancer Reveals Novel Immune Cell Subsets. BioRxiv; 2019.
351. Poczobutt JM, De S, Yadav VK, Nguyen TT, Li H, Sippel TR, et al. Expression Profiling of Macrophages Reveals Multiple Populations with Distinct Biological Roles in an Immunocompetent Orthotopic Model of Lung Cancer. Journal of Immunology. 2016;196(6):2847-59.
352. Kumar V, Cheng PY, Condamine T, Mony S, Languino LR, McCaffrey JC, et al. CD45 Phosphatase Inhibits STAT3 Transcription Factor Activity in Myeloid Cells and Promotes Tumor-Associated Macrophage Differentiation. Immunity. 2016;44(2):303-15.
353. Zuo H, Yang DB, Yang QW, Tang HD, Fu YX, Wan YH. Differential regulation of breast cancer bone metastasis by PARP1 and PARP2. Nature Communications. 2020;11(1):10.
354. Cheng PY, Corzo CA, Luetteke N, Yu B, Nagaraj S, Bui MM, et al. Inhibition of dendritic cell differentiation and accumulation of myeloid-derived suppressor cells in cancer is regulated by S100A9 protein. Journal of Experimental Medicine. 2008;205(10):2235-49.
355. Goncalves A, Charafe-Jauffret E, Bertucci F, Audebert S, Toiron Y, Esterni B, et al. Protein profiling of human breast tumor cells identifies novel biomarkers associated with molecular subtypes. Molecular & Cellular Proteomics. 2008;7(8):1420-33.
356. Liu Y, Kosaka A, Ikeura M, Kohanbash G, Fellows-Mayle W, Snyder LA, et al. Premetastatic soil and prevention of breast cancer brain metastasis. Neuro-Oncology. 2013;15(7):891-903.
357. Mahley RW. Apolipoprotein E: from cardiovascular disease to neurodegenerative disorders. Journal of Molecular Medicine-Jmm. 2016;94(7):739-46.
358. Yin CJ, Ackermann S, Ma Z, Mohanta SK, Zhang CK, Li YF, et al. ApoE attenuates unresolvable inflammation by complex formation with activated C1q. Nature Medicine. 2019;25(3):496-+.
359. Caron M, St-Onge P, Sontag T, Wang YC, Richer C, Ragoussis I, et al. Single-cell analysis of childhood leukemia reveals a link between developmental states and ribosomal protein expression as a source of intra-individual heterogeneity. Scientific Reports. 2020;10(1):12.
360. Athanasiadis EI, Botthof JG, Andres H, Ferreira L, Lio P, Cvejic A. Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. Nature Communications. 2017;8:11.
361. Warner JR. The economics of ribosome biosynthesis in yeast. Trends in Biochemical Sciences. 1999;24(11):437-40.

362. Garcion E, Halilagic A, Faissner A, ffrench-Constant C. Generation of an environmental niche for neural stem cell development by the extracellular matrix molecule tenascin C. Development. 2004;131(14):3423-32.
363. Hongu T, Pein M, Insua-Rodriguez J, Meier J, Decker K, Descot A, et al. Perivascular tenascin C triggers sequential activations of macrophages and endothelial cells to generate a pro-metastatic vascular niche in the lungs. bioRxiv; 2021.
364. Kobayashi M, Oshima S, Maeyashiki C, Nibe Y, Otsubo K, Matsuzawa Y, et al. The ubiquitin hybrid gene UBA52 regulates ubiquitination of ribosome and sustains embryonic development. Scientific Reports. 2016;6:11.
365. Spenle C, Gasser I, Saupe F, Janssen KP, Arnold C, Klein A, et al. Spatial organization of the tenascin-C microenvironment in experimental and human cancer. Cell Adhesion & Migration. 2015;9(1-2):4-13.
366. Degen M, Brellier F, Kain R, Ruiz C, Terracciano L, Orend G, et al. Tenascin-W is a novel marker for activated tumor stroma in low-grade human breast cancer and influences cell behavior. Cancer Research. 2007;67(19):9169-79.
367. Klose RJ, Kallin EM, Zhang Y. JmjC-domain-containing proteins and histone demethylation. Nature Reviews Genetics. 2006;7(9):715-27.
368. Kruidenier L, Chung CW, Cheng ZJ, Liddle J, Che KH, Joberty G, et al. A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. Nature. 2012;488(7411):404-8.
369. Lochmann TL, Powell KM, Ham J, Floros KV, Heisey DAR, Kurupi RIJ, et al. Targeted inhibition of histone H3K27 demethylation is effective in high-risk neuroblastoma. Science Translational Medicine. 2018;10(441):12.
370. Rath BH, Waung I, Camphausen K, Tofilon PJ. Inhibition of the Histone H3K27 Demethylase UTX Enhances Tumor Cell Radiosensitivity. Molecular Cancer Therapeutics. 2018;17(5):1070-8.
371. Li YN, Zhang MY, Sheng MY, Zhang P, Chen ZZ, Xing W, et al. Therapeutic potential of GSK-J4, a histone demethylase KDM6B/JMJD3 inhibitor, for acute myeloid leukemia. Journal of Cancer Research and Clinical Oncology. 2018;144(6):1065-77.
372. Donas C, Neira J, Osorio-Barrios F, Carrasco M, Fernandez D, Prado C, et al. The demethylase inhibitor GSK-J4 limits inflammatory colitis by promoting de novo synthesis of retinoic acid in dendritic cells. Scientific Reports. 2021;11(1):14.
