

Multi-Omics Analysis and Modelling of T Cell Immunity and Interferon Signalling Pathways for Engineered T Cell Therapy

Clara Eléonore Pavillet

Keble College



Medical Sciences Division

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2022

Copyright © 2022 by Clara E. Pavillet
All Rights Reserved

Pour Dr. Axel Pavillet

Donner une base solide dans laquelle se circonscrire, encourager une curiosité scientifique perpendiculaire à la sienne, puis accompagner son enfant à prendre de la hauteur, c'est peut-être pour le père géomètre le moyen de poser la base d'un nouveau tétraèdre qui révèlera, j'espère, à son tour, de nouvelles propriétés.

Merci pour cette démonstration.

Merci Papit.

All models are wrong, but some are useful.

— George E. P. Box

Acknowledgements

My deepest gratitude to my supervisors, Dr. Tudor A. Fulga, Prof. Francesca M. Buffa and Prof. Paul Klenerman. Their immense knowledge in many facets of genomics and immunology guided me throughout my research. Thank you to my thesis advisors, Prof. Tao Dong, Prof. Vincenzo Cerundolo, Prof. Helen M. Byrne and Prof. Alan Garfinkel. I cannot thank my thesis committee enough for helping me bridge multiple disciplines and shaping me intellectually.

I am also extremely grateful of the funding received towards my DPhil studies from the Wellcome Trust and the Medical Sciences Doctoral Training Centre at the University of Oxford. Thank you to the Department of Biochemistry, Sir William Dunn School of Pathology and MRC Weatherall Institute of Molecular Medicine for all the training I received.

A special thank you to my wonderful collaborators and colleagues, in particular Prof. Adrian L. Harris.

Beyond medical research, I extend my sincere thanks to those who broadened my horizon, from software engineering to healthcare investing, during my doctoral years.

Thank you to my family and friends around the globe.

To my parents, Axel and Marie-Annick, and to my siblings, Vanessa, Alexandra and Geoffroy – despite being scattered across the world we have managed to maintain strong family values and have provided invaluable support to each other throughout the years, for which I am forever grateful.

To Couscous, I am sure you know more about T cells by now than any other golden retriever. Thank you for being so patient with me, laying by my side all these long hours of writing.

Finally, to Harsh, thank you for being here since Matriculation Day.

Abstract

This thesis examines the nature of T cells through the lens of genomics. It delves into type I interferon (IFN)-mediated signalling in the tumour microenvironment (TME), focusing on the GMP-AMP synthase (cGAS)-stimulator of IFN genes (STING) pathway, and explores novel strategies to optimise T cell-based immunotherapy for triple-negative breast cancer (TNBC). In the first chapter, gene expression patterns of distinct T cell subtypes are studied to gain insight into the diverse T cell landscape. Cell type annotation methods are evaluated for their ability to identify closely-related T cell populations in single-cell RNA sequencing (scRNA-seq) data. In the following chapter, single-cell and spatial transcriptomics methods, mainly scRNA-seq, single-cell immune profiling, Slide-seq, in situ sequencing and Molecular Cartography enable the comprehensive analysis of immune infiltrates in tissue biopsies from patients with TNBC. The advantages and limitations of emerging spatial transcriptomics technologies are put into context. Tumour-infiltrating T cells were characterised at varying levels of resolution and *cGAS* was found to be most highly expressed in proliferating T cells. Tissue-wide spatial patterns revealed a strong relationship between programmed cell death-1/programmed cell death ligand-1 and *cGAS* but not *STING*. In addition, gene transcripts encoding the downstream protein kinase TANK-binding kinase 1 were found to spatially associate with hypoxia markers. Finally, T cells genetically engineered to express chimeric antigen receptors (CARs) are investigated using a systems biology approach. An agent-based model is developed to explore the effect of CAR tuning on on-target off-tumour toxicity and therapy efficacy. Gene regulatory networks are used to present readily testable hypotheses for combination strategies using immune checkpoint inhibitors, oncolytic viruses, and cGAS-STING-targeting treatments to potentiate anti-tumour responses. The work presented therein will feed into a novel computational framework to enable the modelling of patient-specific responses to T cell-based immunotherapy and synergistic combinations.

Table of Contents

List of Tables vi

List of Figures viii

List of Abbreviations xiii

1 Introduction 1

1.1 Motivation . . . 1

1.2 Aims and Objectives . . . 3

1.3 Outline . . . 3

2 Literature Review 5

2.1 T Cell Immunology . . . 5

2.1.1 Stages of the T Cell Life Cycle . . . 5

2.1.2 T Cell Receptor Repertoire . . . 7

2.1.3 T Cell Subtypes and States . . . 9

2.2 Tumour Microenvironment . . . 10

2.2.1 Cancer Genome Evolution . . . 11

2.2.2 Multicellular Composition . . . 12

2.2.3 Immunosuppressive Milieu . . . 13

2.3 Interferon Signalling and Inflammatory Responses . . . 15

2.4 Rewiring T Cells with Genome Editing and Synthetic Biology . . . 19

2.5 Multi-Omics and Imaging Technologies for Therapeutic Insights . . . 23

3 Comparative Analysis of T Cell Identification Methods for Single-Cell RNA Sequencing Data 30

3.1 Overview . . . 30
3.2 Introduction . . . 31
3.3 Materials and Methods . . . 33
3.3.1 Single-Cell RNA Sequencing T Cell Datasets . . . 33
3.3.2 Combining T Cell Subtypes . . . 35
3.3.3 Identifying Gene Expression Signatures . . . 36
3.3.4 Cell Type Identification Methods . . . 37
3.3.5 Performance Evaluation . . . 41
3.4 Results . . . 48
3.4.1 T Cell Overlaps in Low-Dimensional Signal Space . . . 48
3.4.2 T Cell Subtypes Show Unique Expression Patterns . . . 50
3.4.3 CD4 Helper T Cells Require Further Demultiplexing . . . 53
3.4.4 Evaluating T Cell Detection . . . 55
3.4.5 Performance Comparison for T Cell Subtype Identification . . . 58
3.5 Discussion . . . 70
3.6 Conclusion . . . 74

4 Multi-Omics Analysis of T Cell-Intrinsic Interferon Signalling Pathways in the Tumour Microenvironment 75

4.1 Overview . . . 75
4.2 Introduction . . . 76
4.3 Materials and Methods . . . 79
4.3.1 Gene Set Construction . . . 79
4.3.2 Patient Samples . . . 79
4.3.3 10x Genomics Single-Cell RNA Sequencing . . . 81
4.3.4 10x Genomics Single Cell Immune Profiling . . . 83
4.3.5 Spatial Transcriptomics . . . 83
4.3.6 Processing Chromium Single-Cell Data . . . 92
4.3.7 Analysis of Single-Cell T Cell Receptor Sequencing . . . 92
4.3.8 Analysis of Single-Cell RNA Sequencing . . . 93
4.3.9 Analysis and Visualisation of Spatial Transcriptomics . . . 98

4.4 Results . . . 103
4.4.1 Patient Samples and Characteristics . . . 103
4.4.2 Immune Repertoire . . . 108
4.4.3 Transcriptional Landscape of T Cell Clusters . . . 112
4.4.4 Paired Gene Expression and TCR Repertoire . . . 117
4.4.5 Assessing T Cell Activation and Dysfunction . . . 118
4.4.6 Examining IFN Signalling Pathways in T Cells . . . 121
4.4.7 Pseudotemporal Ordering of T Cells . . . 128
4.4.8 Mapping T Cell Interactions . . . 130
4.4.9 Transcriptome-Wide Spatial Profiling Captures Cellular Organisation . . . 135
4.4.10 High-Resolution Exploration of Subcellular Transcriptomic Activity . . . 139
4.5 Discussion . . . 164
4.6 Conclusion . . . 176

5 Multi-Scale Modelling of Chimeric Antigen Receptor T Cell Therapy in Search for the Goldilocks T Cell Design 177
5.1 Overview . . . 177
5.2 Introduction . . . 178
5.3 Materials and Methods . . . 181
5.3.1 Experimental Data . . . 181
5.3.2 Modelling CAR T Cell-Tumour Interactions . . . 184
5.3.3 Modelling CAR T Cell Gene Regulatory Network . . . 198
5.3.4 Combining Modelling Paradigms . . . 203
5.4 Results . . . 208
5.4.1 Comparison of Tumour Growth Rates . . . 208
5.4.2 CAR T Agent Activation Tuning . . . 210
5.4.3 Influence of cGAS-STING Activation on T Cell Stable States 211
5.4.4 Tumour Cell Therapeutic Stable State Responses . . . 211
5.5 Discussion . . . 212

5.6 Conclusion 217

6 Conclusions **218**

6.1 Identifying T Cells with Single-Cell RNA Sequencing 218

6.2 Multi-Omics Analysis of T Cells and cGAS-STING 219

6.3 Modelling T Cell Immunotherapy 219

6.4 Concluding Remarks 220

Bibliography **222**

Appendix A Predicting Base-Calling Errors from the Oxford Nanopore Sequencer **250**

A.1 Overview 250

A.2 Introduction 250

A.3 Materials and Methods 253

A.3.1 MinION-Sequenced Data 253

A.3.2 Analysing Miscalled Positions & Regions 253

A.3.3 Developing Predictive Model 255

A.4 Results and Discussion 257

A.4.1 Overlapping Signal-Spaces Confuse the Base-Caller 257

A.4.2 Local Secondary Structures Affect 3' End Base-Calling 258

A.4.3 Predicting Error-Prone Positions 258

Appendix B Comparing T Cell-Intrinsic Interferon Profiles in Cancer and Infectious Diseases **261**

B.1 Overview 261

B.2 Introduction 262

B.3 Materials and Methods 264

Appendix C Supplementary Material for Chapter 3 **266**

C.1 T Cell Signatures 267

Appendix D Supplementary Material for Chapter 4 **269**

D.1 Tissue Processing . . . 269
D.2 Probes for Spatial Transcriptomics . . . 270
D.3 Scope of Work . . . 277
D.4 Biosafety Declaration . . . 282
D.5 Tender Exemption . . . 286
D.6 Purchase Order . . . 288
D.7 Material Transfer Agreement . . . 289
D.8 Sample Positioning . . . 295
D.9 Cell Ranger Metrics . . . 296
D.10 Cell Segmentations . . . 304
D.11 Cluster Markers . . . 306
D.12 TCR Clones . . . 369
D.13 Pseudotime Analysis . . . 372
D.14 Cellular Interactions . . . 374
D.15 Slide-Seq Pucks . . . 379
D.16 Reference for Spatial Transcriptomics . . . 392
D.17 Colocalisation Matrices . . . 393

Appendix E Supplementary Material for Chapter 5 **398**
E.1 Multi-Agent System . . . 398

Appendix F Source Code **399**

List of Tables

3.1 Single-Cell RNA Sequencing T Cell Datasets 34
3.2 Confusion Matrix Schematic 42
3.3 Performance Metrics for SingleR 61
3.4 Performance Metrics for Azimuth 64
3.5 Performance Metrics for Downsampled Azimuth 66
3.6 Performance Metrics for ProjecTILs 68

4.1 Clinical Profiles of TNBC Samples 80
4.2 Histological Assessment of Cell Type Proportions 85
4.3 Array Cell Type Proportions and Spatially Variable Features 137
4.4 Molecular Cartography Spatial Metrics for BRO18 145
4.5 Molecular Cartography Spatial Metrics for BRO20 148
4.6 Molecular Cartography Spatial Metrics for BRO23 153
4.7 Molecular Cartography Spatial Metrics for TRN1 154
4.8 Molecular Cartography Spatial Metrics for TRN2 155
4.9 Molecular Cartography Spatial Metrics for TRN3 156
4.10 Molecular Cartography Spatial Metrics for TRN4 159

5.1 Truth Table of the Transition Functions 200
5.2 Stable States of CAR T Cell Gene Regulatory Network 211
5.3 Stable States of Tumour Cell Δp53 Gene Regulatory Network 212

A.1 Pore Model Provides Expected Signal Distributions 257

D.1 Panel of Gene-Specific Probes for CARTANA 270
D.2 Panel of Gene-Specific Probes for Molecular Cartography 274

D.3 Mosaic of Cell Segmentations 304
D.5 Top Ten TCR Clones in BRO20 370
D.6 Top Ten TCR Clones in BRO23 371

List of Figures

2.1 T Cell Receptor Genes V(D)J Recombination 8
2.2 Multicellular Composition of the Tumour Microenvironment 13
2.3 Obstacles Posed to T cells in the Tumour Microenvironment 14
2.4 DNA-Sensing Mediates IFN Signalling 16
2.5 A Vicious Cycle of Inflammation in Cancer Patients 18
2.6 Generations of CAR Designs 20
2.7 Types of Immunotherapy 23
2.8 Schematic of the Omics Cascade 24
2.9 Single Molecule Fluorescence In Situ Hybridization 27

3.1 Bootstrapping T Cell Predictions 47
3.2 Projected T Cell Subtypes 49
3.3 T Cell Subtypes Heatmap of Gene Expression for Selected Markers . 52
3.4 Demultiplexing CD4 Helper T Cells 54
3.5 Confusion Matrices for T Cell Detection 57
3.6 Percent T Cells Detected Across Methods 58
3.7 T Cell Subtype Predictions with SingleR 60
3.8 Bootstrap Resampling for SingleR Predictions 61
3.9 T Cell Subtype Predictions with Azimuth 63
3.10 Bootstrap Resampling for Azimuth Predictions 64
3.11 T Cell Subtype Predictions with Downsampled Azimuth 65
3.12 T Cell Subtype Predictions with ProjecTILs 67
3.13 Bootstrap Resampling for ProjecTILs Predictions 68
3.14 Performance Metrics Summary 69

4.1 Processing of Breast Cancer Clinical Samples 81
4.2 Gel Beads in Emulsion Schematic 83
4.3 Chromium Single Cell V(D)J 83
4.4 H&E Staining of Tumour Tissue Samples 89
4.5 Whole Slide Images and Regions of Interests 91
4.6 10x Genomics Cell Ranger Pipeline for GEX 92
4.7 Ensemble Method for T Cell Isolation 95
4.8 Cell Segmentation 101
4.9 Quality Control of Single-Cell RNA Sequencing Datasets 103
4.10 Tissue Dissociation Method and Assay Chemistry Comparison 105
4.11 Clustering Analysis of Single-Cell RNA Sequencing Datasets 106
4.12 Marker-Based Labelling of Cell Clusters 107
4.13 Overview of Patients TCR Repertoire 111
4.14 Isolating and Refining T Cells 113
4.15 Transcriptional Profile of Isolated T Cells 114
4.16 Heatmap of Gene Expression Across T Cell Clusters 115
4.17 Reference-Based T Cell Predictions 116
4.18 TCR Repertoire Projected Onto Transcriptionally-Defined T Cell Space and Trajectories. 117
4.19 Clonotypes Across T Cell Groups 118
4.20 T Cell Activation and Dysfunctional State Scores 120
4.21 IFN-Induced Signature Characteristic of T Cell Cluster 121
4.22 Heightened T Cell Anti-Viral Immune Response 122
4.23 Mapping to IFN Signalling Pathways 125
4.24 Interferon Signalling Scores Across T Cell Clusters 126
4.25 Correlation Matrix of IFN Genes and Exhaustion Markers 127
4.26 Pseudotemporal Ordering of T Cells 129
4.27 Heatmap of Interaction Log Counts 130
4.28 Predicted Interaction Counts for T Cell Populations 131
4.29 Pathways Driving Anti-Inflammatory Effects 132

4.30 Pathways Driving Immune Responses . . . 133
4.31 Circle Plots of Canonical CTL and IFN Signature Interactions . . . 134
4.32 Slide-seq Arrays Bead Counts . . . 135
4.33 Transferring scRNA-Seq Cell Type Annotations Onto Arrays . . . 136
4.34 T Cell Subtypes Transferred Onto Arrays . . . 138
4.35 Spatial Distribution of IFN-Mediated Signalling Genes . . . 138
4.36 Tissue-Wide Colocalisation of In Situ Sequencing Probes . . . 140
4.37 Tissue-Wide Colocalisation of Molecular Cartography Probes . . . 141
4.38 T Cell Markers Overlaying Tissue Sample BBP1467 . . . 143
4.39 T Cell Markers Overlaying Tissue Sample BBP1467 . . . 144
4.40 Spatial Organisation of Cell Populations for Sample BRO18 . . . 147
4.41 Spatial Organisation of Cell Populations for Sample BRO20 . . . 152
4.42 Spatial Organisation of Cell Populations for Sample BRO23 . . . 154
4.43 Spatial Organisation of Cell Populations for Sample TRN2 . . . 156
4.44 Spatial Organisation of Cell Populations for Sample TRN3 . . . 158
4.45 Spatial Organisation of Cell Populations for Sample TRN4 . . . 163

5.1 Lentiviral Vector Encoding Tunable Anti-NY-ESO-1 CAR . . . 181
5.2 Gene Expression Heatmap of CAR Tuning . . . 182
5.3 CAR Expression-Activation Relationship . . . 183
5.4 Visual Representation of an Agent . . . 184
5.5 Flow of Information . . . 186
5.6 Mosaic of Model Topologies . . . 190
5.7 Schematic Representation of the ABM . . . 192
5.8 Flow Chart of the ABM Simulation Algorithm . . . 195
5.9 Wild-Type Agent Step Function . . . 196
5.10 Tumour Agent Step Function . . . 196
5.11 CAR T Agent Step Function Step Function . . . 197
5.12 Schematic Diagram of Boolean State Space . . . 200
5.13 Protein-Protein Interaction Network of cGAS-STING Components . . . 201
5.14 CAR T Agent Gene Regulatory Network . . . 202

5.15 MicroC Environment 204
5.16 Voxel Cube Representation 205
5.17 Modelling T Cell-Based Immunotherapy 206
5.18 Tumour Agent Gene Regulatory Network 207
5.19 Tumour Growth Curves 209
5.20 Activation Marker in Response to CAR Tuning 210

A.1 Oxford Nanopore Technologies DNA Sequencing 251
A.2 Signal Events for Thymidine and Bromodeoxyuridine 252
A.3 Overlapping Signal-Spaces Confuse Base-Caller. 259
A.4 Sequence Minimum Free Energy Secondary Structure Probabilities . . 260

C.1 Fold Changes of Genes Enriched in T Cell Subtypes 267
C.2 Volcano Plots of Differential Gene Expression 268

D.1 T Cell Trajectories 372
D.2 Genes Varying Along Pseudotime 373
D.3 T Cell Subpopulations and Major Cell Types 374
D.3 Circle Plots of T Cell Interactions 378
D.4 Total Counts per Bead for Puck 1 379
D.5 Top Spatially Variable Features for Puck 1 380
D.6 Transferring scRNA-Seq Cell-Type Annotations Onto Puck 1 381
D.7 Total Counts per Bead for Puck 2 382
D.8 Top Spatially Variable Features for Puck 2 383
D.9 Transferring scRNA-Seq Cell-Type Annotations Onto Puck 2 384
D.10 Total Counts per Bead for Puck 3 385
D.11 Top Spatially Variable Features for Puck 3 386
D.12 Transferring scRNA-Seq Cell-Type Annotations Onto Puck 3 387
D.13 Total Counts per Bead for Puck 4 388
D.14 Top Spatially Variable Features for Puck 4 389
D.15 Transferring scRNA-Seq Cell-Type Annotations Onto Puck 4 390

D.16 Heatmap Displaying Average Expression of Spatial Probe Targets in scRNA-Seq 392
D.16 Molecular Cartography Colocalisation Matrices 396
D.16 CARTANA Colocalisation Matrices 397

E.1 Multi-Agent System Flow Chart 398

List of Abbreviations

ABM Agent-Based Model.

ACAD Activated Cell Autonomous Death.

ACE2 Angiotensin-Converting Enzyme 2.

ACT Adoptive Cell Transfer.

AICD Activation-Induced Cell Death.

APC Antigen-Presenting Cell.

ATAC-seq Assay for Transposase-Accessible Chromatin using Sequencing.

BATF Basic Leucine Zipper ATF-Like Transcription Factor.

BiTE Bispecific T Cell Engagers.

BrdU Bromodeoxyuridine.

CAF Cancer-Associated Fibroblast.

CAR Chimeric Antigen Receptor.

CCL CC Chemokine Ligand.

CCLE Cancer Cell Line Encyclopedia.

CDR Complementarity-Determining Region.

CDR3 Complementary Determining Region 3.

cGAS Cyclic GMP-AMP Synthase.

ChIA-PET Chromatin Interaction Analysis by Paired-End Tag Sequencing.

ChIP-seq Chromatin Immunoprecipitation Sequencing.

circRNA Circular RNA.

CITE-seq Cellular Indexing of Transcriptomes and Epitopes by Sequencing.

CLP Common Lymphoid Progenitor.

CLR C-Type Lectin Receptor.

CNV Copy Number Variations.

COVID-19 Coronavirus Disease 2019.

cPAL Combinatorial Probe-Anchor Ligation.

CpG 5'-C-phosphate-G-3'.

CRS Cytokine Release Syndrome.

CTA Cancer Testis Antigen.

CTL Cytotoxic T Cell.

CTLA-4 Cytotoxic T Lymphocyte Antigen 4.

D Diversity.

DAMP Damage-Associated Molecular Pattern.

DAPI 4 6-Diamidino-2-Phenylindole.

DEG Differentially Expressed Gene.

DN Double Negative.

DP Double Positive.

dsRNA Double-Stranded RNA.

EC50 Half Maximal Effective Concentration.

ECM Extracellular Matrix.

EMT Epithelial-To-Mesenchymal Transition.

ER Oestrogen Receptor.

FACS Fluorescence-Activated Cell Sorting.

FFPE Formalin-Fixed Paraffin-Embedded.

FGF2 Fibroblast Growth Factor 2.

FN False Negative.

FP False Positive.

GEM Gel Bead-In-Emulsion.

GEO Gene Expression Omnibus.

GINSim Gene Interaction Network Simulation.

GMT Gene Matrix Transposed.

GO Gene Ontology.

GRN Gene Regulatory Network.

GSEA Gene Set Enrichment Analysis.

HER2 Human Epidermal Growth Factor Receptor 2.

HGNC HUGO Gene Nomenclature Committee.

HMM Hidden Markov Model.

HPCA Human Primary Cell Atlas.

HSC Haematopoietic Stem Cell.

HUGO Human Genome Organisation.

HVG Highly Variable Gene.

ICAM-1 Intercellular Adhesion Molecule 1.

IDO Indoleamine 2,3-Dioxygenase.

IFITM Interferon-Induced Transmembrane.

IFN Interferon.

IL Interleukin.

IRF Interferon Regulatory Factor.

IRF4 Interferon Regulatory Factor 4.

ISG Interferon-Stimulated Gene.

ISH In Situ Hybridisation.

ISS In Situ Sequencing.

ITAM Immunoreceptor Tyrosine-Based Activation Motifs.

J Joining.

JS Jensen-Shannon.

KEGG Kyoto Encyclopedia of Genes and Genomes.

KL Kullback-Leibler.

lncRNA Long Non-Coding RNA.

m7G 7-Methylguanosine.

MAD Median-Absolute-Deviation.

MAIT Mucosal-Associated Invariant T.

MAPK Mitogen-Activated Protein Kinase.

MC Molecular Cartography.

MDSC Myeloid-Derived Suppressor Cells.

MFE Minimum Free Energy.

MFI Median Fluorescence Intensity.

MHC Major Histocompatibility Complex.

MIF Macrophage Migration Inhibitory Factor.

miRNA MicroRNA.

miSFIT MicroRNA Silencing-Mediated Fine-Tuner.

MR1 MHC Class I Related-1 Molecule.

MRCA Most Recent Common Ancestor.

MRE MicroRNA Response Elements.

mRNA Messenger RNA.

MS Mass Spectrometry.

MSigDB Molecular Signatures Database.

MT Mitochondrial.

NCBI National Center for Biotechnology Information.

NF-kB Nuclear Factor Kappa B.

NGS Next Generation Sequencing.

NIH National Institutes of Health.

NK Natural Killer.

NKT Natural Killer T Cells.

NLR NOD-Like Receptor.

NMR Nuclear Magnetic Resonance.

NUPACK Nucleic Acid Package.

ONT Oxford Nanopore Technologies.

OOP Object-Oriented Programming.

OV Oncolytic Virus.

PBMC Peripheral Blood Mononuclear Cell.

PBN Probabilistic Boolean Network.

PCA Principal Component Analysis.

PD-1 Programmed Cell Death Protein 1.

PDGF Platelet-Derived Growth Factor.

PI3K Phosphoinositide 3-Kinase Pathway.

PR Progesterone Receptor.

PRR Pattern Recognition Receptor.

PSGL-1 P-Selectin Glycoprotein Ligand 1.

QC Quality Control.

RAG Recombination Activating Gene.

RCA Rolling Circle Amplification.

RIG-I Retinoic Acid-Inducible Gene I.

RISC RNA-Induced Silencing Complex.

RLR RIG-I-Like Receptors.

RNA-seq RNA Sequencing.

ROI Region of Interest.

RPCA Reciprocal Principal Component Analysis.

rRNA Ribosomal RNA.

RSS Recombination Signal Sequence.

RT Reverse Transcription.

sAPCs Synthetic Antigen Presenting Cell System.

SARS-CoV-2 Severe Acute Respiratory Syndrome Coronavirus 2.

scBCR-seq Single-Cell BCR Sequencing.

scFv Single Chain Variable Fragment.

scRNA-seq Single-Cell RNA Sequencing.

scTCR-seq Single-Cell TCR Sequencing.

SDG Spatially Differentiated Gene.

sgRNA Single Guide RNA.

smFISH Single Molecule Fluorescence In Situ Hybridization.

SNN Shared Nearest Neighbour.

SNP Single-Nucleotide Polymorphism.

snRNA-seq Single-Nucleus Sequencing.

SP Single Positive.

ssDNA Single-Stranded DNA.

ssGSEA Single-Sample GSEA.

SSR Simple Sequence Repeat.

ssRNA Single-Stranded RNA.

ST Spatial Transcriptomics.

STG State Transition Graph.

STING Stimulator of Interferon Genes.

STRING Search Tool for the Retrieval of Interacting Genes/Proteins.

SVM Support Vector Machine.

SynNotch Synthetic Notch-Based Systems.

TAA Tumour Associated Antigen.

TAM Tumour-Associated Macrophage.

TCM Central Memory T Cell.

TCR T Cell Receptor.

TEM Effector Memory T Cell.

TEMRA Effector Memory Cells T Cells Re-Expressing CD45RA.

TF Transcription Factor.

Tfh T Follicular Helper.

Th T Helper.

TIL Tumour Infiltrating Lymphocyte.

TLR Toll-Like Receptor.

TLR2 Toll-Like Receptor 2.

TM Transmembrane Domain.

TME Tumour Microenvironment.

TN True Negative.

TNBC Triple-Negative Breast Cancer.

TP True Positive.

TPM Peripheral Memory T Cells.

Treg T Regulatory.

TRM Tissue-Resident Memory T Cells.

tRNA Transfer RNA.

TSCM Memory Stem T Cells.

TSO Template Switch Oligo.

UMAP Uniform Manifold Approximation and Projection.

UMI Unique Molecular Identifier.

V Variable.

VCAM-1 Vascular Cell Adhesion Molecule 1.

VEGF Vascular Endothelial Growth Factor.

VM Vasculogenic Mimicry.

VST Variance Stabilizing Transformation.

WES Whole Exome Sequencing.

WGS Whole Genome Sequencing.

WNN Weighted-Nearest Neighbour.

WSI Whole Slide Image.



CHAPTER 1

Introduction

1.1 Motivation

Thymus-derived lymphocytes, T cells, are central players of the adaptive immune system that drive humoral and cellular responses to protect the body against disease and infection. T cells can be classified into different subsets based on their T cell receptor (TCR) chains, the extracellular and intracellular markers they express, or epigenetic, metabolic and genetic programme. In a disease like cancer, T cells become dysfunctional owing to an immunosuppressive network that extends even beyond the tumour microenvironment (TME). Signalling of immunomodulatory cytokines, interferons (IFNs), which highly influence T cell priming, expansion and acquisition of effector functions, has also been found altered in cancer but it's role in disease progression and treatment resistance has been understudied.

With the immune system recognised as a key orchestrator of the TME and subsequent therapy response, there has been a growing interest in cancer treatments that specifically target immune components to potentiate natural anti-tumour defences. Advancements in genome editing and synthetic biology paved the way for novel immunotherapies such as T cells engineered to express synthetic tumour-targeting receptors on their surface, called chimeric antigen receptors (CARs). CAR T cell therapy has shown promising results in haematological malignancies but limited success against solid tumours due to additional obstacles posed by the TME that hinder trafficking, survival and tumour detection. Localised anti-tumour immune responses rely on continuous communication with the pe-

riphery and proper coordination between immune cell types, thus the importance of proper intracellular and intercellular signalling. Investigating major events occurring in the TME that might induce T cell dysfunctional states, such as impaired IFN signalling pathways, could help guide therapeutic strategies to enhance CAR T cell function or other types of redirected T cells. Moreover, tumour development is known to alter systemic immunity and progressively change the composition of peripheral T cells. As was observed during the coronavirus disease 2019 (COVID-19) pandemic, patients with cancer are at higher risk of severe illness when facing secondary immune challenges and are less likely to mount an effective immune response to vaccination, albeit being treatment-naive. Similarly, intact functional peripheral immune mechanisms have been found to be essential to coordinate local anti-tumour responses for immunotherapeutic efficacy. Uncovering how tumour-driven disruptions influence the T cell landscape at the intratumoural and systemic level could help further tackle susceptibility to infections and poor response to immunotherapy in cancer patients.

Omics technologies, including single-cell RNA sequencing (scRNA-seq), cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) and spatial transcriptomics (ST) are bringing in a wealth of granular data to dig deeper into the depths of disease. The correct integration and interpretation of these generated datasets can offer valuable insights to optimise immunotherapeutic strategies by helping dissect immune landscapes and cell dynamics at a single-cell resolution. High-throughput biochemical assays quantifying the genome, epigenome, transcriptome, proteome, and metabolome of biological systems are facilitating the creation of multi-scale models for in silico experimenting to test hypotheses and formulate predictions. These have the potential to drive and accelerate biological discoveries in areas like precision medicine by providing new insights into disease mechanisms, biomarkers and drug target identification. Models that capture multi-scale molecular organisations can be particularly valuable to inform the design of new therapies or optimised strategies. Engineering cancer-fighting T cells requires a good understanding of the link between genotype and phenotype, and response to these therapies varies greatly across patients and tumours due to differences in omics profiles. Therefore, comprehensive models able to simulate T cell immunity within the broader context of

the TME, and faithfully recreate patient-specific cellular and molecular interactions could assist researchers in both immuno-oncology and biotechnology to better design future immunotherapies or uncover synergistic combinations.

1.2 Aims and Objectives

The aim of this thesis is to investigate the interplay between type I IFN pathways and T cell function using a systems biology approach with a particular focus on the cyclic GMP-AMP synthase (cGAS)-stimulator of interferon genes (STING) signalling pathway. Multi-scale models guided by the analysis of multi-omics datasets will be developed to predict optimised T cell-based immunotherapy strategies that exploit this interplay, along with novel genome engineering and synthetic biology methods.

The following questions will be investigated during the course of this research:

- Can gene expression patterns be used to accurately dissect the T cell landscape in genomics data? (Chapter 3)
- Is there a relationship between cGAS-STING signalling and T cell dysfunction in the TME? (Chapter 4)
- Is there an optimal number of CAR expression on T cells and how can type I IFN pathways be exploited to enhance CAR T cell efficacy in solid tumours? (Chapter 5)

1.3 Outline

The remainder of this thesis is organised as follows:

Chapter 2 — gives a general context and background for the work presented in this thesis, laying the foundation for subsequent chapters.

Chapter 3 — evaluates and compares the performance of currently available cell identification methods for T cell subtype identification in scRNA-seq data.

Chapter 4 — integrates single-cell gene expression, single-cell immune profiling

and spatial transcriptomics to provide a detailed characterisation of T cell profiles and IFN signalling pathways in the TME of breast cancer patients.

Chapter 5 — leverages modelling paradigms to investigate the effect of tuning CAR surface expression on CAR T cell on-target off-tumour toxicity and explores strategies targeting IFN signalling pathways to counteract T cell dysfunction in solid tumours.

Chapter 6 — concludes and reflects on the research conducted, listing key contributions and discussing potential for future research.

Appendix A — investigates base-calling errors and maps nucleic acid sequence properties to Oxford Nanopore sequencing read accuracy.

Appendix B — introduces an ongoing study comparing T cell-intrinsic IFN signatures across disease conditions (breast cancer, sepsis, influenza A, varying severities of COVID-19) and healthy controls.

Appendix C — provides supplementary material to Chapter 3.

Appendix D — provides supplementary material to Chapter 4.

Appendix E — provides supplementary material to Chapter 5.

Appendix F — links to the source code.

CHAPTER 2

Literature Review

2.1 T Cell Immunology

T cells are specialised lymphocytes and central regulators of the immune response. They derive their name from the organ where they mature, the thymus. A highly diverse repertoire of T cell receptors (TCRs) mediates their ability to recognise foreign or altered-self antigens as they survey and migrate throughout the body. The T cell landscape consists of functionally discrete subtypes that can display a variety of cell states in response to intracellular and extracellular cues.

2.1.1 Stages of the T Cell Life Cycle

Haematopoietic stem cells (HSCs) in the bone marrow give rise to oligopotent cells, common lymphoid progenitors (CLPs), which can differentiate into T lineage cells, B lineage cells or natural killer (NK) cells. Progenitors that travel through blood vessels to the thymus, a bi-lobed gland located behind the sternum and above the heart, commit themselves to the T cell lineage under the influence of the hormone thymosin. Extravasation into the thymus is regulated by thymic endothelial cells expressing P-selectin which bind P-selectin glycoprotein ligand 1 (PSGL-1) on progenitors, thereby promoting interaction with intercellular adhesion molecule 1 (ICAM-1) and vascular cell adhesion molecule 1 (VCAM-1). This multistep adhesion cascade relies on the rapid formation and breakage of bonds to mediate progenitor tethering and rolling on the endothelial surface under physiological flow for subsequent arrest and diapedesis (Ivetic et al., 2019).

T cell development occurs in stages and starts upon arrival of the T cell precursors in the subcapsular cortical areas where they undergo a period of proliferation. As they migrate through distinct thymic microenvironments and their expression of cell surface markers changes, the now called thymocytes go through a series of selection processes (Kumar et al., 2018). The most immature subset is termed double negative (DN), expressing neither the CD4 nor CD8 surface molecules, and can be further subdivided into four subsets based on expression of CD44, CD25 and CD117 — DN1: CD44+, CD25-, CD117+; DN2: CD44+, CD25+, CD117+; DN3: CD44-, CD25+, CD117-; DN4: CD44-, CD25-, CD117- (Y. Wei et al., 2017).

Rearrangement of the *TCR*β, *TCR*γ and *TCR*δ gene loci occurs during the DN3 stage which coincides with $\alpha\beta$- or $\gamma\delta$-lineage choice decision. Gene rearrangement at these sites occurs through a process called V(D)J recombination which will be described in more detail in Section 2.1.2. During β-selection, a successfully rearranged TCRβ chain is paired with an invariant pre-TCRα and CD3 subunits to produce a pre-TCR complex. Pre-TCR ligand-independent signalling will induce commitment to the $\alpha\beta$ T cell lineage, promote additional expansion and enable the cell to proceed to the next stage of the maturation process, eliminating those with a non-productively rearranged TCRβ via apoptosis (N. Zhang et al., 2005). Rearrangement of the γ and δ loci occurs in a single step, resulting in a $\gamma\delta$TCR which will self-oligomerise to initiate intracellular signalling pathways committing to the $\gamma\delta$ T cell lineage. A dense expression of Notch ligands on cortical thymic epithelium drives T lineage commitment as thymocytes transition through DN stages and influences lineage specification. Importantly, inactivation of Notch 1 has been shown to severely impair $\alpha\beta$ T cell development due to severe Vβ-DJβ rearrangement impairment (Brandstadter & Maillard, 2019). Cells expressing $\gamma\delta$TCR, however, survive and expand in the absence of Notch, maturing past the DN3 stage. There is much left to be learned about $\gamma\delta$ T cell development but it is worth noting that, unlike $\alpha\beta$ thymocytes, most remain DN and thus do not progress through double positive (DP). In contrast, as TCRβ expressing thymocytes develop to the DP stage, gaining expression of both CD4 and CD8, rearrangement of the TCRα chain begins. Once the TCR is fully assembled, thymocytes undergo positive and negative selection which provides survival signals to those with a

moderate affinity to self-MHC-peptide complexes and eliminates ones with high affinity. Thymocytes that fail to receive survival signals can further edit their receptor to produce a new TCR. As they migrate from the thymic cortex to the medulla, they commit to either CD4 or CD8 expression becoming single positive (SP). Lineage-specific transcriptional programmes and epigenetic regulations induced by external stimuli and TCR signalling cooperate to mediate co-receptor choice, with transcription factors ThPOK and Runx3 being main drivers of CD4 and CD8 lineage, respectively. In the medulla, thymocytes get exposed to diverse self-antigens expressed on antigen-presenting cells (APCs) to test and eliminate potentially autoreactive cells before letting them exit the thymus via blood vessels at the corticomedullary junction as mature T cells (Kieran et al., 2021).

Travelling through blood vessels and the lymphatic vascular system, mature T lymphocytes migrate between peripheral tissues and secondary lymphoid organs surveying for specific antigens. $\alpha\beta$ T cells will acquire effector functions upon activation in the periphery, whereas most $\gamma\delta$ T cells will already have acquired them during development in the thymus. Antigen exposure and external stimuli promote metabolic reprogramming and genome-wide changes in chromatin interactions, triggering differentiation into several potential lineages or states which are covered in Section 2.1.3. Activated T cells migrate through blood vessel walls and non-lymphoid tissues to exert their effector functions at sites of infection, malignancies or tissue damage. At the end of an immune response, antigen-activated T cells which typically peak between 7-15 days post-stimulation, undergo a contraction phase wherein 90-95% of them are rapidly eliminated via activation-induced cell death (AICD) or activated cell autonomous death (ACAD). AICD results from TCR restimulation whilst ACAD involves activation of extrinsic pathways mediated by death receptors (Green et al., 2003). A small fraction of activated T cells survives to become memory cells that can rapidly be converted into large numbers of effector cells upon re-exposure to their target antigen.

2.1.2 T Cell Receptor Repertoire

The TCR is a protein complex embedded in the membrane of T cells which consists of polypeptide chain pairs, $\alpha\beta$ or $\gamma\delta$. It mediates the ability of T cells to recognise anti-

gens displayed on cellular surfaces with TCR$\alpha\beta$ recognising peptides bound to major histocompatibility complex (MHC) class I or II, and TCR $\gamma\delta$ recognising unprocessed antigens (Roth, 2014). This heterodimeric protein is also associated with the following invariant signalling molecules: CD3$\epsilon\delta$, CD3$\gamma\epsilon$, CD247$\zeta\zeta$. Upon binding, immunoreceptor tyrosine-based activation motifs (ITAMs) on the cytoplasmic tails get phosphorylated by tyrosine kinases, triggering a signalling cascade that propagates downstream. Clonally distinct receptors arise during development in the thymus as recombination events occur in loci encoding the TCR chains, located on chromosomes 7 (β at 7q32-35 and γ at 7p15) and 14 (α and δ at 14q11-12). Recombination activating genes (RAGs) play a central role in initiating the rearrangement of the variable (V), diversity (D) and joining (J) segments (Fig. 2.1). By cleaving the DNA at recombination signal sequences (RSSs), a series of chromosomal DNA breakage and segment ligation starts (Lovely et al., 2015). This permutation process uses a finite gene set to give rise to unlimited specificities. Target specificity of the final TCR product is defined by the six complementarity-determining region (CDR) loops, CDR3 being the most diverse.

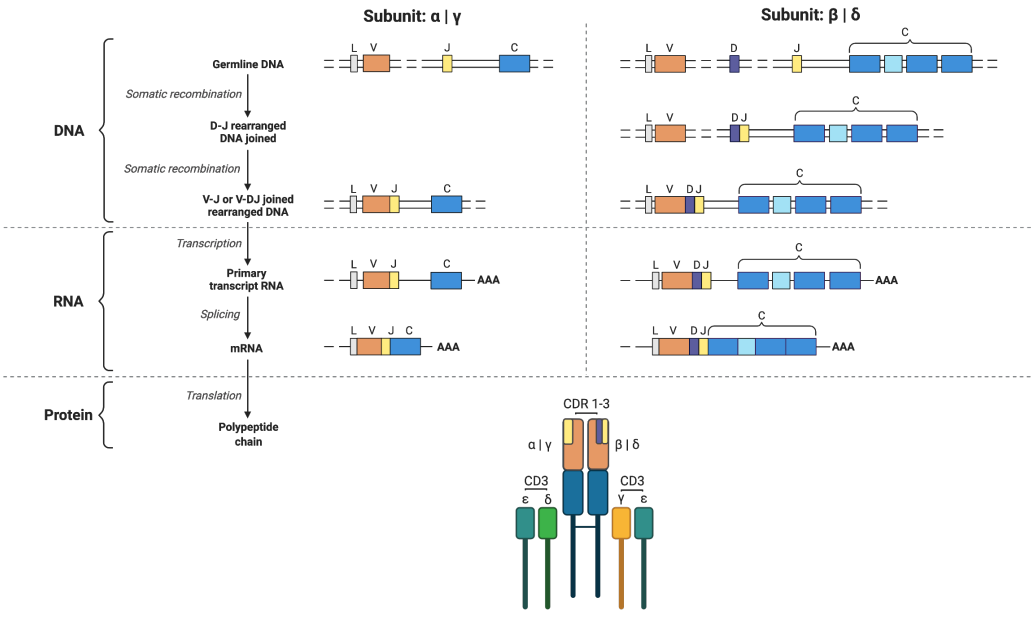


Figure 2.1: T Cell Receptor Genes V(D)J Recombination.
Rearrangement of the variable (V), diversity (D) and joining (J) gene segments to produce functional TCR chains. Adapted from "B Cell Receptors (Light and Heavy Chains)" by BioRender, 2020.

2.1.3 T Cell Subtypes and States

Conventional T cells are those defined as having a highly variable $\alpha\beta$TCR and co-receptor CD4 or CD8. They emerge from the thymus in a resting/naive state (CD45RA$^+$ CD45RO$^-$ CD62L$^+$ CCR7$^+$) and migrate through the peripheral blood, lymph nodes and tissues. Moreover, they recognise short peptide fragments bound to polymorphic cell-surface major histocompatibility complex (MHC) molecules enabling the detection of a myriad of foreign or altered-self antigens. Activation of $\alpha\beta$ T cells requires three distinct signals: 1) TCR interaction with antigenic peptide in the context of self-MHC 2) CD28 receptor binding to B7-1 (CD80) or B7-2 (CD86) molecules 3) cytokine signalling (e.g. IL-2, IL-12, type I interferons). A T cell that has been appropriately activated, undergoes clonal expansion followed by metabolic reprogramming and dramatic chromatin state changes. A network of transcription factors (TFs) orchestrates chromatin accessibility to regulate gene expression and establish enhancer-promoter interactions. Activated T cells (CD45RO$^+$) also increase their uptake of glucose and amino acids to enhance metabolic pathways such as glycolysis for promoting distinct differentiation profiles. They can differentiate into several functionally distinct lineages (Golubovskaya & Wu, 2016). CD4$^+$ T cells can differentiate into different T helper (Th) subtypes: Th1, Th2, Th9, Th17, Th22, T regulatory (Treg) or T follicular helper (Tfh) cells. Similarly, CD8$^+$ T cells will differentiate into cytotoxic T cells (CTLs) and, though less characterised, distinct subsets may arise analogous to their CD4 counterpart: Tc1, Tc2, Tc9, Tc17, Tc22 and CD8$^+$ Treg (Mittrucker et al., 2014; Mishra et al., 2021).

As described in Section 2.1.1, memory T cells arise following contraction of the effector pool at the end of a primary response. Memory T cells have a lower activation threshold than their naive counterpart, a higher proliferative capacity and can more easily access secondary lymphoid organs and peripheral tissues (Pennock et al., 2013). Providing immunosurveillance are effector memory (TEM), central memory (TCM), tissue-resident memory (TRM), peripheral memory (TPM), effector memory cells re-expressing CD45RA (TEMRA) and memory stem T cells (TSCM), each displaying their own migratory patterns (Martin & Badovinac, 2018). For instance, TCM are primarily found in lymphoid

organs and bone marrow, whereas TEM monitor peripheral tissues and TRM settle within organs (e.g. lungs, skin). Of particular interest to this thesis are dysfunctional states, mainly anergic and exhausted, which can arise at different stages of the T cell life cycle. The absence of CD28 co-stimulation or high exposure to co-inhibitory signals during the initial antigen-presentation process results in incomplete activation. This induces an anergic phenotype, a hyporesponsive state with low IL-2 production (Crespo et al., 2013). T cells can also become exhausted in response to chronic antigen stimulation. The gradual loss of T cell effector functions and upregulation of inhibitory receptors (PD-1, TIM-3, LAG-3, CTLA-4, TIGIT) are characteristic of exhaustion (Z. Zhang et al., 2020).

Unconventional T cells include natural killer T cells (NKT), $\gamma\delta$ T cells and mucosal-associated invariant T (MAIT) cells. Contrary to conventional ones, these cells express semi-invariant TCRs recognising non-peptide antigens and possess both innate and adaptive characteristics. NKT recognise glycolipid antigens bound with the non-classical MHC class I homolog CD1d. They can be further subdivided into invariant (iNKT) and diverse (dNKT) subsets that use an invariant TCRα (Vα24-Jα15) with limited TCRβ (Vβ11) chains or those with a more diverse repertoire, respectively (Krijgsman et al., 2018). Similarly, $\gamma\delta$ T cells which account for only 0.5-5% of all T lymphocytes can be classified on the basis of the δ chain they express, V$\delta1^{+}$, V$\delta2^{+}$ or V$\delta3^{+}$ (Manna et al., 2020). They are mainly found in the gut mucosa and do not require intracellular antigen processing, instead they directly recognise soluble or membrane-anchored antigens. Abundant in tissues, MAIT cells bind vitamin B-based antigens presented by the non-polymorphic MHC class I related-1 molecule (MR1) and have an intrinsic effector-memory phenotype, CD45RA^{-} CD45RO^{+} CD95^{Hi} CD62L^{Lo} CD44^{Hi} (Hinks & Zhang, 2020).

2.2 Tumour Microenvironment

The tumour microenvironment (TME) surrounding solid tumours is a complex highly heterogeneous milieu dominated by tumour-induced interactions. It plays a key role in cancer initiation, progression, metastasis, and drug resistance by providing a protective and growth-permissive niche.

2.2.1 Cancer Genome Evolution

Cellular integrity, growth, proliferation, structural and functional characteristics are determined by a complex world of biological interactions. A network of molecular entities composed of genes, proteins, transcripts and small molecules governs different aspects of the cellular organisation including metabolism, gene-regulation and signal transduction. In multi-cellular organisms, cells are an integral part of tissue and organ dynamics and development, thus, disturbances at the cellular level can extend to disease phenotype. Cancer is understood to be a disease of the genome. Tumour cells arise from a single most recent common ancestor cell (MRCA) which accumulated genomic and epigenomic changes driven by a Darwinian evolutionary process (Gerstung et al., 2020). Somatic aberrations like small insertions and deletions, genomic rearrangements, point mutations, copy number alterations and epigenetic alterations are passed on from generation to generation as passenger or driver mutations either tagging along or conferring a selective advantage on the cells. Gene defects causing loss-of-function mutations occur in tumour suppressor genes responsible for maintaining genome integrity, broadly classified as gatekeepers genes such as cell cycle regulators (e.g. PTEN, p14, p53) or caretaker genes such as DNA repair components (e.g. BRCA1/2, ATM). Conversely, alterations in oncogenes cause gain-of-function or dominant effects with a single mutation in genes like PI3K/AKT1 and EGFR being sufficient to stimulate uncontrolled cellular growth. Deregulated microRNAs (miRNAs) have also been found to contribute to the vast array of defects in cell survival, differentiation, division and DNA-damage responses (Si et al., 2019). These are small, 20-25 nucleotides in length, non-coding RNAs that regulate gene expression by targeting complementary messenger RNAs (mRNAs) and guiding RNA-induced silencing complex (RISC) to mediate their degradation or translational repression. Depending on their target genes they can be classified as oncogenes or tumour suppressors. As cells further accumulate mutations and genetic differences, a vast spectrum of variations and multiple cell lineages can arise. Cancer is thus a multifactorial, heterogeneous disease, whereby a single tumour can possess billions of cells varying drastically at the molecular level in terms of type and mutational burden.

2.2.2 Multicellular Composition

As cancer cells interact with non-transformed cells they give rise to a complex organ-like structure defined above as the TME (Fig. 2.2). Coordinating together, they sustain tumour growth by hijacking pre-existing vascular systems (angiogenesis) or by inducing the formation of new vasculogenic networks (vasculogenesis) for the supply of nutrients and oxygen. This also facilitates cancer cell dissemination into the circulation via the cardiovascular or lymphatic system and can lead to the establishment of metastatic colonies in secondary sites which accounts for 90% of cancer deaths. In addition to tumour cells, the TME itself encompasses a heterogeneous, dynamic and highly interactive network of stromal cells and immune cells co-evolving together.