373. Pan YC, Wang JL, Xue YX, Zhao JJ, Li D, Zhang SL, et al. GSKJ4 Protects Mice Against Early Sepsis via Reducing Proinflammatory Factors and Up-Regulating MiR-146a. Frontiers in Immunology. 2018;9:11.
374. Rodriguez-Cruz A, Vesin D, Ramon-Luing L, Zuniga J, Quesniaux VFJ, Ryffel B, et al. CD3(+) Macrophages Deliver Proinflammatory Cytokines by a CD3-and Transmembrane TNF-Dependent Pathway and Are Increased at the BCG-Infection Site. Frontiers in Immunology. 2019;10:17.
375. Allen JN, Dey A, Nissly R, Fraser J, Yu S, Balandaram G, et al. Isolation, Characterization, and Purification of Macrophages from Tissues Affected by Obesity-related Inflammation. Jove-Journal of Visualized Experiments. 2017(122):13.
376. Cossarizza A, Chang HD, Radbruch A, Acs A, Adam D, Adam-Klages S, et al. Guidelines for the use of flow cytometry and cell sorting in immunological studies (second edition). European Journal of Immunology. 2019;49(10):1457-973.
377. Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Global quantification of mammalian gene expression control. Nature. 2011;473(7347):337-42.
378. Saito M, Iwawaki T, Taya C, Yonekawa H, Noda M, Inui Y, et al. Diphtheria toxin receptor-mediated conditional and targeted cell ablation in transgenic mice. Nature Biotechnology. 2001;19(8):746-50.
379. Van Hove H, Martens L, Scheyltjens I, De Vlaminck K, Antunes ARP, De Prijck S, et al. A single-cell atlas of mouse brain macrophages reveals unique transcriptional identities shaped by ontogeny and tissue environment. Nature Neuroscience. 2019;22(6):1021-35.

380. Ochocka N, Segit P, Walentynowicz KA, Wojnicki K, Cyranowski S, Swatler J, et al. Single-cell RNA sequencing reveals functional heterogeneity of glioma-associated brain macrophages. Nature Communications. 2021;12(1):14.
381. Wolf T, Jin WJ, Zoppi G, Vogel IA, Akhmedov M, Bleck CKE, et al. Dynamics in protein translation sustaining T cell preparedness. Nature Immunology. 2020;21(8):927-+.
382. Agger K, Cloos PAC, Christensen J, Pasini D, Rose S, Rappsilber J, et al. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature. 2007;449(7163):731-U10.
383. De Santa F, Totaro MG, Prosperini E, Notarbartolo S, Testa G, Natoli G. The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing. Cell. 2007;130(6):1083-94.
384. Lan F, Bayliss PE, Rinn JL, Whetstine JR, Wang JK, Chen SZ, et al. A histone H3 lysine 27 demethylase regulates animal posterior development. Nature. 2007;449(7163):689-U3.
385. Li J, Hardy K, Olshansky M, Barugahare A, Gearing LJ, Prier JE, et al. KDM6B-dependent chromatin remodeling underpins effective virus-specific CD8(+) T cell differentiation. Cell Reports. 2021;34(11):21.
386. Dawson SJ, Rueda OM, Aparicio S, Caldas C. A new genome-driven integrated classification of breast cancer and its implications. Embo Journal. 2013;32(5):617-28.
387. Newman AM, Liu CL, Green MR, Gentles AJ, Feng WG, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nature Methods. 2015;12(5):453-+.
388. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. Nature Communications. 2015;6:11.
389. Cox DR. REGRESSION MODELS AND LIFE-TABLES. Journal of the Royal Statistical Society Series B-Statistical Methodology. 1972;34(2):187-+.
390. Umansky V, Blattner C, Gebhardt C, Utikal J. The Role of Myeloid-Derived Suppressor Cells (MDSC) in Cancer Progression. Vaccines. 2016;4(4):16.
391. Veglia F, Sanseviero E, Gabrilovich DI. Myeloid-derived suppressor cells in the era of increasing myeloid cell diversity. Nature Reviews Immunology. 2021;21(8):485-98.
392. Yang YH, Li CY, Liu T, Dai XF, Bazhin AV. Myeloid-Derived Suppressor Cells in Tumors: From Mechanisms to Antigen Specificity and Microenvironmental Regulation. Frontiers in Immunology. 2020;11:22.
393. Sinha P, Okoro C, Foell D, Freeze HH, Ostrand-Rosenberg S, Srikrishna G. Proinflammatory S100 proteins regulate the accumulation of myeloid-derived suppressor cells. Journal of Immunology. 2008;181(7):4666-75.
394. Okabe Y, Medzhitov R. Tissue biology perspective on macrophages. Nature Immunology. 2016;17(1):9-17.
395. Ehrchen JM, Sunderkotter C, Foell D, Vogl T, Roth J. The endogenous Toll-like receptor 4 agonist S100A8/S100A9 (calprotectin) as innate amplifier of infection, autoimmunity, and cancer. Journal of Leukocyte Biology. 2009;86(3):557-66.
396. Wang SW, Song R, Wang ZY, Jing ZC, Wang SX, Ma J. S100A8/A9 in Inflammation. Frontiers in Immunology. 2018;9:14.
397. Yang H, Wang B, Wang T, Xu L, He C, Wen H, et al. Toll-Like Receptor 4 Prompts Human Breast Cancer Cells Invasiveness via Lipopolysaccharide Stimulation and Is Overexpressed in Patients with Lymph Node Metastasis. Plos One. 2014;9(10).
398. Giblin SP, Schwenzer A, Midwood KS. Alternative splicing controls cell lineage-specific responses to endogenous innate immune triggers within the extracellular matrix. Matrix Biology. 2020;93:95-114.