The tumour stroma consists of specialised connective tissue cells and an extracellular matrix (ECM) composed of proteoglycans, polysaccharides, fibrous proteins (e.g. type I collagen) and glycosylated proteins providing the physical scaffolding and biochemical support for cellular constituents (Baghban et al., 2020). Stromal cells include endothelial cells, adipocytes, fibroblasts, stellate cells and pericytes that secrete pro-tumourigenic factors such as TGFβ, fibroblast growth factor 2 (FGF2), vascular endothelial growth factor (VEGF) and platelet-derived growth factor (PDGF). Cancer-associated fibroblasts (CAFs) arise as a result of FGF2 and play a central role in mediating ECM remodelling, influencing tumour progression by altering structural and signalling interactions. The release of TGFβ further promotes epithelial-mesenchymal transition resulting in a loss of epithelial cell polarity and adhesion. This helps them gain the ability to migrate and invade other tissues. Transformed cells can exit the primary tumour site and enter the circulation through newly formed blood vessels. Signalling proteins like VEGF promote the remodelling of tumour microvasculature by stimulating a highly organised sequence of cellular events which triggers the proliferation of endothelial cells and orchestrates their reassembly into tube-like structures forming the vessel lumen. Pericytes are recruited for structural support around the vascular tube.

Leukocytes and soluble mediators of the innate and adaptive immune system play multifaceted, inhibitory or tumour-promoting roles during cancer progression. The level of spe-

cific immune populations infiltrating the tumour can positively or negatively correlate with cancer patient prognosis. Tumour infiltrating lymphocytes (TILs) such as $CD8^+$ CTLs or $CD4^+$ Th1 cells exert or promote beneficial cytotoxic anti-tumour functions. Moreover, infiltrate levels have been found to predict favourable prognosis across multiple cancer types (Shimizu et al., 2019). Macrophages recruited to the TME on the other hand, evolve into M2-like tumour-associated macrophages (TAMs) that exert immune suppressive functions and produce stabilising ECM. TAM infiltration has been reported to correlate with worse patient prognosis and higher risk of metastasis (X. Qiu et al., 2022). Likewise, T regulatory cells which maintain immune homeostasis and inhibit anti-tumour immunity also correlate with poor prognosis in various cancers (Kim et al., 2020).

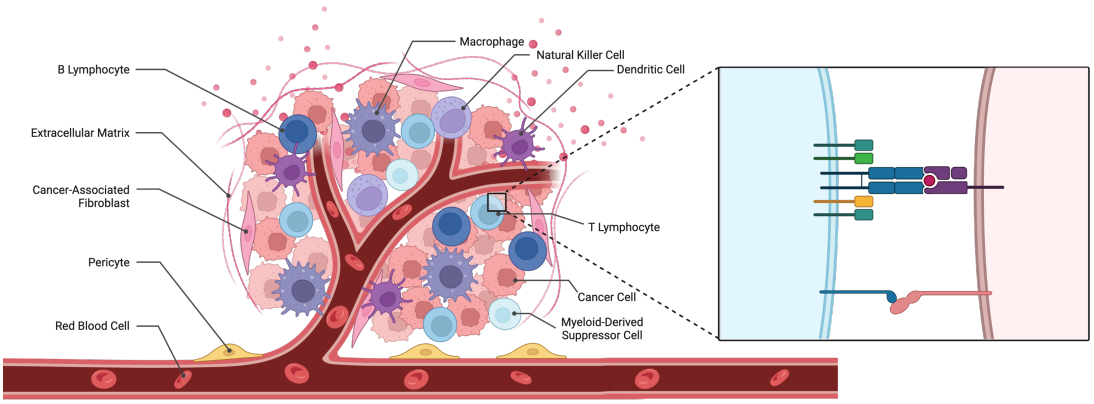


Figure 2.2: Multicellular Composition of the Tumour Microenvironment. An illustration showing the various cellular components of the tumour microenvironment with a focus on T cell-tumour cell interaction. Adapted from *Tumor Microenvironment* by BioRender.com.

2.2.3 Immunosuppressive Milieu

The TME is a highly immunosuppressive milieu that uses multiple strategies to escape host immune control. The process of cancer immunoediting is divided into three distinct phases: elimination, equilibrium and escape (Fridman, 2018). During the elimination phase, the innate and adaptive immune system coordinate together to identify and clear tumour cells. Surviving tumour clones help establish a subsequent equilibrium phase, balancing anti-tumour immune responses and tumour promoting mechanisms. As accu-

mulating genetic mutations further enable immune evasion, tumour cells persist into the escape phase, gradually gaining the upper hand over the immune system. Whilst malignant cells continue to grow and thrive in the inhospitable conditions they promote, the growing tumour mass poses many obstacles to immune cells, in particular T cells (Fig. 2.3).

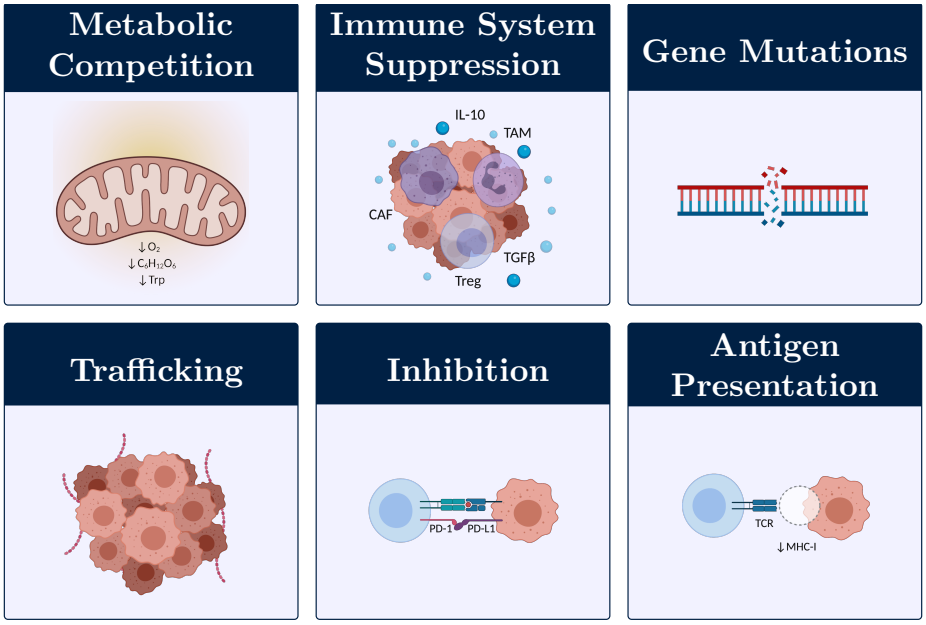


Figure 2.3: Obstacles Posed to T cells in the Tumour Microenvironment. Major challenges faced by T cells on the tumour battleground include metabolic competition with tumour cells, exposure to immunosuppressive molecules and tumour-promoting cell types, physical barriers restraining trafficking, direct inhibition through the stimulation of immune checkpoints and downregulation of antigen presentation-associated pathway proteins on tumour cells. Icons from Biorender.com.

Hypoxic regions, areas of reduced oxygen levels, and metabolite scarcity, including tryptophan and glucose, have detrimental effects on T cells and dampen their effector functions (Labani-Motlagh et al., 2020). Furthermore, local conditions within tumours promote the accumulation of lactate and acidosis. Acidic pH suppresses T cell functions and blocks T cell activation and proliferation (Bellone et al., 2013). Also hindering their survival are mechanisms of immune system suppression and direct inhibition through immune checkpoint engagement. Tumour cells promote the production of immunosuppressive cytokines

and cell types such as TGF-β, TAMs and T regulatory cells, as described in Section 2.2.2. In turn, these enhance expression of inhibitory receptors on T cells to promote self-tolerance. T cells expressing programmed cell death protein 1 (PD-1) on their surface can engage with its ligand, PD-L1, on immunosuppressive cell types or cancer cells. Upon binding, the PD-1/PD-L1 axis suppresses T cell activation through the TCR, survival and effector functions (Sharpe & Pauken, 2018). It promotes transition to an exhausted or anergic state and induction of the caspase cascade resulting in apoptosis. In addition, the high density of extracellular matrix and the increased interstitial fluid pressure act as physical barriers to T cell trafficking through the tumour. Similarly, T cell migration is guided by a gradient of chemoattractants. Molecular factors produced by tumour cells such as VEGF can block the induction of chemokines thereby reducing T cell infiltration. Moreover, genetic mutations can modulate expression of neoantigens on tumour cells and influence the level of intratumoural genomic heterogeneity, a major obstacle for targeted therapies as will be described in Section 2.4. Another mechanism commonly used by tumour cells for immune evasion consists of downregulating expression of MHC-I thereby minimising antigen presentation and avoiding immunosurveillance (Dhatchinamoorthy et al., 2021).

2.3 Interferon Signalling and Inflammatory Responses

Interferons (IFNs) consist of a family of monomeric cytokines classified into three types: type I (α, β, δ, ϵ, κ, τ, ω), type II (γ) and type III (λ). Mainly studied in the context of viral infections, they play a central part in the anti-viral immune response by inhibiting viral replication, driving infected cells into apoptosis, and inducing an anti-viral state in adjacent cells. By bridging innate and adaptive immune responses they modulate the survival, maturation, differentiation and migration of diverse immune cell types including T cells. Importantly, these signalling molecules induce expression of interferon-stimulated genes (ISGs) capable of modulating T cell immunity to direct not only anti-viral but also anti-tumour functions. The focus of this thesis is on type I IFNs which are generally produced in response to pathogen-derived motifs such as double-stranded RNA (dsRNA), single-stranded RNA (ssRNA), 5'-C-phosphate-G-3' CpG DNA, and viral glycoproteins.

Self-nucleic acids derived from damaged or dying cells can also stimulate their production.

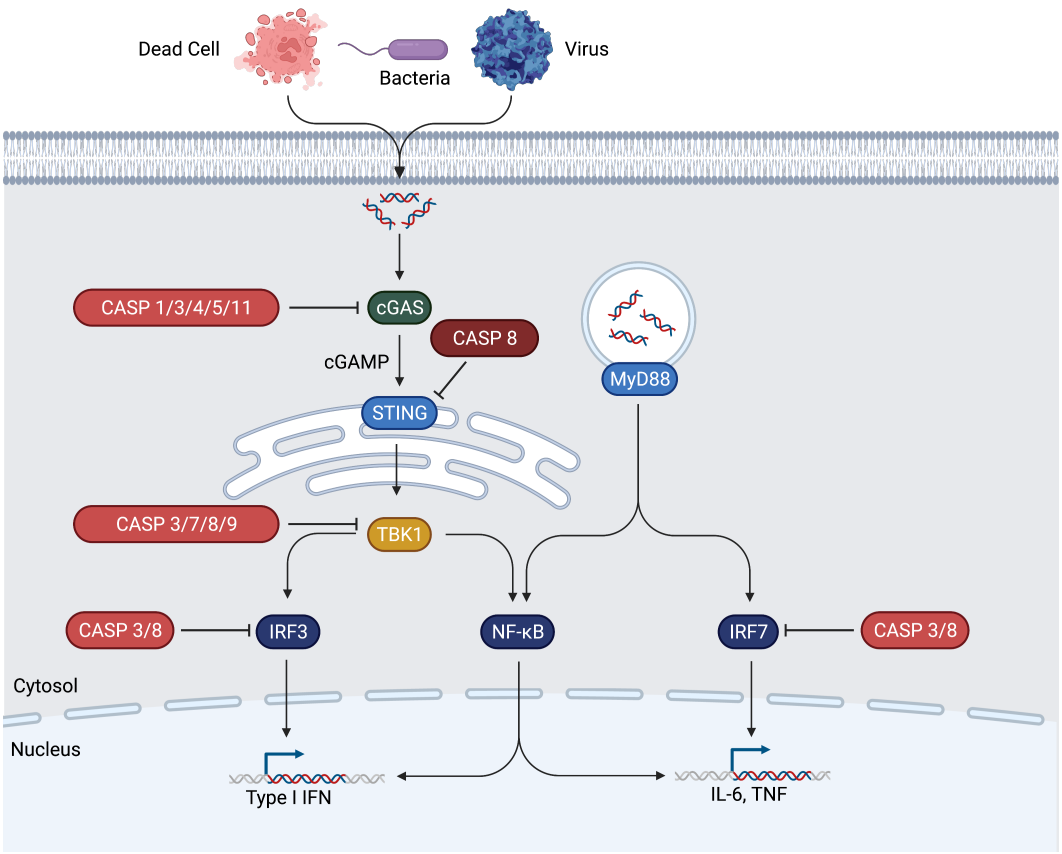


Figure 2.4: DNA-Sensing Mediates IFN Signalling.
Illustration depicting dsDNA-induced activation of the cGAS-STING signalling pathway and regulation by apoptotic caspases. Adapted from *cGAS-STING DNA Detection* by BioRender.com.

Type I IFN responses can be initiated through a number of pathways following the activation of retinoic acid-inducible gene I (RIG-I)-like receptors (RLRs), Toll-like receptors (TLRs), C-type lectin receptors (CLRs), nucleotide-binding domain and leucin-rich repeat-containing receptors (NLRs) or cyclic GMP-AMP synthase (cGAS). In recent years, there has been a particular interest in targeting the cGAS-stimulator of IFN genes (STING) signalling pathway (Fig. 2.4) for immunotherapy due to its central role in mediating inflammation and orchestrating innate and adaptive immunity in the context of cancer, infectious diseases and autoimmune and inflammatory diseases (Decout et al., 2021). Specifically, cytosolic DNA is detected by cGAS which will create a phosphodiester bond between the 2'-OH of GMP and 5'-phosphate of AMP producing cGAMP (Kwon &

Bakhoum, 2020). Upon binding to STING, also known as TMEM173, cGAMP triggers a conformational change that enables the recruitment of the serine/threonine-protein kinase TBK1. Once phosphorylated, STING acts as a scaffold protein for the subsequent phosphorylation of the type I IFN transcription factor IRF3. Type I IFNs that are produced will bind to IFN receptors (IFNARs) in a paracrine or autocrine manner, acting on the same cell or adjacent cells. Caspases, a family of cysteine proteases responsible for the initiation of apoptosis, are capable of negatively regulating this pathway by robustly cleaving multiple of its components.

Agonists of cGAS-STING pathway have been proposed as vaccine adjuvants to optimise the immunogenicity and efficacy of subunit vaccines (Herck et al., 2021). Similarly, clinical trials are underway to evaluate their use as anticancer immunostimulatory agents, alone or in combination with other therapies to potentiate therapeutic effects (Naour et al., 2020). For instance, DNA damage induced by radiotherapy will activate cGAS-STING signalling which plays a central role in coordinating radiation-induced tumour antigen-specific T cell responses. However, cGAS and STING deficiency are commonly reported in various cancers impacting their type I IFN-inducing capacity. As such, STING agonists are being tested for their ability to evoke favourable additive anti-tumour effects and have already been shown to promote infiltration of T cells into the TME. Consistently, the STING agonist DMXAA was found by N. Xu et al. (2021) to promote trafficking and persistence of engineered T cells in solid tumours of mouse tumour models (Section 2.4). It acts by generating a T cell-attracting chemokine milieu and altering the balance of immunosuppressive and immunostimulatory cell types. Paradoxically, chronic activation of this pathway can also have immune suppressive functions (Zhu et al., 2019). Tumour cells treated with STING agonists have been reported to upregulate expression of indoleamine 2,3-dioxygenase (IDO) and PD-L1, molecules that exert important tumour-related immunosuppressive functions to help immune escape (Lemos et al., 2016; M. Petrovic et al., 2021). Recent evidence suggests that chronic STING stimulation may also promote inflammation-aggravated cancer and tumour metastasis (Jiang et al., 2020). The mechanism through which STING activation promotes metastasis has yet to be elucidated. However, it is clear that inflammation and tumour development are inextricably linked, with the

deregulation of local inflammatory parameters being both a result and a cause of tumourigenesis that can drive systemic pro-metastatic inflammation (Fig. 2.5). Most importantly, though little is known about the role of intrinsic cGAS-STING in T cells, STING activation has been reported to induce T cell death (Larkin et al., 2017). Another study found that constitutively active STING impairs T cell proliferation by inducing mitotic errors (Cerboni et al., 2017). Of note, the cGAS-STING pathway also activates nuclear factor kappa B (NF-kB) leading to NF-kB-dependent pro-inflammatory cytokine expression. Although this transcription factor is known to be critical for T cell function and survival after TCR engagement, a recent study investigating the intrinsic function of NF-kB in tumour infiltrating T cells found that it promoted a dysfunctional CTL phenotype (C. Lu et al., 2020). More research is needed to reconcile conflicting studies. There is a clear need to delineate the beneficial or detrimental effects of cGAS-STING signalling on different cell types, particularly T cells, as the co-administration of cancer therapies with STING agonists continues to gain traction.

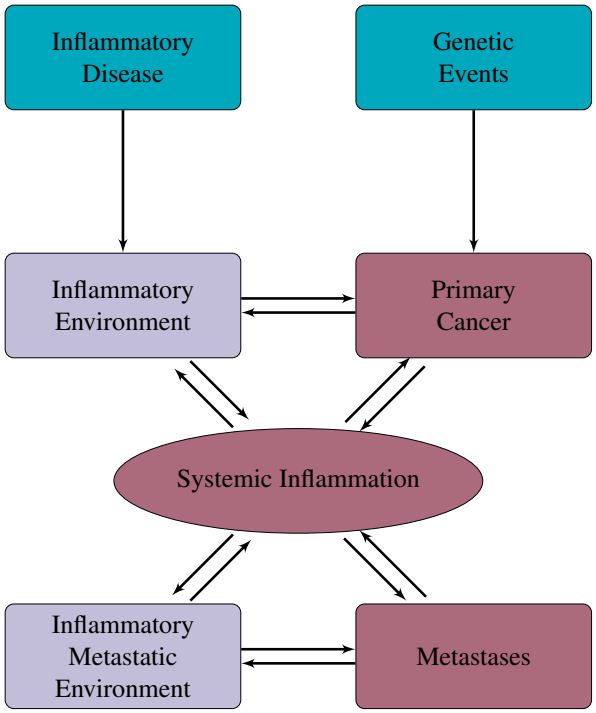


Figure 2.5: A Vicious Cycle of Inflammation in Cancer Patients.
Local and systemic inflammatory environments arise in cancer patients as a result of a sequence of reciprocal cause and effect between tumour promoting inflammation and inflammation predisposing and further promoting tumourigenesis. Adapted from Chechlinska et al. (2010).

2.4 Rewiring T Cells with Genome Editing and Synthetic Biology

The fields of genome engineering and synthetic biology have led to significant advances in biomedicine, paving the way for novel immunotherapies including patient-tailored reprogrammed immune cells. Adoptive cell transfer (ACT) initially relied on the isolation and ex vivo expansion of tumour-specific TILs, and their reinfusion into patients. The new era of ACT-based cancer therapies relies on engineering or reprogramming immune cells to yield more sophisticated cancer-targeting agents. Genome engineering has been successfully applied to rewire the behaviour of a variety of cell types, including dendritic cells, natural killer cells, and T cells. For instance, cytotoxic T cells can be redirected specifically to target cancer cells through the expression of highly specific tumour-recognising chimeric antigen receptors (CARs). A common strategy consists of targeting antigens abundantly expressed on tumour cells but minimally expressed in healthy tissue, coined tumour-associated antigens (TAAs), which will trigger a signalling cascade that induces cytotoxic effects upon binding. CAR T cell therapies like CD19-targeted tisagenlecleucel have been granted FDA approval for their proven efficacy against certain haematological cancers such as B cell acute lymphoblastic leukaemia (Maude et al., 2018). However, these therapeutic modalities have shown limited efficacy against solid tumours. As described in Section 2.2.3, the highly complex TME supporting and surrounding solid tumours both displays a large degree of antigen heterogeneity, and acts itself as a physical barrier to cellular infiltration, making it more difficult for engineered cells to find, diffuse, and survive in the tumour. A better understanding of the biology underlying CAR T cell interactions with the TME will be critical to overcome these hurdles. Furthermore, genetic variations between patients translate into different levels of therapy efficacy and toxicity. Engineering designs must primarily ensure their safety and effectiveness. Current research aims to address the multiple barriers engineered cells need to overcome when transplanted into a patient: ensuring proper trafficking to the tumour site, efficient tumour recognition and killing, proliferation and persistence for clinical efficacy, resistance to the immunosuppressive TME, and implementing control mechanisms to avoid off-tumour effects. Cytokine

release syndrome (CRS), a life-threatening systemic inflammatory response induced by elevated IFN-γ and TNF-α is the most common adverse event (Yunshuo et al., 2021). Integrating immune engineering design with personalised oncology, which tailors drugs to the genetic characteristics of a patient's tumour, will certainly lead to more precise and effective treatments in the future. High throughput omics methods which will be discussed in Section 2.5 hold much promise for precision medicine and optimised cell-based therapies.

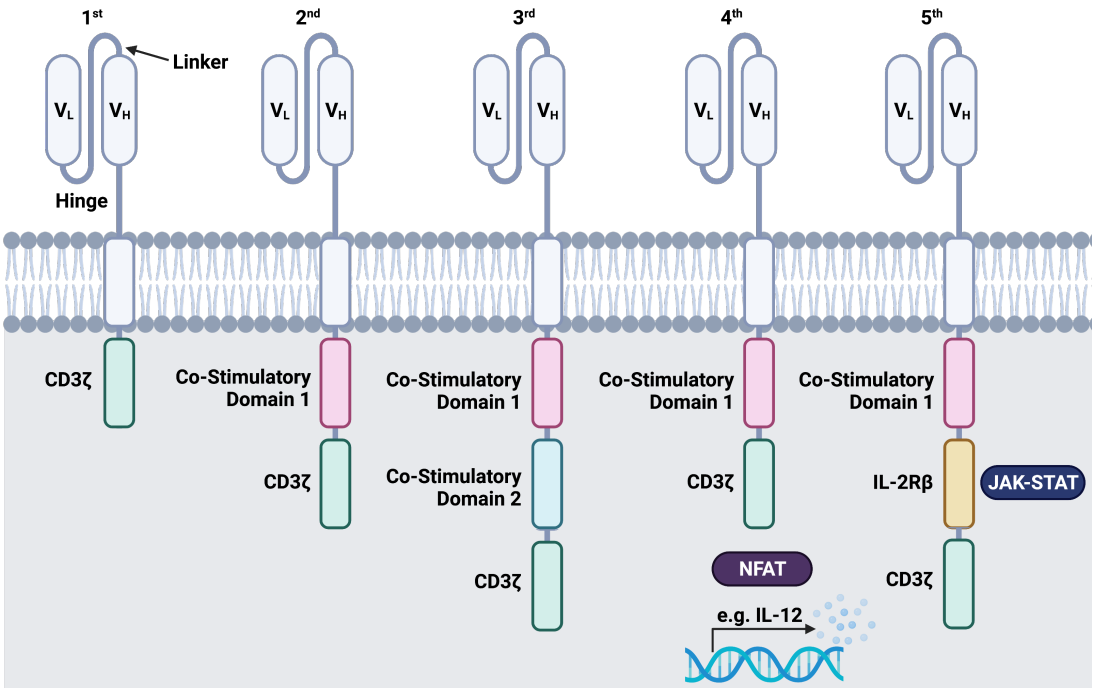


Figure 2.6: Generations of CAR Designs
Five different generations of CARs exist at the time of writing. Across generations changes were mainly made to the intracellular signalling domain. The first generation consists of the CD3ζ alone. The second generation added the co-stimulatory domain of either CD28 or 4-1BB whilst the third generation combined the two. The fourth generation also called T cells redirected for antigen-unrestricted cytokine-initiated killing (TRUCKs), is armoured with transgenic cytokines regulated by a NFAT-inducible expression cassette. The most recent generation contains the truncated intracellular domain of a cytokine receptor to induce JAK-STAT signalling. Created with BioRender.com.

Immune cell engineering is a rapidly evolving field which has led to the design of an arsenal of synthetic receptors including, but not limited to, multiple generations of CARs,

bispecific T cell engagers (BiTEs) and synthetic notch-based systems (SynNotch). Successive generations of CAR T cells differ in their intracellular signalling domains with additional costimulatory domains being added to improve T cell proliferation, persistence and potency (Fig. 2.6). More sophisticated T cells like SynNotch use gating strategies to improve precision of therapeutic T cell responses. For example, SynNotch AND-gate circuits can be designed to respond to a pair of antigens instead of a single target (Roybal et al., 2016). Specifically, a T cell will get activated and kill the target cell if and only if both target antigens are present. Conversely, BiTEs consists of antibody constructs with two binding domains joined by a flexible peptide linker (S. Zhou et al., 2021). They redirect T cells to specific target cells by simultaneously binding molecules expressed on the surface of each. Once the immunological synapse is formed, activated T cells will secrete granzymes and perforin to induce tumour lysis. At the time of writing, blinatumomab is the only BiTE that has been granted FDA approval (Tian et al., 2021).

T cells can be further engineered using CRISPR-Cas9 gene editing to have more sophisticated intracellular circuits with synthetic functionalities. At the forefront of genome editing technologies is the Fulga Lab (University of Oxford). CRISPR-Cas9 technology paved the way for genome editing in mammalian cells. This RNA-programmable system can be used to generate gene knockouts or knockins through the generation of a single guide RNA (sgRNA) which directs the Cas9 nuclease to the target site. Guide RNAs are generally transcribed by RNA polymerase III under the control of a constitutive human U6 promoter. This, however, restricts gRNA expression, greatly limiting the flexibility of the CRISPR/Cas9 system. Transfer RNAs (tRNA) are well-conserved RNA molecules with an intrinsic polymerase III promoter. The RNA polymerase III transcribed tRNA precursor is cleaved by Ribonuclease P (RNase P) and Ribonuclease Z (RNase Z) to generate mature 5' and 3' ends, respectively. Endogenous tRNA-processing systems are being increasingly adapted for gRNA production and have been designed for multiplex genome editing. Using a polycistronic-tRNA-gRNA strategy, multiple gRNAs can be expressed under the same promoter. To account for the lack of spatial and temporal control of expressed gRNAs, the Fulga Lab engineered tRNA variants that allow specific RNA polymerase II transcription whilst retaining moderate processing abilities. Transcription

through RNA polymerase II results in a more flexible CRISPR/Cas9 system with guide RNAs being expressed in a more precise and controlled manner. (Knapp et al., 2019). In addition, the Fulga Lab developed a powerful technology for modulating gene expression using synthetic microRNA response elements (MREs) called microRNA silencing-mediated fine-tuners (miSFITs). The miSFIT system was used to explore the impact of antigen expression levels on tumour selection, both in vitro and in vivo (Michaels et al., 2019). Modulating expression level of the immunogenic protein ovalbumin was found to affect the anti-tumour immune response in that higher expression levels correlated significantly with T cell activation and tumour control. As will be seen in Chapter 5, this new technology can also be applied to optimise the level of synthetic receptor expression on engineered T cells.

Engineered T cells can be further combined to other therapies to produce synergistic effects (Fig. 2.7). In order to enhance their efficacy, particularly in the context of solid tumours, checkpoint inhibitors such as anti-PD-1 antibodies can either be combined or incorporated into CAR T cells (Hosseinkhani et al., 2020). Recent studies have looked at combining CAR T cells with oncolytic viruses (Rezaei et al., 2022). Naturally existing or genetically engineered oncolytic viruses selectively replicate in cancer cells. Lysis of infected cells causes the release of danger signals which can have an immunostimulatory effect that facilitates CAR T cell trafficking, proliferation and persistence (Guedan & Alemany, 2018). Similarly, the release of tumour-associated antigens from dying tumour cells can induce anti-tumour responses that mitigate tumour escape by antigen loss. To counter together the hostile environment of the TME, oncolytic viruses can be armed with therapeutic transgenes to support CAR T cells. For example, one study armed CAR T cells with leptin to provide metabolic support for tumour infiltrating T cells against glucose starvation (Rivadeneira et al., 2019). Furthermore, to circumvent the lack of tumour-specific antigens, Evgin et al. (2022) designed CAR T cells specific to both tumour and virally encoded epitopes, thereby boosting their proliferation in the TME when used in conjunction with oncolytic viruses. In tumour cells with an intact cGAS-STING cascade, the free DNA released upon cell lysis could trigger the cGAS-STING signalling pathway introduced in Section 2.3 to enhance CAR T cell trafficking (N. Xu et al., 2021).

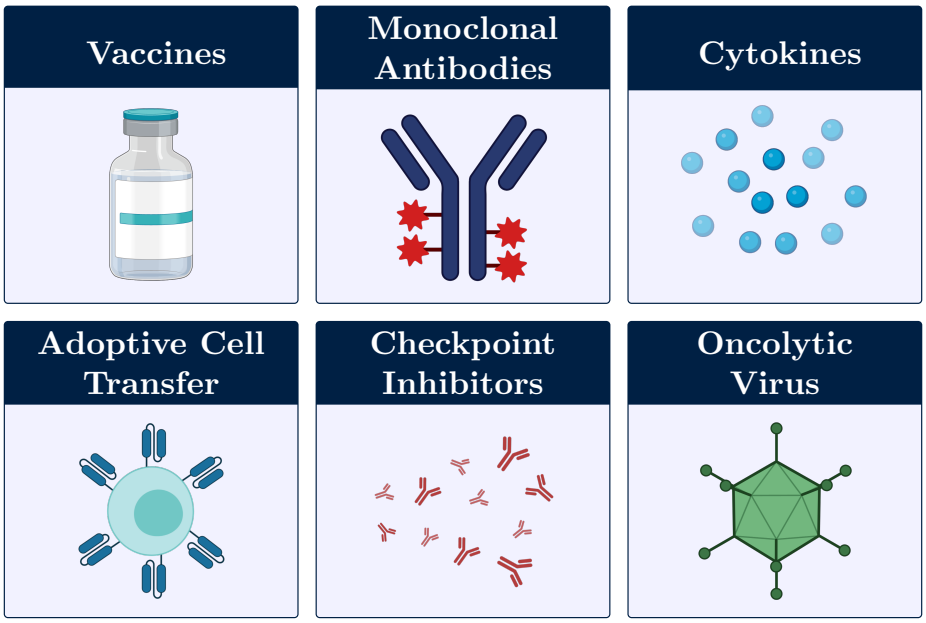


Figure 2.7: Types of Immunotherapy.
Cancer therapies that harness the immune system include vaccines, monoclonal antibodies, cytokines, adoptive cell transfer, checkpoint inhibitors and oncolytic viruses. Icons from Biorender.com.

2.5 Multi-Omics and Imaging Technologies for Therapeutic Insights

Advances in omics technologies have enabled the quantification of biological molecules at multiple levels of the central dogma including the genome, epigenome, transcriptome, proteome, and metabolome (Fig. 2.8). Omics layers combined can provide valuable insights into the flow of biological information to better navigate the complex intracellular world. Imaging technologies are used in cell biology to interrogate cellular landscape and architecture. Driving the next generation of research are spatial omics and multiplexed imaging technologies which aim to bridge the gap between quantified expression levels and the spatial context of tissue sections. The integration of multiple dimensions of information is bound to advance the understanding of cellular crosstalk and how different biological layers contribute to the emergence and development of malignancies, helping guide improved therapeutic interventions.

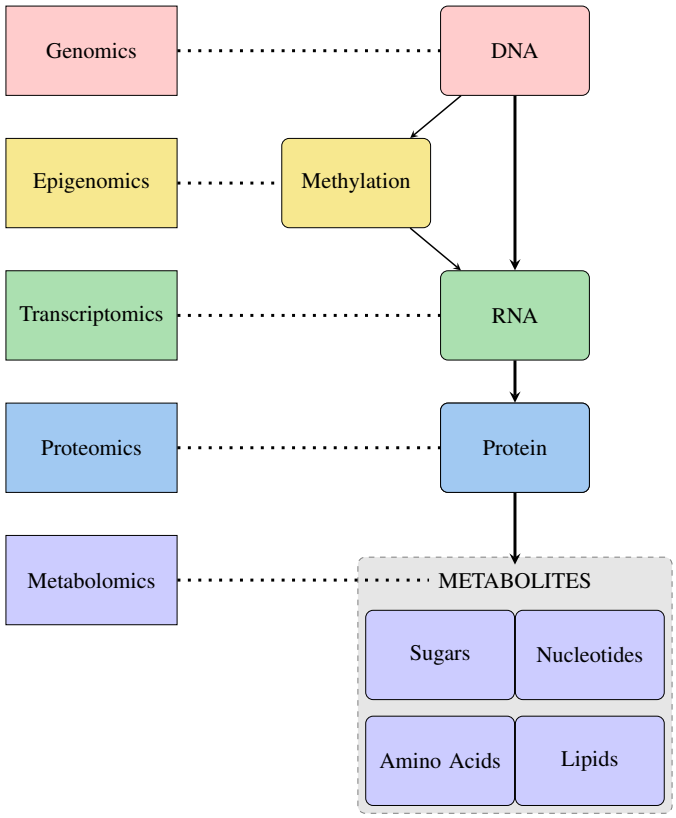


Figure 2.8: Schematic of the Omics Cascade.
Diagram showing the flow of information from DNA to Metabolites and associated omics disciplines.

The haploid human genome consists of a long string of approximately 3 billion DNA bases (A, C, T, G) packaged into 23 chromosomes (Nurk et al., 2022). Our ability to read nature's genetic blueprint is revolutionising medical research by helping decipher disease mechanisms at a much greater level of granularity than ever before. Starting at the DNA level, next generation sequencing (NGS) technologies like whole genome sequencing (WGS) can be used to determine the sequence of nucleotides in a cell (Yin et al., 2019). Single-nucleotide polymorphisms (SNPs), simple sequence repeat (SSR) markers and structural variations such as copy number variations (CNVs), translocations, inversions, duplications, deletions and insertions can be detected. Depending on the application, more cost-effective alternatives such as whole exome sequencing (WES) or targeted sequencing panels will restrict sequencing to the exons ($\sim$1% of the genome) or selected regions within genes, respectively (Seaby et al., 2016). If the goal is to determine whether

a mutation for a particular disease is present, DNA microarray can also be used and has become relatively inexpensive. Real-time analysis of DNA or RNA sequences is possible using novel platforms like Oxford Nanopore which will be the focus of Appendix A.

Epigenetic regulations such as DNA methylation and histone modifications act as molecular switches turning on or off expression of certain genes. Bisulfite genomic sequencing can detect methylation occurring on the fifth carbon position of cytosine residues, predominantly found at CpG dinucleotide-rich regions in promoters (Yuanyuan & Tollefsbol, 2011). Treatment with sodium bisulfite converts unmethylated cytosine residues to uracil, leaving methylated ones intact. PCR amplification will convert all uracil residues to thymidine before DNA methylation status can be interpreted from sequencing analysis. Histone modifications and transcription factor DNA binding sites can be studied by chromatin immunoprecipitation sequencing (ChIP-seq) which involves isolating protein-bound sheared DNA (Furey, 2012). Formaldehyde is used for DNA-protein crosslinking and DNA fragments are generated through sonification-induced hydrodynamic shearing. Chromatin accessibility on the other hand is commonly studied using assay for transposase-accessible chromatin using sequencing (ATAC-seq). As the name suggests, ATAC-seq uses the transposase enzyme to fragment accessible DNA and simultaneously ligate short oligonucleotide adapters prior to sequencing (Buenrostro et al., 2015). Long-range chromatin interactions and spatial genome organisation can be identified using chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) or the chromosome conformation capture technique Hi-C (X. Li et al., 2017).

The main focus of this thesis is on the transcriptome which consists of the complete set of RNA molecules at any one time (e.g. mRNA, rRNA, tRNA, miRNA, lncRNA, circRNA). Traditionally, RNA microarray has been used to profile predefined gene transcripts through hybridisation whilst bulk RNA-seq provides full sequencing of the whole transcriptome, thereby allowing for novel RNA and RNA variants to be captured. More recently, powerful techniques have been developed to examine the transcriptome at a finer-scale. Whereas RNA-seq averages gene expression across a population of cells, single-cell RNA-seq (scRNA-seq) measures transcript expression at the level of individual cells

(Haque et al., 2017). It is ideal to study heterogeneity and can be extended to study full-length BCR (scBCR-seq) and TCR (scTCR-seq) sequences of B and T cells found in a sample (Valkiers et al., 2022). Another novel methodology consists of single-nucleus (snRNA-seq) sequencing which profiles the transcriptome at the level of isolated nuclei (Fischer & Ayers, 2021). Cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq), on the other hand, uses antibodies conjugated to oligonucleotides to couple the measurement of surface proteins with scRNA-seq, adding an additional layer of valuable information.

Proteomics, the study of the protein repertoire, is closer to molecular phenotype than transcriptomics with post-translational modifications having a vital role in diversifying protein functions (Vistain & Tay, 2021). In addition, there is an ongoing debate on whether transcript levels correlate with protein levels which can further vary across cell types and tissues. Though it is steadily gaining momentum, proteomics continues to lag behind other omics disciplines. High-throughput techniques only recently started to be developed due to the extreme complexity of proteomes and methods are not yet available for whole-proteome amplification. The field relies on chromatography techniques coupled to mass spectrometry (MS) or nuclear magnetic resonance (NMR) to isolate and quantify abundance, modifications and interactions (Aslam et al., 2017). Like for proteomics, metabolomics, the study of small molecules typically less than 2 kDa (e.g. amino acids, nucleotides, lipids, sugars), relies on MS and NMR approaches. Metabolites include the building blocks of all other comics disciplines and the substrates used to power biochemical pathways.

Complementing omics technologies are state-of-the-art microscopy systems that can be used to decipher cellular behaviour and architecture. Multi-dimensional visual representations of biological systems can be obtained using a range of imaging modalities. From imaging individual cells at a nanoscopic resolution to biological tissues and anatomical structures at a microscopic level, these provide insights into spatial and temporal aspects of biological processes. Cellular biology greatly benefits from advances in microscopy techniques, ranging from light-based, super-resolution (e.g. TIRF, SIM, STORM, PALM) and

confocal microscopy, to electron-based microscopy (e.g. SEM, TEM, Cryo-EM). These allow for the detection and analysis of cellular components at high-resolution, and can be combined to fluorescent techniques to better examine cellular behaviour and responses to internal and external cues. State-of-the-art technologies include a range of motorised imaging techniques and labelling methods. Imaging relies on the creation of contrast, at a molecular level this can be achieved using fluorescent probes or antibodies to depict structures of interest. Immunofluorescence microscopy is a particularly robust technique which gives rise to a high resolution of the cellular entities being studied. These can be targeted in live or fixed cells with antibodies linked to a fluorophore or by genetically encoding fluorescent tags to endogenous proteins. Similarly, single-molecule fluorescent in situ hybridisation (smFISH), uses short, fluorescently labelled probes for the detection of single RNA transcripts (Fig. 2.9). Thus, oligonucleotides that hybridise to their complementary sequence, allow for the visual quantification of RNA levels giving insight into the transcriptional regulation of gene expression (Maekiniemi et al., 2020). These labelling methods can be combined to imaging technologies such as super-resolution microscopy that break the diffraction limit of regular microscopes, or spinning disk confocal microscopy which enable fast acquisition of spatio-temporal information. By providing a detailed visualisation of biological processes including protein and transcript location and their association, these methods can be used to accentuate the effect of varying genetic profiles on cellular dynamics and architecture.

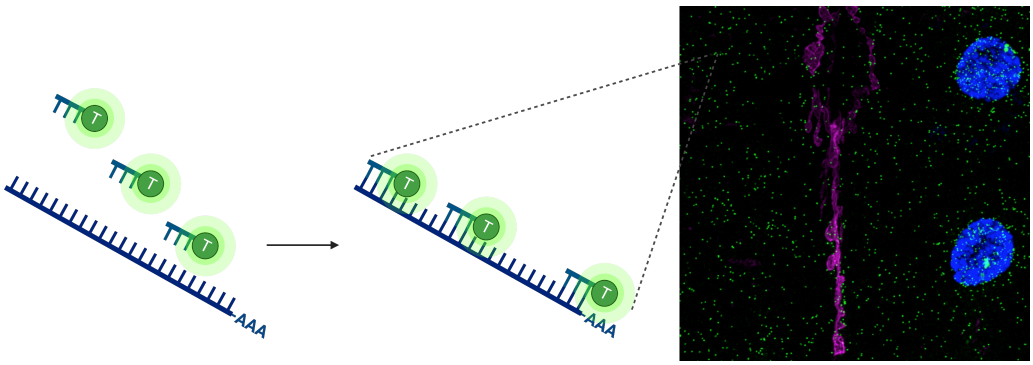


Figure 2.9: Single Molecule Fluorescence In Situ Hybridisation.
Short fluorescent-labelled DNA probes hybridise to target RNAs in situ. Schematic created using a smFISH image of the neuromuscular junction taken by Dr. Joshua Titlow (Davis Lab) and part of the author's first year rotation project in the Davis Lab (Department of Biochemistry).

At the intersection of omics and imaging came spatial transcriptomics, a range of methods for the high-dimensional assessment of gene expression in a tissue (Marx, 2021). In recent years, multiple platforms have been developed to assign cell types to spatial locations by measuring the number of RNA transcripts in tissue whilst preserving information on positional organisation. Broadly, spatial transcriptomics techniques fall into one of two categories: sequencing-based or imaging-based (Rao et al., 2021). The former consists of recording spatial location at the point of RNA capture prior to NGS. This can be done by overlaying the tissue onto spatially-barcoded microarray slides that add a unique positional molecular barcode to polyadenylated mRNAs upon their release from the tissue before undergoing reverse transcription. Similarly, microdissection techniques use gene-specific photocleavable probes to iteratively release mRNAs from regions of interest using UV light. Following NGS, the barcode of each read is used to map back each transcript to its original location in the tissue. Conversely, imaging-based spatial transcriptomics commonly relies on the use of fluorescently-labelled complementary probes for in situ hybridisation (ISH) prior to microscopy, akin to smFISH. Other imaging-based methods include in situ sequencing (ISS) which uses rolling circle amplification (RCA) by attaching padlock probes to targeted transcripts and subsequently catalysing the formation of a circularised fluorescently-labelled DNA (H. Lee et al., 2022). As will be seen in Chapter 4, all methods suffer from their own technical limitations and trade-offs. For instance, transcriptome-wide techniques using NGS tend to have lower capture efficiencies than ones targeting specific genes with a panel of probes. On the other hand, techniques like highly multiplexed smFISH can target a limited number of genes due to optical crowding. Other factors to consider when selecting a spatial transcriptomics platform include spatial resolution, capture area, tissue type compatibility and cost.

All omics and imaging experiments are critically dependent on computational methods for their correct manipulation, integration and interpretation. They generate large-scale complex datasets that require cutting-edge analysis techniques to generate biological knowledge. However, the lack of gold standard datasets incorporating multiple omics continues to challenge the systematic evaluation of computational methods. Despite the obvious need for standardised strategies, the field of biology continues to be fragmented in its

methods, which poses a well-known challenge for big data biology and partially fuels the reproducibility crisis. Of importance to this thesis, genomics is a very fast-moving field, and computational approaches are continuously devised to work with imaging or sequencing data. For instance, imaging-based spatial transcriptomics leverages classical image-processing tools to process raw images of fluorescent spots, match them to genes and assign them to cells. Imaging analysis pipelines generally comprise a series of steps to correct images for acquisition artefacts before subsequent cell segmentation and characterisation tasks can be undertaken. Digital images are comprised of pixel values, along with embedded metadata such as bitdepth, image type, and pixel sizes which can be read into computational algorithms as numerical data. Pixels collectively represent objects or features of images, their spectral properties, size, shape, local homogeneity, and their spatial relationship with surrounding pixels can be used to recognise underlying information in images. Objects in biological images such as cellular organelles or fluorescent probes representing transcripts in the case of spatial transcriptomics, are thus represented by a collection of pixels. In order to visualise nuclei cells are by convention stained with a blue-fluorescent DNA dye, 4',6-diamidino-2-phenylindole (DAPI), that preferentially binds AT regions of dsDNA. Moreover, DAPI staining can be used to infer cell locations and boundaries during cell segmentation. For biological objects to be detected, algorithms are devised to look for local features within the image domain that distinguish them from their immediate surrounding (e.g. corners, edges, colour, intensity). These patterns within subsets of the image domain help describe the regions of interest (ROIs), and can be used to find similar areas in other images, a trivial concept for automated object recognition in both supervised and unsupervised image analysis. Likewise, the correct spatial reconstruction of scRNA-seq data and cell-type deconvolution through the identification of spatially variable genes are critical steps in the analysis of sequencing-based spatial transcriptomics. In all areas of genomics, computational sciences plays a key role in translating biological data into meaningful scientific findings.

CHAPTER 3

Comparative Analysis of T Cell Identification Methods for Single-Cell RNA Sequencing Data

3.1 Overview

Single-cell RNA sequencing is a powerful approach to investigate biological systems by dissecting gene expression at a single-cell resolution. The resulting transcriptome which is a measure of the relative number of mRNAs in a cell at any one time correlates with cellular traits and changes in cellular states. T cells can be categorised into subtypes, each having their own characteristic expression profile. As T cells encounter their cognate antigen, or as they get exposed to extracellular stimuli, gradual changes in gene expression can trigger transitions between states. A bottleneck in scRNA-seq analysis continues to be cell type identification, particularly when distinguishing between closely related populations. In silico cell labelling methods ranging from correlation-based annotation and supervised classification to marker gene-based annotation are continuously being developed. Benchmarking studies have compared the performance of annotation methods primarily on major cell labels. Comprehensive evaluations of these methods on more refined labels such as T cell subsets are lacking. In this chapter, selected methods are evaluated for their ability to assign T cell subtype identities. The knowledge gathered therein will be used in the next chapter to explore the heterogeneity of T cell populations in clinical genomics data.

3.2 Introduction

Advances in single-cell sequencing technologies have made it possible to uncover transcriptional signatures in health and disease and study cell-to-cell variation at much higher resolution than conventional methods. In the field of immunology, this translates into more comprehensive studies of the human immune system through the characterisation of the many cell types and states underlying immune responses. For instance, single-cell RNA sequencing has been increasingly used in infectious disease research to understand responses to vaccination and differences in disease severity. During the COVID-19 pandemic, single-cell multi-omics studies were performed on PBMCs from large cohorts of patients with varying severities of COVID-19. Research findings resulted in the identification of IFN-related transcriptional signatures, altered cellular profiles and immune cell proportions characteristic of severe disease (J. S. Lee et al., 2020; Stephenson et al., 2021). Other studies featuring scRNA-seq investigated risk factors such as age, gender and smoking. Lewis et al. (2021) reported age-associated exacerbation of IFN signalling accompanied by T cell hyperactivation and increased exhaustion. In addition, scRNA-seq enabled Muus et al. (2021) to demonstrate tissue- and cell-type-specific changes in expression levels of angiotensin-converting enzyme 2 (*ACE2*) and accessory proteases that positively correlate with being a male, smoking and aging. Moreover, scRNA-seq was used to compare humoral and cellular responses to the Pfizer/BioNTech (BNT162b2) and Moderna (mRNA-1273) mRNA-based vaccines (Arunachalam et al., 2021; Sureshchandra et al., 2021). Many such studies were reviewed by the author for the University of Oxford Immunology COVID-19 Literature Initiative (Oxford COVID-19 Literature Initiative, 2020) attesting to the growing use of single-cell genome sequencing in immunology research. Likewise, scRNA-seq is providing unprecedented insights into the heterogeneity of the tumour microenvironment and is promising to play a pivotal role in precision medicine with the tremendous potential to help revolutionise cancer treatment.

Despite the rapid expansion of scRNA-seq studies and the substantial advances in experimental methods, significant analytical challenges remain. Identifying cell types and states is a crucial step in scRNA-seq analysis and traditionally involves the unsupervised

clustering of cells followed by the manual annotation of clusters based on genes that are differentially expressed between them. This approach is known to be time-consuming, error-prone and has limited reproducibility. Furthermore, it heavily relies on the assumption that dimensionality reduction techniques such as uniform manifold approximation and projection (UMAP) accurately preserve local and global cell relationships. As shown by Chari et al. (2021), UMAP results in extensive distortions that can impact the downstream quantitative and qualitative analysis of cell relationships and lead to contradictory interpretations. Additional challenges beyond the curse of dimensionality include the inherent stochasticity and sparsity of scRNA-seq data (Lahnemann et al., 2020). Confounding factors such as technical and biological artefacts, and gene dropout events in which only a fraction of the transcriptome of each cell is captured need to be correctly addressed. Low capture efficiency can result in discrepancies between cells of the same type with some having disproportionately a large number of genes that fail to be detected but are found to be expressed at a certain level in others. On the other hand, recent papers are suggesting to leverage dropout events for cell type identification since dropout patterns seem to be shared by genes in the same pathway (P. Qiu, 2020). Gene expression matrices are also highly heteroscedastic meaning that counts vary more for highly expressed genes than lowly expressed genes. As such, a change in expression value from 0 to 100 is more significant than a change between 1,000 and 1,100 (Ahlmann-Eltze & Huber, 2021). The intrinsic sparsity and heteroscedasticity of scRNA-seq data can hinder downstream analyses if not handled appropriately by data analysis models and methods.

Single-cell genomics is a fast-moving field which remains in its infancy and has yet to be standardised and gold-standard analysis platforms developed. In order to accurately deconvolve immune system heterogeneity, alternative cell type identification methods are an active area of research along with the development of cell type atlases and reference datasets. One approach for cell type annotation consists of using multiple correlation measures between a reference dataset with known cell labels and the query dataset. Another approach automates the traditional process of mapping gene-expression markers to cell types using basic scoring systems. More sophisticated methods rely on supervised or semi-supervised learning, training classifiers on labelled reference scRNA-seq datasets.

Every method has advantages and limitations which prompted the recent publication of many benchmarking studies evaluating and comparing cell type classification methods (Abdelaal et al., 2019; Pasquini et al., 2021; Huang et al., 2021; Xie et al., 2021). However, the focus of these studies has been on their ability to discriminate between major cell types instead of closely-related cell populations which is known to be a significantly more challenging task. In this chapter, some of the most popular cell identification methods and references are evaluated on publicly available scRNA-seq datasets for their ability to identify T cell subtypes. This work will feed into the subsequent chapter to help supplement traditional cell identification methods and more accurately navigate the T cell landscape.

3.3 Materials and Methods

3.3.1 Single-Cell RNA Sequencing T Cell Datasets

Publicly available scRNA-seq datasets of known or experimentally isolated T cell subtypes (Table 3.1) were gathered from either Gene Expression Omnibus (GEO) or 10x Genomics. All cells had been isolated from healthy human donors. Gene expression data consists of sparse matrices, $X(m \times n)$, containing the number of RNA molecules for each feature detected in each cell. The raw sequencing data had already been processed and aligned to give count matrices. Additional filtering was implemented to the ones in the original publications. Cells with less than 500 genes and genes expressed in less than 3 cells were excluded. A quality control (QC) threshold of 10% mitochondrial (MT) transcripts was used. These additional thresholds were decided upon examining the distributions of QC covariates for each dataset. However, it was not possible to further filter out cells based on the fraction of MT counts from the MAIT dataset because all MT genes had been filtered out by the authors once a 20% threshold had been applied (Vorkas et al., 2022) and the raw data could not be found. From each of the nine T cell populations, 1,000 cells were randomly selected to construct the final T cell object with an equal proportion of each subtype.

Table 3.1: Single-Cell RNA Sequencing T Cell Datasets.
Publicly available T cell scRNA-seq datasets used to test T cell identification methods.

Category	Subtype	Source
Conventional	$CD8^+$ Naive $CD8^+$ Cytotoxic $CD4^+$ Helper $CD4^+$ Regulatory $CD4^+$ Naive $CD4^+$ Memory	Zheng et al. 2017
Unconventional	$\gamma\delta$ NKT MAIT	GSE128223 GSE128243 GSE182239

Conventional T Cells

All conventional T cell subtypes were downloaded from the 10x Genomics website and had been deposited by Zheng et al. (2017). $CD4^+$ Naive, $CD4^+$ Helper, $CD4^+$ Regulatory, $CD4^+$ Memory, $CD8^+$ Naive and $CD8^+$ Cytotoxic T cells were all enriched from fresh PBMCs by fluorescence-activated cell sorting (FACS) yielding $>95\%$ pure populations. Cells were sequenced on Illumina NextSeq 500 High Output with reads per cell ranging from 19,000 to 28,600. Alignment and processing of FASTQ files had been completed using Cell Ranger (v1.1.0) and human genome reference GRCh37 (hg19) (Single Cell Gene Expression Dataset by Cell Ranger 1.1.0, 10x Genomics, 2017).

Unconventional T Cells

Preprocessed data from three scRNA-seq studies (GSE128223, GSE128243, GSE182239) were downloaded from the GEO database (Barrett et al., 2013). $\gamma\delta$ T cells from Pizzolato et al. (2019) were FACS-purified from three healthy individuals and consist of TCRVδ1 and TCRVδ2 lymphocytes. Only cells from the two cytomegalovirus negative donors were used. The cells had been sequenced using 10x Genomics Single-Cell 3' V2 Chemistry and the FASTQ files aligned on the GRCh38 transcriptome. NKT cells were isolated from the PBMCs of three healthy donors by L. Zhou et al. (2020). The authors used autoMACS (Miltenyi Biotec) with anti-PE microbeads before FACS sorting with anti-hCD3

and hCD1d PBS-57 antibodies. The stimulated NKT cell population had been stimulated with PMA/Ionomycin prior to sorting. NKT and FACS-sorted MAIT cells from Vorkas et al. (2022) had been sequenced and aligned. The MAIT dataset consists of both $CD4^+$ and $CD8^+$ populations which were either left unstimulated or stimulated with 5-OP-RU or anti-CD3/CD28 in the presence of IL-2 or IL-2/TGF-β.

3.3.2 Combining T Cell Subtypes

Integration algorithms which are used to counter technical variability and batch effects require some level of overlap to exist between datasets to guide data alignment. Whilst merging datasets together results in a combined raw count matrix that does not account for confounding variables, integration methods run the risk of overcorrecting and removing biological heterogeneity which can impede the accuracy of cell type identification. Though the aim of this chapter is to compare the performance of automated cell identification methods that act on the unnormalised RNA counts, data integration enabled the visualisation of the cell populations in a common space based on the assumption that T cell subtypes must share some commonalities that would be used for anchoring. In this context, it was hypothesised that this method could be used to compare the level of similarity between subtypes, blurring out differences between more similar populations and making those with a more dissimilar profile stand out. The standard Seurat (v4.0.6) integration workflow (Butler et al., 2019) in R (v4.1.2) was used with the parameters suggested for large datasets by Satija Lab (2022). Briefly, individual gene expression matrices were natural-log normalised such that feature counts are divided by the sum of the counts for each cell and multiplied by a scale factor of 10,000 before being log1p transformed. A subset of 2,000 genes exhibiting high variability across cells was selected using the variance stabilizing transformation (VST) method. Specifically, a line of best fit between the logarithm of variance and logarithm of mean is calculated using local polynomial regression. The feature values are standardised using the observed mean and variance from which the feature variance is computed as a measure of single-cell dispersion. A ranked list of features is constructed and those with the highest median across datasets are selected as the integration features. The features were then used to find cross-dataset pairs of cells that

share a common biological state, known as anchors, using reciprocal PCA (RPCA) with five neighbours (k.anchor) and 50 dimensions (dims). The RPCA method was chosen over canonical correlation analysis because it is less prone to overcorrecting biological signals and is more robust to datasets with non-overlapping populations. Using these anchors points or mutual nearest neighbours, the datasets were iteratively integrated together with 100 neighbours and 50 dimensions used in the anchor weighting procedure. This resulted in the construction of a common corrected data matrix. A more detailed explanation of the integration workflow can be found in the original publication by Butler et al. (2019). The features were scaled and centered before computing a principal component analysis (PCA) on the integrated matrix. Finally, the top 50 principal components were used for visualisation with UMAP and graph-based clustering with shared nearest neighbour (SNN). Clustering trees were plotted using clustree (v0.4.4) to test cluster stability across a range of resolutions. The final resolution was set to 0.4 which resulted in 10 clusters.

3.3.3 Identifying Gene Expression Signatures

In order to account for the sequencing depth in each cell, the RNA counts, X, were normalised as follows using Seurat (v4.0.6):

$$X = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \dots & \dots & \dots & \dots \\ c_{m1} & c_{m2} & \dots & c_{mn} \end{pmatrix}$$

$$l_{ij} = \frac{c_{ij}\, 10^4}{\sum_{i=1}^{m} c_{ij}}$$

$$\log(1 + l_{ij}) = \log\left(1 + \frac{c_{ij}\, 10^4}{\sum_{i=1}^{m} c_{ij}}\right)$$

To identify differentially expressed genes (DEGs) for each T cell subtype the non-parametric Wilcoxon Rank Sum Test (Mann-Whitney U test) was used on the non-integrated log-

transformed RNA measurements. The different T cell subtypes consist of independent samples and the gene expression measurements in each population is not assumed to be normally distributed. Non-parametric tests have been shown to have better control over false discovery rate and are more robust to outliers (Habiger & Pena, 2011). Statistically significant differences in gene expression across groups were calculated. The null hypothesis being that gene expression levels are not differentially expressed, the resulting p-values represent the probability of seeing the observed changes in expression magnitudes. As such, rather than the focus being on the actual observed expression values, it is the ranks of genes in a gene set that are compared. Thereafter, the same normalised counts were used to calculate the average expression of selected genes in each identity class. Dynamic changes in gene expression values were measured to order CD4 T cells in pseudotime using Slingshot (v2.2.0) by Street et al. (2018). Briefly, new clusters are generated from the low-dimensional embeddings using Gaussian mixture models with the assumption that Euclidean distances reflect biological differences. The clusters are subsequently connected using a minimum spanning tree. Principal curves are fitted simultaneously and orthogonal projections are used to overlay pseudotime for the detected lineage. A more detailed explanation of the method can be found in the original publication.

3.3.4 Cell Type Identification Methods

Five commonly used scRNA-seq cell type annotation methods were selected during the literature search based on popularity at the time the research was conducted. Each method differs with respect to the underlying algorithm and the reliance on prior knowledge. All methods were tested in R (v4.1.2) using the settings provided in the accompanying vignettes or suggested in the corresponding papers. Prediction models were trained on the reference datasets provided by the authors, unless otherwise stated below. In order to enforce a consistent vocabulary between reference labels and the query dataset, predicted labels for T cell subtypes were binned into each category class based on biological relevance.

SingleCellNet

SingleCellNet (v0.1.0) implements a multi-class Random Forest classifier of 1,000 trees (Tan & Cahan, 2019). It leverages prior biological knowledge to learn cell type-specific gene pairs and predict cell identity at the single-cell level. The classifier was trained on scRNA-seq data from the mouse *Mus musculus* (The Tabula Muris Consortium, 2018). An ortholog table from the HUGO Gene Nomenclature Committee (HGNC) Comparison of Orthology Predictions (HCOP) was used to convert *Homo sapiens* gene symbols from the query to the mouse equivalent (Eyre et al., 2007). SingleCellNet had been previously reported by Abdelaal et al. (2019) and Xie et al. (2021) to perform well on accuracy, F1-score and speed relative to other methods when identifying major cell types. Therein, performance was only assessed at the T cell detection level but not on T cell subtype identification due to the lack of granularity in the training dataset.

scPred

scPred (v1.9.2) uses support vector machine (SVM) with radial basis function kernel to enhance classification. Similar to SingleCellNet, scPred has been reported to perform accurately across different datasets. The algorithm was trained on a collection of PBMCs from 28 unrelated individuals (Senabouth et al., 2020) and used the query dataset PCA cell embeddings for cell classification. More details on the method can be found in the original publication by Alquicira-Hernandez et al. (2019). As above, this predictor was only tested for T cell detection because of the reference dataset containing broadly defined $CD8^+$ and $CD4^+$ T cell populations.

Azimuth

The Azimuth (v0.4.3) workflow from Seurat (v4) developed by Hao et al. (2021) maps scRNA-seq to high-quality reference datasets using a weighted-nearest neighbour (WNN) model-based approach. Since the pipeline requires the query data to be normalised along with the reference, unnormalised RNA raw counts were used. Azimuth was found to require a minimum of 100 cells to be present in the query dataset. The human PBMC atlas

from Stuart et al. (2019) was used as reference. For Section 3.4.5, the original reference was modified to generate a T cell subtype reference by filtering out other cell types before being integrated to generate the WNN representation required by Azimuth. For the down-sampled reference the same steps were followed after adjusting the number of cell types to generate a more balanced dataset. Sets of anchors between the query object and the references were used to transfer cell type labels, as described in the original publication.

The following T cell subtypes were present in the reference and were binned into the corresponding class labels shown to the right:

- MAIT $\Rightarrow$ MAIT
- $\gamma\delta \Rightarrow$ Gamma Delta
- $CD8^+$ Naive $\Rightarrow$ CD8 Naive
- $CD8^+$ Proliferating $\Rightarrow$ CD8 Cytotoxic
- $CD8^+$ Central Memory $\Rightarrow$ CD8 Cytotoxic
- $CD8^+$ Effector Memory $\Rightarrow$ CD8 Cytotoxic
- $CD4^+$ Naive $\Rightarrow$ CD4 Naive
- $CD4^+$ Proliferating $\Rightarrow$ CD4 Helper
- $CD4^+$ CTL $\Rightarrow$ CD4 Helper
- $CD4^+$ Central Memory $\Rightarrow$ CD4 Memory
- $CD4^+$ Effector Memory $\Rightarrow$ CD4 Memory
- T Regulatory Cells $\Rightarrow$ CD4 Regulatory

SingleR

SingleR (v1.8.0) uses the non-parametric Spearman rank correlation to score each cell in the dataset against each cell type in a reference (Aran et al., 2019). SingleR relies only on genes that vary between cell types to perform the analysis. Reference expression data from the microarray-based Human Primary Cell Atlas (HPCA) and bulk RNA-seq from

Monaco et al. (2019) were used. The references were combined and only genes shared between the two were kept.

The following T cell labels binned into their corresponding classes were present in the final reference:

- Naive $CD8^+$ T cells $\Rightarrow$ CD8 Naive
- Central memory $CD8^+$ T cells $\Rightarrow$ CD8 Cytotoxic
- Central memory $CD4^+$ T cells $\Rightarrow$ CD4 Memory
- Effector memory $CD8^+$ T cells $\Rightarrow$ CD8 Cytotoxic
- Effector memory $CD8^+$ T cells re-expressing CD45RA $\Rightarrow$ CD8 Cytotoxic
- Effector memory $CD4^+$ T cells $\Rightarrow$ CD4 Memory
- Terminal effector $CD8^+$ T cells $\Rightarrow$ CD8 Cytotoxic
- MAIT cells $\Rightarrow$ MAIT
- Vd2 $\gamma\delta$ T cells $\Rightarrow$ Gamma Delta
- Non-Vd2 $\gamma\delta$ T cells $\Rightarrow$ Gamma Delta
- $\gamma\delta$ T cells $\Rightarrow$ Gamma Delta
- Follicular helper T cells $\Rightarrow$ CD4 Helper
- T regulatory cell $\Rightarrow$ CD4 Regulatory
- Th1 cells $\Rightarrow$ CD4 Helper
- Th1/Th17 cells $\Rightarrow$ CD4 Helper
- Th17 cells $\Rightarrow$ CD4 Helper
- Th2 cells $\Rightarrow$ CD4 Helper
- Naive $CD4^+$ T cells $\Rightarrow$ CD4 Naive
- Terminal effector $CD4^+$ T cells $\Rightarrow$ CD4 Helper

ProjecTILs

A computational framework developed by Andreatta et al. (2021), called ProjecTILs, enables the projection of scRNA-seq data onto reference atlases without altering the embedding space. ProjecTILs v2.0.0 was used with the default tumour-infiltrating T lymphocytes atlas they provide and that is described in the original paper. The ProjecTILs analysis workflow was specifically created for T cell subtype and state predictions. Since only conventional T cells are present in the reference atlas, however, performance evaluation was restricted to the following classes: CD8 Naive, CD4 Naive, CD8 Cytotoxic, CD4 Regulatory and CD4 Helper. The CD4 Memory class was also omitted because the reference lacks a clearly defined CD4 memory group. At the time this work was conducted, no paper in the literature had been found to assess its performance, likely due to it being relatively new and for a very niche use case.

The reference labels were binned into the following classes:

- CD8 NaiveLike ⇒ CD8 Naive
- CD8 Tex ⇒ CD8 Cytotoxic
- CD8 Tpex ⇒ CD8 Cytotoxic
- CD8 EffectorMemory ⇒ CD8 Cytotoxic
- CD8 EarlyActiv ⇒ CD8 Cytotoxic
- CD4 NaiveLike ⇒ CD4 Naive
- Treg ⇒ CD4 Regulatory
- Tfh ⇒ CD4 Helper
- Th1 ⇒ CD4 Helper

3.3.5 Performance Evaluation

All methods were evaluated at the major cell type and T cell subtype prediction level, unless otherwise stated in Section 3.3.4. All calculations were performed in R (v4.1.2)

using metrics commonly used to test multi-class classifiers, as reviewed by Grandini et al. (2020). A confusion matrix was constructed to summarise the performance of each cell identification method in which columns represent true cell labels and rows represent the instances of predicted labels (Table 3.2).

Table 3.2: Confusion Matrix Schematic
Schematic of a confusion matrix. Each column represents the instances of true labels whilst each row represents the instances of predicted labels.

		Actual Cell Class				
	Cell	L1	L2	L3	L4	Total
Predicted	L1	5	1	3	1	10
	L2	2	6	0	0	8
	L3	0	4	8	0	12
	L4	1	0	1	6	8
	Total	8	11	12	7	76

Given the following definitions:

- True Positive: TP
- True Negative: TN
- False Positive: FP
- False Negative: FN
- Accuracy: Acc
- Precision: Prn
- Sensitivity: Sns
- Specificity: Spc
- F1-Score: F_1
- Balanced Accuracy: BA
- Error Rate: ER

The equations listed below were used to measure performance:

Accuracy:

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}.$$

Precision:

$$Prn = \frac{TP}{TP + FP}.$$

Sensitivity:

$$Sns = \frac{TP}{TP + FN}.$$

Specificity:

$$Spc = \frac{TN}{TN + FP}.$$

Balanced Accuracy:

$$BA = \frac{Sns + Sps}{2}.$$

F1-Score:

$$F_1 = \frac{2}{\frac{1}{Prn} + \frac{1}{Sns}} = \frac{2\,TP}{2\,TP + FP + FN}.$$

Error Rate:

$$ER = 1 - Acc.$$

Precision, F1-score, sensitivity (true positive rate), specificity (true negative rate) and balanced accuracy per class were measured in a one-vs-rest manner. Total TP, TN, FN and FP were used to calculate overall accuracy, error rate, sensitivity and specificity.

True positive (TP), true negative (TN), false positive (FP) and false negative (FN) values were all directly computed from the resulting confusion matrix as described below.

In the first instance, Table 3.2 can be represented as a square matrix:

$$C_M = \begin{pmatrix} 5 & 1 & 3 & 1 \\ 2 & 6 & 0 & 0 \\ 0 & 4 & 8 & 0 \\ 1 & 0 & 1 & 6 \end{pmatrix}.$$

From which the total number of cells correctly classified is calculated as:

$$c = \sum_{1}^{4} C_{M_{i,i}} = \text{tr}(C_m).$$

The total of cells in column j is given by:

$$C_j = \sum_{i=1}^{4} C_{M_{i,j}}.$$

The total of cells in row i is given by:

$$R_i = \sum_{j=1}^{4} C_{M_{i,j}}.$$

To evaluate performance on a per class basis, the following steps were taken, generalised to a square confusion matrix (n, n):

$$C_m = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \dots & & \dots & \\ c_{i1} & c_{i2} & \dots & c_{in} \\ \dots & & \dots & \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{pmatrix}.$$

A column vector containing all TP values was obtained from the diagonal elements:

$$T_p = \text{diag}(C_m).$$

In order to calculate F_p, a vector where each component represents the sum of values per

row was obtained using the $(n, 1)$ column vector $V = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ such that:

$$\begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \dots & & \dots & \\ c_{i1} & c_{i2} & \dots & c_{in} \\ \dots & & \dots & \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} c_{1j} \\ \sum_{j=1}^{n} c_{2j} \\ \vdots \\ \sum_{j=1}^{n} c_{ij} \\ \vdots \\ \sum_{j=1}^{n} c_{nj} \end{pmatrix} = C_m V.$$

In other words, $C_m V$ is a $(n, 1)$ column vector giving the sum of the rows of C_m which can then be used to calculate F_p, by subtracting the vector containing the TP values in $T_p = \text{diag}(C_m) = \begin{pmatrix} c_{11} & c_{22} & \dots & c_{nn} \end{pmatrix}^T$:

$$F_p = C_m V - T_p.$$

Similarly, to get the sum of the columns of C_m and calculate the number of FN we can

compute $C_m^T V$ to get:

$$C_m^T V = \begin{pmatrix} \sum_{i=1}^{n} c_{i1} \\ \sum_{i=1}^{n} c_{i2} \\ \vdots \\ \sum_{i=1}^{n} c_{ij} \\ \vdots \\ \sum_{i=1}^{n} c_{in} \end{pmatrix}$$

and subsequently subtract T_p giving:

$$F_n = C_m^T V - T_p.$$

To get the sum of all terms in the confusion matrix the $(1,1)$ matrix is computed in order to get the number s:

$$\Big(s\Big) = V^T C_m V = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} c_{ij} \right).$$

This number is used to get the column vector $s\,V$.

Thereafter, the vector T_n can be calculated by subtracting T_p, F_n and F_p from this vector:

$$T_n = s\,V - T_p - F_n - F_p.$$

Now, it is clear that, because $V^T C_m V$ is $(1,1)$, we have:

$$V^T C_m V = \left(V^T C_m V\right)^T = V^T C_m^T V.$$

Therefore, because $F_p = C_m\,V - T_p$, the sum of all components of vector F_p is given by:

$$F_p^T\,V = \left(C_m\,V - T_p\right)^T\,V = V^T C_m^T\,V - T_p^T\,V.$$

Replacing $\mathrm{V}^T C_m^T V$, we get $F_p^T V = V^T C_m V - T_p^T V$. Right factoring out V and transposing subsequently yields:

$$F_p^T V = \left(C_m^T V - T_p\right)^T V = F_n^T V$$

which shows that the sum of all components in vector F_p will always equal to the sum of all components in vector F_n. This is because one label's FP is always another's FN.

Bootstrap resampling was used to estimate confidence intervals and test whether sample size affects the variation in the estimates (Hesterberg, 2011). Samples of size k were repeatedly drawn with replacement from the collection of n cells (Fig. 3.1). Each sample was used to assess the performance of the T cell subtype predictor being evaluated and sampling distributions were generated over 500 repetitions. The total number of possible combinations as described in Davison and Hinkley (1997) is given by:

$$\left(\!\binom{n}{k}\!\right) = \frac{(n+k-1)!}{k!(n-1)!} = \binom{n+k-1}{k}.$$

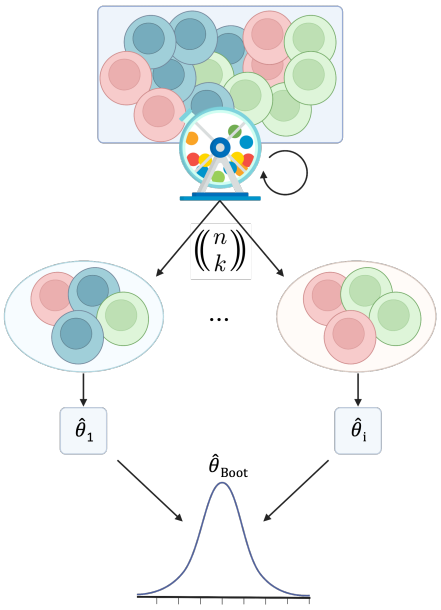


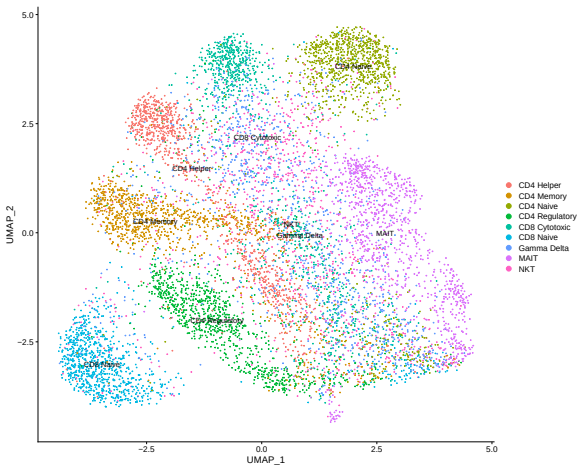
Figure 3.1: Bootstrapping T Cell Predictions.
Schematic of the bootstrapping procedure used to evaluate performance of cell identification methods for T cell predictions. To estimate $\hat{\theta}$, samples of size k were drawn with replacement from the collection of n cells. Created with BioRender.com and icon by Freepik from Flaticon.com.

3.4 Results

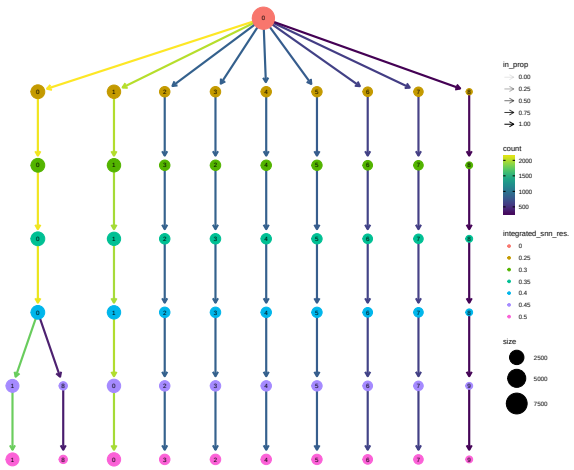
3.4.1 T Cell Overlaps in Low-Dimensional Signal Space

As described in Section 3.3.2, the query dataset was created by randomly selecting 1,000 cells from each T cell class and integration methods were used to combine the multiple transcriptomic subpopulations. Nonlinear dimensionality-reduction was used for visualisation (Fig. 3.2A). Clusters were found to be relatively stable across resolutions (Fig. 3.2B). A clustering resolution of 0.4 was selected as it was found to yield an equal number of clusters as subtypes. The justification behind using integration is that all cells being T cells, there should be enough commonality between subtypes to guide cell alignment. It was thus expected that there would be a certain level of overlap between all datasets, as seen in cluster 0 (Fig. 3.2C). For the most part, the resulting cell population segregated into distinct clusters corresponding to each subtype (Fig. 3.2D). Interestingly, clustering resulted in three classes, CD8 Cytotoxic, Gamma Delta and NKT entirely overlapping in signal space but not in others (CD4 Helper, CD4 Memory, CD4 Naive, CD4 Regulatory, CD8 Naive and MAIT). Biologically it has been recognised that $\gamma\delta$ T cells share cytotoxic hallmarks of $CD8^+$ cytotoxic T cells and display NK-cell like innate reactivities (Tosolini et al., 2017). Furthermore, natural killer T cells have a similar transcriptomic profile to NK cells. Together, this could explain the shared clustering across these three cell types. The distinct clusters that arise with the other cells could be representative of strong characteristic molecular features that distinguish these subtypes from the others. Though integration methods attempt to correct for batch effects, it is important to remember that since these cells come from different studies, intrinsic confounding factors may still be driving clustering. The overlap in signal space between closely related subsets suggests that cell-type assignment methods relying on clustering and marker gene detection may not always be a suitable option. This will be particularly apparent when trying to distinguish between subtypes present in the same batch.

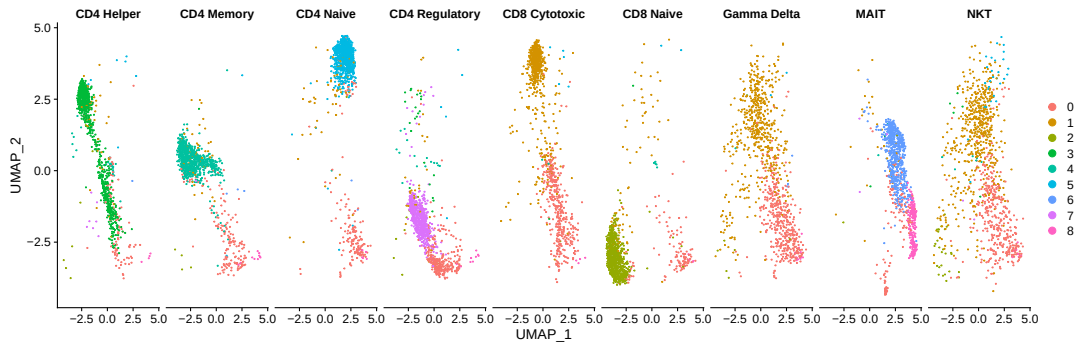
A.



B.



C.



D.

Cluster	CD4 Helper	CD4 Memory	CD4 Naive	CD4 Regulatory	CD8 Cytotoxic	CD8 Naive	Gamma Delta	MAIT	NKT
0	90.0	153.0	166.0	302.0	393.0	117.0	423.0	130.0	370.0
1	35.0	49.0	62.0	32.0	593.0	77.0	552.0	27.0	554.0
2	8.0	5.0	1.0	9.0	1.0	793.0	6.0	2.0	27.0
3	818.0	1.0	0.0	22.0	1.0	1.0	1.0	3.0	5.0
4	29.0	775.0	1.0	15.0	3.0	5.0	1.0	3.0	6.0
5	8.0	3.0	762.0	5.0	5.0	2.0	2.0	1.0	29.0
6	1.0	3.0	2.0	0.0	0.0	0.0	4.0	626.0	0.0
7	7.0	1.0	0.0	606.0	0.0	0.0	0.0	3.0	2.0
8	4.0	10.0	6.0	9.0	4.0	5.0	11.0	205.0	7.0

Figure 3.2: Projected T Cell Subtypes.
A. UMAP projection of integrated T cell subtypes. **B.** Visualisation for interrogating resulting clusters at different resolutions (0, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5). **C.** UMAP clusters split by T cell subtype. **D.** Number of cells from each T cell subtype per cluster.

3.4.2 T Cell Subtypes Show Unique Expression Patterns

T cell subtypes were compared to one another and the most highly variable genes in each group were identified (Fig. C.1A). Each T cell class showed characteristic gene expression patterns which were cross-checked and found to be in agreement with the literature (Figs. C.2A–C.2I). Accordingly, the $\gamma\delta$ T cell group highly expressed TCRδ (*TRDV1/2*) and TCRγ (*TRGC1/2*) genes when compared to the rest of the population (Fig. 3.3). Cytotoxic genes were also enriched, including *GNLY*, *GZMH* and *FGFBP2*. NKT cells were found to neither express CD4 nor CD8, making them double negative. The analysed population was found to be rich in CC chemokine ligand (CCL) genes such as *CCL20*, *CCL3*, *CCL4* and *CCL3L3* which would mediate the migration of cells expressing their corresponding receptors. Cytotoxic granules (*GZMB*, *GZMK*), *PDCD1* (*PD-1*) and genes characteristic of natural killer cells like *KLRD1*, *CD69* and *NKG7* were also highly expressed. This is consistent with the view that NKT exhibit features of both T cells and NK cells. Interestingly, both *IL-4* and *IFN*γ were highly transcribed. Large amounts of *IL-4* have been previously reported to be produced after NKT activation and may be acting in an autocrine manner to directly influence the level of *IFN*γ. MAIT cells on the other hand were expressing *CD8A* and *KLRB1* (*CD161*) high. They were the only cells expressing the invariant TCRα chain *TRAV1-2* paired with β chain *TRBV20-1* (Fig. 3.3). The MAIT cells exhibited high levels of *CCR6* and *CXCR6*, and predominantly expressed cytolytic molecules *GNLY* and *GZMB*. In addition, several genes involved in type 1 IFN signalling were upregulated (*ISG15*, *OAS1*, *MX1*, *IFI6*, *IFITM3*). They showed highly downregulated expression of *MALAT1*, a previously described hallmark of T cell activation. All these are indicative of their activated state induced by multiple stimulation conditions as described in the original paper. The CD8 T cell population was split into naive and effector (cytotoxic) populations. Naturally, both populations highly expressed *CD8A* and *CD8B* subunits. Whilst the CD8 cytotoxic (CTL) group highly expressed cytotoxic granules (*GZMH*, *NKG7*, *GZMM*, *GNLY*) representative of their effector functions, their naive counterpart expressed high *CD7* and *CCR7*. The CD4 naive and memory groups differed in their expression of *CD27* and *CCR7* with naive cells positive for both and memory cells

CD27^{-}*CCR7*$^{-}$. Resting naive cells also expressed *SELL*, *SARAF* (*TMEM66*), *LEF1*. The most differentially expressed gene in the memory group was *S100A4*, a marker of known to be exclusively expressed in this cell type. Interferon-induced transmembrane (*IFITM*) proteins 1 and 2 were also enriched. T regulatory cells expressed the well recognised *FOXP3* gene, a master regulator of their phenotype and immunosuppressive functions, along with *IL32* and *TIGIT*. The MHC class II genes, *HLA-DRB1* and *-B5* were identified in this population and chemokine receptor *CCR10*. The conventional T cell subsets had been bead-enriched, as described in the original publication, and other than the CD4 Helper group all showed expression patterns characteristic of their corresponding cell type. The CD4 Helper class was found to highly express a variety genes such as *LTB*, *CCR7*, *SELL*, *TMEM66* and *LDHB*, however, there was no clear signal distinguishing it from other CD4 cell types, warranting a more refined analysis. The summary metrics provided by 10x Genomics for this dataset described them as CD4 T Helper 2 Cells (Single Cell Gene Expression Dataset by Cell Ranger 1.1.0, 10x Genomics, 2017). However, in the methods section of Zheng et al. it is mentioned how the approach used for subtype classification resulted in a CD4 Helper class containing all isolated CD4 T cells that could not confidently be binned into one of the more refined classes (Naive, Memory or Regulatory) which explains the overlapping signals.

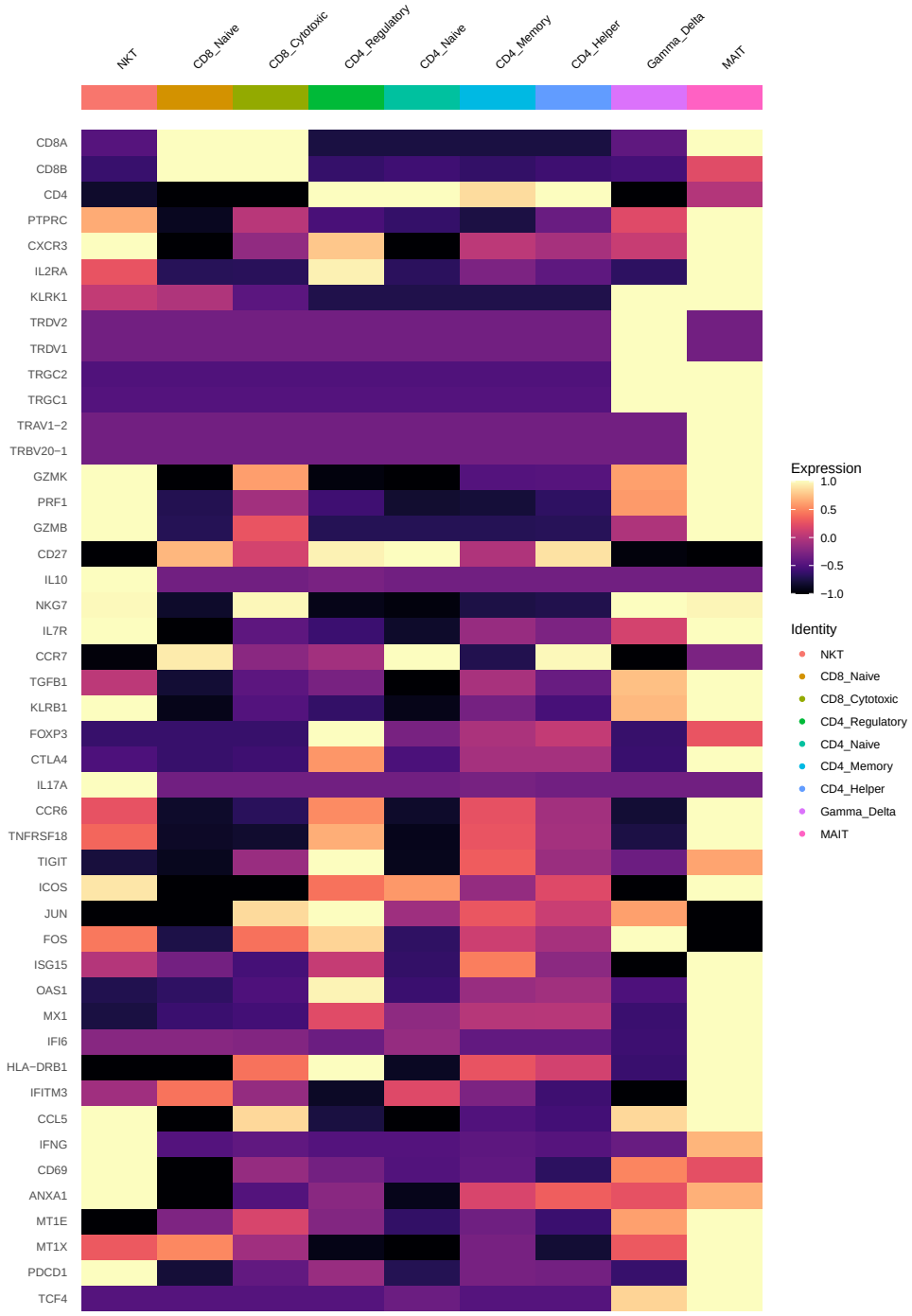


Figure 3.3: T Cell Subtypes Heatmap of Gene Expression for Selected Markers. Heatmap showing averaged RNA expression values for selected genes across T cell subtypes.

3.4.3 CD4 Helper T Cells Require Further Demultiplexing

Upon further examination and re-clustering of the initial pre-downsampled CD4 T cell helper dataset (7,278 cells), the population was split into 5 distinct groups (Fig. 3.4A). Effector $CD4^+$ T cells normally comprise of different T helper phenotypes, namely Th1, Th2 and Th17. Genes characteristic of each subset were expressed across the four clusters suggesting that this dataset likely consists of a mixed population of functional subsets. To test whether cells clustered based on differentiation status instead, signature genes defining naive and memory subsets from Cano-Gamez et al. (2020) were used for cluster labelling (Fig. 3.4B). The gene profile of cluster 0 closely resembled those of naive T cells (TN) with markers *LRRN3*, *CCR7*, *NPM1* and *SELL*, and was labelled as such. Similarly, the transcriptional profile of cluster 3 was consistent with the ones observed in the paper for the highly effector (TEMRA) memory T cells (*CCL4*, *GZMA*, *GZMH*). Meanwhile, cluster 1 was found to express markers of effector T cells (TEM) such as *TIMP1*, *LGALS1*, *USP10*. Like for TN, *SELL* is expressed in cluster 3, which was subsequently labelled as central memory T cells (TCM). Cluster 4 is characterised by high expression of *GNLY*. It represents a very small subset of the population that shares properties of cluster 2 (TEM) and cluster 3 (TEMRA), closely resembling CD4 CTL cells. The single-cell trajectory of the CD4 population inferred by Slingshot placed clusters 0 -> 1 -> 3 -> 4 or TN -> TEM -> TEMRA -> CTL along a path (Fig. 3.4C). Central memory T cells were not part of the main trajectory but found themselves between TEM and TEMRA when ordered in pseudotime (Fig. 3.4D).

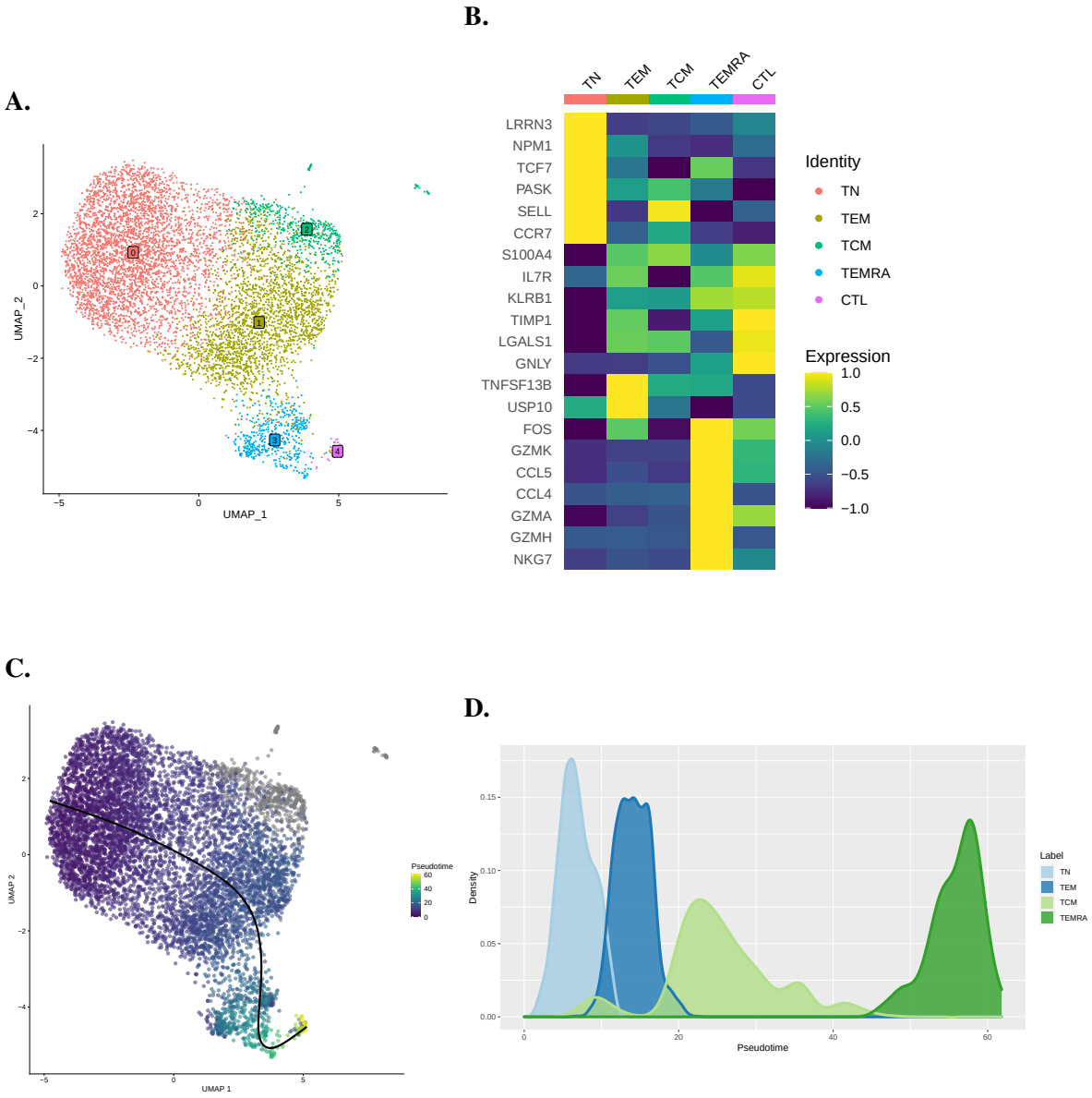
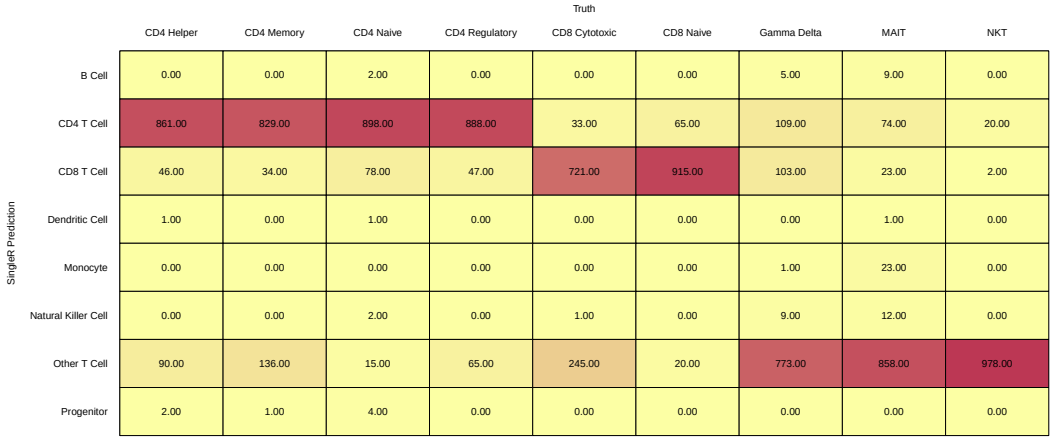


Figure 3.4: Demultiplexing CD4 Helper T Cells.
A. UMAP of CD4 T cell clusters. **B.** Heatmap showing averaged RNA expression values for CD4 differentiation status markers. **C.** Trajectory path from cluster 0 to cluster 4 based on temporal gene expression patterns. **D.** Pseudotime ordering of CD4 T cell clusters.

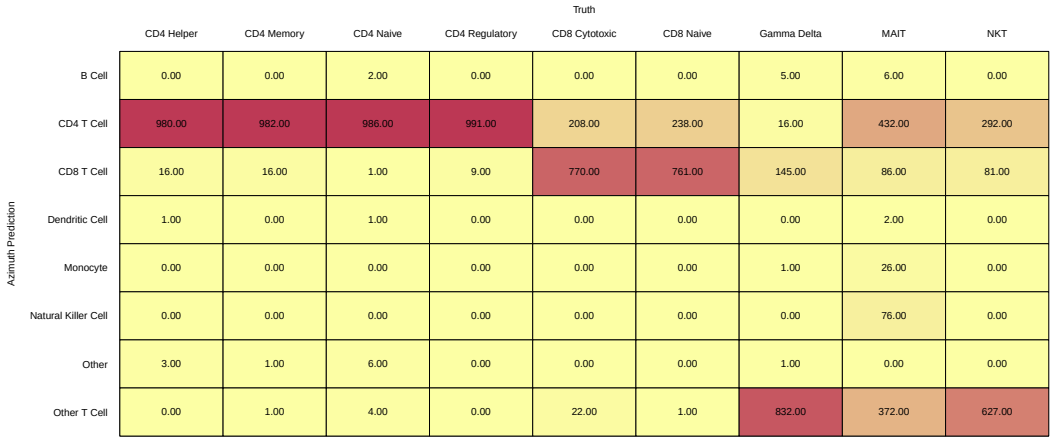
3.4.4 Evaluating T Cell Detection

The T cell subtypes were used to test the ability of SingleR, SingleCellNet, scPred, ProjecTILs and Azimuth to correctly identify T cells in a one-vs-rest manner turning it into a binary classification problem. All predictions were made at the single-cell level on the raw RNA counts of the integrated dataset generated in Section 3.4.1 comprised of 9,000 cells across 9 classes (CD4 Helper, CD4 Memory, CD4 Naive, CD4 Regulatory, CD8 Cytotoxic, CD8 Naive, Gamma Delta, MAIT, NKT). All T cell subtypes in the query were assigned to the T cell class and no additional cell classes were added. Therefore, only true positives (TP) and false negatives (FN) were tested, not false positives (FP) nor true negatives (TN). Four of the five predictors (SingleR, SingleCellNet, scPred, Azimuth) had been trained on reference datasets containing multiple class labels. ProjecTILs, on the other hand, was specifically developed for refining T cell states by projecting cells onto a reference T cell atlas. However, it contains a pre-filtering method called scGate that uses a gating strategy on a set of markers, akin to flow cytometry analysis, to filter out non-T cells. Confusion matrices are used to summarise the prediction results for this task (Figs. 3.5A–3.5E). If the references contained multiple levels of granularity, only the major cell labels were used. SingleR and Azimuth split T cells into three classes (CD4, CD8, others), and ProjecTILs and scPred into two (CD4, CD8). Conversely, SingleCellNet does not discriminate between major T cell subtypes and consists of just one T cell class membership. For each predictor, the multiple major T cell classes were combined, where applicable, to sum the total count of correctly detected T cells. The recall for each of the five predictors was calculated. As expected for this task, all performed well overall, particularly SingleR, ProjecTILs and Azimuth which accurately detected T cells 99.18%, 99.16% and 98.54% of the time, respectively (Fig. 3.6). A relatively lower performance was attributed to scPred (86.31%) and SingleCellNet (78.27%).

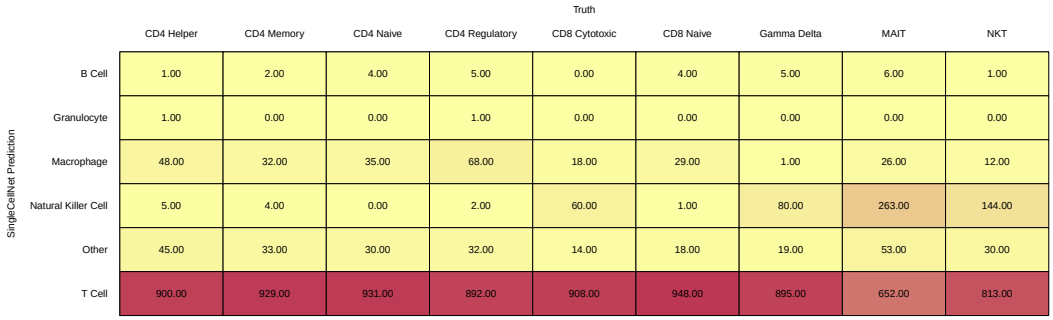
A.


Truth
CD4 Helper
CD4 Memory
CD4 Naive
CD4 Regulatory
CD8 Cytotoxic
CD8 Naive
Gamma Delta
MAIT
NKT
SingleR Prediction
B Cell 0.00 0.00 2.00 0.00 0.00 0.00 5.00 9.00 0.00
CD4 T Cell 861.00 829.00 898.00 888.00 33.00 65.00 109.00 74.00 20.00
CD8 T Cell 46.00 34.00 78.00 47.00 721.00 915.00 103.00 23.00 2.00
Dendritic Cell 1.00 0.00 1.00 0.00 0.00 0.00 0.00 1.00 0.00
Monocyte 0.00 0.00 0.00 0.00 0.00 0.00 1.00 23.00 0.00
Natural Killer Cell 0.00 0.00 2.00 0.00 1.00 0.00 9.00 12.00 0.00
Other T Cell 90.00 136.00 15.00 65.00 245.00 20.00 773.00 858.00 978.00
Progenitor 2.00 1.00 4.00 0.00 0.00 0.00 0.00 0.00 0.00

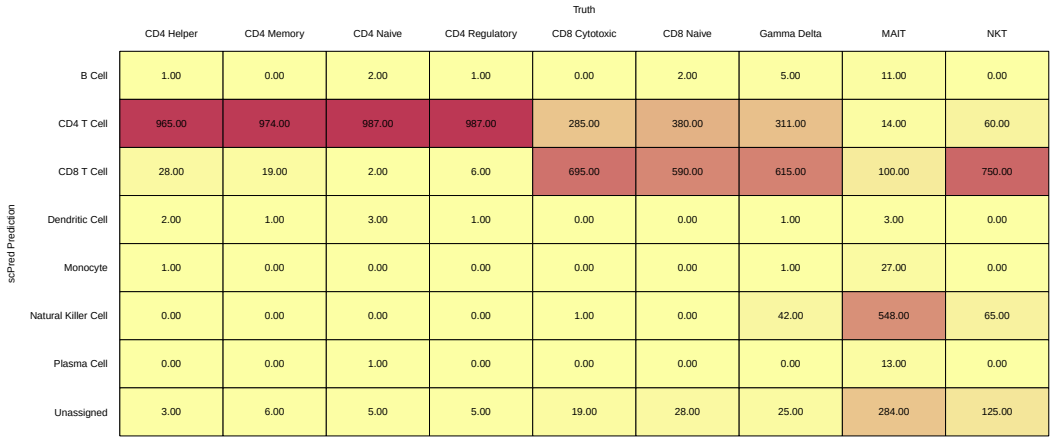
B.


Truth
CD4 Helper
CD4 Memory
CD4 Naive
CD4 Regulatory
CD8 Cytotoxic
CD8 Naive
Gamma Delta
MAIT
NKT
Azimuth Prediction
B Cell 0.00 0.00 2.00 0.00 0.00 0.00 5.00 6.00 0.00
CD4 T Cell 980.00 982.00 986.00 991.00 208.00 238.00 16.00 432.00 292.00
CD8 T Cell 16.00 16.00 1.00 9.00 770.00 761.00 145.00 86.00 81.00
Dendritic Cell 1.00 0.00 1.00 0.00 0.00 0.00 0.00 2.00 0.00
Monocyte 0.00 0.00 0.00 0.00 0.00 0.00 1.00 26.00 0.00
Natural Killer Cell 0.00 0.00 0.00 0.00 0.00 0.00 0.00 76.00 0.00
Other 3.00 1.00 6.00 0.00 0.00 0.00 1.00 0.00 0.00
Other T Cell 0.00 1.00 4.00 0.00 22.00 1.00 832.00 372.00 627.00

C.


Truth
CD4 Helper
CD4 Memory
CD4 Naive
CD4 Regulatory
CD8 Cytotoxic
CD8 Naive
Gamma Delta
MAIT
NKT
SingleCellNet Prediction
B Cell 1.00 2.00 4.00 5.00 0.00 4.00 5.00 6.00 1.00
Granulocyte 1.00 0.00 0.00 1.00 0.00 0.00 0.00 0.00 0.00
Macrophage 48.00 32.00 35.00 68.00 18.00 29.00 1.00 26.00 12.00
Natural Killer Cell 5.00 4.00 0.00 2.00 60.00 1.00 80.00 263.00 144.00
Other 45.00 33.00 30.00 32.00 14.00 18.00 19.00 53.00 30.00
T Cell 900.00 929.00 931.00 892.00 908.00 948.00 895.00 652.00 813.00

D.



E.

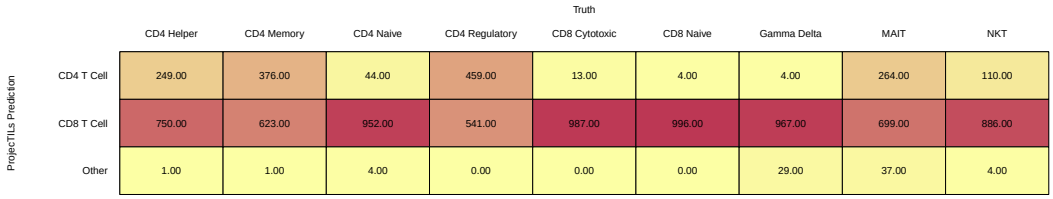


Figure 3.5: Confusion Matrices for T Cell Detection.
Confusion matrices summarising results for T cell detection where columns represent the true labels in the query data and rows the predicted labels. **A.** SingleR. **B.** Azimuth. **C.** SingleCellNet. **D.** scPred. **E.** ProjecTILs.

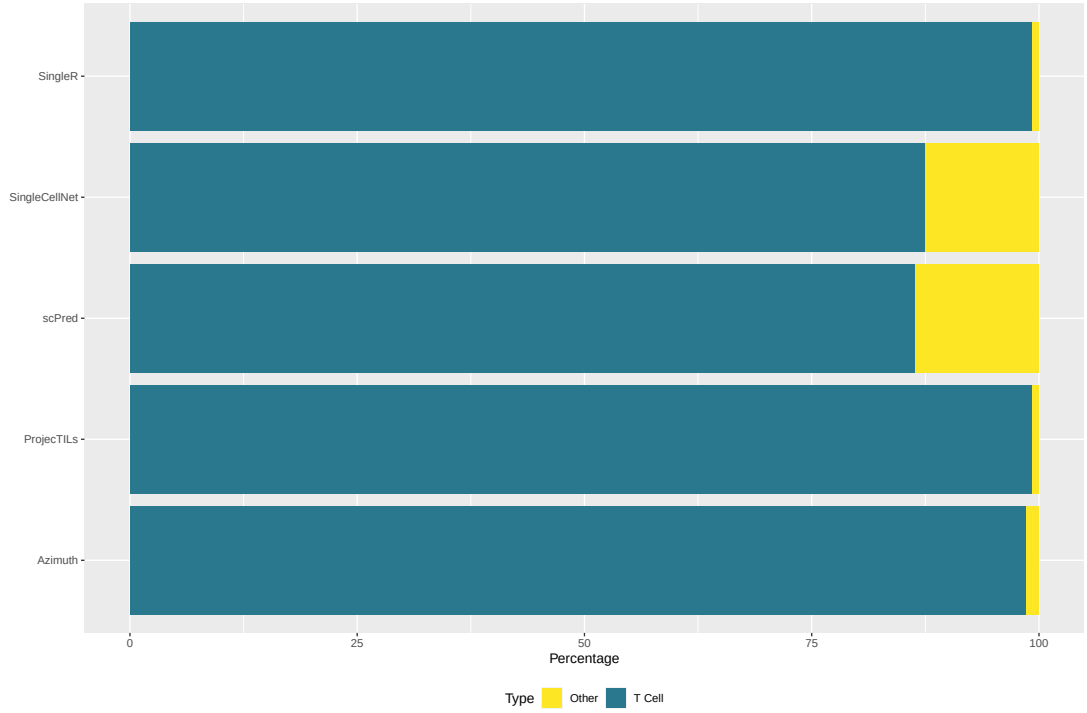


Figure 3.6: Percent T Cells Detected Across Methods.
Stacked bar chart showing percent identified T cells (blue) versus non-T cells (yellow).

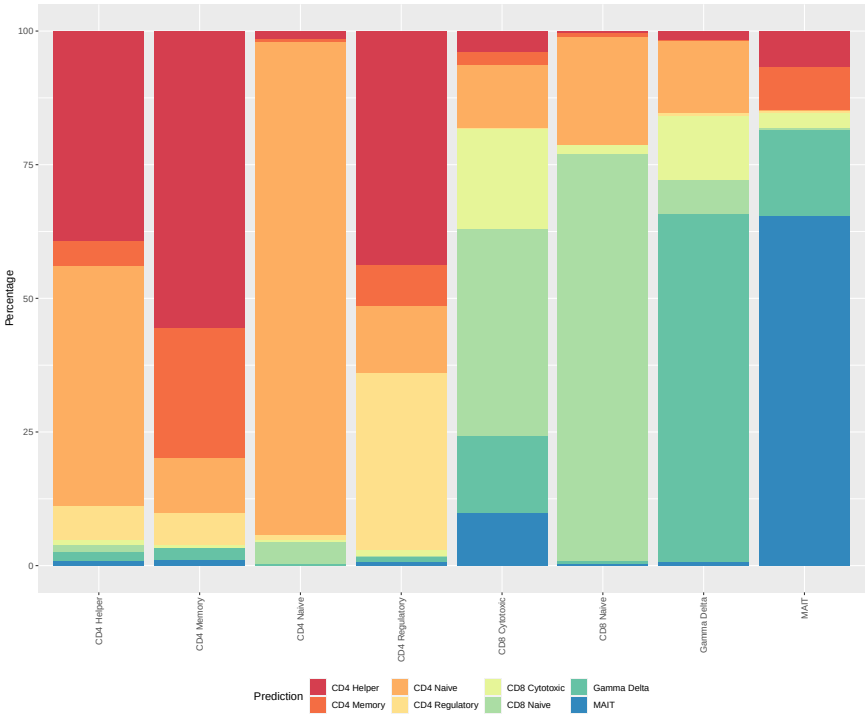
3.4.5 Performance Comparison for T Cell Subtype Identification

It is well known that distinguishing between highly similar cell types can be particularly challenging for classifiers. Unlike the simplistic task presented in Section 3.4.4, closely related cell types, including T cell subtypes, tend to share a similar transcriptomic profile making it more difficult to discriminate between them than more distant cell types. As such, SingleR, Azimuth and ProjecTILs were evaluated for their ability to identify T cell subtypes at a more granular level. SingleCellNet and scPred were not tested as they had not been trained on more refined T cell labels. Natural killer T cells were omitted for this section since none of the method was trained on reference datasets containing them. Similarly, ProjecTILs with the built-in reference can only be used to discriminate between conventional T cell subtypes, hence MAIT and $\gamma\delta$ T cells were also removed to assess its performance. All cell types other than T cells were removed from the reference datasets and predicted labels were binned into the corresponding classes, as described in Section 3.3.4. Like in Section 3.4.4, predictions were made at the single-cell level. As expected,

all methods showed a significant drop in performance classifying fine T cell labels rather than detecting T cells at the major cell type level.

The labels assigned to each T cell class by SingleR are shown in Figure 3.7A. Upon visual inspection, none of the classes were correctly predicted 100% of the time, as depicted by the non-homogeneous colouring in all categories. Nevertheless, more than 50% of MAIT, $\gamma\delta$, CD8 Naive and CD4 Naive T cells were correctly assigned to their corresponding class. CD4 Helper, CD4 Memory, CD4 Regulatory and CD8 Cytotoxic classes appeared to be more challenging for SingleR to predict. The confusion matrix (Fig. 3.7B) was used to quantitatively assess performance (Table 3.3). Examination of the per class metrics confirms that the classifier performed relatively better for some classes than others. For instance, CD4 Naive has both the highest sensitivity (92%) and balanced accuracy (88%) whilst the highest specificity scores (98%) were associated with the CD4 Regulatory and MAIT subtype. The least performing class prediction with the lowest precision (49%), F1-score (27%), sensitivity (19%) and balanced accuracy (58%) is CD8 Cytotoxic followed by CD4 Helper. The overall performance across all classes was calculated and gave an accuracy of 88%, sensitivity of 52% and specificity of 93%. The recall had dropped from 99.18% reported in Section 3.4.4 to 52%. Bootstrap resampling was used to further evaluate model accuracy and the effect of sample size on overall performance metrics (Fig. 3.8). Three different runs consisted of 500 iterations with sample sizes 50, 100 or 200 selected with replacement. The accuracy and true negative rate mean and median were comparable though the true positive rate of nBoot 50 was 9.6% lower than the overall sensitivity.

A.



B.

Truth

SingleR Prediction	CD4 Helper	CD4 Memory	CD4 Naive	CD4 Regulatory	CD8 Cytotoxic	CD8 Naive	Gamma Delta	MAIT
CD4 Helper	394.0	556.0	16.0	438.0	39.0	4.0	18.0	67.0
CD4 Memory	46.0	244.0	5.0	77.0	24.0	8.0	1.0	81.0
CD4 Naive	449.0	102.0	922.0	124.0	119.0	201.0	135.0	3.0
CD4 Regulatory	64.0	59.0	9.0	331.0	1.0	0.0	5.0	2.0
CD8 Cytotoxic	8.0	7.0	4.0	13.0	188.0	18.0	119.0	29.0
CD8 Naive	14.0	0.0	41.0	1.0	387.0	761.0	65.0	3.0
Gamma Delta	16.0	22.0	3.0	10.0	143.0	6.0	651.0	162.0
MAIT	9.0	10.0	0.0	6.0	99.0	2.0	6.0	653.0

Figure 3.7: T Cell Subtype Predictions with SingleR.
A. Stacked bar chart showing percent SingleR predicted labels per T cell class. **B.** Confusion matrix of SingleR predicted counts to true T cell labels.

Table 3.3: Performance Metrics for SingleR.
Summary statistics of SingleR T cell subtype predictions.

	TP	FP	TN	FN	Precision	F1-Score	Sensitivity	Specificity	Balanced Accuracy
CD4 Helper	394	1138	5862	606	0.26	0.31	0.39	0.84	0.62
CD4 Memory	244	242	6758	756	0.5	0.33	0.24	0.97	0.6
CD4 Naive	922	1133	5867	78	0.45	0.6	0.92	0.84	0.88
CD4 Regulatory	331	140	6860	669	0.7	0.45	0.33	0.98	0.66
CD8 Cytotoxic	188	198	6802	812	0.49	0.27	0.19	0.97	0.58
CD8 Naive	761	511	6489	239	0.6	0.67	0.76	0.93	0.84
Gamma Delta	651	362	6638	349	0.64	0.65	0.65	0.95	0.8
MAIT	653	132	6868	347	0.83	0.73	0.65	0.98	0.82

Accuracy	Error Rate	Sensitivity	Specificity
0.88	0.12	0.52	0.93

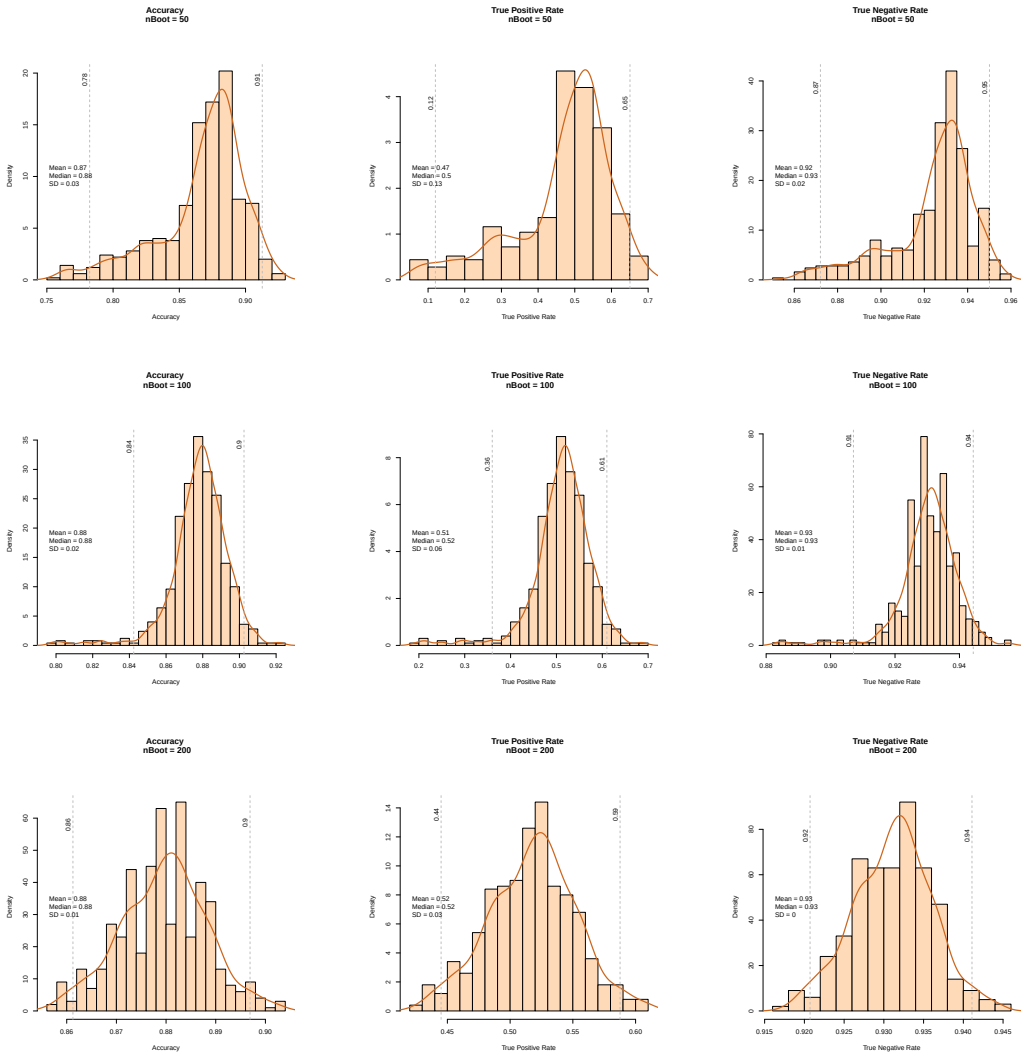
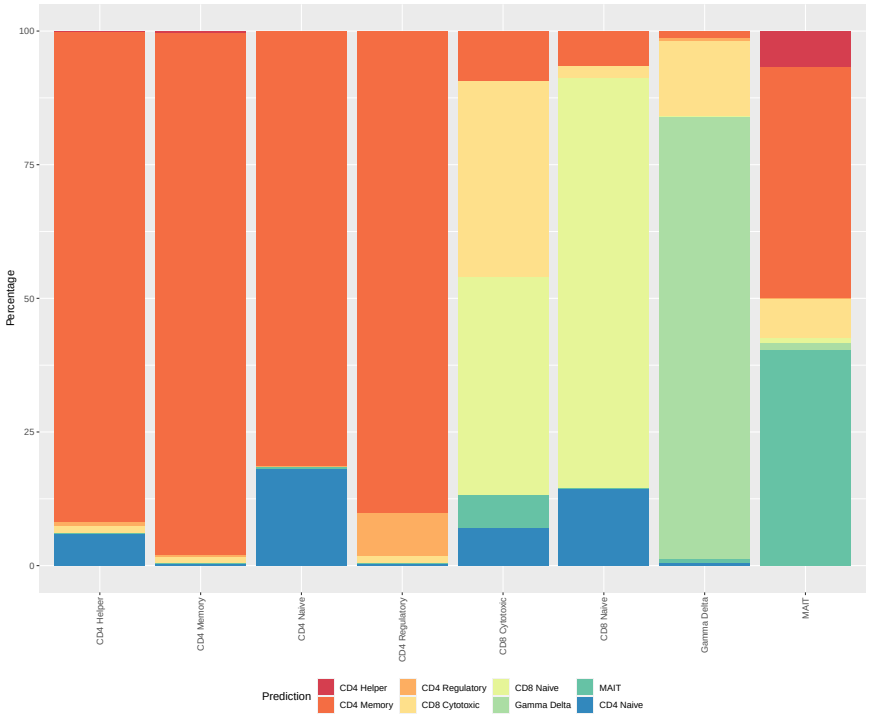


Figure 3.8: Bootstrap Resampling for SingleR Predictions.
Distributions of bootstrap resampling results for accuracy, true positive rate and true negative rate at nBoot 50, 100 and 200. Confidence intervals are shown in grey.

At first sight, Azimuth is found to have a strong bias towards CD4 Memory (Figs. 3.9A–3.9B). Compared to SingleR, Azimuth had slightly lower performance metrics with an overall sensitivity of 52% and a specificity of 93% (Table 3.4). However, Azimuth was better than SingleR at predicting the Gamma Delta class with a 55%, 38%, 28%, 5% and 14% higher precision, F1-score, sensitivity, specificity and balanced accuracy, respectively. The strong bias towards labelling CD4 positive cells as CD4 Memory brings the class precision down to 23% and specificity to 54%. Upon further investigation, it was found that the PBMC reference it is trained on is heavily biased towards the CD4 Memory class which consists of 62% of the total labels. Nevertheless, the prediction bias was expected for the CD4 Helper class for which the reference does not have a clearly defined label (cf. Section 3.3.4). In addition, as reported in Section 3.4.3, this dataset consists of both Naive and Memory cells. During bootstrapping it was also found that Azimuth requires at least 100 samples, as such nBoot 100 and 200 were used (Fig. 3.10). The heavily skewed training set makes the overall calculated accuracy of 86% not a reliable measure of model performance. To reduce this bias, the reference dataset was downsampled to have equal labels in each class for model retraining. Doing so did help reduce the CD4 Memory bias and improved Azimuth performance (Figs. 3.11A–3.11B). Azimuth's sensitivity and specificity increased to 58% and 94%, respectively, resulting in a better overall performance than SingleR (Table 3.5). In turn, CD4 Memory class specificity went from 54% to 80% and precision from 23% to 38% but at a cost of sensitivity which went from 98% to 84%. Downsampling the Azimuth reference dataset also increased the sensitivity of the MAIT class by 58% bringing Azimuth on par with SingleR for detecting this T cell subtype.

A.



B.

Truth

Azimuth Prediction	CD4 Helper	CD4 Memory	CD4 Naive	CD4 Regulatory	CD8 Cytotoxic	CD8 Naive	Gamma Delta	MAIT
CD4 Helper	2.0	4.0	0.0	0.0	1.0	0.0	0.0	67.0
CD4 Memory	917.0	977.0	815.0	902.0	93.0	67.0	14.0	433.0
CD4 Naive	59.0	3.0	181.0	4.0	72.0	143.0	6.0	0.0
CD4 Regulatory	7.0	3.0	2.0	81.0	0.0	0.0	5.0	1.0
CD8 Cytotoxic	14.0	12.0	0.0	12.0	367.0	21.0	140.0	74.0
CD8 Naive	0.0	0.0	0.0	0.0	407.0	768.0	2.0	10.0
Gamma Delta	0.0	0.0	0.0	0.0	0.0	0.0	827.0	12.0
MAIT	1.0	1.0	2.0	1.0	60.0	1.0	6.0	403.0

Figure 3.9: T Cell Subtype Predictions with Azimuth.
A. Stacked bar chart showing percent Azimuth predicted labels per T cell class. **B.** Confusion matrix of Azimuth predicted counts to true T cell labels.

Table 3.4: Performance Metrics for Azimuth.
Summary statistics of Azimuth T cell subtype predictions.

	TP	FP	TN	FN	Precision	F1-Score	Sensitivity	Specificity	Balanced Accuracy
CD4 Helper	2	72	6928	998	0.03	0	0	0.99	0.5
CD4 Memory	977	3241	3759	23	0.23	0.37	0.98	0.54	0.76
CD4 Naive	181	287	6713	819	0.39	0.25	0.18	0.96	0.57
CD4 Regulatory	81	18	6982	919	0.82	0.15	0.08	1	0.54
CD8 Cytotoxic	367	273	6727	633	0.57	0.45	0.37	0.96	0.66
CD8 Naive	768	419	6581	232	0.65	0.7	0.77	0.94	0.85
Gamma Delta	827	12	6988	173	0.99	0.9	0.83	1	0.91
MAIT	403	72	6928	597	0.85	0.55	0.4	0.99	0.7

Accuracy	Error Rate	Sensitivity	Specificity
0.86	0.14	0.45	0.92

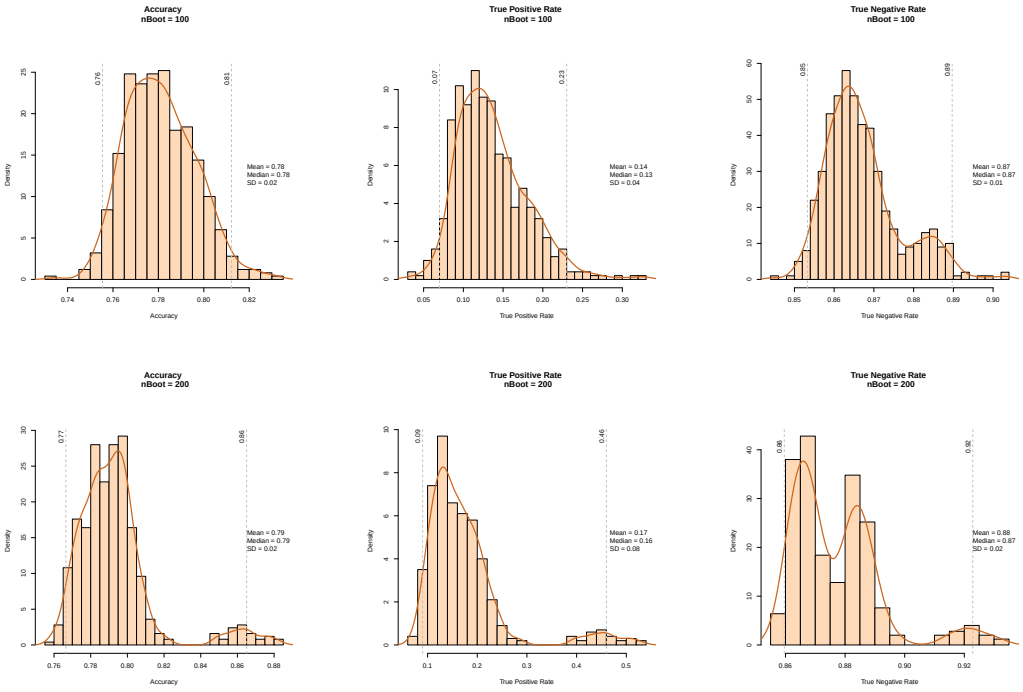
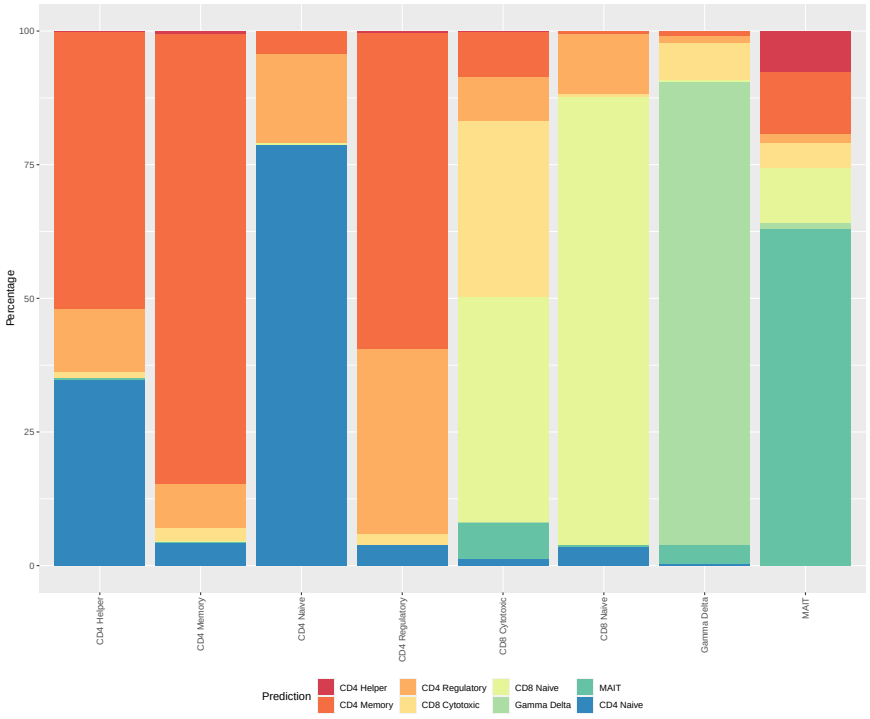


Figure 3.10: Bootstrap Resampling for Azimuth Predictions.
Distributions of bootstrap resampling results for accuracy, true positive rate and true negative rate at nBoot 100 and 200. Confidence intervals are shown in grey.

A.



B.

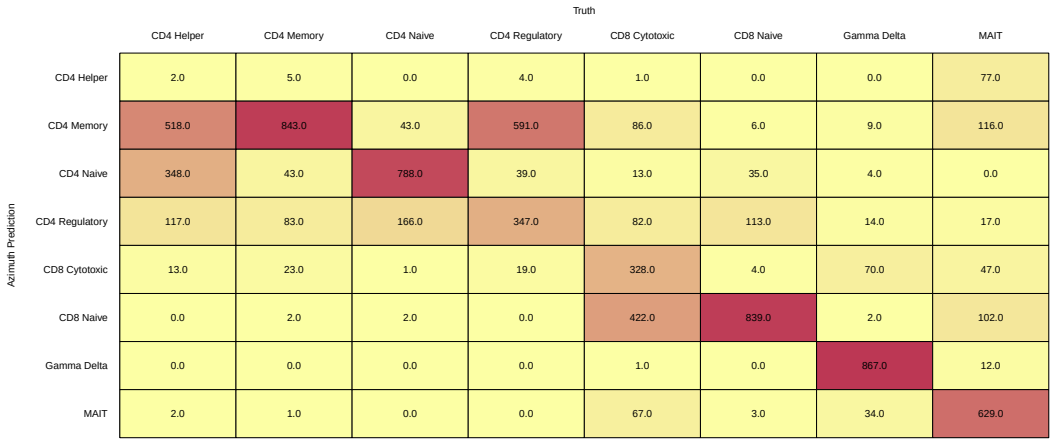


Figure 3.11: T Cell Subtype Predictions with Downsampled Azimuth.
A. Stacked bar chart showing percent retrained Azimuth predicted labels per T cell class. **B.** Confusion matrix of retrained Azimuth predicted counts to true T cell labels.

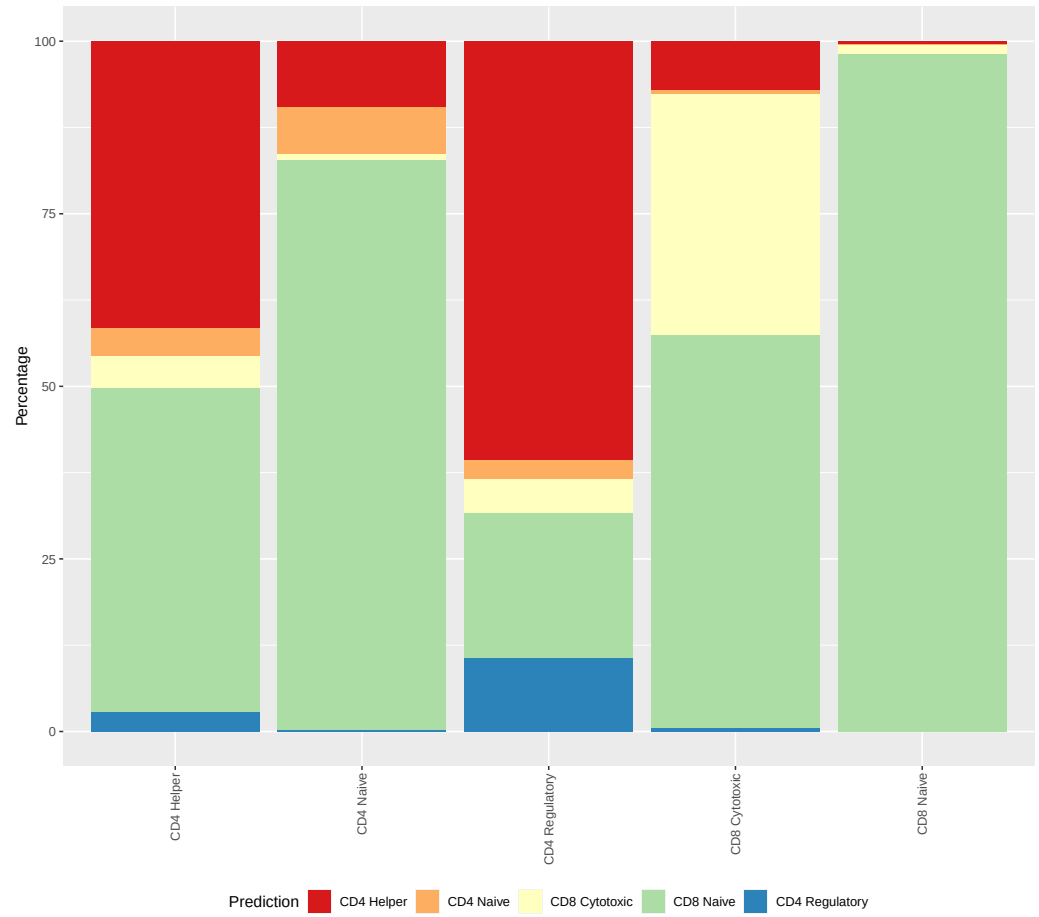
Table 3.5: Performance Metrics for Downsampled Azimuth.
Summary statistics of retrained Azimuth T cell subtype predictions.

	TP	FP	TN	FN	Precision	F1-Score	Sensitivity	Specificity	Balanced Accuracy
CD4 Helper	2	87	6913	998	0.02	0.00	0.00	0.99	0.49
CD4 Memory	843	1369	5631	157	0.38	0.52	0.84	0.80	0.82
CD4 Naive	788	482	6518	212	0.62	0.69	0.79	0.93	0.86
CD4 Regulatory	347	592	6408	653	0.37	0.36	0.35	0.92	0.63
CD8 Cytotoxic	328	177	6823	672	0.65	0.44	0.33	0.97	0.65
CD8 Naive	839	530	6470	161	0.61	0.71	0.84	0.92	0.88
Gamma Delta	867	13	6987	133	0.99	0.92	0.87	1.00	0.93
MAIT	629	107	6893	371	0.85	0.72	0.63	0.98	0.81

Accuracy	Error Rate	Sensitivity	Specificity
0.9	0.1	0.58	0.94

ProjecTILs was specifically developed to predict T cell subtypes and states with a reference trained only on conventional T cells. Nonetheless, it did not outperform SingleR nor Azimuth for neither CD4 nor CD8 subtype identification (Figs. 3.12A–3.12B). The CD4 Naive class was majorly mistaken for CD8 Naive T cells. As shown in Table 3.6, CD8 Naive specificity was only 48% but represented the class with the highest sensitivity (98%). Conversely, there was low sensitivity for CD4 subtypes, mainly Naive and Regulatory. With an accuracy of 75%, sensitivity of 38%, specificity of 85% and an error rate of 25%, ProjecTILs performance on this dataset significantly lagged behind both Azimuth and SingleR. Bootstrapping resulted in a bimodal distribution for all performance metrics and very large 95% CIs (Fig. 3.13). This suggests that ProjecTILs performance is highly influenced by the proportion of CD4 to CD8 T cells which differs at different sample sizes and bootstrap samples.

A.



B.

ProjecTILs Prediction \ Truth	CD4 Helper	CD4 Naive	CD4 Regulatory	CD8 Cytotoxic	CD8 Naive
CD4 Helper	416.0	96.0	607.0	71.0	5.0
CD4 Naive	41.0	68.0	28.0	6.0	1.0
CD4 Regulatory	29.0	2.0	106.0	5.0	0.0
CD8 Cytotoxic	45.0	9.0	49.0	350.0	12.0
CD8 Naive	469.0	825.0	210.0	568.0	982.0

Figure 3.12: T Cell Subtype Predictions with ProjecTILs.
A. Stacked bar chart showing percent ProjecTILs predicted labels per T cell class. **B.** Confusion matrix of ProjecTILs predicted counts to true T cell labels.

Table 3.6: Performance Metrics for ProjecTILs.
Summary statistics of ProjecTILs T cell subtype predictions.

	TP	FP	TN	FN	Precision	F1-Score	Sensitivity	Specificity	Balanced Accuracy
CD4 Helper	416	779	3221	584	0.35	0.38	0.42	0.81	0.61
CD4 Naive	68	76	3924	932	0.47	0.12	0.07	0.98	0.52
CD4 Regulatory	106	36	3964	894	0.75	0.19	0.11	0.99	0.55
CD8 Cytotoxic	350	115	3885	650	0.75	0.48	0.35	0.97	0.66
CD8 Naive	982	2072	1928	18	0.32	0.48	0.98	0.48	0.73

Accuracy	Error Rate	Sensitivity	Specificity
0.75	0.25	0.38	0.85

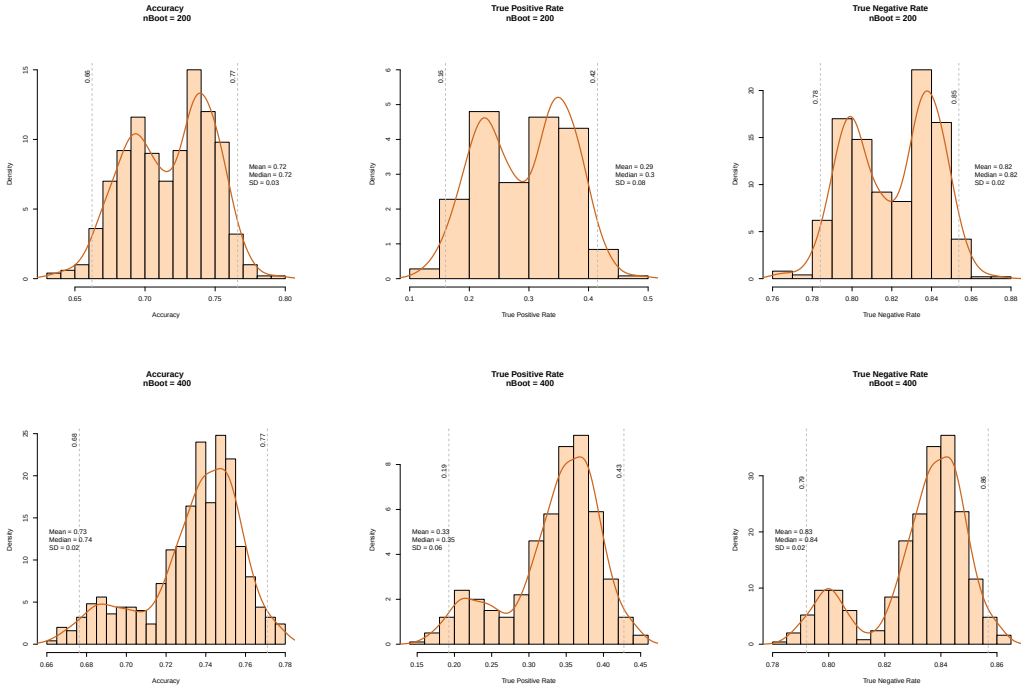


Figure 3.13: Bootstrap Resampling for ProjecTILs Predictions.
Distributions of bootstrap resampling results for accuracy, true positive rate and true negative rate at nBoot 200 and 400. Confidence intervals are shown in grey.

Per class performance for all predictors is summarised in Figure 3.14. Performance was found to highly vary on a per class basis. For the detection of Gamma Delta cells, Azimuth Downsampled outperformed other predictors across all metrics with performance comparable to SingleR for the MAIT class. For the CD8 Naive class, ProjecTILs was found to be the most sensitive but least specific resulting in Azimuth Downsampled showing the highest balanced accuracy and F1-score. None of the classifiers had good sensitivity to CD8 Cytotoxic, with ProjecTILs and Azimuth scoring the highest balanced accuracy for this group. Similarly, CD4 Regulatory were poorly detected, ProjecTILs was the least sensitive and the highest balanced accuracy was reached by SingleR. Azimuth Downsampled outperformed the default Azimuth across all metrics and classes with few exceptions. One such exception was for the CD4 Memory class, Azimuth outperformed SingleR and Azimuth Downsampled on sensitivity. CD4 Helper cells were only detected using SingleR or ProjecTILs but both had an average performance.

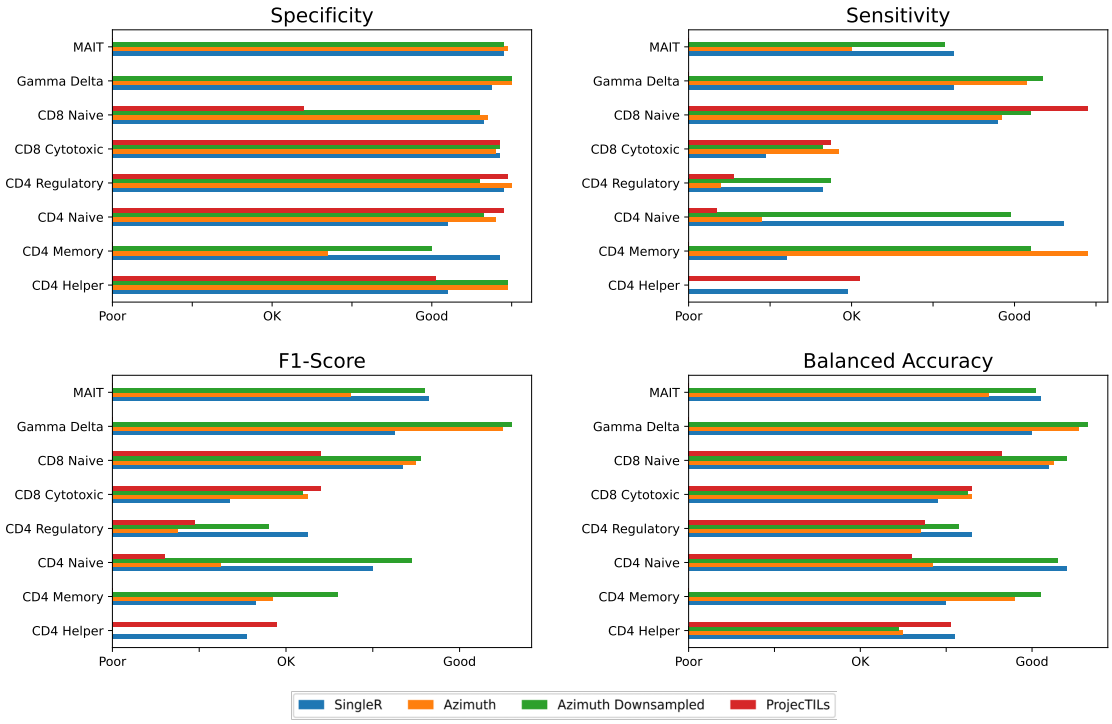


Figure 3.14: Performance Metrics Summary.
Summary of the performance of all classifiers during T cell subtype identification. Specificity, sensitivity, F1-score and balanced accuracy are shown for each class and whether values represent good, intermediate or poor performance.

3.5 Discussion

Single-cell RNA sequencing (scRNA-seq) is a popular and powerful technique to measure the transcriptome of individual cells. The accurate classification of these single cells into cell types represents an ongoing challenge, in particular when distinguishing between closely related profiles. This critical step can influence the downstream interpretation of the data, thus both computational and biological aspects need to be carefully considered. There is no consensus, however, on what the best methods are and new ones are continuously being developed. Moreover, the lack of gold-standard datasets makes it difficult to comprehensively assess the performance of different methods. Benchmarking studies have been published but are found to primarily focus on the ability of annotation methods to discern between major cell types. In order to decipher the T cell transcriptomic landscape in this thesis, it was first essential to identify the advantages and limitations of available methods used to differentiate between T cell subsets. T cells consist of a diverse group of phenotypically and functionally distinct subpopulations. It is common for them to express shared marker genes across subsets, and fine distinctions may not be apparent transcriptionally or be hidden below the detection threshold of the methods.

The traditional cell-type identification method involves the clustering of cells using an unsupervised approach followed by the the manual annotation of clusters based on differentially expressed genes. Clustering algorithms will have tunable parameters that will highly influence the resulting clusters. For instance, when using the common graph-based clustering approach in Section 3.4.1, the number of nearest neighbours and the clustering resolution need to be carefully chosen. Finding the right balance between under- and over-clustering is an iterative process and it can be difficult to assess the quality of the resulting clusters. Clustering of the publicly available scRNA-seq datasets representing experimentally-isolated or validated T cell subtypes revealed that certain subtypes were more likely to be grouped together than others. CD8 Cytotoxic T cells were found to most highly overlap with $\gamma\delta$ T cells and NKT, reflective of their shared biological features. One of the resulting clusters consisted of a homogeneous mix of T cell subtypes. Taking into account that the T cell datasets come from different sources, and that integration methods

rely on shared cell types to handle batch effects, non-biological variance must be driving some of the clustering. As such, the subtype overlaps identified in this analysis are expected to be more significant when addressing inter-dataset populations. Interestingly, much after this work was conducted X. Wang et al. (2022) published a study which used the same bead-enriched T cell subsets from Zheng et al. (2017) to conduct a similar but more comprehensive analysis. After integrating the different T cell subtypes, they also found that cell populations identified when clustering scRNA-seq are more homogeneous than classically defined subsets. Their results indicated that CD4 and CD8 naive cells for the most part split into distinct clusters and could be identified accordingly. Conversely, non-naive subtypes were more likely to mix with each other. They found a high overlap between unconventional T cells (MAIT and NKT) and CD8 CTLs. Importantly, T cell subtypes that were matched to different categories were found to have different expression patterns, indicative of clustering by cell states or gene-expression gradients rather than subtype. Thus, conflicting annotations within a cluster that arise when using traditional cell-type identification methods can be driven by the underlying biology and parameter tuning may not successfully untangle this.

Canonical markers used to annotate cell groups are often few, and may not be applicable to the scRNA-seq experiment. A number of studies have reported poor correlation between protein levels and RNA expression, yet markers are frequently derived from characteristic cell-surface proteins. In addition to the high drop-out rate of the technology that can influence the detection of these markers, different tissues or disease states may also influence the measured expression levels. Gene signatures characteristic of each T cell subtype were identified from the scRNA-seq data for use in the next chapter. All T cell populations were found to have distinguishing features consistent with the literature. However, the CD4 Helper class had no clear characteristic signal which prompted a more refined analysis to understand whether it contained subpopulations of CD4 T cells. In addition, it was important to know whether the composition of the CD4 Helper group could influence the downstream performance of the T cell classifiers. Subclustering analysis identified five subgroups of CD4 T cells, the majority of which were classified as naive. Trajectory inference revealed a progression from CD4 naive (TN), to effector memory (TEM),

to effector memory T cells re-expressing CD45RA (TEMRA), ending on CD4 cytotoxic T lymphocytes (CTLs). Interestingly, TEMRA have been previously described as highly enriched for CD4-CTLs (Patil et al., 2018), which could explain the path from TEMRA to CTLs.

SingleR, SingleCellNet and scPred were popular cell annotation tools at the time the research was conducted and had previously been tested on major cell types by Abdelaal et al. (2019). Azimuth and ProjecTILs became available on bioRxiv in 2020 before their publication in Hao et al. (2021) and (Andreatta et al., 2021), respectively. All five were tested on the scRNA-seq datasets with known labels. False positive rates were not tested because this was not the primary focus of this study, and had already been evaluated by others (Abdelaal et al., 2019; Andreatta et al., 2021; Xie et al., 2021; Huang et al., 2021). As expected, all methods scored a high true positive rate. Performance ranked from best to worst: SingleR, ProjecTILs, Azimuth, SingleCellNet, scPred. SingleCellNet was found to have misclassified 16% of unconventional T cells, mainly MAIT and NKT, as NK cells. Previous studies had found that it's accuracy highly depends on whether a class is well represented in the reference dataset. This suggests that the reference dataset provided by SingleCellNet is likely to only consist of conventional T cells, though more refined labels are not available to confirm this. Similarly, MAIT followed by NKT cells were the primary classes that scPred failed to detect as T cells, assigning them to the NK or unassigned class. Per class performance was subsequently assessed for those which had appropriate reference datasets, mainly SingleR, Azimuth and ProjecTILs. Multiple evaluation criteria were used, and performance highly varied on a per class basis for the different metrics. For the identification of unconventional T cells, excluding NKT, Azimuth was found to outperform SingleR based on balanced accuracy and F1-Score. CD8 and CD4 T cells could be identified using all methods. CD8 Cytotoxic cells were best identified using ProjecTILs. which achieved the highest sensitivity, F1-score and balanced accuracy. Cells from the CD4 Regulatory class were relatively poorly identified based on sensitivity and F1-score, SingleR being the top performer. This was surprising considering that traditional marker-based methods can rely on well defined markers for this class, including CD4, FOXP3 and CTLA4. The most sensitive method for CD4 Naive was SingleR, whilst

Azimuth was significantly better at detecting CD4 Memory. CD4 Helper cells were not identified by Azimuth which mainly classified them as CD4 Memory. This is not a good representation of its performance on this class since it was shown in Section 3.4.3 that this class is composed of a mixture of CD4 T cells including a large portion of memory cells. Overall SingleR and Azimuth were the preferred methods but their performance varied across T cell subtypes. Balancing the reference of Azimuth significantly improved its performance. It isn't clear, however, whether this will be preferred when applying it to clinical samples. Whilst statistically justified in the context of a well-balanced query, the original unbalanced reference may be more relevant to the frequency of cell types in tissue. Balancing the reference would therefore mask important biological signals.

In recent years, a burgeoning literature has emerged on cell type identification (Mao et al., 2022; Ianevski et al., 2022; Nguyen & Griss, 2022). In future work, it will be important to develop a systematic benchmarking approach to keep up with the fast-growing set of methods and streamline the evaluation of T cell subtype identification. Optimally, gold standard datasets of isolated T cell subtypes should be generated in the laboratory. The performance of the methods is highly influenced by the quality of the reference datasets. It will also be important to ensure sufficient diversity representing the different subtypes. A limitation in this study was that all methods had been trained on different reference datasets. It will be important to re-train all methods on the same reference dataset to properly compare the underlying algorithms. All methods ranging from the manual annotation of clusters to the automatic classification of cell types had their own strengths and weaknesses. With clustering unlikely to result in pure clusters and the variable performance of automatic methods across T cell subtypes, complementary annotation approaches should be used. The evidence gathered in this chapter indicates that in the context of T cell subtype classification an ensemble approach is likely to achieve better predictive performance.

3.6 Conclusion

In this chapter, five methods commonly used to annotate cell types from scRNA-seq data were evaluated for their ability to identify T cells or more refined T cell subtypes. Discriminating between closely related cells based on gene expression profile alone is a significantly more challenging task than identifying major cell types. Moreover, clustering was found to blur differences between T cell subtypes, thus the importance of using alternative approaches in addition to the commonly used method of classifying cells based on canonical surface markers. It was shown that SingleR, Azimuth, ProjecTILs, scPred and SingleCellNet all perform relatively well when discerning T cells from other cell populations, SingleR and Azimuth being the top performers. When faced with having to classify T cell subtypes, however, classification performance significantly degraded and highly varied on a per class basis. Taken together, by highlighting the strengths and weaknesses of the different methods, the work presented in this chapter suggests that an ensemble approach may be favourable for T cell identification.

CHAPTER 4

Multi-Omics Analysis of T Cell-Intrinsic Interferon Signalling Pathways in the Tumour Microenvironment

4.1 Overview

Interferons (IFNs) are important drivers of the inflammatory response that influence T cell functional states. Cancer and inflammation are inextricably linked, and tumour cells are commonly found to have impaired IFN signalling. Similar defects have been identified in the peripheral blood lymphocytes of cancer patients and are believed to have confounding effects on immunotherapeutic approaches. However, little is known about cancer-associated defects in T cell-intrinsic IFN cascades and how these may influence the functional states of those that reside in the tumour microenvironment (TME). In this chapter, spatially resolved transcriptomics, single-cell RNA-sequencing and single cell T cell receptor sequencing are integrated to investigate T cell-intrinsic IFN signalling pathways within the TME of triple-negative breast cancer patients. T cell subsets are identified and IFN signalling pathways are explored at a single-cell resolution whilst retaining spatial context. Moreover, this chapter dives into the spatial relationship between type I IFN signalling pathways and hypoxic neighbourhoods. This detailed multi-omics characterisation of T cell profiles and type I IFN signalling in the TME provides insight into immunotherapeutic strategies that could help improve T cell therapy for solid tumours.

4.2 Introduction

Triple-negative breast cancer (TNBC) is an aggressive breast cancer subtype with a highly immunosuppressive tumour microenvironment (TME) characterised by low tumour oxygenation. The lack of oestrogen (ER) and progesterone (PR) receptors, and the absence of human epidermal growth factor receptor 2 (HER2) overexpression results in limited treatment options for TNBC patients. Moreover, a large number of them fail to respond to T cell-mediated immunotherapies. T cells on the tumour battleground can enter a dysfunctional or exhausted state, defined by the upregulation of inhibitory receptors, profound metabolic deficiencies, and changes in cytokine signalling pathways that impair their antitumour ability. The release of damage-associated molecular patterns (DAMPs) as a result of cell death and stress occurring in the TME can stimulate inflammatory responses through the production of interferons (IFNs). IFNs are a family of pleiotropic cytokines that have been extensively studied in the context of viral infections but are also known to have potential implications in cancer control and immunotherapy response. They play a central role in regulating T cell immunity by either promoting or inhibiting their activation, proliferation, differentiation, survival and associated functions. Impaired IFN-signalling cascades are commonly observed in tumour cells, including breast cancer cell lines, and are being exploited by cancer therapies such as oncolytic viruses and cyclic GMP-AMP synthase (cGAS)/stimulator of interferon genes (STING) agonists. Defects in IFN signalling and downstream functional responses have also been identified in the peripheral T cells of breast cancer patients, both in early- and late-stage disease (Critchley-Thorne et al., 2009). In a recent study, W. Li et al. (2020) found the type I interferon-inducing cGAS-STING pathway introduced in Section 2.3 to be substantially impaired in peripheral blood $CD8^+$ T cells of cancer patients compared to healthy volunteers. Cell surface IFN receptors are essential for proper IFN-induced signal transduction. Correspondingly, loss of IFNAR1 on T cells has been shown to play part in establishing immune-privileged niches within the TME, and influence therapeutic responses (Katlinski et al., 2017). Likewise, Zanker et al. (2021) reported that ablation of IFNAR1 impairs the NK cell-mediated regulation of tumourigenesis and metastasis in TNBC. The mechanism through which cancer

rewires or hijacks downstream signalling cascades that control both the expression levels and the effect of endogenous IFNs remains poorly understood. In addition, promoting the establishment of an immunosuppressive environment is a complex interplay between hypoxia and the regulation of type I IFN signalling (Arnaiz & Harris, 2022).

Multiple strategies are emerging to further overcome T cell exhaustion in the context of T cell-based immunotherapy and improve clinical outcomes. Targeting T cell intrinsic pathways is an attractive way to counteract the immunosuppressive effect of the TME on T cells. Advances in synthetic biology and cell engineering have made possible the rewiring of gene circuitry to reprogram T cells for enhanced survival. Prosser et al. (2012) engineered T cells capable of converting the inhibitory signal induced by PD-L1 into an activating signal through expression of a PD-1:CD28 chimera. Likewise, chimeric cytokine receptors can be engineered to redirect immunosuppressive cytokine signals into another. For example, Y. Wang et al. (2019) fused the ectodomain of interleukin (IL)-4R to the endodomain of IL-7R in order to activate the STAT3 instead of STAT6 signalling pathway in response to IL4 stimulation, thus promoting Th17 polarisation over tumour-promoting Th2 (Xin et al., 2019). Combinatorial strategies can also be used to promote or inhibit specific signalling pathways that sensitise tumour cells or directly improve the efficacy of chimeric antigen receptor (CAR)-modified T cells. A classical example is the use of checkpoint blockade monoclonal antibodies that prevent binding of PD-L1 to PD-1, thus reducing T cell anergy or apoptosis. Harnessing IFN-related pathways has also been proposed to combat T cell exhaustion. Seo et al. (2021) found that overexpression of basic leucine zipper ATF-like transcription factor (BATF) prevents progressive exhaustion of $CD8^+$ CAR T cell via its interaction with interferon regulatory factor 4 (IRF4). Another type of pattern recognition receptor (PRR), toll-like receptor 2 (TLR2), was incorporated into CAR T cells by Lai et al. (2018) to yield more robust treatment efficacy. Of interest to this thesis is the targeting of the cGAS-STING pathway. A recent paper investigated the effect of activating T cell-intrinsic STING on tumour-infiltrating CD4 T cell differentiation and effector functions (Benoit-Lizon et al., 2022). They reported that STING activation enhanced both Th1 and Th9 differentiation with more robust anti-cancer functions. According to W. Li et al. (2020), the cGAS-STING cascade plays an important role

in the anti-tumour effect of adoptively transferred T cells. Mainly, therapeutic $CD8^+$ T cells require both cGAS and STING to sustain their proliferative potential and cytokine production, particularly TNF-α and IFN-γ. This is consistent with the anti-tumourigenic action of IFNs that has been shown to correlate with tumour-infiltrating T cell infiltration levels through direct stimulation of clonal expansion and increased T cell viability. Most importantly, W. Li et al. (2020) also found that the differentiation and maintenance of T cell memory subsets upon chronic tumour antigen stimulation relies on the autocrine type I IFN signalling mediated by the cGAS-STING pathway. Furthermore, the authors provide evidence that CAR T cells could be used in conjunction with STING agonists to improve their clinical efficacy. An older study by Larkin et al. (2017), had also highlighted how STING agonists boost T cell type I IFN production in vivo. Their in vitro data, however, suggested that STING activation is a strong inducer of T cell apoptosis. These conflicting results highlight the potential multifaceted effect of STING signalling and the need to further investigate T cell-intrinsic cGAS-STING pathway. W. Li et al. (2020) hypothesise that varying degrees and duration of activation of the endogenous pathway could lead to different outcomes. This further extends to the opposing roles of type I IFNs which can paradoxically exert both protective and deleterious effects on T cells (Boukhaled et al., 2021). Thus, achieving a balanced and appropriate response will be crucial to maximise the anti-tumour immune response.

Cancer-associated defects in T cell-intrinsic interferon pathways, particularly those in the TME, warrant further investigation as potential targets to overcome immunotherapeutic resistance in cancer patients. Additional strategies are required to drive T cell therapy towards solid tumours. Combinatorial approaches or additional modifications that exploit IFN signalling pathways offer a potential avenue. In order to better harness the opposing roles of type I IFNs to improve cancer therapy it is essential to elucidate the underlying biology. This chapter dissects T cell-inherent IFN signalling pathways at a single-cell resolution in the spatial context of the TME of TNBC patients using single-cell RNA sequencing, single cell T cell receptor sequencing and spatial transcriptomics. The effect of hypoxia on interferon-stimulated gene expression patterns is also investigated, with a particular focus on the cGAS-STING pathway.

4.3 Materials and Methods

4.3.1 Gene Set Construction

Sets of genes to be used in subsequent analyses were put together using the Human Genome Organisation (HUGO) Gene Nomenclature Committee (HGNC), Kyoto Encyclopedia of Genes and Genomes (KEGG), Search Tool for the Retrieval of Interacting Genes/Proteins (STRING), GeneCards Knowledgebase, Reactome Pathway Database and by conducting comprehensive searches of the literature. Each gene was added with its name, any of its aliases, Entrez ID, Ensembl ID and a note stating which pathway of interest or process it belongs to. The primary focus was on genes either defining or involved in the following processes:

- T cell immunity
- Interferon-mediated signalling
- Immune checkpoints
- Tumour antigen targets for TNBC
- Major cell types in the TME

In addition, a list of housekeeping genes generated using the Cancer Cell Line Encyclopedia (CCLE) was provided by the Buffa Lab. All gene sets can be found in the corresponding repository for this chapter (Appendix F).

4.3.2 Patient Samples

Fresh TNBC tissue samples were collected from the Churchill Hospital (Oxford, UK) and Horton General Hospital (Banbury, UK) by Dr. Karen Sayal (Buffa Lab) who had received ethical approval. Intraoperative biopsies (4 mm in diameter) were paired with peripheral blood samples. Patient sample recruitment occurred between the end of January 2020 to March 2020, stopping prematurely because of the COVID-19 pandemic. As a result, the study was limited to three patients though many more were in the process of

being recruited (n=15). Patient samples, referred therein with prefix BRO, were collected during primary breast surgery from a treatment naive patient cohort with ER negative, PR negative, HER2 negative invasive breast carcinoma (Table 4.1). Tissue samples were frozen within the operating theatre with liquid nitrogen and 2-methylbutane, ensuring the maintenance of the spatial structure by preventing the formation of ice crystals or cracks. For one patient (BRO20), two separate needle biopsies were taken from different parts of the tumour area. Samples were stored in -80°C at the Old Road Campus Research Building (Oxford, UK) as the first lockdown started. The collection of blood samples was lost in July 2020 when the freezer failed. Fortunately, the fresh frozen tissue biopsies had recently been moved to another freezer and were saved.

Table 4.1: Clinical Profiles of TNBC Samples.

ID	Clinical Stage	Hormone Receptors	Treatment	scRNA-Seq	scTCR-Seq	Slide-Seq	ISS	ISH
BRO18	pT2 50 mm pN1a (1/2) M0	ER 2/8 PR 0/8 $HER2^-$	Naive	✓	✗	✓	✗	✓
BRO20	pT4 110 mm pN3a (33/33) M1	ER 0/8 PR 0/8 $HER2^-$	Naive	✓	✓	✓	✗	✓
BRO23	pT3 55 mm pN1a (3/18) M0	ER 3/8 PR 0/8 $HER2^-$	Naive	✓	✓	✓	✗	✓
BBP1467	pT2 N0	ER 0/8 PR 0/8 $HER2^-$	Naive	✗	✗	✗	✓	✗
BBP1557	pT2 pN2	ER 0/8 PR 0/8 $HER2^-$	Naive	✗	✗	✗	✓	✗
TRN1	*	*	*	✗	✗	✗	✗	✓
TRN2	*	*	*	✗	✗	✗	✗	✓
TRN3	*	ER 0/8 PR 0/8 $HER2^-$	*	✗	✗	✗	✗	✓
TRN4	*	*	*	✗	✗	✗	✗	✓

*Not available.

As described in subsequent subsections, the fresh tissue samples were sent for scRNA-seq, scTCR-seq and scBCR-seq, whilst the fresh frozen tissues were kept for spatial profiling

experiments using Slide-seq and in situ hybridisation (ISH) with Resolve Biosciences' Molecular Cartography platform (Fig. 4.1). Additional TNBC samples were provided by the Oxford Radcliffe Biobank, namely formalin-fixed paraffin-embedded (FFPE) and fresh frozen tissue for the in situ sequencing (ISS) and ISH experiments, respectively (Table 4.1).

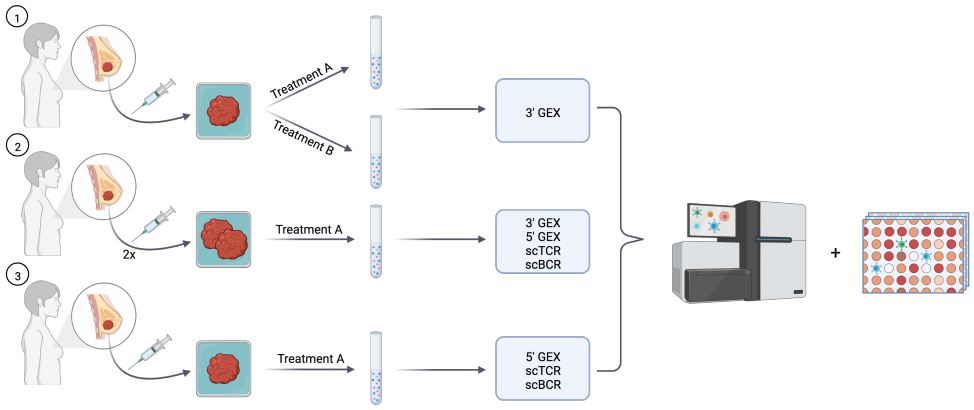


Figure 4.1: Processing of Breast Cancer Clinical Samples.
Overview of patient sample processing and data collected through 3'/5' scRNA-seq, scTCR-seq, scBCR-seq and spatial transcriptomics. Created with BioRender.com.

4.3.3 10x Genomics Single-Cell RNA Sequencing

On the same day as they were collected, the samples were processed for scRNA-seq by Dr. Karen Sayal, Dr. Dawn O'Reilly (Buffa Lab) and Dr. Thomas Carroll (Lu Lab) with the help of Dr. Neil Ashley (MRC WIMM Single Cell Genomics Facility). The solid tumour tissues were dissociated using enzymatic techniques and two different mixtures were compared:

- Treatment A — Miltenyi Biotec Tumor Dissociation Kit, human (order no.: 130-095-929).
- Treatment B — An unpublished protocol optimised for the dissociation of clinical pancreatic tumour samples and pancreatic organoids was provided by Professor David Tuveson from the Cold Spring Harbor Laboratory (Appendix D.1). It uses the enzymes Collagenase D, Liberase DL and DNase I from Sigma-Aldrich.

Since downstream analysis revealed that the treatment choice did not significantly influ-

ence resulting cell populations, the more cost-effective dissociation mixture, treatment A, was used for all samples that followed sample ID BRO18. Fluorescence-activated cell sorting (FACS) is a commonly used technique to purify and enrich for specific cell populations. In this case, cells were not FACS-sorted in order not to bias the cellular landscape. Single cell partitioning and barcoding through droplet generation was completed using the 10x Genomics Chromium Controller at the MRC WIMM Single Cell Facility. Cells loaded onto a microfluidic chip were encapsulated within individual oil droplets containing lysis buffer, reverse transcription (RT) reagents and hydrogel beads conjugated to barcoded primers (Fig. 4.2). Upon cell lysis within each droplet, the released mRNA transcripts undergo RT. Following the 10x Genomics Single Cell Protocols, the gene expression libraries were prepared with either Chromium Next Gel Bead-in-emulsion (GEM) Single Cell 3' v3.1 (BRO18, BRO20) or Chromium Next GEM Single Cell 5' Kit v2 (BRO20, BRO23). Depending on which kit is used, different ends of the mRNA are captured. Reverse transcription starts once the polydT sequence anneals to the 3' poly(A) tail. In the 3' assay, this sequence is attached to the bead and a template switch oligo (TSO) is supplied in the RT primer. The TSO hybridises with the cytosine nucleotides generated during RT of the 5' 7-methylguanosine (m7G) cap. In the 5' assay, on the other hand, the bead contains the TSO sequence and the polydT is supplied separately. Through this process, cDNAs tagged with a cell barcode and unique molecular identifier (UMI) were generated. The strands were recovered through demulsification and bead purification prior to being PCR-amplified with adapter sequences added to the 5' and 3' ends upon fragmentation in preparation for sequencing with Illumina HiSeq 4000. Since the polymerase moves along the cDNA fragments in the 3' to 5' direction, synthesising the new strand 5' to 3', the relative position of the 10x barcode decides which end is sequenced. As such, the final product either consists of the 3' end (barcode adjacent to poly(A) tail) or the 5' end (barcode adjacent to TSO). One of the sample (BRO20) was subjected to both assays to see how this influences the measured gene expression profile. The 5' library has the additional benefit of enabling the sequencing of the full length V(D)J sequence for immune repertoire profiling, as will be described in Section 4.3.4.

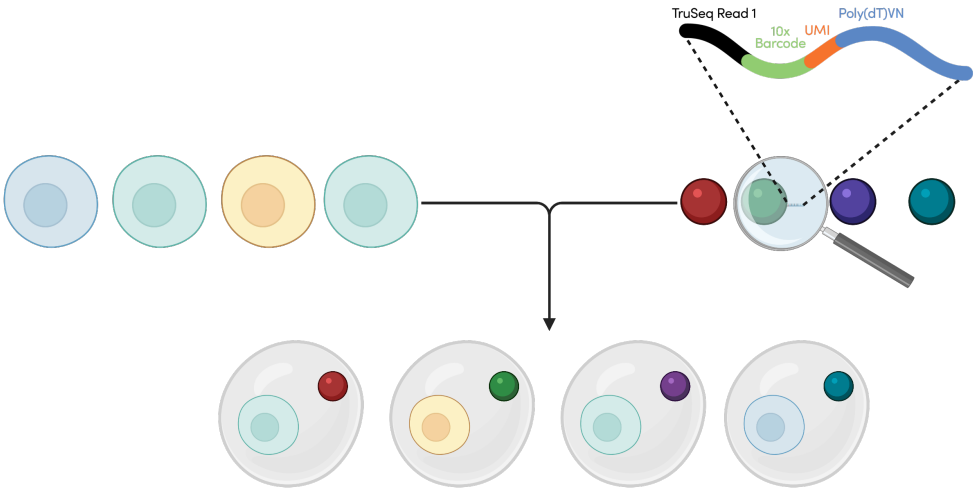


Figure 4.2: Gel Beads in Emulsion.
Schematic of oligonucleotide-conjugated gel beads being combined with oil and forming a reaction vesicle containing a single cell. Created with BioRender.com using the single cell 3' gel bead sequence from 10xgenomics.com (Accessed November 2021).

4.3.4 10x Genomics Single Cell Immune Profiling

In parallel to the above, scTCR-seq and scBCR-seq were applied to those that underwent 5' GEX, BRO20 and BRO23. Enrichment of barcoded cDNA *TCR* and *BCR* V(D)J segments (Fig. 4.3) enabled the identification of the alpha and beta chains, and heavy and light chains, respectively.



Figure 4.3: Chromium Single Cell V(D)J
Schematic of cDNA resulting from targeted enrichment of V(D)J segments using the 10x Genomics' Chromium single cell V(D)J library. Created with BioRender.com.

4.3.5 Spatial Transcriptomics

Samples BRO18, BRO20 and BRO23 were sent for spatial transcriptomics (ST) by Dr. Karen Sayal who aimed to establish an optimised experimental protocol for different technologies to be used on breast cancer tissue as part of her DPhil with the help of Dr. Esther Bridges (Simmons Lab) and Dr. Dawn O'Reilly. Thereafter, the spatial experiments were

to be analysed in this DPhil thesis to explore the spatial T cell landscape in TNBC. In addition, FFPE preserved samples, BBP1467 and BBP1557, were also used. The following two techniques were selected based on their cost, availability of equipment in Oxford and suitability for use on either fresh frozen or FFPE preserved breast tissue:

- Ex situ sequencing with Slide-seq on fresh frozen tissue
- In situ sequencing (ISS) with CARTANA on FFPE tissue

Subsequently, fresh frozen samples BRO18, BRO20, BRO23, TRN1, TRN2, TRN3 and TRN4 were sent for ST using the Resolve Biosciences Molecular Cartography technology by the author with the help of Dr. Helen Sheldon (Buffa Lab) and Dr. Esther Bridges.

Ex Situ Sequencing with Slide-seqV2

Slide-seq was found to be a cost-effective alternative to 10x Genomics' Visium and was conducted through a collaborator at the Broad Institute, the Macosko Lab (Stickels et al., 2021). The most recent protocol, termed Slide-seqV2, was used in this project. Briefly, the Macosko Lab provided profiled spatial slides (i.e. pucks). These slides contain randomly positioned 10 µm-diameter barcoded-beads, akin to the ones described in Section 4.3.3, representing a location rather than a cell, with rounds of microscopy used to record spatial coordinates. Of note, this technology provides an improved spatial resolution compared to the commercially available 10x Genomics Visium platform which has a feature size of 55 µm-diameter. In Oxford, the clinical biopsies were first mounted onto a cryostat chuck with minimum OCT medium before being carefully sectioned by Dr. Esther Bridges (Buffa Lab) and placed onto the pucks. Subsequently, the tissue was permeabilised to enable RNA diffusion and hybridisation to the spatial beads. Sequencing libraries were prepared and sequencing took place at the Oxford Genomics Centre using NextSeq with the help of Dr. Thomas Carroll. The capture efficiency of Slide-seqV2 is said to be around 3,000 reads per bead, thus hundreds to thousands of genes can be detected per 10 µm pixel (Stickels et al., 2021). Although Slide-seq enables transcriptome-wide detection using beads smaller in size than cells, it does not provide true single-cell resolution since beads may span multiple cells. Furthermore, this approach has been found to have limited power

for detecting lowly expressed genes. Importantly, hematoxylin and eosin (H&E) staining had been performed on the samples by Dr. Alistair Easton (Department of Oncology) and Dr. Karen Sayal prior to mounting. The approximated cell type proportions can be found in Table 4.2 and highlight a histologically diverse collection of breast cancer biopsies.

Table 4.2: Histological Assessment of Cell Type Proportions.
H&E images of the biopsies were collected prior to Slide-seq and cell type proportions were approximated. Courtesy of Dr. Karen Sayal (Buffa Lab).

Biopsy ID	Tumour (%)	Immune (%)	Fibroblast (%)
BRO18	30	55	15
BRO20 (1)	80	10	10
BRO20 (2)	45	5	50
BRO23	30	10	60

In Situ Sequencing with CARTANA

Two of the TNBC biopsies provided by Dr. Stephanie Jones (Oxford Radcliffe Biobank) were used for ST with CARTANA, specifically FFPE preserved ones. The samples were originally sent from the Buffa Lab to ReadCoor technologies (Boston, USA) prior to the pandemic. Due to the pandemic, this project was placed on hold for a year and their acquirement by 10x Genomics resulted in the samples being transferred to CARTANA (Sweden). Unlike Slide-seq, CARTANA in situ sequencing (ISS) consists of a targeted approach requiring the selection of a panel of genes (J. H. Lee et al., 2014). A panel of 200 pre-selected genes was provided by CARTANA and a list of 48 additional genes related to hypoxia and epithelial markers was compiled by members of the Buffa Lab (Table D.1). CARTANA ISS relies on rolling circle amplification (RCA) to amplify the signal from the gene barcodes prior to in situ sequencing using combinatorial probe-anchor ligation (cPAL). By increasing the signal-to-noise, this reduces the required microscopy magnification whilst enabling the examination of large tissue areas. Furthermore, it provides high subcellular level resolution albeit with a capture efficiency comparable to other sequencing methods due to inefficiencies associated with RCA preventing the detection of non-amplified transcripts.

Molecular Cartography with Resolve Biosciences

Resolve Biosciences' Molecular Cartography platform was chosen to run experiments specific to this chapter's research questions because of it's ability to resolve subcellular patterns in situ at a higher resolution than transcriptome-wide methods. It extends upon single molecule fluorescence in situ hybridization (smFISH) and a proprietary imaging technology co-developed with ZEISS. By employing multiple rounds of hybridisation, it detects mRNA transcripts more efficiently than ISS. However, it is important to find the right balance between the number of genes profiled and the number of hybridisation rounds necessary to prevent potential compounded errors (Williams et al., 2022). As with other in situ hybridisation (ISH) methods, it is not only limited by optical crowding but is subject to a detection error rate that grows with each round (K. H. Chen et al., 2015).

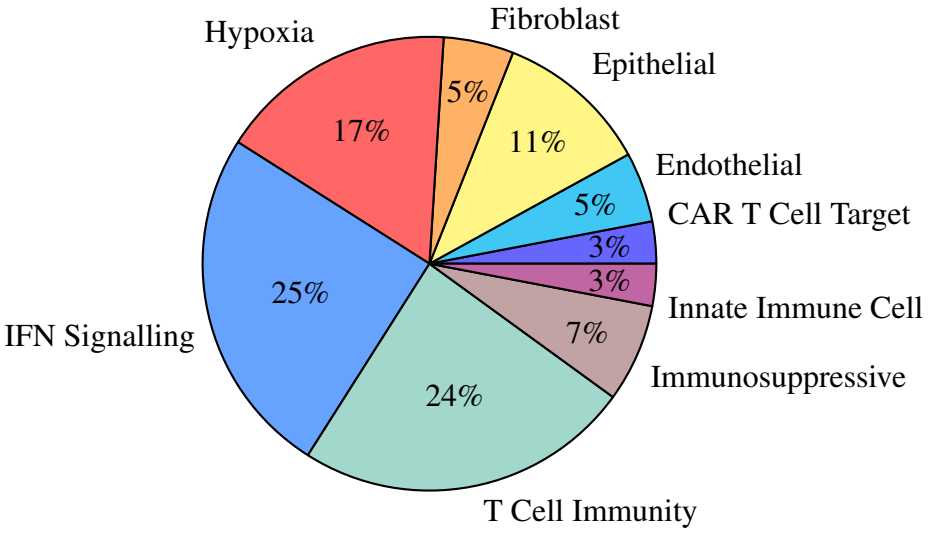
Probes were designed using the gene sets described in Section 4.3.1 and Resolve's probe design panel to ensure that the following criteria were met:

- Target RNA length $>$ 500 nucleotides for probes to anneal
- Removal of highly expressed genes to prevent optical crowding

A maximum of 100 genes could be selected for the experiment and the steps taken were as follows:

1. Define target cell types, pathways and regions of interest
2. Filter genes in gene sets from Section 4.3.1
3. Study expression levels and distributions in TNBC scRNA-seq samples
4. Study expression levels in bulk RNA-seq from The Cancer Genome Atlas
4. Score using Resolve Bioscience Probe Designer
5. Cross-check genes targeted in CARTANA spatial samples

The final transcript panel is listed in Table D.2 and consists of the following proportions:

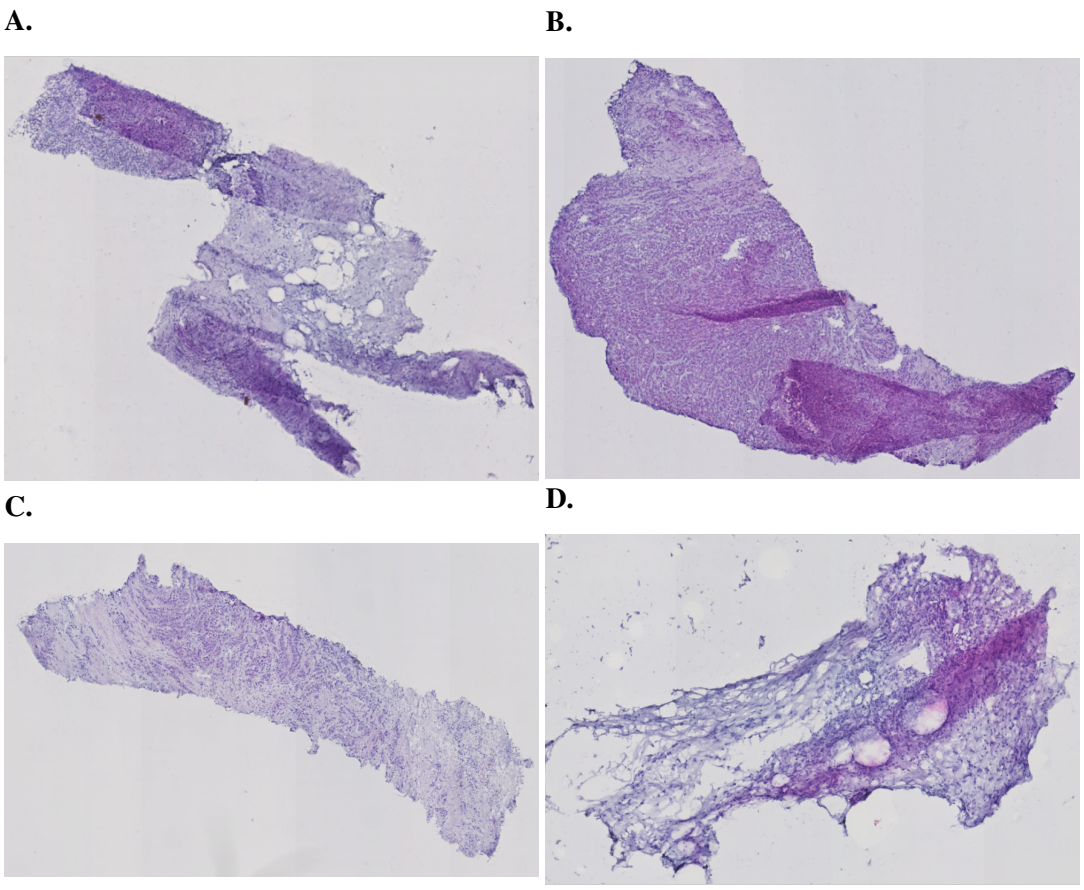


Once the panel of probes was ready, significant delays were experienced due to limited laboratory access caused by COVID-19, additional regulations imposed by Brexit and difficulties encountered when shipping the tissue samples to Germany. All documents pertaining to this project including the scope of work, biosafety level declaration, tender exemption request form, purchase order and material transfer agreement can be found in Appendix D.

Microscopy slides with eight wells are used for the experiment. A total of eight samples coming from seven patients with prefix BRO and TRN in Table 4.1 were used. According to Dr. Jeroen Aerts (Resolve Biosciences), the platform has low technical variability. As such, the number of patient samples was favoured over having biological replicates. Regardless, true replicates were not possible as they would have consisted of cuts made at different sites of the tissue. Cross-sections of all eight freshly frozen tissue samples were made by Dr. Esther Bridges with the help of Dr. Helen Sheldon. Sectioning was performed within a cryostat and H&E staining was used for section selection (Fig. 4.4). Sample sections 10 μm thick were transferred onto the Molecular Cartography glass slides provided by Resolve Biosciences (Appendix D.8).

The specimens were sent from Oxford (UK) to Monheim am Rhein (Germany) via FedEx with additional dry ice. Fortunately they arrived intact despite delays at custom due to strikes. Bright-field whole slide images (WSIs) of the human breast biopsies were taken by

Resolve Biosciences. They requested that the regions of interest (ROIs) be drawn, choosing up to 300 tiles out of the blue superimposed gridlines (Fig. 4.5). This step was kindly completed by a trained clinical pathologist, Dr. Alistair Easton. Resolve Biosciences ran the experiment using the panel of 100 transcript-specific probes which hybridised to the corresponding RNA species in situ. Their proprietary technology sequentially colourises and decolourises the probes, running several imaging rounds to locate and quantify the barcoded fluorescent spots.



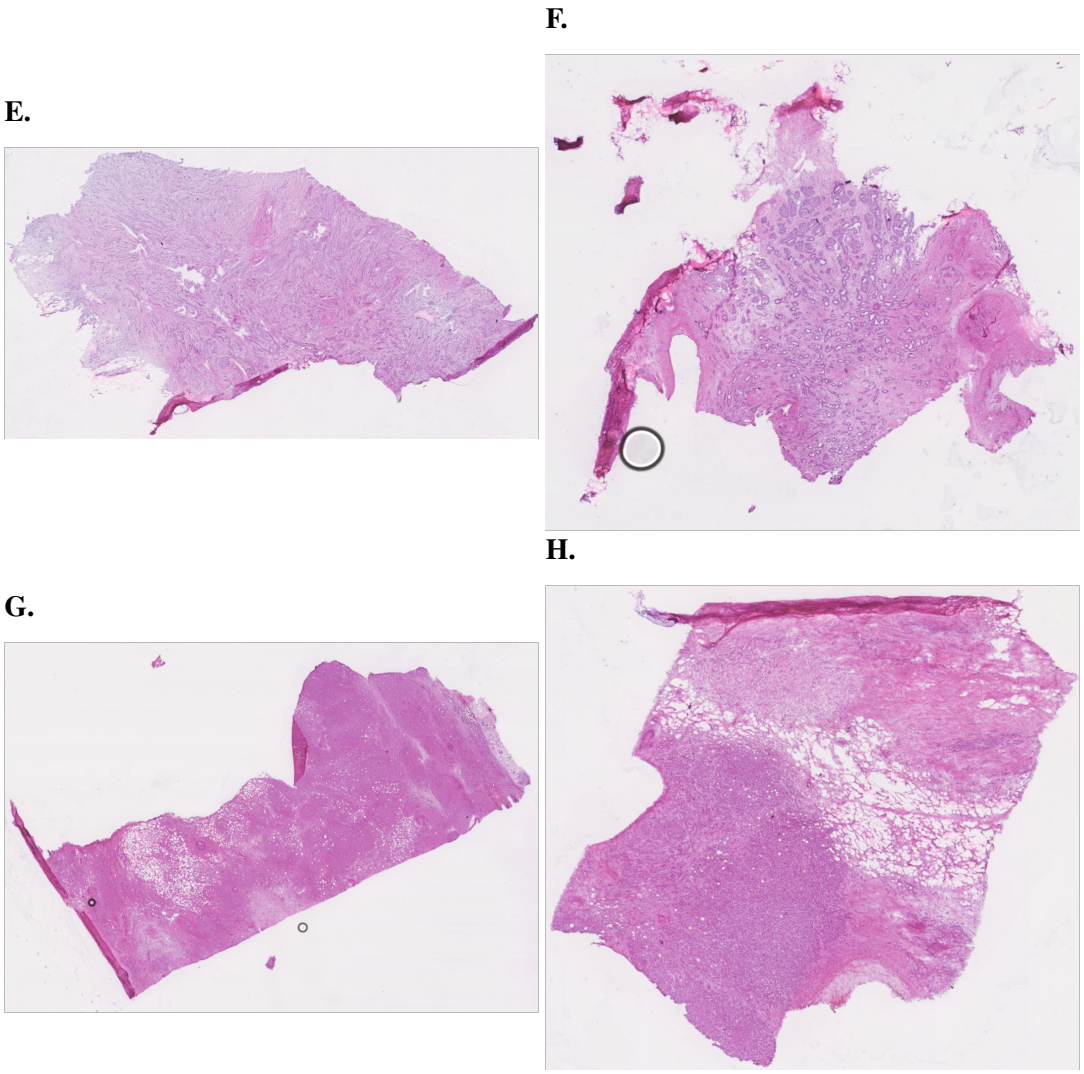
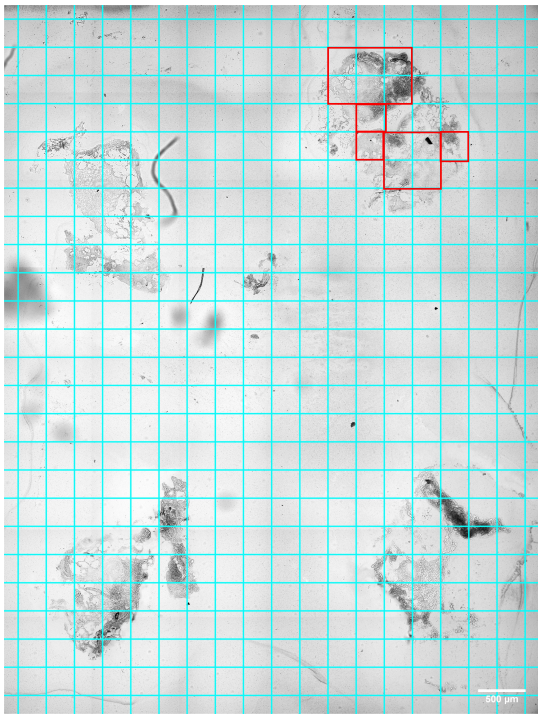
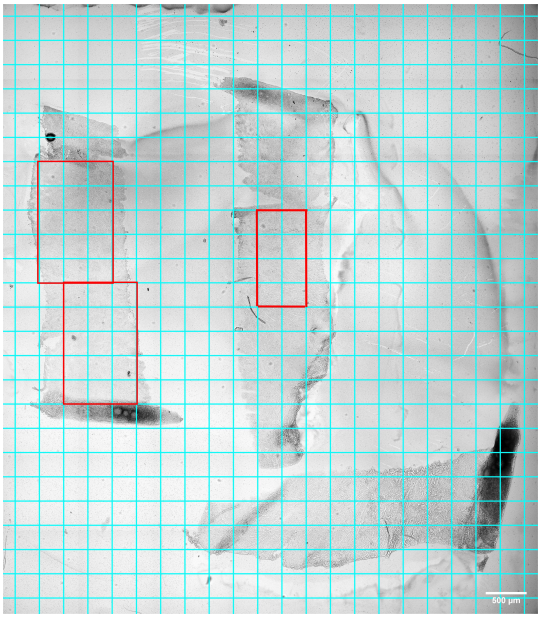


Figure 4.4: H&E Staining of Tumour Tissue Samples.
H&E staining was used to examine the patient biopsies for sample selection.
A. BRO18. **B.** BRO20. **C.** BRO20. **D.** BRO23. **E.** TRN1. **F.** TRN2. **G.** TRN3. **H.** TRN4.

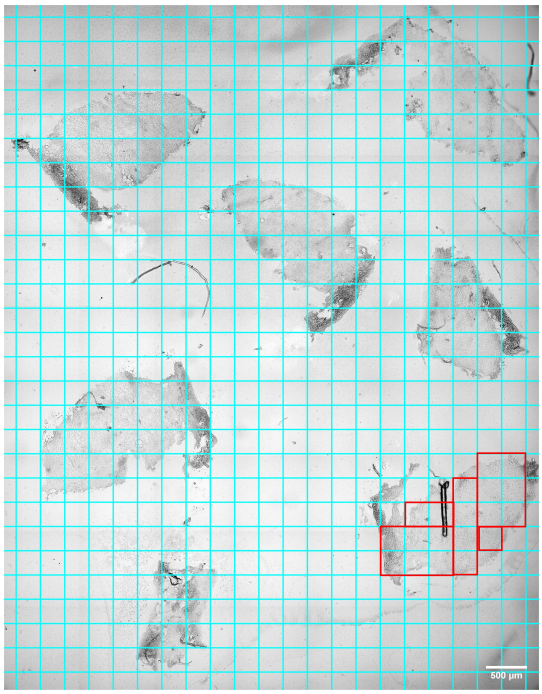
A.



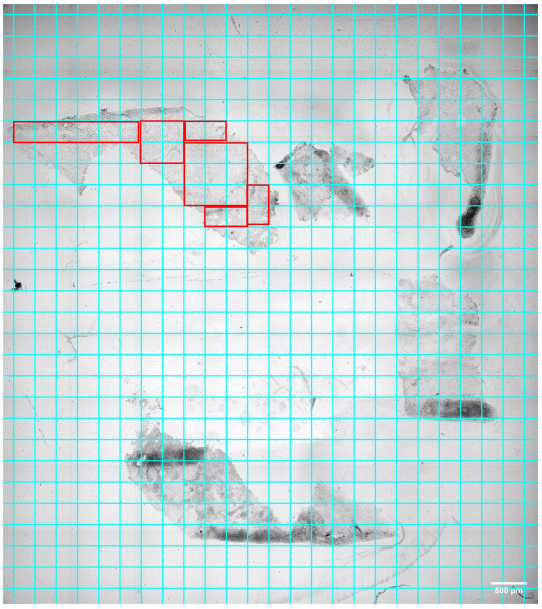
B.



C.



D.



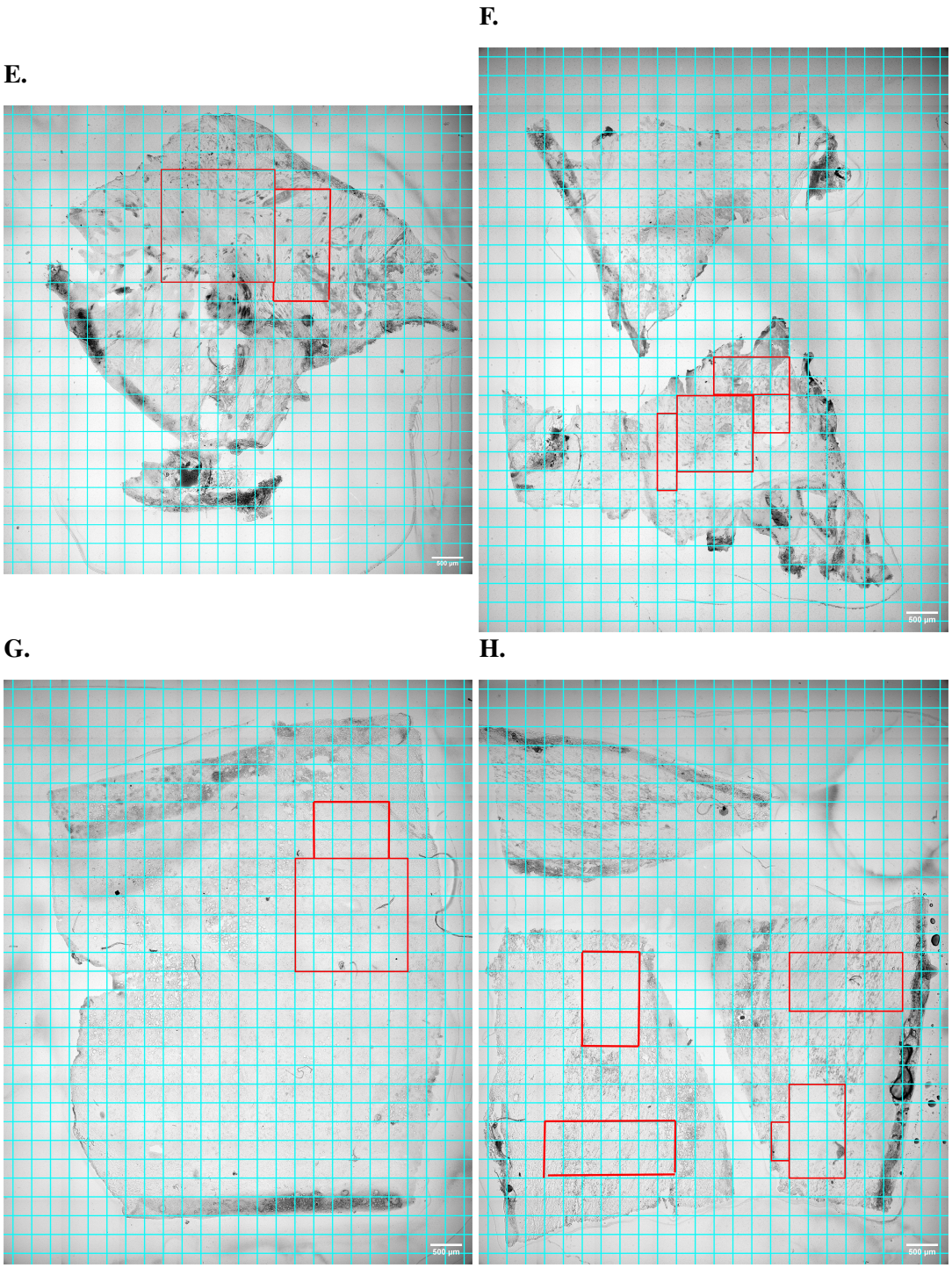


Figure 4.5: Whole Slide Images and Regions of Interests.
Whole slide images of the eight breast tumour samples with red lines depicting selected regions of interests for the spatial transcriptomics experiment. **A.** BRO18. **B.** BRO20. **C.** BRO23. **D.** BRO20. **E.** TRN1. **F.** TRN2. **G.** TRN3. **H.** TRN4.

4.3.6 Processing Chromium Single-Cell Data

Sequencing results generated by the Illumina sequencer for the scRNA-seq, scTCR-seq and scBCR-seq runs were processed using 10x Genomics Cell Ranger (v5.0.0) (Zheng et al., 2017) and aligned to Human reference GRCh38 (v3.0.0). The Medical Sciences Division Biomedical Research Computing cluster was used for this step. Briefly, the raw base call (BCL) files were demultiplexed to generate FASTQ files (cellranger mkfastq). Subsequently, feature-barcode matrices for the scRNA-seq runs were generated (Fig. 4.6) following sequence alignment, filtering, barcode and UMI counting (cellranger count). Likewise, the cellranger vdj pipeline was used on the data produced from Chromium Single Cell 5' V(D)J libraries. The aggregate functionality (cellranger aggr) was used to combine the two different runs of BRO18 that had been treated with either Miltenyi or CSHL, normalising them to the same sequencing depth and recomputing the gene expression matrix. Cell Ranger summary metrics describing sequencing quality for each run can be found in Appendix D.9.

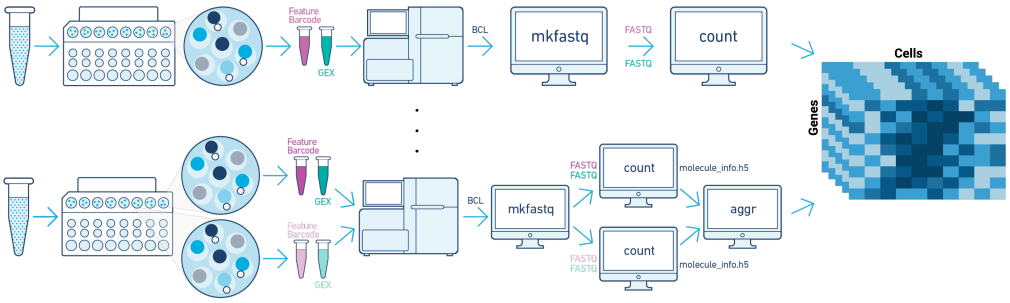


Figure 4.6: 10x Genomics Cell Ranger Pipeline for GEX
Schematic of the Cell Ranger pipeline used to generate feature-barcode matrices. Adapted from 10x Genomics using BioRender.com.

4.3.7 Analysis of Single-Cell T Cell Receptor Sequencing

Whilst the assembled BCR sequences were just used downstream to identify B cells in the scRNA-seq data, an in-depth analysis of the TCR repertoire was conducted. First, the identified CDR3 and TCR genes were analysed and compared across the two patients to give an overall view of the TCR repertoire. A clonotype is defined therein by the nucleotide sequence of the CDR3 region and paired VDJC genes comprising the TCRA and

TCRB chains. Subsequently, the scTCR-seq data was combined with the scRNA-seq to conduct a joint analysis and filter out ones that had not been annotated as members of the isolated T cell group (Section 4.3.8). Visualisations were made using both immunarch (v0.6.8) by ImmunoMind (2019) and scRepertoire (v1.3.5) by Borcherding et al. (2020). Clonotypes were ranked by frequency and binned to visualise the relative space they occupy. The Morisita index, a dispersion measure, was used to calculate the level of clone overlap between clusters. Diversity across clusters was estimated using five metrics (Shannon, inverse Simpson, Chao1, abundance-based coverage estimator, Pielou's evenness) and downsampling with 100 bootstraps. Shannon and inverse Simpson estimate baseline diversity whilst the others sample richness.

4.3.8 Analysis of Single-Cell RNA Sequencing

Quality Control

All analysis presented therein was conducted in R (v4.1.2). Ensuring the quality of the data being analysed is an essential first step. The raw count matrices were used to calculate the total number of detected molecules (UMIs) and features in each cell. Mitochondrial proportion (mtDNA) was calculated as the percent of features with prefix 'MT-'. All three metrics were used to filter out low-quality cells. A threshold of 20% mtDNA was agreed to be the most appropriate upon discussion with clinical oncologists in the Buffa Lab. Threshold values for the other two metrics were decided upon careful examination of the distributions. In addition, outliers were defined as those 2 median-absolute-deviation (MAD) away from the median once transformed to the log10 scale:

$$\text{MAD} = \frac{\sum |x_i - \bar{x}|}{n}$$

Dataset Integration

In order to correct for batch effects and technical variations that may be obscuring biological signal, datasets were harmonised using the standard Seurat (v4.0.6) integration workflow (Butler et al., 2019; Satija Lab, 2022), similar to what was described in Section

3.3.2. Individual gene expression matrices were normalised using regularised negative binomial regression as described in Hafemeister and Satija (2019). The features were ranked by residual variance (i.e. the unexplained or error variance) and the 3,000 top ones were selected for each dataset. A subsequent ranking is used to find integration features, filtering them by the number of datasets they are variable in. These features were then used to find cross-dataset pairs of cells that are predicted to share a common biological state, known as anchors. Using these anchor points or mutual nearest neighbours, the datasets were iteratively integrated together with 100 neighbours and 50 dimensions used in the anchor weighting procedure. This resulted in a common corrected data matrix being constructed. A more detailed explanation of the integration workflow can be found in the original publication by Butler et al. (2019).

Dimensionality Reduction and Clustering

As a mean to reduce the dimensionality of the data, principal component analysis (PCA) was performed on the integrated matrix. The top 30 principal components were used for visualisation with uniform manifold approximation and projection (UMAP) and graph-based clustering with shared nearest neighbour (SNN) with a resolution of 0.5.

Differential Gene Expression Analysis

In order to annotate major cell types, the most highly variable genes (HVGs) across clusters were identified using a Wilcoxon Rank Sum test and by measuring those that are differentially expressed across clusters but conserved across datasets in each cluster. Gene marker p-values and combined p-values were returned. Significant gene markers were manually checked against the literature and the gene sets compiled in Section 4.3.1. These were used to assign cell type labels to the clusters.

Isolating T Cells

T cells were isolated using a voting ensemble approach which was deemed to be more appropriate for reasons highlighted in Chapter 3 (Fig. 4.7). Cells that had been assigned in the last section to a cluster labelled as ‘T Cells’ were given a score of 1. Reference-based

methods, specifically SingleR and scPred, the best available methods at the time this was performed, were used as described in Chapter 3. A score of 1 was added for each T cell call. All those with a score of 2 or 3 were isolated for downstream analysis. By combining vastly different ways of identifying T cells – marker-based, Spearman correlation and SVM – the T cells were isolated with greater confidence.

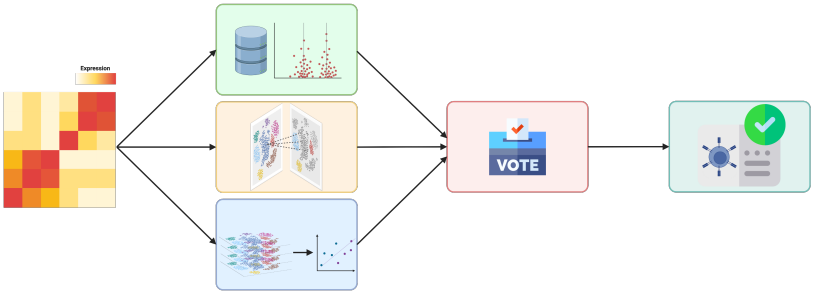


Figure 4.7: Ensemble Method for T Cell Isolation.
Three different cell identification methods that rely on marker-based expert curation, Spearman correlation and SVM were combined to assign a T cell prediction score. All cells with a score of 2 or 3 were isolated for downstream analysis. Created with BioRender.com using icons by Freepik from Flaticon.com.

Characterising T Cells

T Cell Profiles A variety of methods were used to further refine T cell populations and transcriptional profiles. To start, isolated T cells were re-integrated and re-clustered with a resolution of 0.4 as described above. The most differentially expressed genes across the new clusters were identified and an unbiased labelling approach was used. T cell clusters were labelled based on their transcriptional profile, not swaying them towards a specific lineage as conventionally done. To study the potential T cell subtype composition in each cluster, the best performing predictors, Azimuth and SingleR, were used as described in Chapter 3. The Azimuth paper had recently been published at the time this part of the analysis was performed.

TCR Repertoire The scTCR-seq data was projected onto the T cell space to run a joint analysis of the TCR repertoire and gene expression (Section 4.3.7). TCR sequences that did not find a corresponding cell were discarded.

T Cell States The Molecular Signatures Database (MSigDB) was searched for gene

signatures defining T cell dysfunctional states, particularly anergy and exhaustion. These dysfunctional states are traditionally defined by high expression of immune checkpoint markers. A classical PD-1 signature of dysfunction published by Duraiswamy et al. (2011) was downloaded from the MSigDB (GSE26495). However, this cannot be an accurate way of distinguishing between activated phenotypes and dysfunctional ones since TCR-mediated activation is also known to promote the upregulation of inhibitory receptors such as PD-1 and CTLA-4. As such, the literature was searched for alternative signatures that specifically decouple the highly intertwined activation and dysfunction transcriptional profiles. A distinct signature for T cell dysfunction uncoupled from T cell activation had been derived by Singer et al. (2016). The gene modules had been generated using transgenic mice and tested on T cells isolated from human melanoma tissue. Since the genes in the signature were from the organism *Mus musculus*, the National Institutes of Health (NIH) National Center for Biotechnology Information (NCBI) database was searched for the human homologs. If a homolog was not found, the gene was removed. Three distinct methods were used to calculate gene set enrichment due to the absence of a gold standard. The first consisted of using Seurat's AddModuleScore function to average and bin relative expression levels at the single cell level (cf. Tirosh et al. (2016) for more details). In addition, scores were measured using a more recent method based on the Mann-Whitney U test, UCell (Andreattaa & Carmonaa, 2021). An extension of the classical Gene Set Enrichment Analysis (GSEA), single-sample GSEA (ssGSEA), was also used to determine the level of activation and dysfunction in each cell (Yi et al., 2020). Scores of T cell dysfunction calculated using either the Duraiswamy et al. (2011) signature or Singer et al. (2016) gene module were compared.

IFN Signalling One of the isolated T cell cluster had a pronounced IFN signature identified through HVGs. It was further analysed using clusterProfiler (v4.2.1) to find specific Gene Ontology (GO) terms enriched in the cell cluster related to specific pathways or cellular processes (T. Wu et al., 2021). The frequency of fold changes for each gene was also used to identify pathways that are likely to be up-regulated or down-regulated in this group compared to other T cells. Similarly, fold changes were mapped to KEGG pathways, mainly cytosolic DNA-sensing pathway (hsa04623), RIG-I-like receptor signalling

pathway (hsa04622), Toll-like receptor signalling pathway (hsa04620) and NOD-like receptor signalling pathway (hsa04621). IFN signalling scores were calculated using the same methods described above for T cell states. Pathway-specific IFN gene sets were generated from the gene sets in Section 4.3.1. The MSigDB was also searched for IFN-specific signatures. The Gene Matrix Transposed (GMT) files corresponding to the hallmark gene set collection were downloaded and a sub-collection of genes specific to the IFNα response was used (Liberzon et al., 2015). Expression correlations between genes of interest were calculated using Pearson's correlation on the RNA counts.

Trajectory Inference T cell trajectories were inferred using Slingshot (v2.2.0) by Street et al. (2018). Briefly, new clusters are generated from the low-dimensional embeddings using Gaussian mixture models which assume that Euclidean distances reflect biological differences. The clusters are subsequently connected using a minimum spanning tree. Principal curves are fitted simultaneously and orthogonal projections are used to overlay pseudotime for each detected lineage. A more detailed explanation of the method can be found in the original publication. Activation, dysfunction and IFN scores were subsequently projected onto the lineage path.

Cell-Cell Interactions Cellular crosstalk between different cell clusters was investigated using CellPhoneDB (v2.0.0) which brings together several biological databases (e.g. Ensembl, PDB) to assess potential ligand-receptor pairs (Efremova et al., 2020). Ligands and receptors expressed in at least 10% of the cells in a cluster are considered significant. Pairwise comparisons were performed to identify significant ligand-receptor interactions. Briefly, cluster labels get randomly permutated 1,000 times to generate a null distribution from the calculated mean receptor and ligand expression levels. The p-value is measured from the proportion of means that are greater or equal to the actual mean and the false discovery rate is used to adjust the p-values. The pairs with an adjusted p-value higher than 0.05 are filtered out. The total interaction score for each cluster pair consists of the sum of significant ligand-receptor pairs between them. InterCellar (v2.0.0) was used to generate sunburst plots and explore significant functional terms or related biological pathways between pairs (Interlandi et al., 2022).

4.3.9 Analysis and Visualisation of Spatial Transcriptomics

Spatial transcriptomics technologies are still in their early stages. As such, downstream analysis methods have not yet been clearly established which brings additional challenges. Due to the severe delays caused by the COVID-19 pandemic in acquiring the Slide-seqV2, CARTANA ISS and Molecular Cartography datasets for this project, an in-depth analysis could not be completed within the limited time frame of the DPhil as initially planned. This section covers the preprocessing and initial analysis that was possible before this thesis was submitted to the Examination School (University of Oxford).

Sequencing-Based Technologies

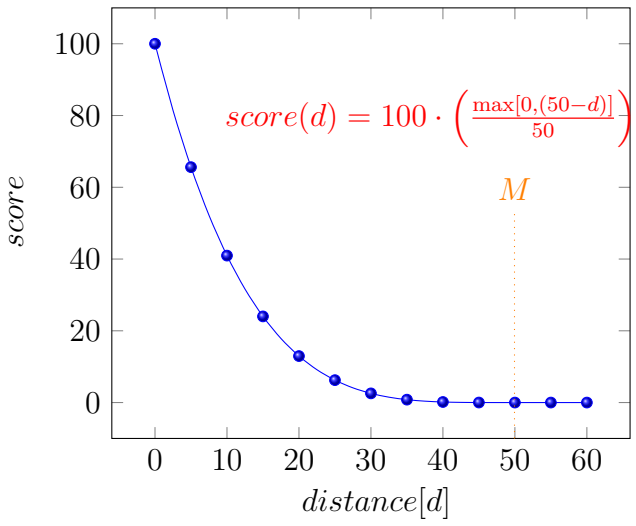
As described earlier, Slide-seqV2 is a sequencing-based technology which relies on NGS rather than microscopy, thereby avoiding image-processing pipelines. The initial processing of the puck data which required mapping the sequencing results back to the bead spatial locations was conducted by the Macosko Lab (Broad Institute) since the pipeline had yet to be made publicly available. Downstream analysis of the four pucks was conducted in R (v4.1.2) using Seurat (v4.0.6). All detected beads were kept regardless of UMI counts. Similar to the workflow described in Section 4.3.8, regularised negative binomial regression was used for normalisation prior to dimensionality reduction using PCA and clustering by constructing a shared nearest neighbour graph. The scRNA-seq cell atlas created in Section 4.3.8 was used to assign cell types to spatial locations. Cell labels from the scRNA-seq reference were transferred using the method described in Stuart et al. (2019). Major cell type labels were transferred in a patient-specific manner. The dataset was projected onto the PCA of the reference with 30 dimensions used in the anchor weighting procedure and 50 neighbours. For every bead on the array, a cell type prediction score was assigned based on gene expression profile similarity and the type corresponding to the highest score was selected as the cell identity. Spatially variable features were calculated using the Moran's I test for spatial autocorrelation implemented by Seurat (v4.0.6) which quantifies the dependence of a feature on location.

Imaging-Based Technologies

Computer-aided enhancement turn raw images into data that can be evaluated, measured, and quantified. Image processing pipelines largely vary depending on the technology used to acquire the images, the type of image, and the user expectation. A variety of operations can be performed on biological images that include noise reduction, contrast enhancement, and quantifying intensities. Diverse software are designed specifically to work with multi-dimensional, heterogeneous biological datasets including QuPath (Bankhead et al., 2017) and Fiji/ImageJ2 (Schindelin et al., 2012; Rueden et al., 2017). These two software were used to process both the Resolve Biosciences and CARTANA datasets as described below. For each of the eight samples sent to Resolve Biosciences, the data received consisted of three kind of TIFF images (brightfield, DAPI, raw) and a matrix of transcript coordinates (X,Y, Z). Similarly, CARTANA provided DAPI TIFF images with (X,Y) coordinates for the two samples. A total of 18 ST images generated by Molecular Cartography and two using CARTANA ISS were analysed.

Tissue-wide colocalisation of targeted transcripts was measured for individual images using a scoring method derived by Resolve Biosciences (unpublished). Briefly, for each detected transcript a linkage score is assigned to all signals within a maximal distance, M. A distance penalty score greater or equal to 0 of p defines the relation of distance to score. Scores between transcript pairs were calculated as follows:

$$100 \cdot \left(\frac{\max[0,(M-d)]}{M}\right)^p$$



Distances are measured in pixel units. A search radius M of 50 and a penalty score p of 4 were used, representing an exponentially decaying relationship between distance and score up to a distance of M. The stepsize was set to 1 unit in X-Y and 3.50 units in Z, where applicable. Since the sums of the scores are biased by transcript abundance, a normalisation step is used for correction, where μ_s is the mean of the linked transcript scores:

$$Enrichment = \log_2 \frac{score}{\mu_s}.$$

A dual-indexing method helped reduce the computational complexity. An all-versus-all approach for n transcript signals would otherwise require $O(n^2)$ computations. Instead, the images were compartmentalised into tiles of length M such that transcripts are assigned to a tile and only distances between objects in the same or adjacent tiles are measured. Scores were scaled to a range between -2 and 2, with higher scores representing closer spatial proximity between gene pairs and -2 characterising anticolocalisation. For each imaging-based technology which represented different gene sets, the colocalisation matrices were averaged across all datasets that passed quality control. Gene pairs present only in a fraction of the datasets were average across the number of samples they were detected in. For transcript visualisation, selected sets of genes were superimposed onto the images using the Python package TissUUmaps (Pielawski et al., 2022). Analysis of the Molecular Cartography images proceeded past this point, to cell segmentation and cell type identification.

Cell segmentation which consists of enclosing cells within an inner and an outer contour for them to be extracted from the image was conducted using QuPath with a pixel size of 0.1380 x 0.1380 μm or 1.38 x 1.38 nm. Digital images are comprised of pixel values and embedded metadata (bitdepth, image type, pixel sizes) which can be read as numerical data. Pixels collectively represent objects or features of images. Meaningful biological regions can be extracted by partitioning the image into meaningful regions or volumes based on pixel intensities and colours, boundary changes and underlying patterns in pixel distributions. Specifically, objects such as cells or fluorescent probes will be represented by a collection of pixels with specific spectral properties, size, shape, local homogene-

ity and spatial relationship with surrounding pixels. For these objects to be detected, the underlying segmentation algorithm will rely on local features within the image domain that distinguish them from their immediate surrounding (e.g. corners, edges, colour, intensity). These patterns within subsets of the image domain help describe the regions of interest (ROIs). Different sigma and cell expansion values were tried to optimise the cell detection process and minimise artefacts for each image with the requested pixel size set to 0.25 μm (Fig. 4.8). Cell segmentation for all other images can be found in Appendix D.10. Boundaries around the detected nuclei were extended to catch the full cell area by increasing the cell expansion value to be between 5-8 μm. Sigma values in the range of 1-2 μm were used for noise reduction with Gaussian smoothing in an attempt to find the right balance between boundary detection and nuclei fragmentation. The segmented ROIs representing individual cells were extracted.

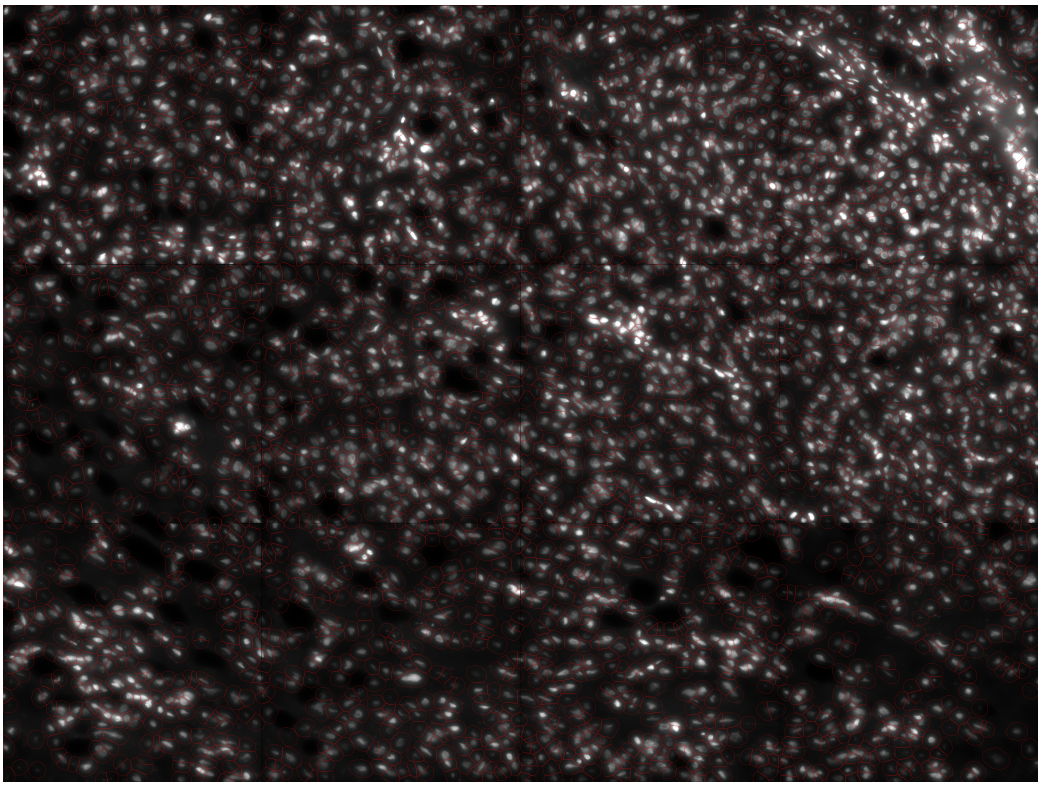


Figure 4.8: Cell Segmentation.
Whole-cell segmentation of tissue images using DAPI-stained nuclei as reference. The image above consists of sample TRN3 in well D1.

Fiji/ImageJ2 also support a multitude of algorithms to explore and process images such as single-channel segmentation, channel subtraction, colour correction, filtering, thresholding, and selecting ROI. These methods help overcome quality limitations and remove acquisition artefacts prior to computing meaningful measurements. DAPI images were adjusted and the Fiji/ImageJ2 plugin, Polylux, provided by Resolve Biosciences was used to overlay gene transcripts using matching pixel coordinates. Coordinate (0,0) is located in the top left corner of the image by convention. Moreover, the extracted ROIs from QuPath were used to locate cells. Sets of genes were split into colour-coded groups in an attempt to highlight regions associated with the categories in Table D.2 (e.g. hypoxic zones) or cell types. Density maps were used to quantify and compare the level of transcripts or gradients across different regions. This feature was also used for quality control, to identify regions with lower concentrations of transcripts that could be a result of cutting artefacts. A minimum threshold of 90 genes detected was used to select images for downstream analysis. The overlayed cell ROIs were used to extract transcript counts per cell for analysis in R using Seurat (v4.0.6), similar to scRNA-seq analysis.

The number of genes per cell and total signal per cell was calculated for each sample. Genes not expressed in at least one cell, and cells not expressing at least one transcript were removed. The counts were log normalised with a scale factor of 1. Variable features were calculated as described in Section 3.3.2. The scaled data underwent PCA dimensionality reduction and a scree plot was examined to determine the number of PCs to retain. Cells were clustered using KNN and SNN as done in Sections 3.3.2 and 4.3.8. Since no spatial information was provided, cells were clustered based on the similarity of their transcriptome for the selected gene panel. These were visualised using UMAP and the most variable features across clusters were calculated using receiver operating characteristic analysis, as recommended by Resolve Biosciences. Cell cluster memberships were exported and used in Fiji/ImageJ2 to colour cells based on cluster ID. As a result, RNA transcripts could be localised in the spatial context of segmented cell ROIs and cluster IDs. To help guide cell identification, the scRNA-seq data was used as reference by calculating the average expression of all 100 transcripts targeted by the panel of spatial probes in each cell group.

4.4 Results

4.4.1 Patient Samples and Characteristics

As described in Section 4.3.3, breast tissue samples from three treatment naive TNBC patients were collected during tumour excision and sent for 10X Genomics Chromium scRNA-seq. Following quality control, single-cell datasets were integrated into a single object for downstream analysis, and a total of 28,292 cells were recovered (Figs. 4.9A–4.9C).

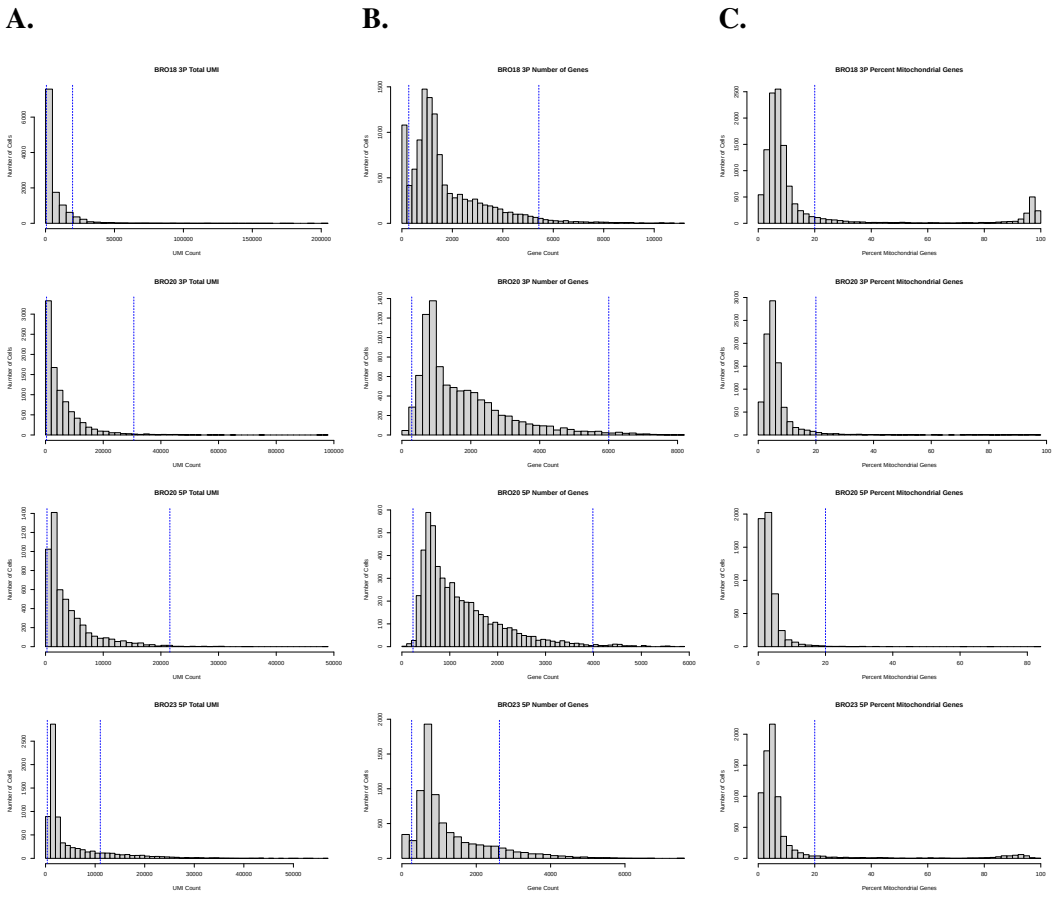


Figure 4.9: Quality Control of Single-Cell RNA Sequencing Datasets.
A. Distributions of UMI counts in pre-filtered scRNA-seq data. Blue lines represent computed outlier thresholds based on median-absolute-deviation. **B.** Distributions of feature counts in pre-filtered scRNA-seq data. Blue lines represent computed outlier thresholds based on median-absolute-deviation. **C.** Distribution of percent mitochondrial genes in pre-filtered scRNA-seq data. Blue lines represent 20% threshold.

Patient sample BRO18 had been subjected to two different tissue dissociation protocols to see whether populations recovered from the resulting single cell suspensions are affected by the choice of treatment. The two isolation protocols yielded comparable results, with no significant differences in cell composition (Fig. 4.10A) and contamination with mitochondrial genes (Fig. 4.10B). However, the BRO18 dataset which had been treated with Miltenyi showed a higher mean total reads per cell prior to depth normalisation but a lower total cell count than the one treated with the CHSL pancreatic protocol. Gene expression libraries (GEX) were constructed using either the 3' or 5' assay, with one sample sent for both (BRO23). Looking at the expression levels of housekeeping genes, it was found that *ALDOA* transcripts were only detected in the 5' assay (Figs. 4.10C–4.10D).

As described in Section 4.3.8, integration methods were used to remove batch effects and technical factors whilst conserving biological variation between datasets. Unsupervised clustering of the integrated data gave rise to 21 clusters present in all patients (Figs. 4.11A–4.11C). The most highly variable genes (HVGs) across clusters were calculated to identify major cell populations based on known marker genes (Appendix D.11). Out of the 10 identified cell types, T cells and fibroblasts represent the largest populations across all datasets (Fig. 4.12A). T cell receptor (TCR) and B cell receptor (BCR) VDJ segments available for two of the patients (BRO20 and BRO23) were projected onto the scRNA-seq space to enable joint analysis of the transcriptome and immune repertoire (Fig. 4.12B). For the most part, TCR and BCR segment projections agreed with cluster labels, and few outliers were detected. Cells can be found to express TCR or BCR gene segments due to technical noise or may be expressing it at the mRNA transcript level but not translating it into proteins.

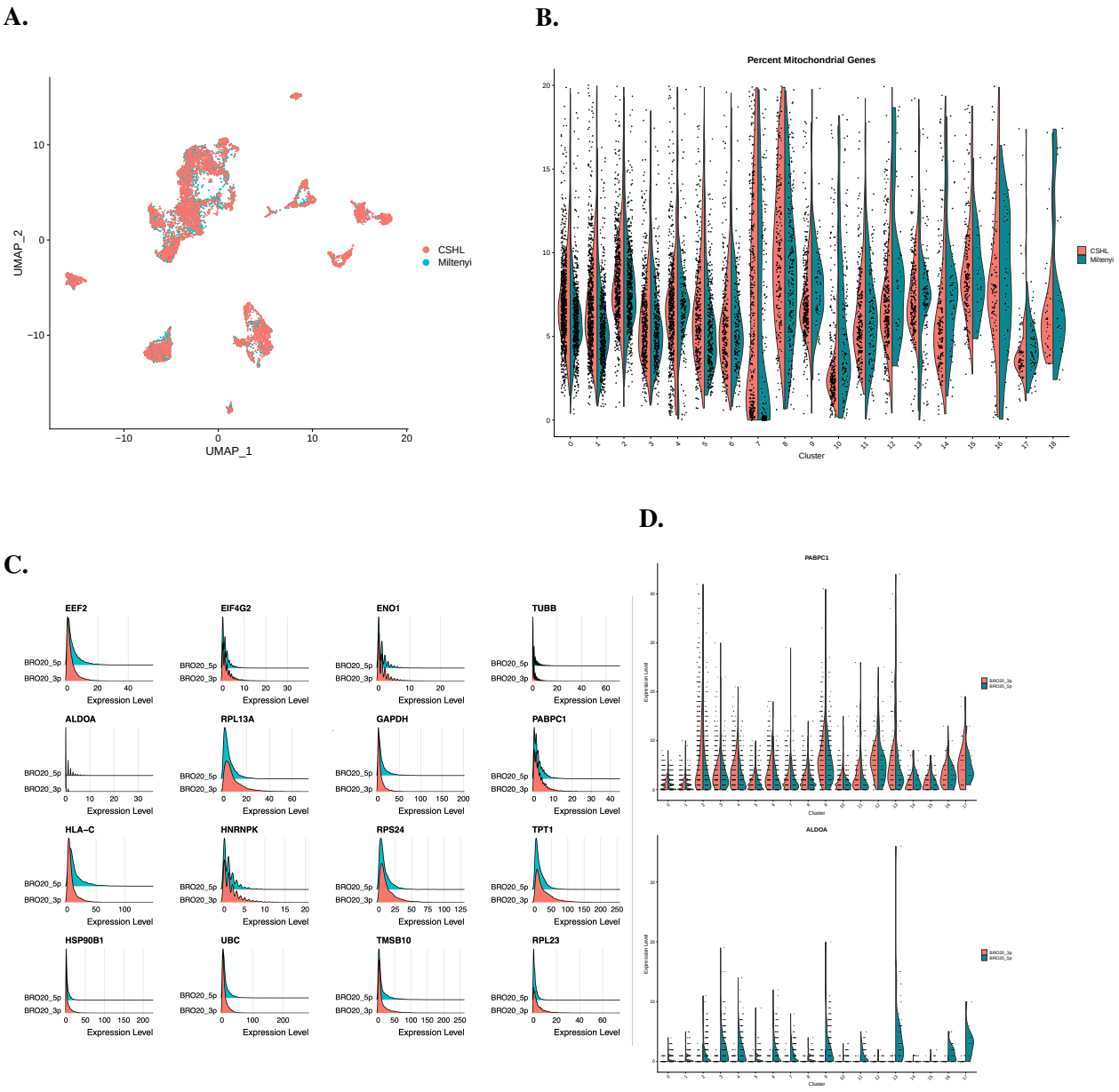
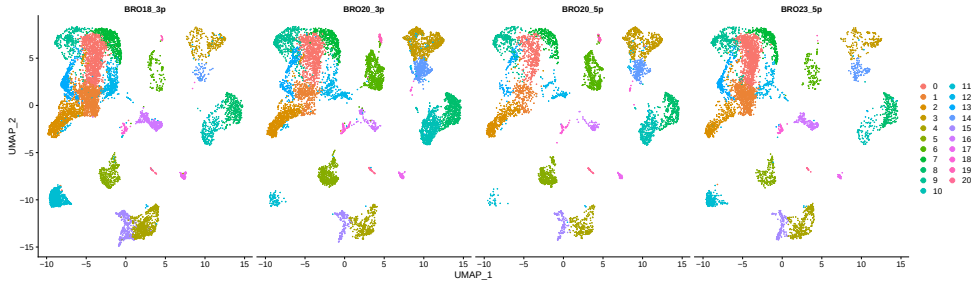
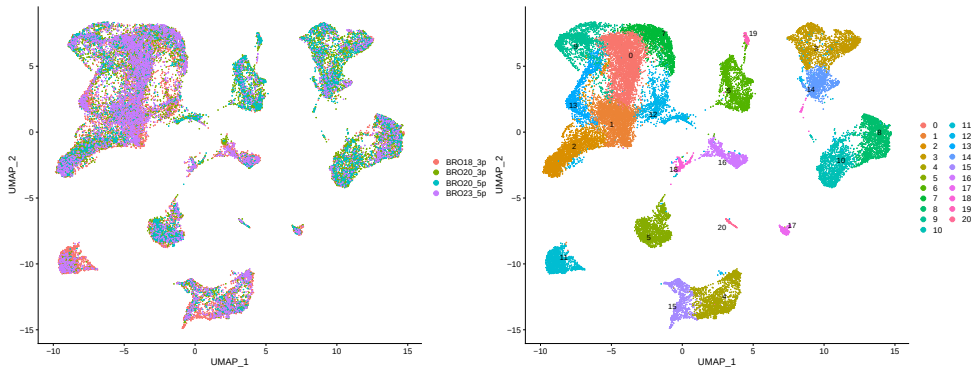


Figure 4.10: Tissue Dissociation Method and Assay Chemistry Comparison. **A.** Reduced dimensionality visualisation (UMAP plot) for sample BRO18 treated with either the Miltenyi or CSHL dissociation protocol. **B.** Violin plot comparing percent mitochondrial genes per cluster for BRO18. **C.** Total expression level of selected housekeeping genes for sample BRO23 which underwent both 3' and 5' GEX. **D.** Violin plots showing expression levels across clusters of *PABPC1* which is detected in both assays and *ALDOA* which is only detected in the 5' assay.

A.



B.



C.

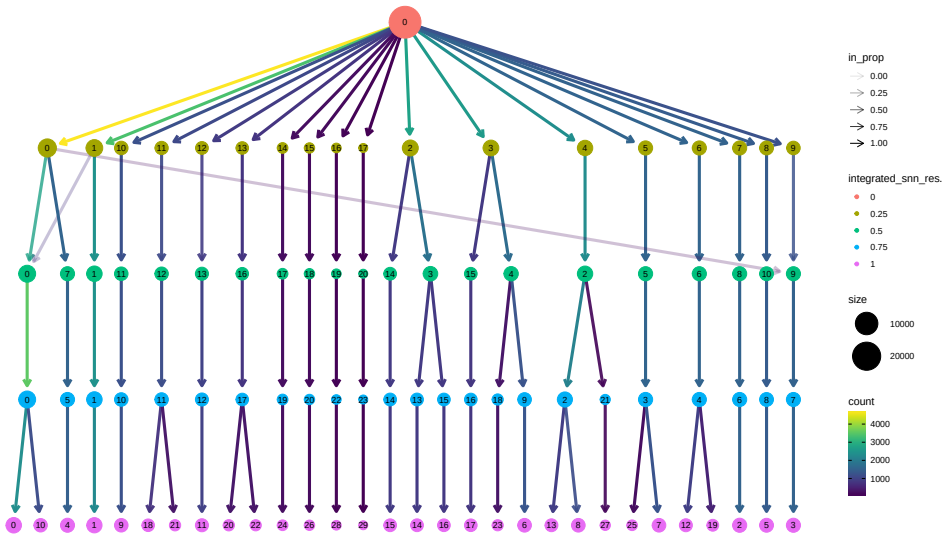


Figure 4.11: Clustering Analysis of Single-Cell RNA Sequencing Datasets. **A.** Clusters across scRNA-seq datasets. **B.** Clusters of integrated scRNA-seq data. **C.** Visualisation for interrogating resulting clusters at different resolutions (0.0, 0.25, 0.5, 0.75, 1.0).

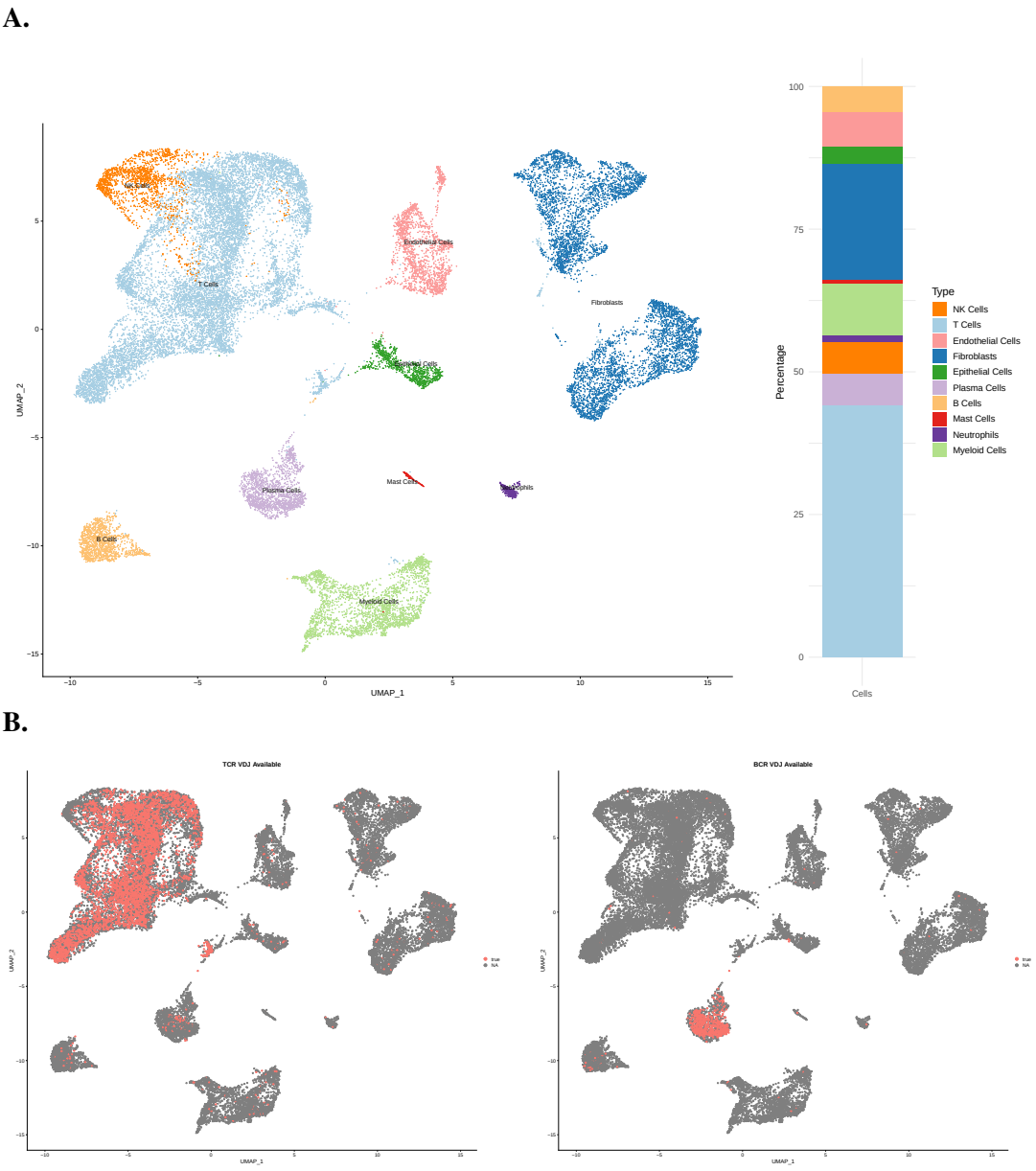
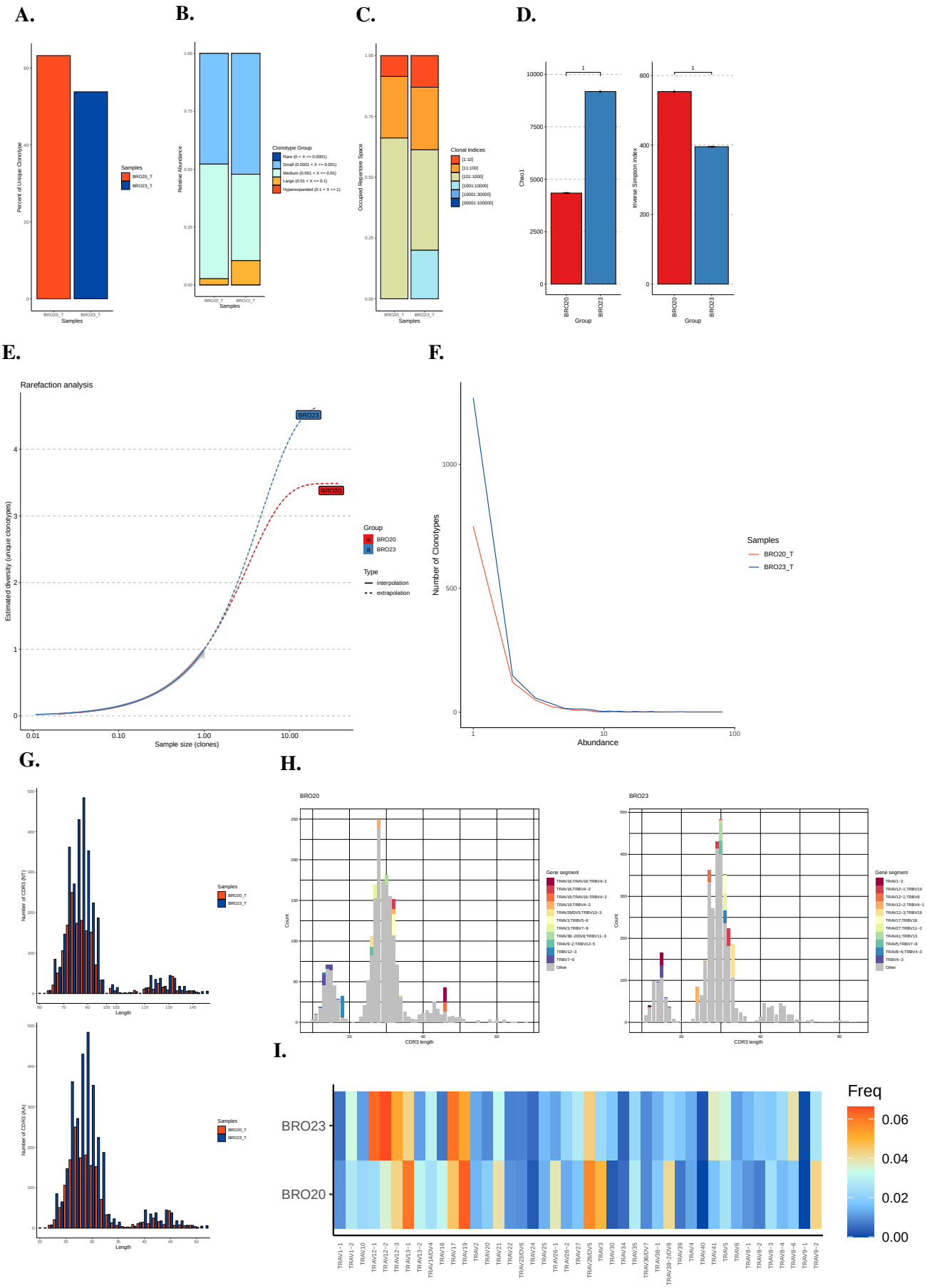


Figure 4.12: Marker-Based Labelling of Cell Clusters.
A. Major cell populations and proportions identified in the integrated datasets through clustering analysis and based on most highly variable genes. **B.** Projected 5' V(D)J data onto UMAP of integrated datasets.

4.4.2 Immune Repertoire

The total number of clonotypes recovered by combining the VDJC genes and the complementary determining region 3 (CDR3) nucleotide sequences were 4,563 (1,569 for BRO20; 2,994 for BRO23). Though sample BRO23 was found to have a slightly larger clonotype repertoire than BRO20, the relative percent of unique clonotypes scaled by the total size of each repertoire was greater for BRO20 (63.22%) than BRO23 (53.81%), as shown in Figure 4.13A. The space occupied by clones were binned into five groups based on their relative proportions: rare (0<X<=0.0001), small (0.0001<X<=0.001), medium (0.001<X<=0.01), large (0.01<X<= 0.1) and hyperexpanded (0.1<X<=1). Using these cutoffs, none of the samples were found to have rare nor hyperexpanded clones (Fig. 4.13B). Small and medium clonotype groups occupied 97.3% and 89.51% of BRO20 and BRO23, respectively, and large clonotypes 2.67% and 10.49%. Clones were also ranked and placed into bins based on their frequency of occurrence (Fig. 4.13C). Total richness and diversity is estimated to be higher for patient BRO23 based on a higher Chao1 score and lower inverse Simpson index (Figs. 4.13D–4.13E). The relative distribution of clonotypes by abundance is shown in 4.13F. Diversity was also assessed based on the CDR3 sequence length distribution (Fig. 4.13G). The frequency of observations was multimodal, with the most common lengths ranging from 81-90 nucleotides or 25-30 amino acids, and there was no significant difference between samples. The gene segment pairs for the most frequent CDR3 length in each sample include *TRAV19*:*TRBV4-2*, *TRAV41*:*TRBV13*, *TRAV5*:*TRBV7-8*, *TRAV12-2*:*TRBV4-1* and *TRAV12-3*:*TRBV19* (Fig. 4.13H). A preferential usage of *TRAV12-1/2/3* was identified in patient BRO23 and a bias towards *TRAV19* in BRO20 (Fig. 4.13I). Single cell level TRA and TRB clonotype analysis revealed that for the V gene, the *TRBV20-1* and *TRBV19* variants were most abundant across the samples, followed by *TRBV20-1* and *TRBV28* (Fig. 4.13J). In addition, *TRBJ2-1*, *TRBJ2-7* and *TRBJ-3* were the most frequent J segments (Fig. 4.13K). A gene usage correlation between the two samples of 0.46 and 0.80 was calculated for segments V and J, respectively. Likewise, the Jensen-Shannon (JS) divergence score was calculated to be 0.46 for V and 0.30 for J. The calculated Jaccard similarity coefficient (0.14 for V; 0.13 for J) suggested

very little overlap between the two samples, and no overlap in the CDR3 pairs. The top 10 most frequent clones in each sample are shown in Tables D.5 and D.6. All CDR3 sequences were compared to sequences in TCR databases, mainly McPAS-TCR and VDJdb. Since the databases predominantly contain pathogen-related antigens over cancer-related ones, the identified closely related CDR3s were biased towards EBV, CMV, Influenza and SARS-CoV-2 antigens. Non-viral CDR3 antigen target matches included TKT, MLANA, SEC24A, IGF2BP2, PMEL, BST2, PLA2G6, FNDC3B, MED13 and p53. None of these corresponded to the top clonotype CDR3 sequences. The amino acid sequence of the p53 epitope was returned and found to match the p53 tumour mutation-derived neoepitope, SCMGGMNWRPILTII, reported by Malekzadeh et al. (2019).


A.
B.
C.
D.
Percent of Unique Clonotype
Samples
BRO20_T
BRO23_T
Relative Abundance
Clonotype Group
Rare (0 < X <= 0.0001)
Small (0.0001 < X <= 0.001)
Medium (0.001 < X <= 0.01)
Large (0.01 < X <= 0.1)
Hyperexpanded (0.1 < X <= 1)
Occupied Repertoire Space
Clonal Indices
[1:10]
[11:100]
[101:1000]
[1001:10000]
[10001:30000]
[30001:100000]
Chao1
Inverse Simpson index
Group
E.
F.
Rarefaction analysis
Estimated diversity (unique clonotypes)
Sample size (clones)
Group
Type
interpolation
extrapolation
Number of Clonotypes
Abundance
G.
H.
I.
Number of CDR3 (NT)
Number of CDR3 (AA)
Length
BRO20
BRO23
Count
CDR3 length
Gene segment
Freq
0.06
0.04
0.02
0.00

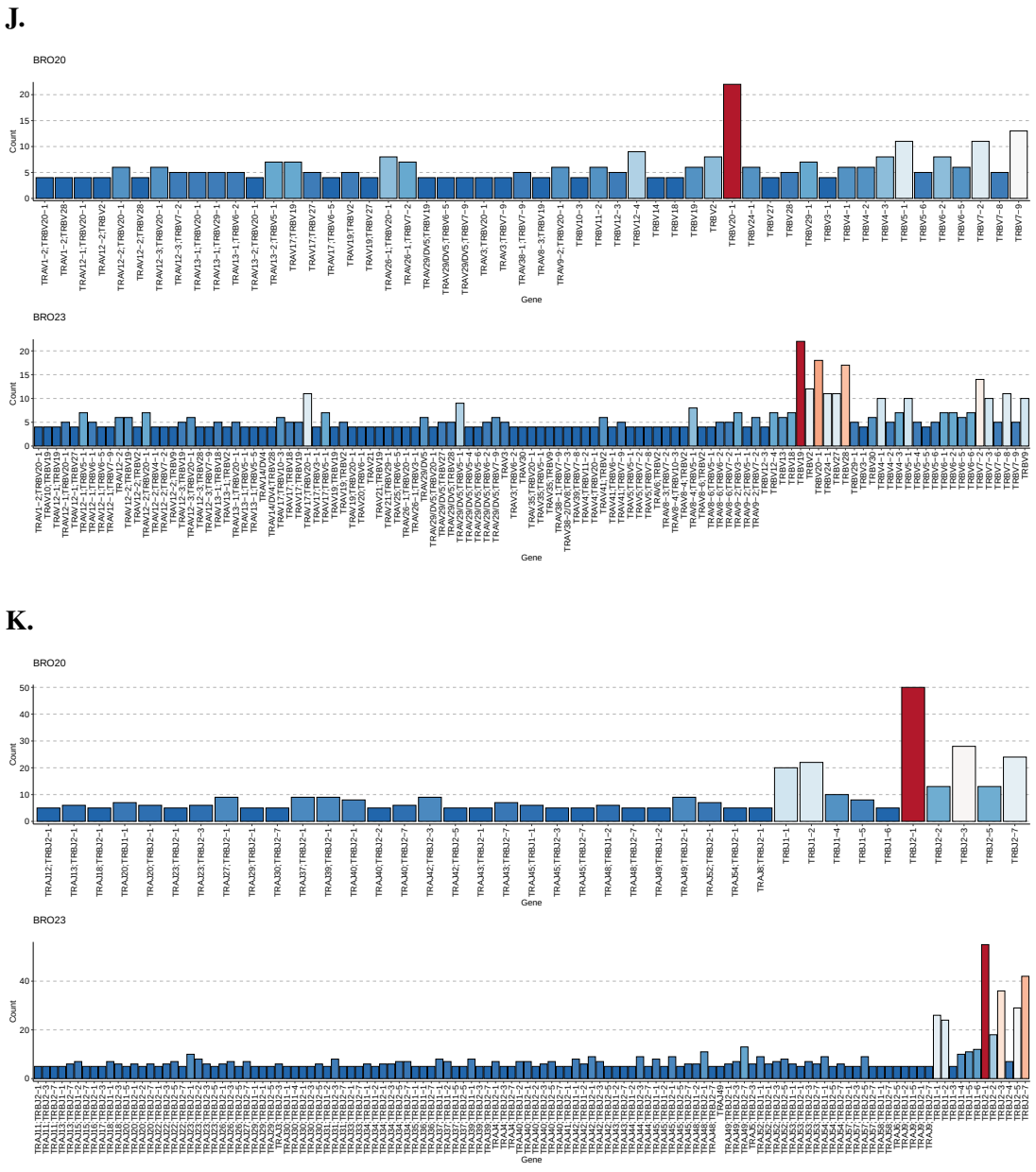


Figure 4.13: Overview of Patients TCR Repertoire.
A. Percent of unique clonotypes based on both TRA and TRB chains, and scaled by total repertoire size. **B.** Space occupied by clones at specific proportions. **C.** Clones ranked by total number and split by frequency of occurrence. **D.** Chao1 richness estimator and inverse Simpson index comparing repertoire diversity. **E.** Rarefaction curve showing number of unique clonotypes to sample size. **F.** Distribution of clonotypes by abundance. **G.** Length distribution of the CDR3 nucleotide and amino acid sequences. **H.** Distributions of clonotypes per CDR3 amino acid sequence lengths paired with V gene segments. **I.** Gene usage frequency for TRAV. **J.**, **K.** Histogram of gene usage for segments V (counts>3) and J (counts>4).

4.4.3 Transcriptional Landscape of T Cell Clusters

T cells were identified using marker-based and reference-based techniques (Figs. 4.14A–4.14D). Unsupervised clustering of 11,772 isolated T cells resulted in twelve clusters that were conserved across patients and methods (Fig. 4.14E). Pairwise differential gene expression and pathway enrichment analyses were performed to characterise each cluster (Figs. 4.15A–4.15B). The largest cluster was found to highly express *IL7R*, *KLRB1* and *CCR6* which is commonly characteristic of pro-inflammatory $CD4^+$ Th17 T cells (Fig. 4.16). Expression of *CCR7* could further suggest a central or effector memory state. Interestingly, *TPT1* was highly expressed, along with ribosomal protein genes. This finding was cross-checked with breast tissue T cell populations in the Human Protein Atlas (Karlsson et al., 2021). Clusters defined as cytotoxic T cells (CTLs) were enriched for cytotoxic markers (*GZMK*, *GZMB*, *PRF1*) and CD8. The third largest cluster expressed conventional $CD4^+$ T regulatory (Treg) markers (*FOXP3*, *TNFRSF18*, *TIGIT*) and high levels of *ICOS*. Other clusters included a group of T cells with an upregulation of type I IFN-inducible genes (*ISG15*, *OAS1*, *MX1*, *IFI6*). The smallest cluster expressed high levels of metallothionein (*MT*) 1 and *MT2*. These markers were later found to be associated with dysfunctional tumour infiltrating T cells (TILs), and are investigated in more detail in Section 4.4.5. In addition to reviewing top differentially expressed genes (DEGs), clusters were mapped to reference datasets of T cell subsets to further refine T cell populations. Top performers from Chapter 3 were used to predict the composition of each cluster (Figs. 4.17A–4.17B). The resulting heterogeneous compositions across all clusters suggest that cells may have clustered based on shared transcriptional signatures instead of lineage markers. For instance, a cluster that was called Cycling due to the high expression of proliferation markers (*TUBB*, *STMN1*, *TOP2A*, *Ki67*), was predicted using both methods to contain CD4 and CD8 lineage cells. The same was seen with clusters defined by high levels of the endogenous anti-inflammatory molecule, *ANXA1*, or high *IFITM3* and *TCF4*, genes regulated by the wnt/β-catenin signalling pathway. Likewise, reference mapping shows that the CD69+ cluster likely consists of both conventional and unconventional (MAIT, $\gamma\delta$, double negative) T cells.

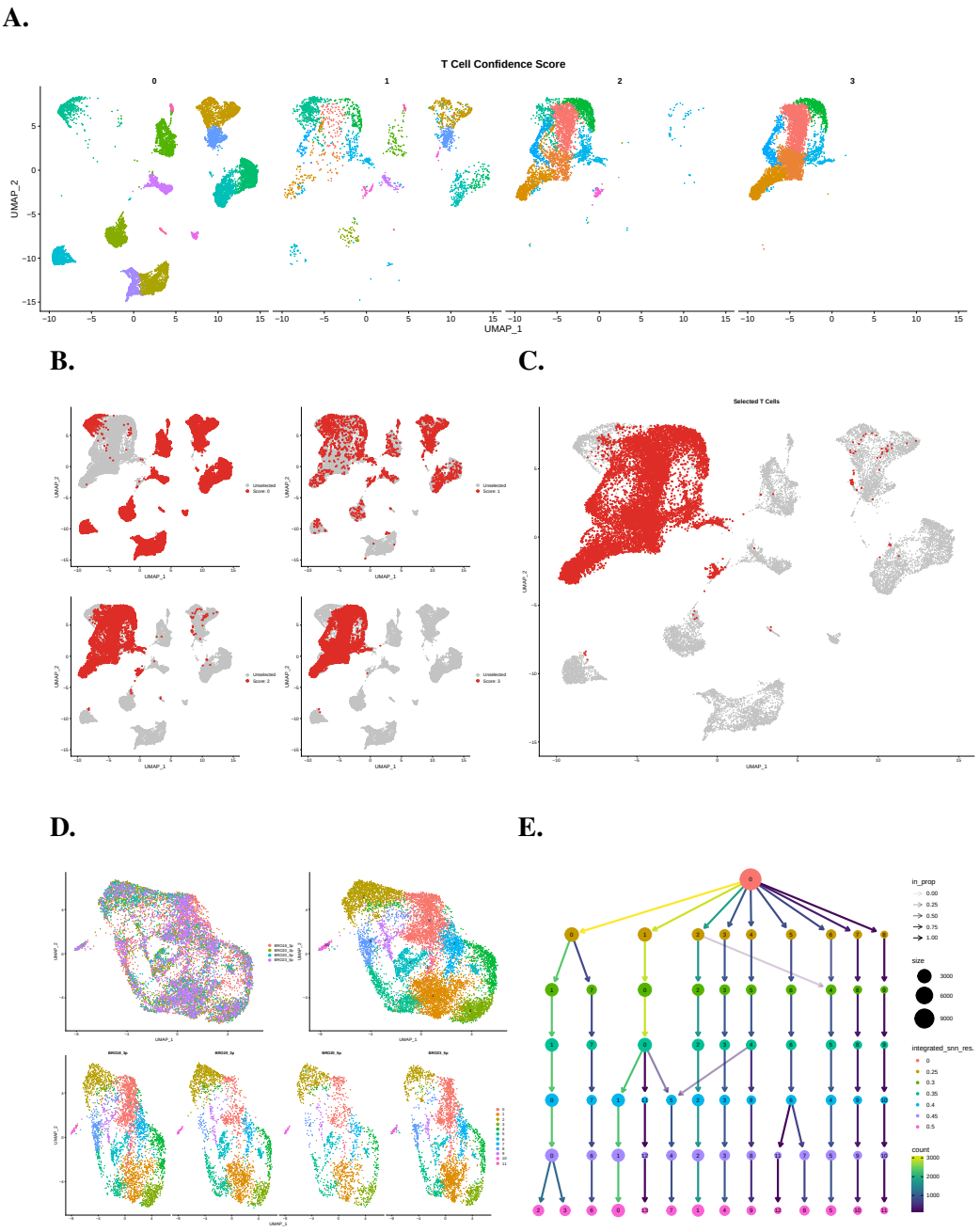


Figure 4.14: Isolating and Refining T Cells.
A. T cell confidence scores ranging from 0 to 3. **B.** UMAP plots showing cells in red with corresponding score. **C.** Cells selected for downstream analysis of T cells. **D.** Clustering analysis of isolated T cells. **E.** Visualisation for interrogating resulting clusters at different resolutions (0.0, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5).

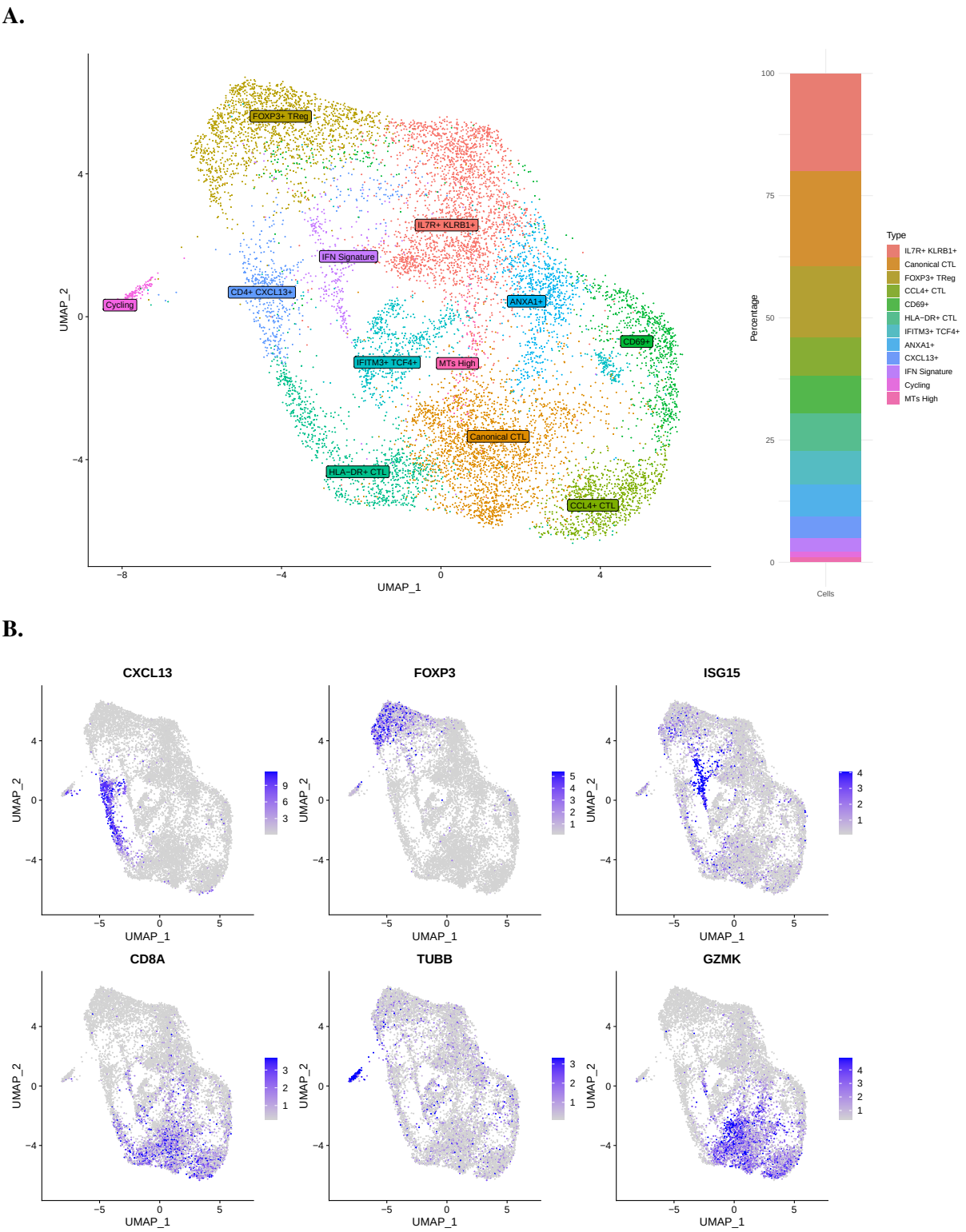


Figure 4.15: Transcriptional Profile of Isolated T Cells.
A. Isolated T cells were split into 12 clusters and labelled based on their characteristic transcriptional profile. The stacked bar graph shows the proportion of cells in each group. **B.** UMAP plots with colour gradient showing the level of expression for six highly variable genes (*CXCL13*, *FOXP3*, *ISG15*, *CD8A*, *TUBB*, *GZMK*).

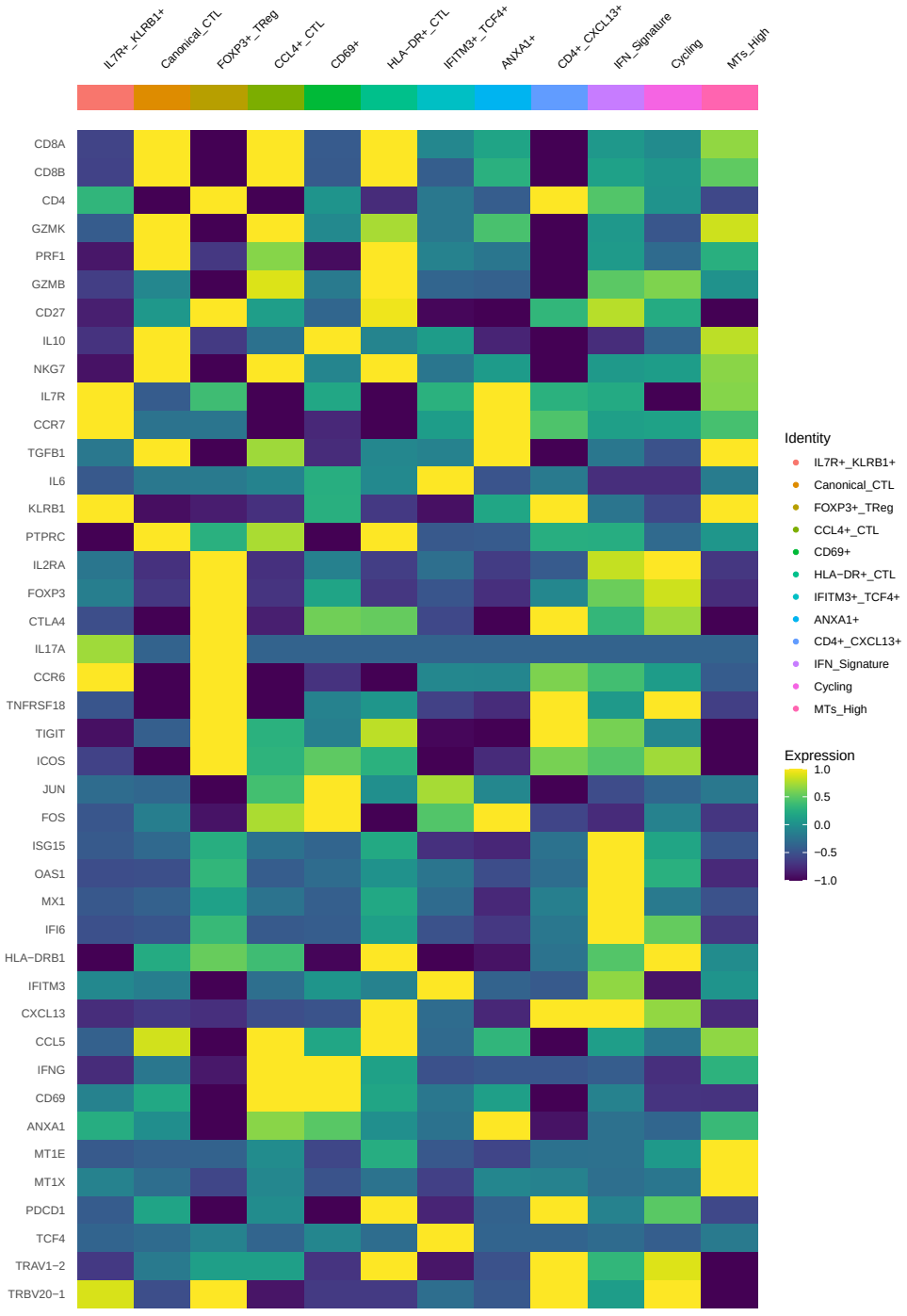
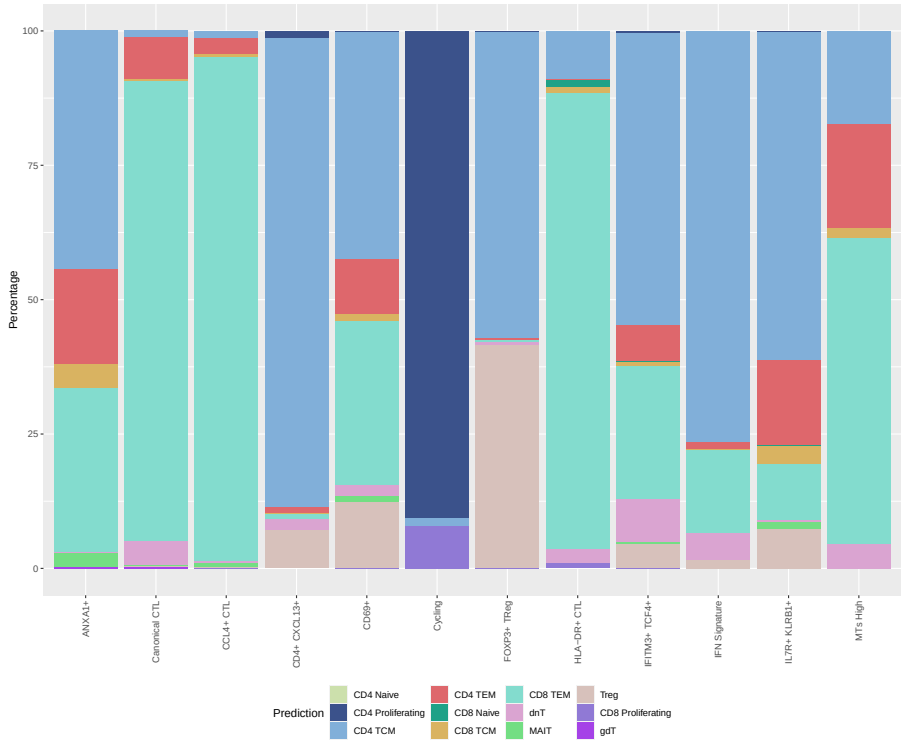


Figure 4.16: Heatmap of Gene Expression Across T Cell Clusters.
Heatmap showing averaged RNA expression values for selected genes across T cell groups.

A.



B.

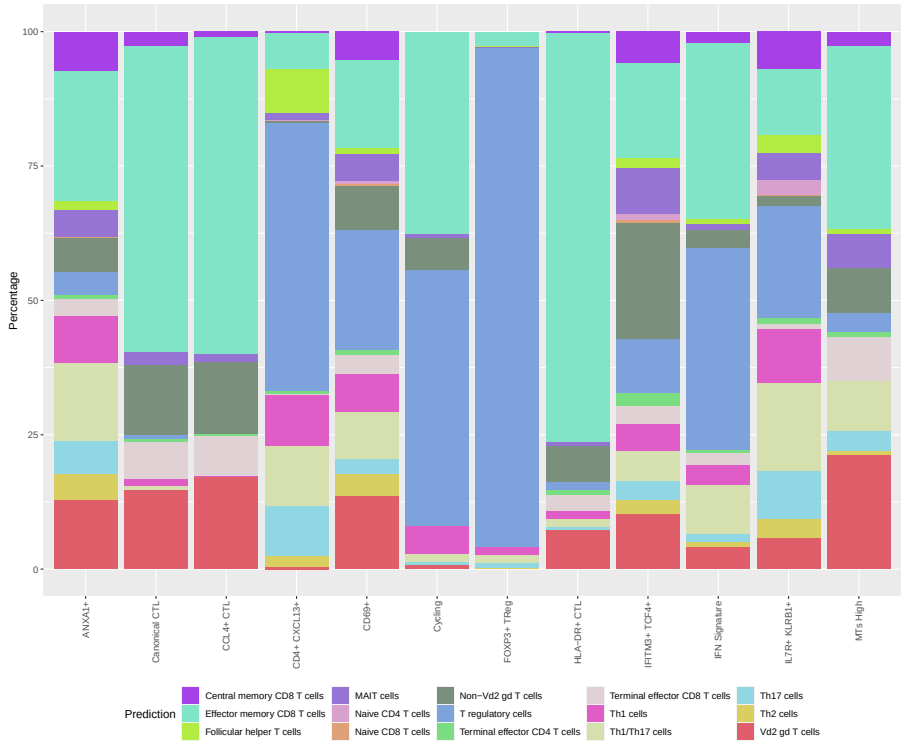


Figure 4.17: Reference-Based T Cell Predictions.
A. Proportion of Azimuth predicted T cell subtypes across transcriptionally-defined T cell clusters. **B.** Proportion of SingleR predicted T cell subtypes across transcriptionally-defined T cell clusters.

4.4.4 Paired Gene Expression and TCR Repertoire

Assembled TCR sequences available for two of the patients were projected onto the T cell space in order to examine the TCR repertoire in the context of the transcriptionally-defined T cells (Fig. 4.18A). The most expanded clones identified in Section 4.4.2 were found to be concentrated in the CTL zone, mainly in clusters Canonical CTL, HLA-DR+ CTL and CCL4+ CTL (Fig. 4.18B). Some of the top clones were also distributed across other clusters including ANXA1+, IL7R KLRB1+, IFN signature, MTs High, CD69+ and IFITM3 TCF4+ but at much lower amounts. None of the top expanded clones were found in clusters Cycling, FOXP3+ TReg and CD4+ CXCL13+ which contained no clones categorised as large (Fig. 4.19A). The top 10 clonotypes in each class (1:10) were the ones least occupying the repertoire space in each cluster (Fig. 4.19B). The repertoire overlap between clusters was evaluated using the Morisita index (Fig. 4.19D). The highest level of overlap was found between CCL4+ CTL and Canonical CTL. Clonotype richness was found to be highest in ANXA1+ and IFN Signature. Conversely, clonotype diversity which includes a measure of both the number and abundance of each clonotype was highest in MTs High and Cycling, and lowest in HLA-DR+ CTL and Canonical CTL (Fig. 4.19C).

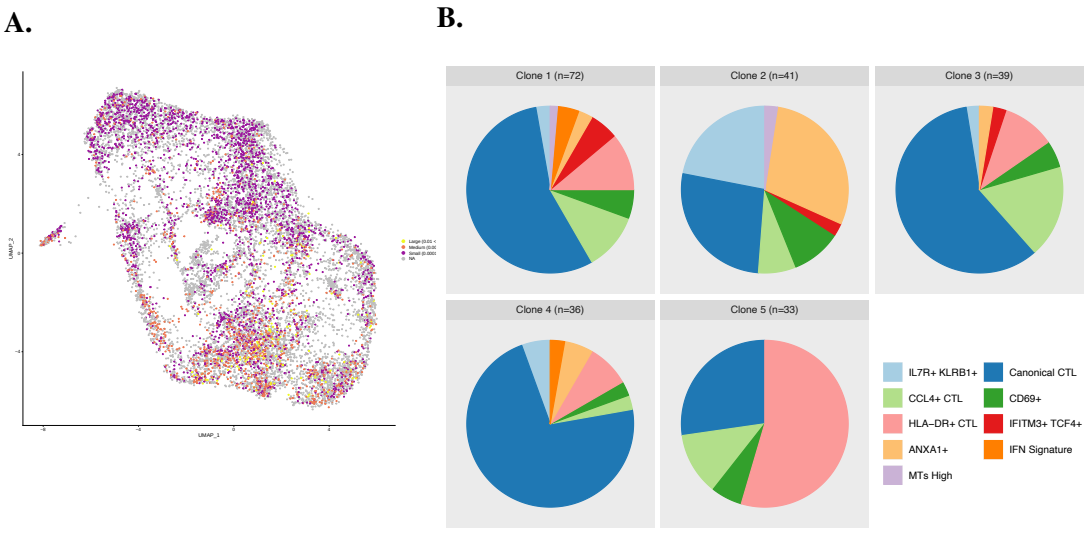


Figure 4.18: TCR Repertoire Projected Onto Transcriptionally-Defined T Cell Space and Trajectories.
A. UMAP plot with projected TCR clonotype membership. **B.** Pie charts depicting the proportion of the top five TCR clones in each isolated T cell cluster.

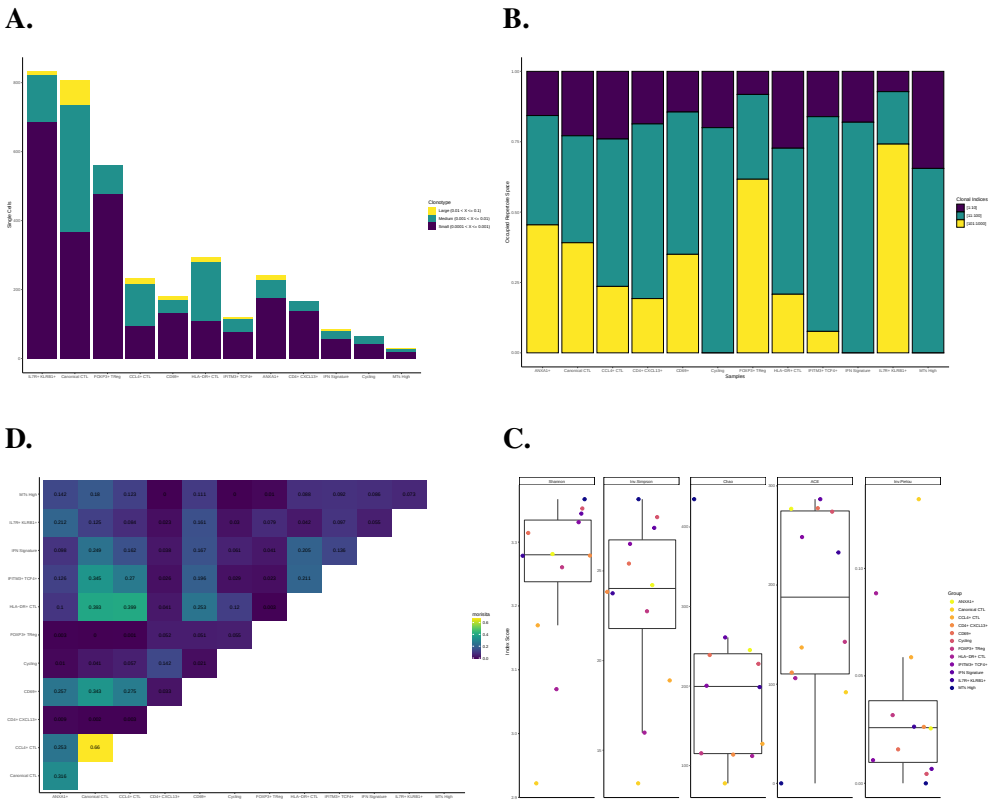


Figure 4.19: Clonotypes Across T Cell Groups.
A. Stacked bar chart showing the proportion of clonotype sizes across T cell groups. **B.** Stacked bar chart of clonal indices representing frequency of occurrence across T cell clusters. **C.** Plot showing the level of overlap for clonotypes across T cell clusters as calculated using the Morisita index of dispersion. **D.** Clonotype diversity measures for each T cell cluster using Shannon, inverse Simpson, Chao, ACE and inverse Pielou.

4.4.5 Assessing T Cell Activation and Dysfunction

The suppressive TME drives T cells toward dysfunctional phenotypes, hindering effective anti-tumour immunity and immunotherapy. In order to estimate the level of T cell activation and dysfunction across T cell clusters and at a single-cell level, gene signatures published by Singer et al. (2016) were used for gene set enrichment analysis. A classical PD-1 signature of dysfunction from the MSigDB published by Duraiswamy et al. (2011) was used for comparison (GSE26495). The MTs High cluster was found to have the highest median score for both the dysfunctional module and PD-1 signature whilst the Cycling cluster had the highest median for the activation score (Fig. 4.20A). Nevertheless, clus-

ters could not be segregated into groups based on their activation or dysfunction status. Interestingly, metallothioneins genes which are characteristic of the MTs High cluster, are not part of the dysfunction module but in Duraiswamy et al. (2011) expression levels were found to positively correlate with T cell acquisition of a dysfunctional phenotype. Upon examining the scores at the single cell level, it was found that clusters are composed of a mix of cells with a range of dysfunction and activation scores. There was a greater overlap between the PD-1 signature and the activation module, making it more difficult to discern between these potential states. Duraiswamy et al. (2011) found the activation and dysfunction module scores to be negatively correlated across cells in their dataset. Indeed, cells that have a high activation score tend to have a low dysfunction score, and vice versa (Figs. 4.20B–4.20C). However, only a very weak negative correlation was calculated across all T cells (r=-0.06). In order to further investigate the relationship between activation and dysfunction scores, the scores were binned by quantile (Top 5%, 5-25%, 25-50%, 50-75%, 75-100%). A strong negative correlation between the activation and dysfunction modules was confirmed for the top 5% of each score (r=-0.71) which fell to moderate (r=-0.45) when considering the top 25%, down to -0.24 for the top 50% and -0.10 for the top 75% scores. This observation suggests that T cells find themselves along a transcriptional gradient and that these gene signatures may be uncoupled only at extreme levels. Differentially expressed genes were calculated for the top 5% dysfunction scores. As expected, the most enriched genes included those in the corresponding module used to calculate the scores. However, additional genes not part of the original signature were also found significantly enriched in this population including: *KLF2*, *NEAT1*, *IFITM1*, *PTGES3*, *MIF*, *PRF1*, *CD83*, *TNFAIP3*, *MT-ATP8* and *DGKZ*.

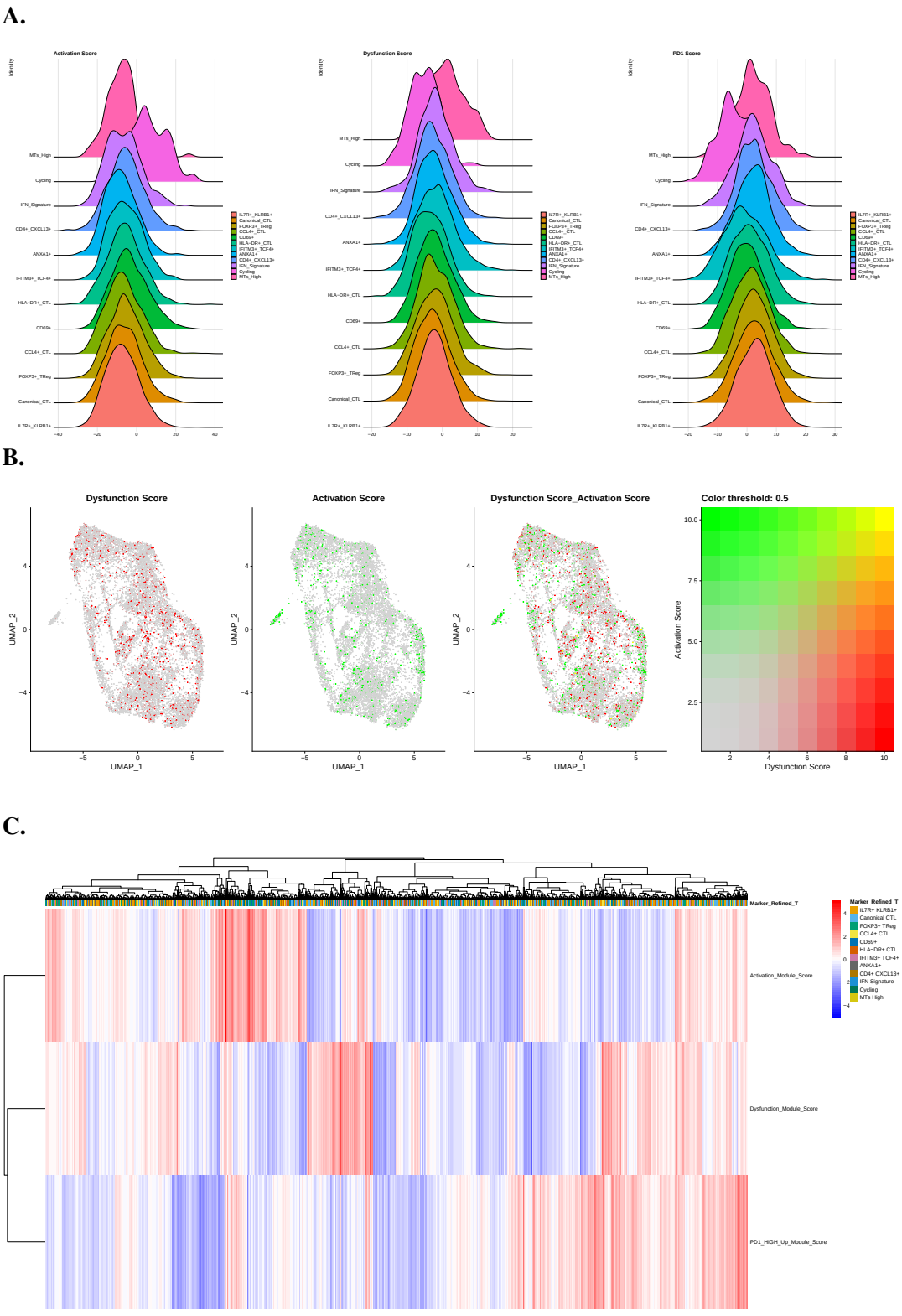


Figure 4.20: T Cell Activation and Dysfunctional State Scores.
A. Ridgeplot of calculated dysfunction, activation and PD-1 scores across T cell clusters. **B.** Dysfunction and activation scores projected onto UMAP plot of T cells **C.** Heatmap of scores across cells for activation, dysfunction and PD-1 signatures.

4.4.6 Examining IFN Signalling Pathways in T Cells

The expression level of genes related to type I IFN signalling pathways were examined, both pathways of IFN production and signatures of IFN response. First, the cluster labelled as IFN Signature was explored in more details. As shown in Section 4.4.3, top differentially expressed features for this T cell group consist of IFN-stimulated genes (Figs. 4.21A–4.21B). Naturally, there is a clear relationship between the top enriched genes in this group, since binding of type I IFN with its receptor induces STAT1 through the JAK-STAT signalling pathway, thereby positively regulating expression of *RSAD2* (viperin), IFIT family genes, and other IFN target genes like *MX1/2* and *ISG15*. Similarly, *LY6E* is an IFN-regulated gene known to modulate T cell activity, though the underlying mechanism has not yet been elucidated. This type I IFN signature was further confirmed through gene set enrichment analysis which showed a strong enrichment for anti-viral immune responses (Figs. 4.22A–4.22D). Several factors in the TME may be recognised by pattern recognition receptors to activate IFN responses similar to viral molecular patterns.

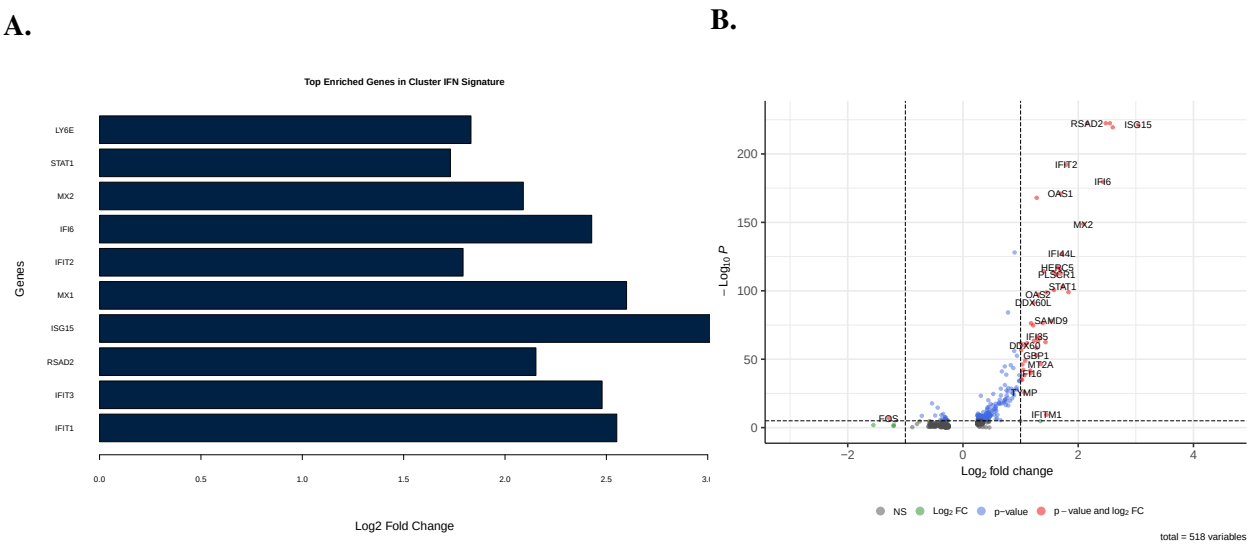


Figure 4.21: IFN-Induced Signature Characteristic of T Cell Cluster.
A. Log 2 fold change of top enriched genes in T cell cluster defined by a type I IFN-induced signature. **B.** Volcano plots showing statistical significance (p-value) versus magnitude of change (log base 2 fold change) of genes differentially expressed in the IFN-induced T cells relative to other T cell populations. Outlined points represent top differentially expressed genes that pass a log 2 fold change cut-off of 1 and a p-value threshold of 10E-14.

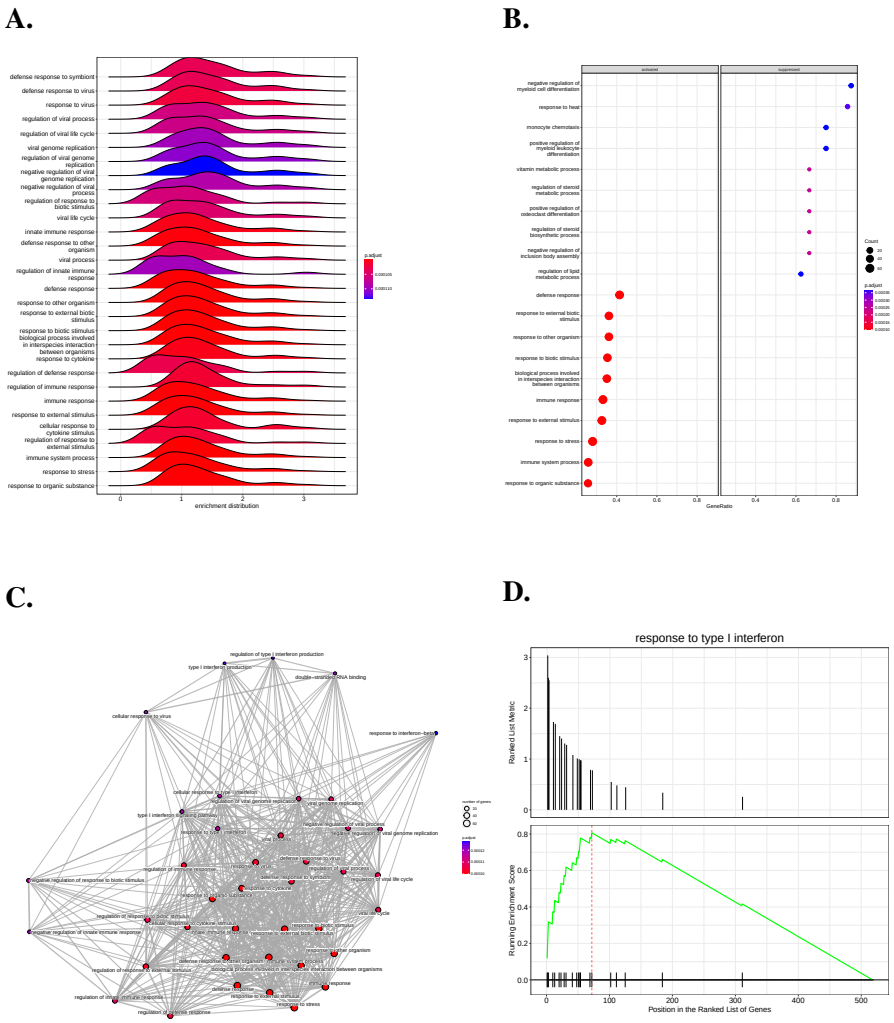
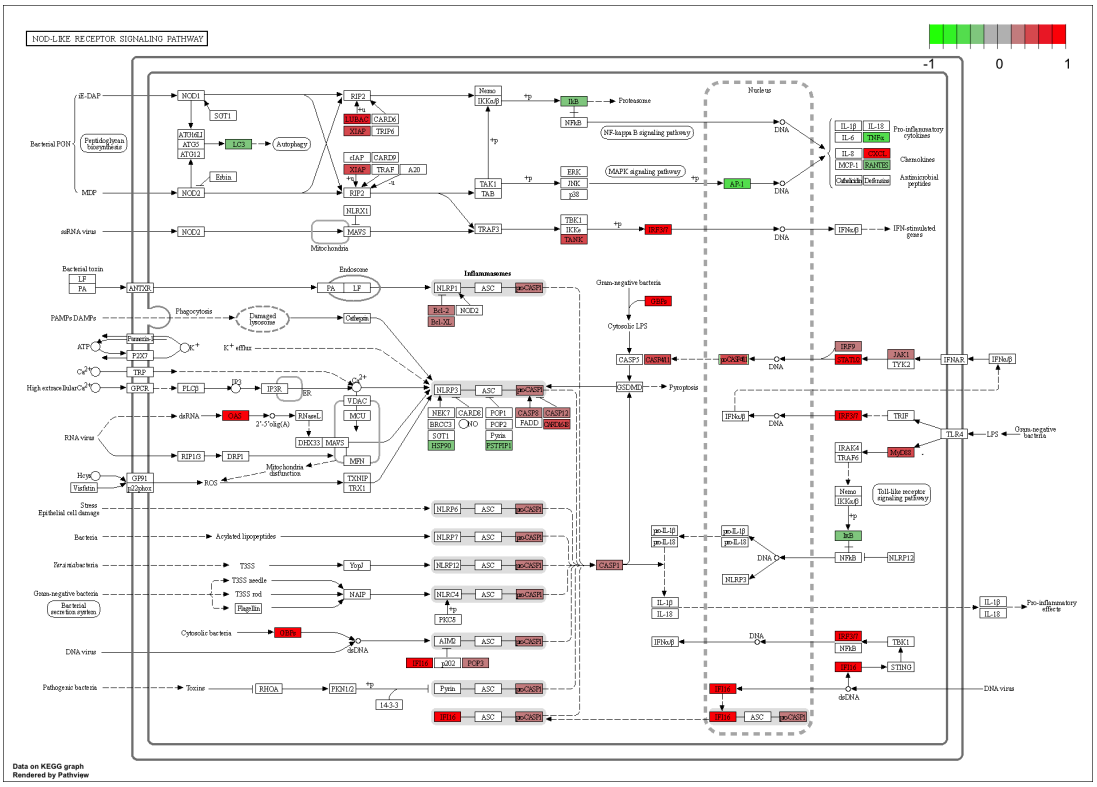


Figure 4.22: Heightened T Cell Anti-Viral Immune Response.
A. Ridge plot showing enrichment distribution of GSEA categories for the IFN Signature T cell cluster. **B.** Dot plot of activated and suppressed GSEA terms with gene ratio and enrichment scores. **C.** Network of enriched terms with edges connecting overlapping gene sets. **D.** Enrichment score (ES) reflecting the degree to which the response to type I IFN is over-represented in a ranked list of genes calculated using a running sum statistic starting at the top of the list. The ES is represented by the red dotted line and consists of the maximum deviation from zero encountered when walking down the list.

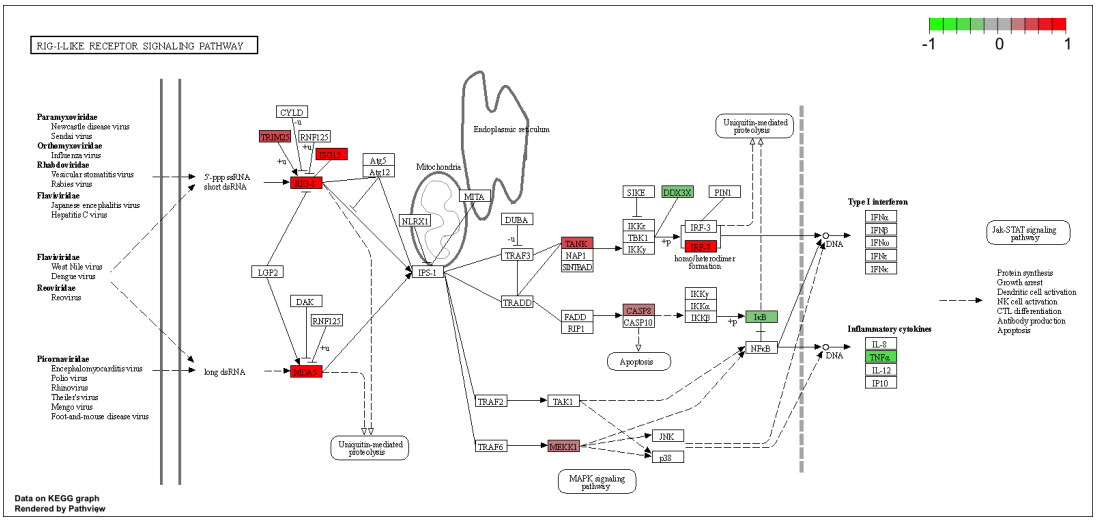
Type I IFNs can exert autocrine and paracrine signalling, as such the calculated fold changes were mapped onto four pathways retrieved from the KEGG database, mainly cytosolic DNA-sensing (hsa04623), RIG-I-like receptor signalling (hsa04622), Toll-like receptor signalling (hsa04620) and NOD-like receptor signalling (hsa04621) to see whether these cells may be contributing to their own IFN signature (Figs. 4.23A–4.23D). Indeed, the transcription factor, *IRF7*, which directly regulates the transcription of IFNα and IFNβ, was found to be upregulated. Upstream components of the RIG-I pathway, *ISG15*, *TRIM25*, *TANK*, *MDA5* and *RIG-I* were also found to be highly expressed. Multiple members of the caspase family involved in activation of inflammatory responses, mainly *CASP8*, *CASP12*, *CARD16-18*, *CASP4/11* and *pro-CASP1* were differentially expressed in this T cell group and could theoretically risk driving these T cells to pyroptosis. Additional nucleic acid sensors like DNA sensing *DAI* and the RNA sensing TLR-adaptor *MyD88* were also found to be relatively more expressed whilst downregulated gene expression values were reported for the pro-inflammatory molecules *TNF*α, *CCL4* and *CCL5* (RANTES).

Next, expression levels of IFN-related genes were compared across T cells clusters. Specifically, the IFNα response signature from Liberzon et al. (2015) was used to measure type I IFN exposure. Unsurprisingly, the IFN Signature cluster had the highest enrichment score for the signature followed by Cycling, HLA-DR+ CTL and CD4 CXCL13+. To infer the activation status of the cGAS-STING, RIG-I, TLR and NLR pathways, gene sets corresponding to each pathway collected in Section 4.3.1 were used to measure enrichment (Figs. 4.24A–4.24C). The Cycling cluster expressed the highest level of cGAS-STING and NLR related genes whilst CCL4+ CTL showed the highest median for the TLR signalling pathway. The RIG-I pathway was highly expressed in the IFN Signature cluster. Since chronic IFN stimulation has been reported to induce T cell exhaustion, correlations between conventional exhaustion markers and IFN-related genes were examined across all T cells (Fig. 4.25). There was a strong correlation between expression of the long non-coding RNA, *MALAT1*, and the negative regulator of type I IFN production, *DDX5*. A slight positive correlation was identified between the following three transcripts: *TMEM173*, *BATF* and *S100A4*. They were negatively correlated with *NKG7* and *CCL5*.

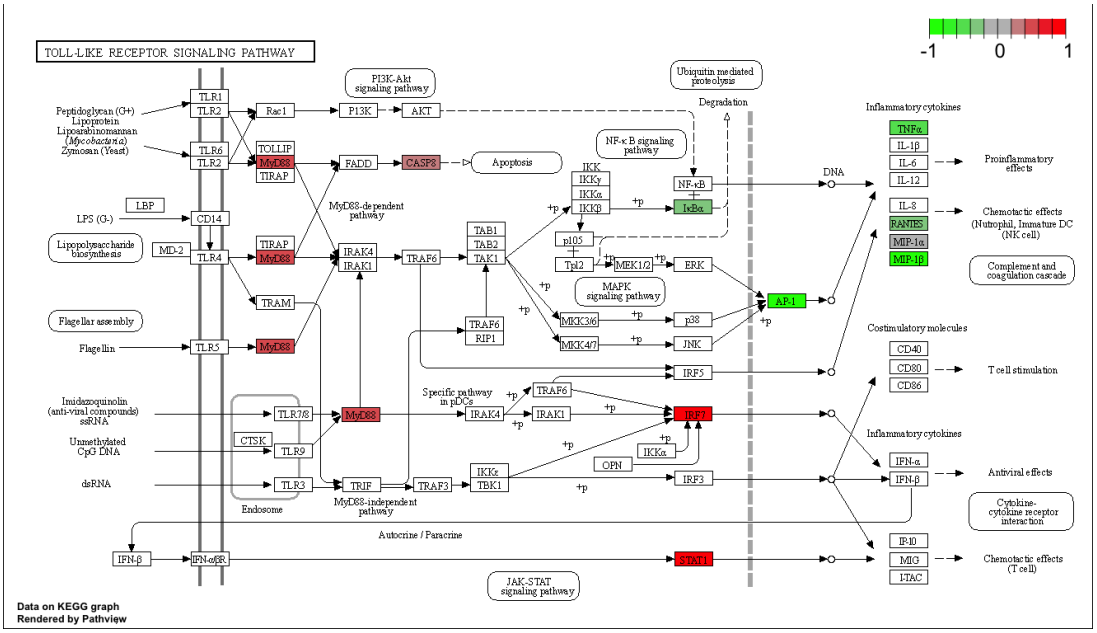
A.


NOD-LIKE RECEPTOR SIGNALING PATHWAY
-1
0
1
Data on KEGG graph
Rendered by Pathview

B.


RIG-I-LIKE RECEPTOR SIGNALING PATHWAY
-1
0
1
Paramyxoviridae
Newcastle disease virus
Sendai virus
Orthomyxoviridae
Influenza virus
Rhabdoviridae
Vesicular stomatitis virus
Rabies virus
Flaviviridae
Japanese encephalitis virus
Hepatitis C virus
Flaviviridae
West Nile virus
Dengue virus
Reoviridae
Reovirus
Picornaviridae
Encephalomyocarditis virus
Polio virus
Rhinovirus
Theiler's virus
Mengo virus
Foot-and-mouse disease virus
Type I interferon
Inflammatory cytokines
Jak-STAT signaling pathway
Protein synthesis
Growth arrest
Dendritic cell activation
NK cell activation
CTL differentiation
Antibody production
Apoptosis
MAPK signaling pathway
Data on KEGG graph
Rendered by Pathview

C.



D.

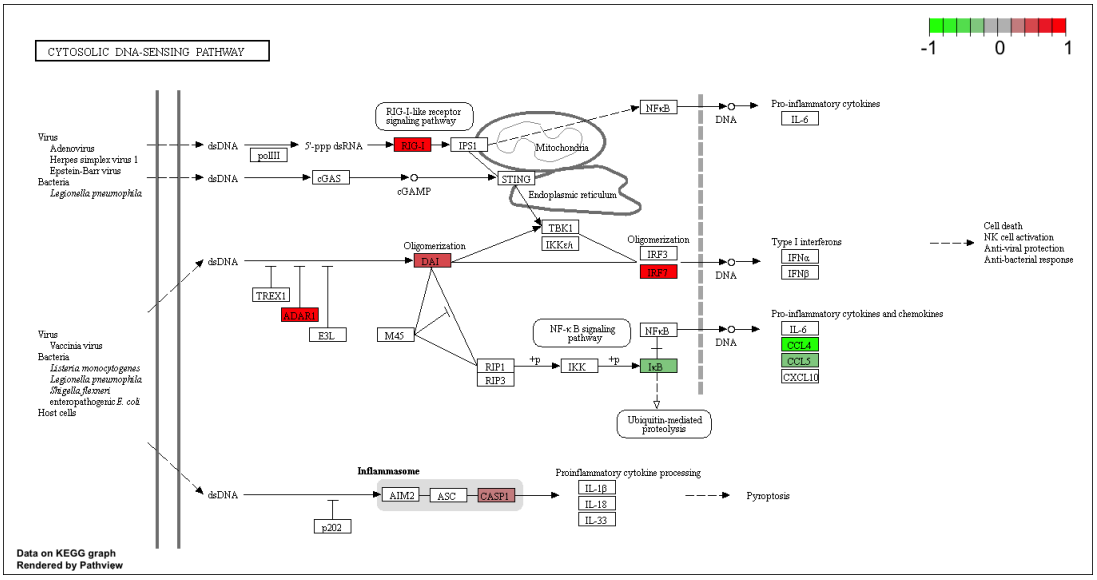


Figure 4.23: Mapping to IFN Signalling Pathways.
Expression fold changes of the IFN Signature T cell population overlayed onto the following KEGG pathway maps: **A.** RIG-I-like receptor signalling pathway (hsa04622). **B.** NOD-like receptor signalling pathway (hsa04621). **C.** Cytosolic DNA-sensing pathway (hsa04623). **D.** Toll-like receptor signalling pathway (hsa04620).

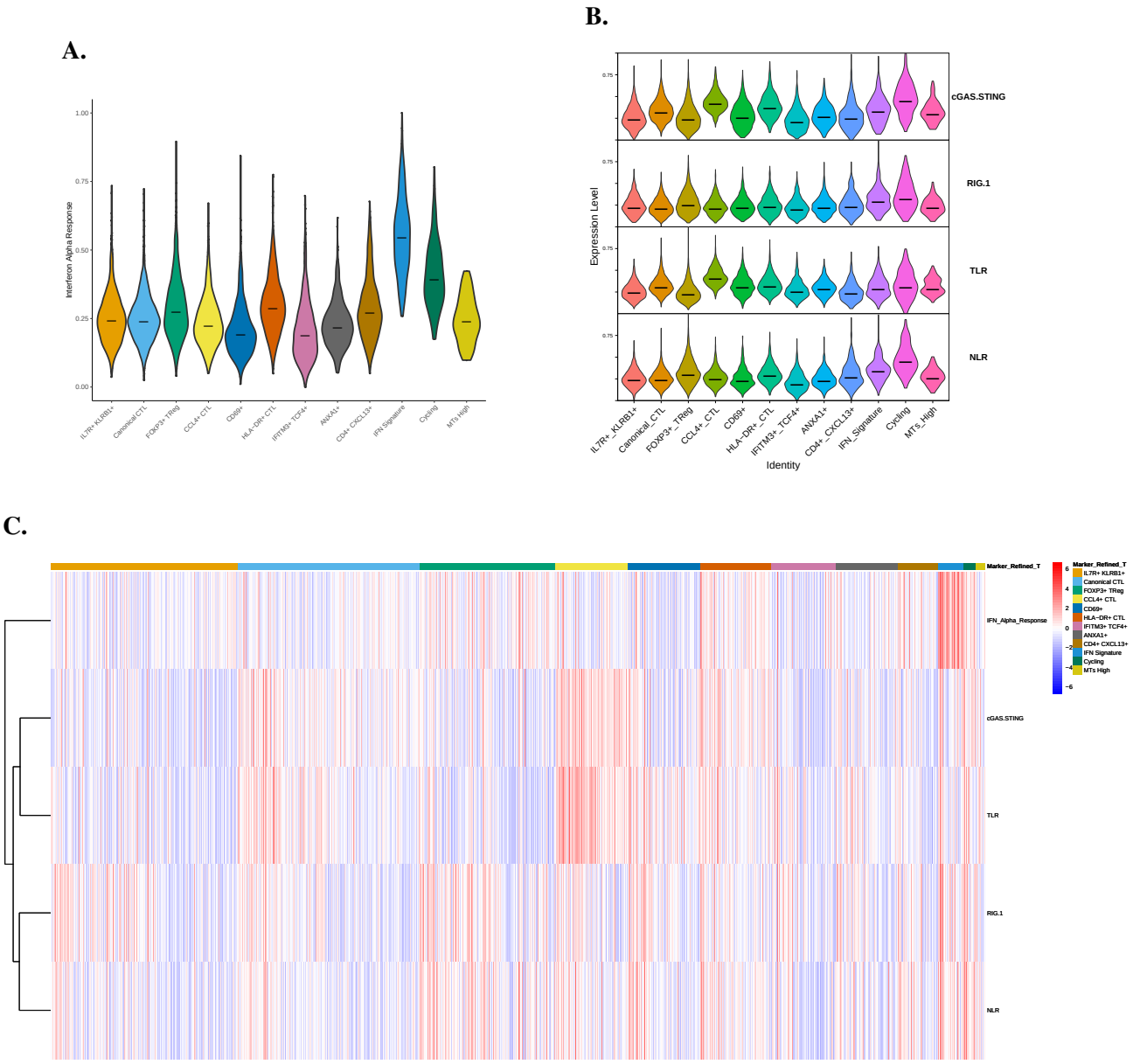


Figure 4.24: Interferon Signalling Scores Across T Cell Clusters.
A. Interferon alpha response enrichment score across T cell clusters calculated using single-sample GSEA and normalised to the range of the individual gene set. **B.** Enrichment scores for IFN-signalling pathways driven by cGAS-STING, RIG-I, TLR and NLR across T cell clusters. **C.** Heatmap grouped by T cell cluster showing enrichment scores across individual cells for the IFN alpha response, cGAS-STING, TLR, RIG-I and NLR pathways.

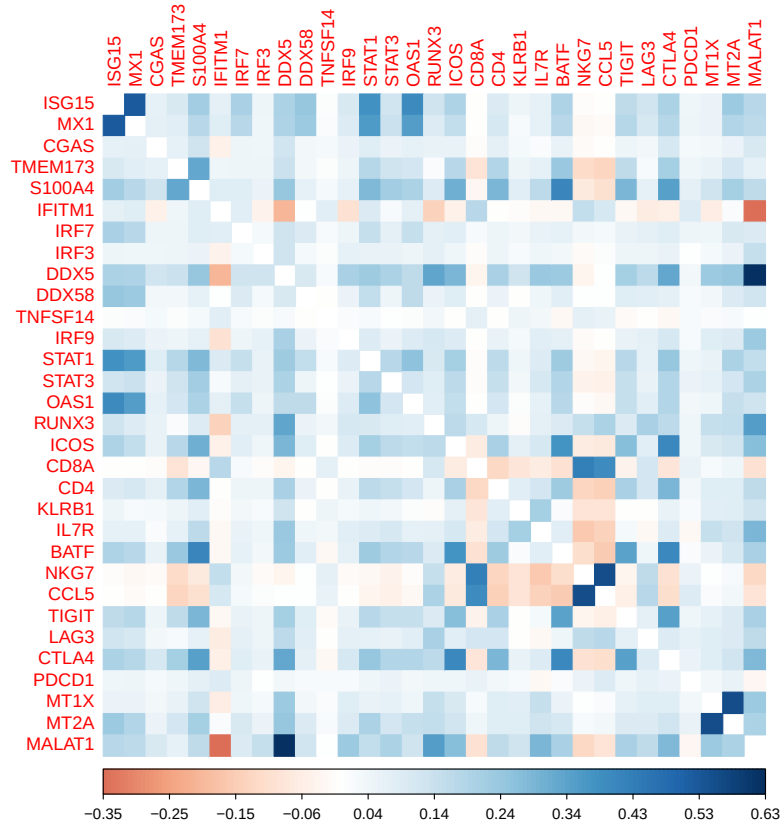
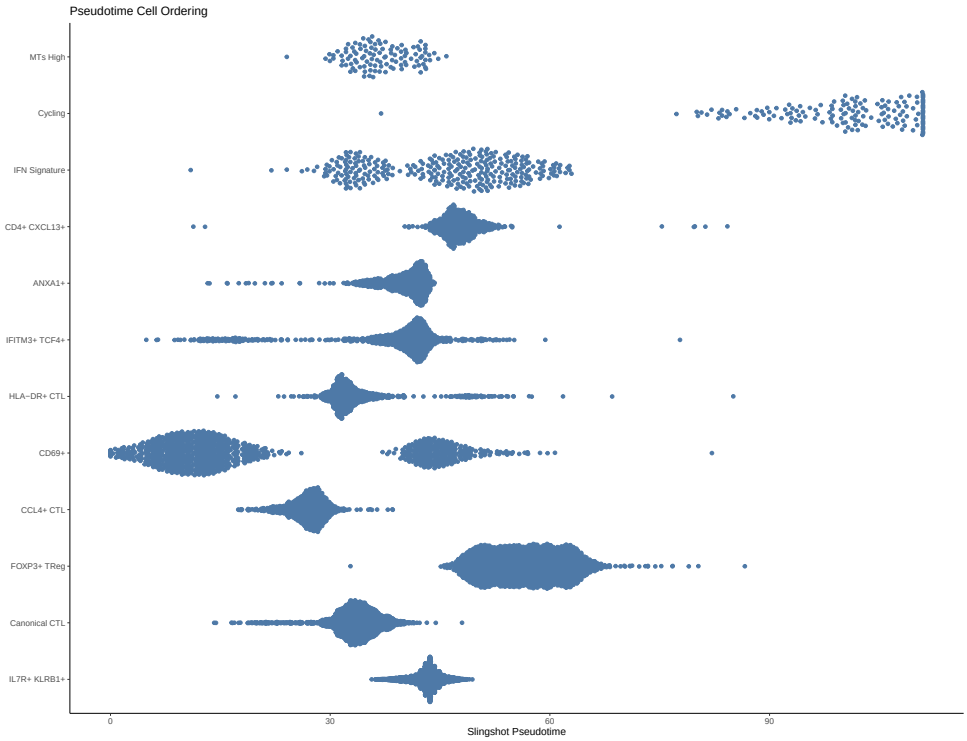


Figure 4.25: Correlation Matrix of IFN Genes and T Cell Exhaustion Markers. Pearson's correlation coefficient calculated across all isolated T cells for a selected set of genes related to IFN signalling and conventionally used to define T cell exhaustion.

4.4.7 Pseudotemporal Ordering of T Cells

T cells from all clusters were ordered in pseudotime to gain insight on their relative transcriptional progression. Single cell profiles can be ordered based on transcriptional similarity along a linear path of continuously placed cells. Generally, this path consists of a reconstructed biological developmental trajectory, as shown in Section 3.4.3. However, the method was used in this section independent of T cell lineage in order to explore any potential continuity in gene patterns that could indicate state transitions – in other words, a trajectory of states. Moreover, as described in Section 4.4.3, each cluster is predicted to contain a mixture of CD4 and CD8 T cell subtypes. To this end, six distinct trajectories were estimated (Fig. D.1). The primary predicted lineage placed CD69+ cells at the origin and Cycling cells at the end (Fig. 4.26A). Figure 4.26B shows the distribution of cell arrangement in each cluster. CD69+ is a marker of early T cell activation and the Cycling cluster consists of highly proliferative cells predicted in Section 4.4.3 to contain memory cells. State scores calculated in Section 4.4.5 were mapped onto the pseudotime-ordered T cells but did not show a logical transition between states along any of the predicted paths (e.g. activated to dysfunctional). As shown in Section 4.4.5, all T cell clusters contain a mix composition of states with most T cells having an intermediate enrichment score for both the activation and dysfunction modules. Expression levels along pseudotime of the most significant time-dependent genes are shown in Figure D.2. Production of both *CCL4* and *CCL5* was found to decrease over pseudotime whilst *IL2RA* and *IL32* gradually increased. Cytotoxic markers including *GZMK* and *NKG7* were highest at earlier time points where CTL clusters were placed and could be indicative of an activated state. At the end of the pseudotime were *S100A4* and tubulin-encoding gene *TUBB* in cluster Cycling which have been identified as markers of tissue resident T effector memory cells by Szabo et al. (2019).

A.



B.

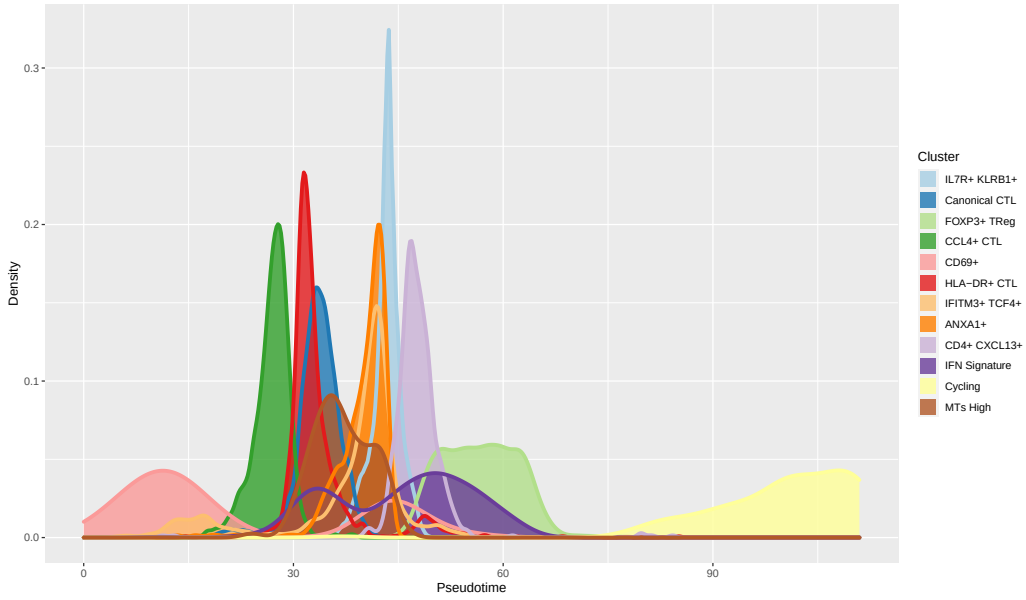


Figure 4.26: Pseudotemporal Ordering of T Cells
A. Diffusion map of T cells across clusters ordered in pseudotime. **B.** Density estimates across pseudotime showing the tendency of different T cell populations.

4.4.8 Mapping T Cell Interactions

Cellular crosstalk mediated by ligand-receptor pairs was assessed between transcriptionally defined T cells and other cell types (Fig. D.3). Across all cell types the most frequent interactions were found to happen between the following five clusters: Neutrophils, Cycling, Fibroblasts, Myeloid Cells and Endothelial Cells (Fig. 4.27). For all T cell populations as well, the top three interactions based on counts were predicted to be with the Myeloid Cells, Endothelial Cells and Fibroblasts clusters (Fig. 4.28). Looking further into these interactions, it was found that the highest interaction prediction scores were attributed to the macrophage migration inhibitory factor (*MIF*) and *CD74* gene pair between T cell populations and myeloid cells. Both endothelial cells and fibroblasts are found to predominantly interact with HLA-DR CTL, IFITM3 TCF4 and CCL4 CTL through COL4A1 and ITGB1 (CD29). There is a strong indication that secretion of CCL4 from the CCL4 CTL group interacts with SLC7A1 on endothelial cells and CCR8 on FOXP3 T regulatory cells.

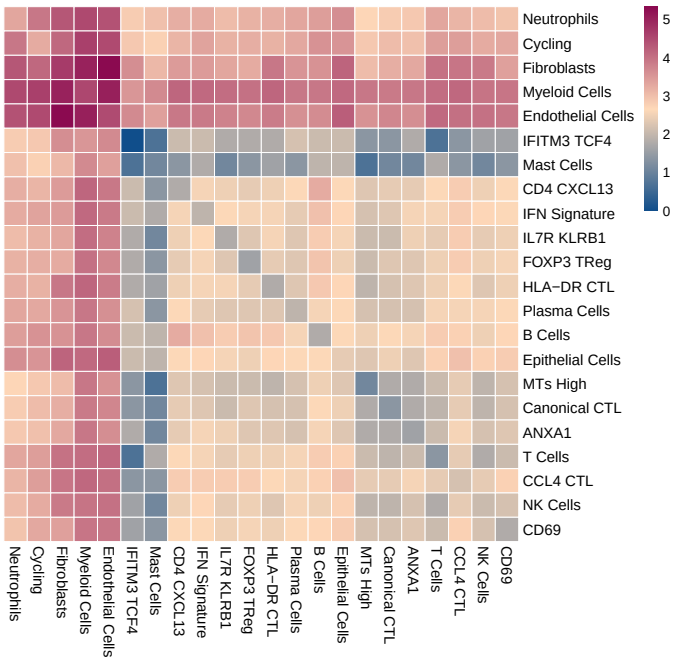


Figure 4.27: Heatmap of Interaction Log Counts.
Heatmap generated by CellPhoneDB showing the log total number of interactions between cell types.

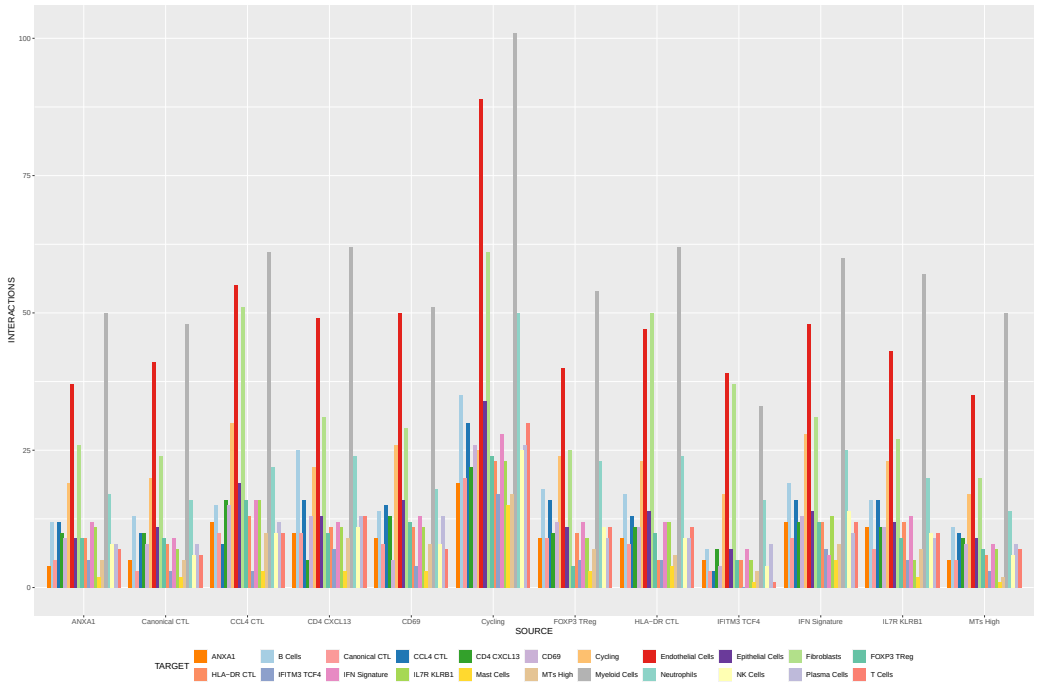


Figure 4.28: Predicted Interaction Counts for T Cell Populations.
Barplot showing predicted total count of molecular interactions between T cells (SOURCE) and all cell populations (TARGET).

Certain interactions were predicted to promote an anti-inflammatory state. Myeloid cells for instance were found to express high levels of *IL-10*, and were responsible for driving significant functional pathways negatively regulating inflammatory responses (Figs. 4.29A–4.29B). For instance, epithelial cells expressing *SAA1* were predicted to interact with FPR2 on myeloid cells. Likewise, myeloid cells were also involved in promoting the PI3K-AKT signalling pathway, along with cells from the fibroblast, endothelial and epithelial cell clusters from gene pairs that include *COL1A1-a1b1* complex, *FGF2-FGFR1* and *VEGFB*/*VEGFA-FLT1* (Fig. 4.29C). Conversely, co-stimulation of T cells is predicted to occur from CD28 binding CD86 or CD80 (B7) on either B cells or myeloid cells (Fig. 4.30A). B cells, CCL4 CTL and CD69 T cells were found to positively regulate the inflammatory response through secretion of the inflammatory cytokine TNF binding to RIPK1 or TNFRSF1A (Fig. 4.30B).



Figure 4.29: Pathways Driving Anti-Inflammatory Effects
Significant functional pathways related to anti-inflammatory effects that were identified from enriched interacting gene pairs between clusters in the inner circle and those in the outer circle. The width represents the relative fraction of weighted interactions. **A.** Negative regulation of the inflammatory response by epithelial cells secreting SAA1 that interacts with FPR2 on myeloid cells. **B.** Anti-inflammatory signalling induced by IL-10 secreted from myeloid cells. **C.** Promotion of the PI3K/Akt pathway mainly by fibroblasts, endothelial cells, myeloid cell and epithelial cells interacting primarily through the COL4A family binding to ITGA1.



Figure 4.30: Pathways Driving Immune Responses
Significant functional pathways related to the immune response that were identified from enriched interacting gene pairs between clusters in the inner circle and those in the outer circle. The width represents the relative fraction of weighted interactions. **A.** Interactions promoting T cell activation. **B.** Interactions promoting the positive regulation of inflammatory responses.

Directed outgoing and incoming putative interactions were analysed in more details for each of the 12 isolated T cell profiles. All T cell populations interact with a variety of cell types. As mentioned above, the highest scored outgoing interaction for all T cells is associated with *MIF* expression pairing with *CD74* on myeloid cells. The canonical CTL cluster receives input from B cells, neutrophils, myeloid cells, endothelial cells and other T cell clusters, mainly cycling, CD4 CXCL13, IFN signature and FOXP3 Treg (Fig. 4.31A). PTPRC (CD45) is predicted to interact with mannose receptor C-type 1 (MRC1) on myeloid cells and CD22 on B cells. They express *CCL5* (RANTES), an inflammatory chemokine that acts as a potent chemoattractant for immune cells. Incoming interactions for both the Canonical CTL and IFN Signature group include CCR7 stimulation by endothelial cells, TNFRSF1B by neutrophils and myeloid cells and importantly a4b1 complex interacting with FN1 on fibroblasts. Circle plots for all T cell clusters can be found in Appendix D.14.

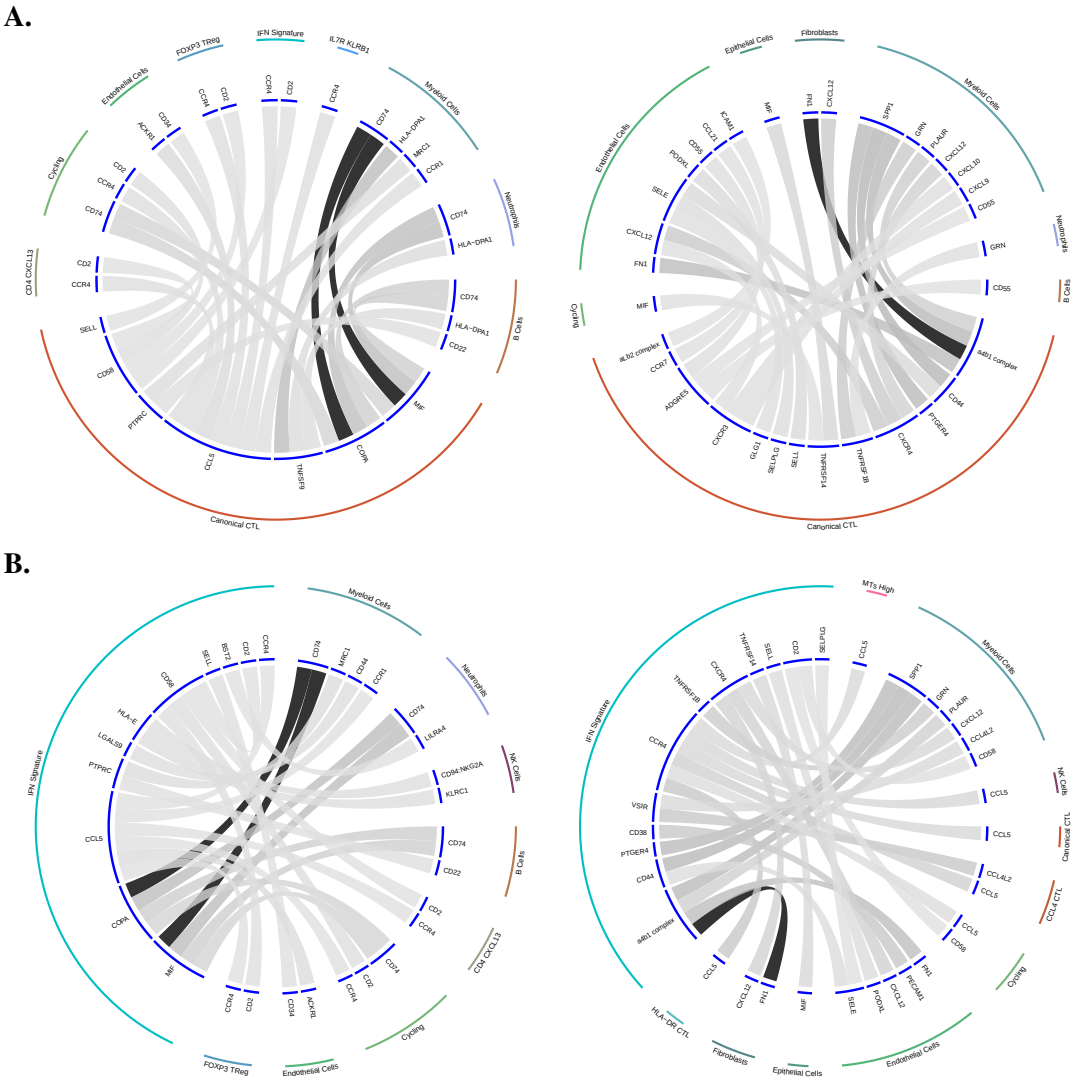


Figure 4.31: Circle Plots of Canonical CTL and IFN Signature Interactions. Outgoing and incoming interactions on the left and right, respectively. Cell clusters are defined around the outer circle with genes in the inner one. Links are drawn between ligand-receptor pairs. Darker links represent highest interaction scores. **A.** Canonical CTL cluster in orange. **B.** IFN Signature cluster in aqua blue.

4.4.9 Transcriptome-Wide Spatial Profiling Captures Cellular Organisation

Gene expression was assigned to a specific location on the Slide-seqV2 pucks by mapping the sequencing data to the pre-sequenced spatial tags. The four pucks were analysed in parallel and similar preprocessing steps to the scRNA-seq experiments were followed (Fig. 4.32). The recovered UMI counts per bead were significantly less than the UMI counts per cells in scRNA-seq. Whereas with scRNA-seq the median UMI counts fell in the range of 1,000 to 3,500, the median UMI counts for Slide-seqV2 was calculated to fall in the range of 50 to 200. However, the variance in counts was extremely high and could be as a result of differences in cell density across the tissue.

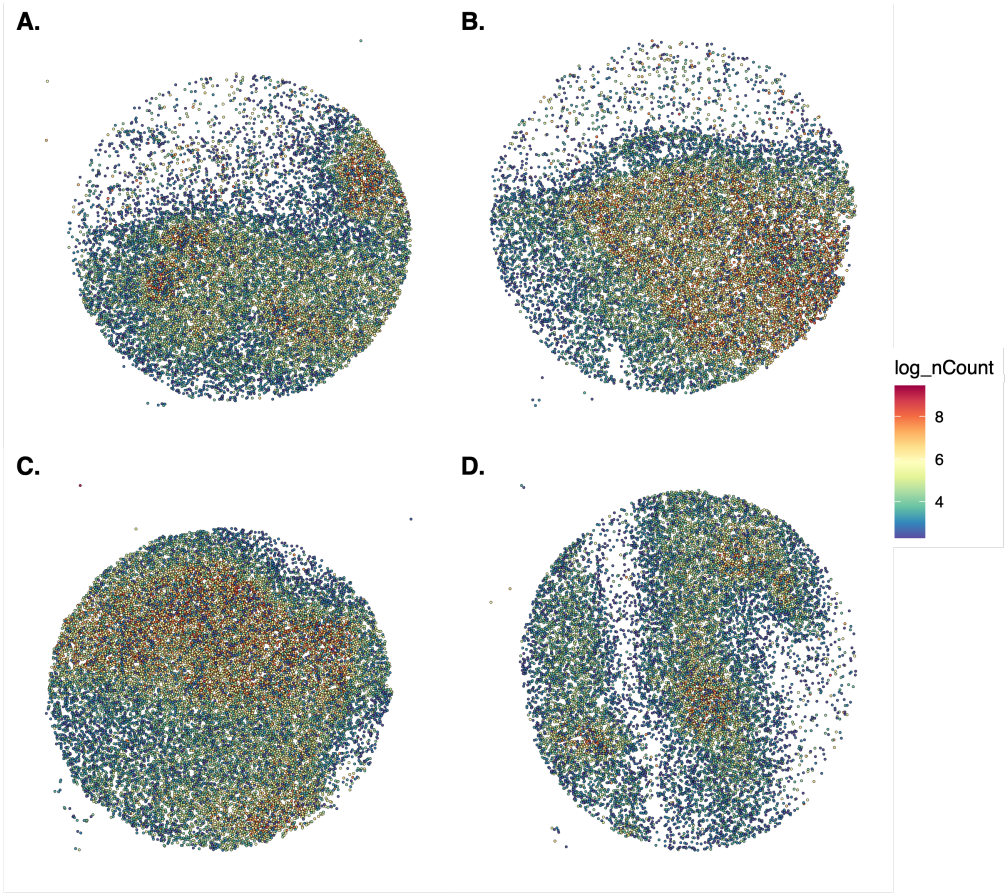


Figure 4.32: Slide-seqV2 Arrays Bead Counts.
Arrays displaying log-transformed bead UMI counts for **A.** Puck 1 (BRO18). **B.** Puck 2 (BRO20). **C.** Puck 3 (BRO20). **D.** Puck 4 (BRO23).

Predicted cell populations from the scRNA-seq data were used to assign detected beads to cell types in a patient-specific manner (Fig. 4.33). The wide variance in cell type proportions identified through H&E staining qualitatively holds true in the scRNA-seq space. Since Slide-seq provides a near-single-cell resolution, this method provides insight into the potential cellular organisation of the populations identified in Section 4.4.1.

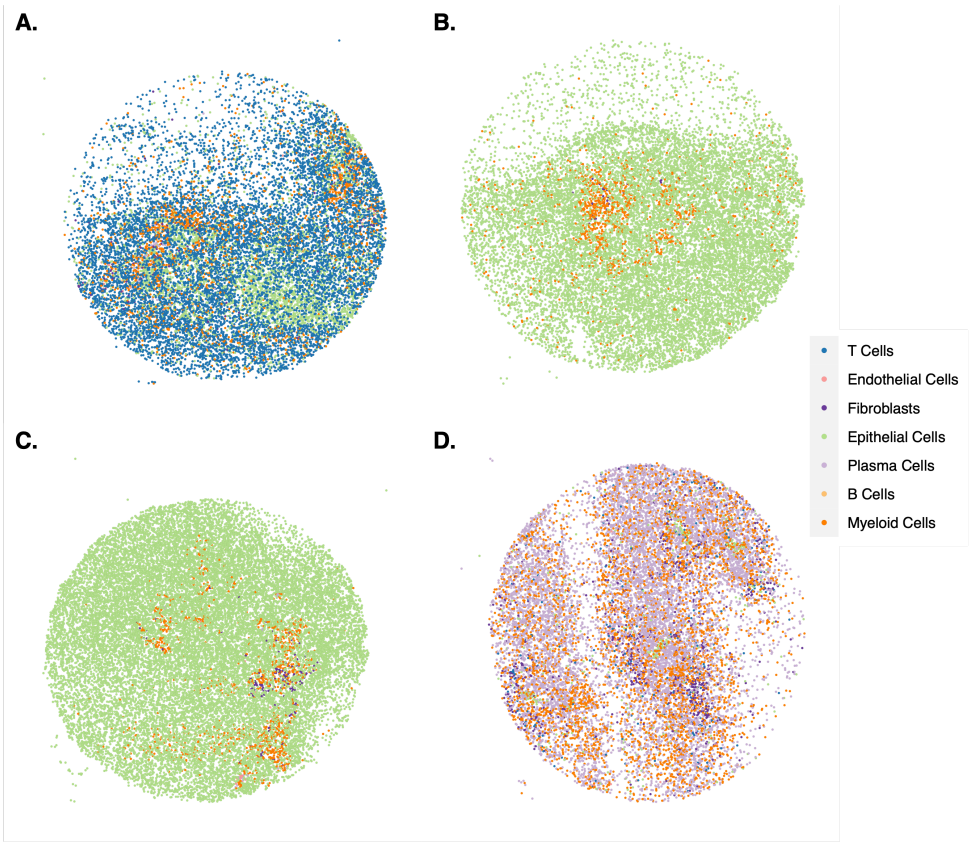


Figure 4.33: Transferring scRNA-Seq Cell Type Annotations Onto Arrays. Arrays displaying all cell types identified using TNBC scRNA-seq as reference for **A.** Puck 1 (BRO18). **B.** Puck 2 (BRO20). **C.** Puck 3 (BRO20). **D.** Puck 4 (BRO23).

The resulting proportions can be found in Table 4.3, along with the top spatially differentiated genes (SDGs) per puck. As shown in Table 4.2, sample BRO18 had been identified to exhibit a high level of lymphocytic infiltrate, in line with the high proportion of immune cells identified using the scRNA-seq as reference. Though the level of identified tumour cells was comparable to the H&E measured proportion (23% vs. 30%), only 1% of the beads were matched to the fibroblast class. As a matter of fact, few beads matched to fibroblasts across the four pucks. The second and third puck were predicted to consist

predominantly of a large region of tumour (93% and 94%) coming from the same patient. The scarce level of immune cells (5%) identified in puck 3 was in agreement with the H%E approximations. Puck 4 was biased towards a high immune proportion due to the high level of plasma cells identified across the tissue. Prior to Slide-seqV2, all four samples had been found to be of variable quality. Importantly, the sample mounted onto puck 4 had demonstrated some limitations, including fresh frozen tissue artefacts and smaller regions of carcinoma. The dependence of features on spatial locations was measured and ranked for each puck. The top SDGs mainly consisted of immunoglobulin genes including *IGKC*, *IGHA1* and *IGLC3*. Collagen-dense regions were identified with the enrichment of collagen type I and III genes in pucks 2 and 3. Puck 2 showed an elevated expression level of *MALAT1* concentrated in the epithelial cells surrounding the area of predicted myeloid cells. Visual representations of SDGs can be found in Appendix D.15.

Table 4.3: Array Cell Type Proportions and Spatially Variable Features.
Summary of findings for all four Slide-seqV2 arrays.

Puck ID	Patient ID	Tumour (%)	Immune (%)	Fibroblasts (%)	Spatially Differentiated Genes (Top 5)
1	BRO18	23	76	1	*IGKC*, *IGHG4*, *IGHG1*, *IGHG3*, *LINC00486*
2	BRO20	93	7	0	*MALAT1*, *IGKC*, *IGHG1*, *IGHG3*, *COL1A2*
3	BRO20	94	5	1	*IGHA1*, *COL3A1*, *IGLC2*, *COL1A2*, *IGKC*
4	BRO23	4	91	5	*IGLC3*, *IGLV7-46*, *IGLV3-16*, *IGLV3-1*, *SCGB2A2*

For the two pucks containing beads matched to T cells, the scRNA-seq T cell subtype atlas created in Section 4.4.3 was used to further refine the T cell space (Fig. 4.34). Patient sample BRO18 mounted onto puck 1 had the highest level of T cell infiltrate, consisting of 65% of the beads whilst only 2% of the beads from BRO23 mapped to T cells (Fig. 4.34). In puck 1, 41% matched to the IL7R+ KLRB1+ subtype, 11% to FOXP3+ TReg+ and 4% to CD69+ with a smaller proportion of beads matching to the other T cell subtypes. In sample BRO23, the largest population was also found to come from the IL7R+ KLRB1+ class (0.83%) followed by IFITM3+ TCF4+ (0.5%) and FOXP3+ TReg (0.25%).

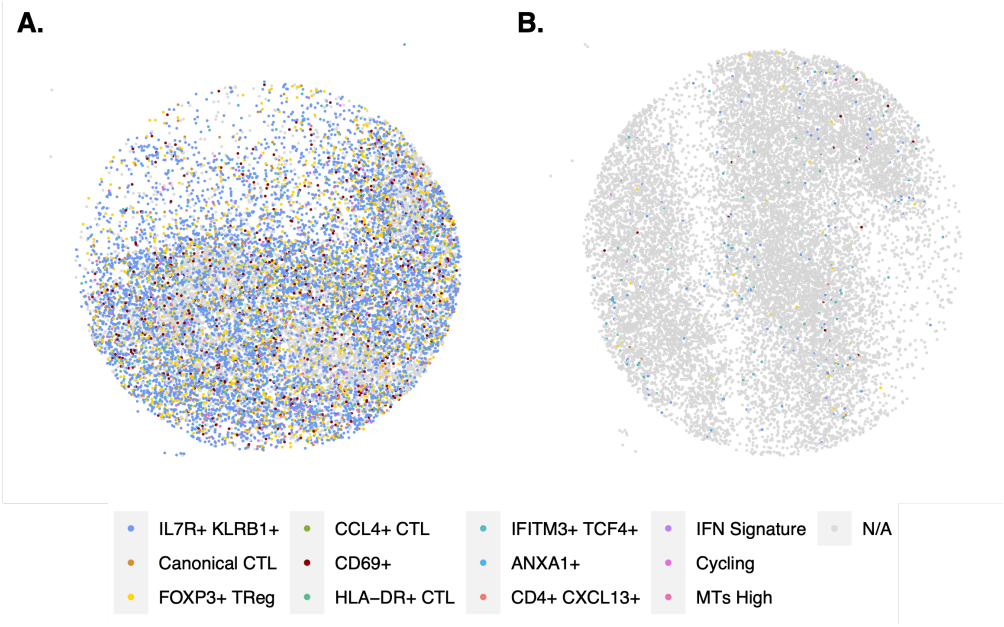


Figure 4.34: T Cell Subtypes Transferred Onto Arrays.
The scRNA-seq T cell subtype atlas was used to further refine the identity of the beads labelled as T cells.

In order to get a sense of the spatial distribution of genes related to IFN signalling relative to T cells, corresponding genes compiled in Section 4.3.1 were used to search the bead space in puck 1. However, a limited number of IFN-related genes had been recovered, likely owing to the high drop-out rate associated with the method. Despite this low capture efficiency, genes *ISG15*, *MX1* and *IRF7* were found to be concentrated in beads identified as epithelial or endothelial cells (Fig. 4.35).



Figure 4.35: Spatial Distribution of IFN-Mediated Signalling Genes.
Spatial differentiation scores for ISG15, MX1 and IRF7 projected onto puck 1.

4.4.10 High-Resolution Exploration of Subcellular Transcriptomic Activity

Contrarily to capture-based methods, spatial imaging technologies detect selected target genes by hybridising complementary fluorescent probes (Molecular Cartography, Resolve Biosciences) or sequencing directly into a fixed tissue section (in situ sequencing, CARTANA), thereby providing subcellular resolution of gene activity. Using 10 clinical tissue samples, a total of 20 images were collected for analysis, 18 through Molecular Cartography (MC) experiments and 2 with ISS. Images that passed quality control were used for colocalisation analysis of spatial probes. Individual images were further analysed to visualise transcriptomic patterns within the spatial context of the tumour microenvironment. Selected sets of genes were overlayed onto the tissue images collected using ISS and a more comprehensive analysis was conducted on the MC data. The scRNA-seq data from Section 4.4.1 was used as reference with the average expression levels of spatial probe targets measured for each major cell population (Fig. D.16). The segmented cell ROIs were used to isolate transcript counts per cell for subsequent cell type prediction. Cell labels were projected onto the spatial space to visualise tissue organisation and architecture. A total of 34 genes were shared between the two probe sets: *ADAM17*, *AREG*, *CA9*, *CD28*, *CD34*, *CD3D*, *CD4*, *CD68*, *CD8A*, *DDIT4*, *EGFR*, *ERBB3*, *EREG*, *ESR1*, *FGF7*, *FGFR1*, *FOXP3*, *HBEGF*, *HK2*, *IDO1*, *KLRB1*, *KRT19*, *KRT7*, *LDHA*, *MS4A2*, *NDRG1*, *P4HA1*, *PECAM1*, *PGK1*, *PTPRC*, *SLC16A1*, *SLC2A1*, *TGFBR1* and *VEGFA*.

Spatial Colocalisation of Gene Transcripts

Tissue-wide colocalisation of gene transcripts was averaged across all datasets that passed quality control for each type of experiment. Colocalisation was averaged across 14 images for the MC experiment and consists of 10,000 gene pairs (Fig. 4.37). On the other hand, the ISS probe colocalisation was averaged across two samples, resulting in the 9,604 gene pairs (Fig. 4.36). Probes were assigned to a class related to the biological role of their target gene. Hierarchical clustering was used on both rows and columns to arrange probes based on the calculated distance between them. Individual matrices can be found in Ap-

pendix D.17. Gene pairs with a high colocalisation score ($\geq$1.5) and those with a score that suggests spatial avoidance ($\leq$-1.5) were searched in the literature, both for validation and exploratory purposes.

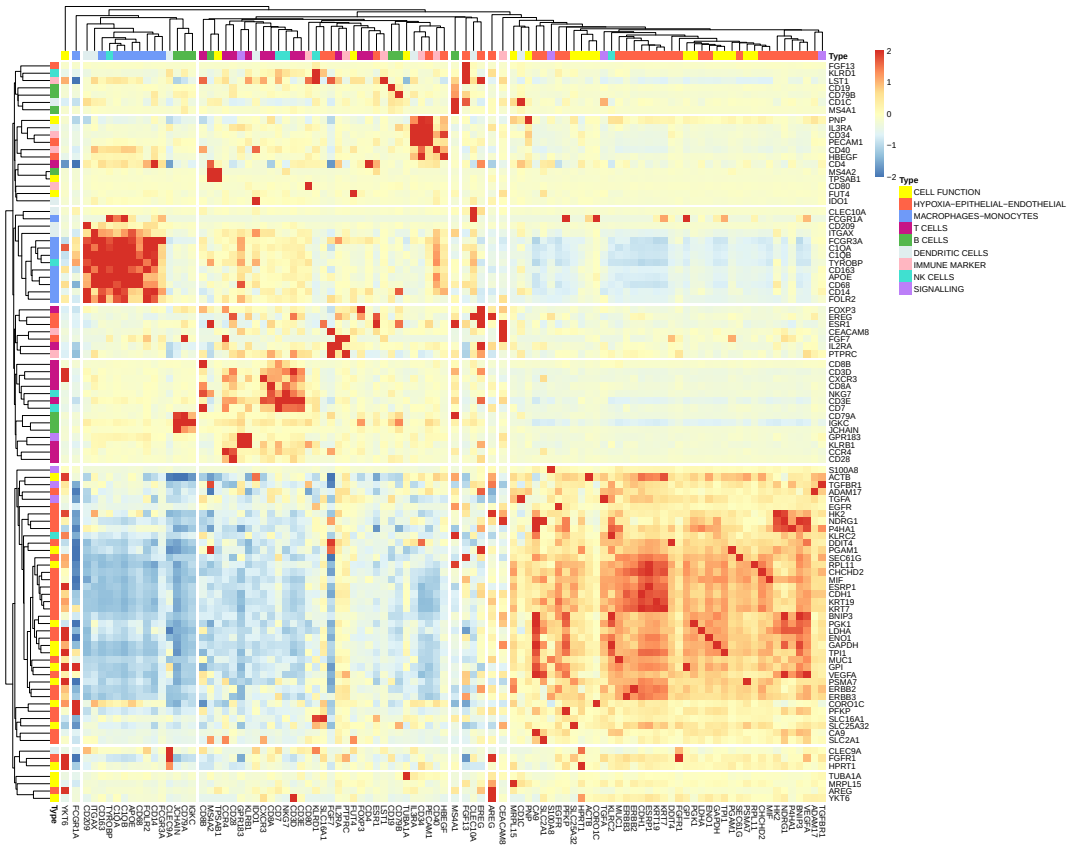


Figure 4.36: Tissue-Wide Colocalisation of In Situ Sequencing Probes.
Clustered heatmap of colocalisation scores for the in situ sequencing padlock probes averaged over 2 tissue samples. Search centre transcripts are given on the x-axis and linked partners are found on the y-axis. A high score is indicative of close spatial proximity and a low score suggests spatial avoidance. Probe class labels are colour-coded and dendrograms obtained using Euclidean distance measure are displayed.

Analysis of the spatial organisation of the ISS probes revealed relationships consistent with the underlying biology. For the most part, probes colocalised with probes from the same set or similar. The Cell Function class is composed of genes providing instructions for making enzymes involved in life-sustaining processes including cellular metabolism, cellular integrity, cell motility and cell cycle progression. Gene probes from this class were found to highly cluster with the class aggregating hypoxia, epithelial and endothelial markers. Gene sets related to immune function, mainly Macrophages-Monocytes,

T Cells, B Cells, Dendritic Cells, NK Cells and Immune Markers had the tendency to cluster together. T cells markers were found to be spatially correlated particularly *CD3D*, *CD8B*, *CD8A* and *NKG7*, similar to *FOXP3* and *CD4*. Marker *KLRB1* was found to highly associate with *GPR183*. Interestingly, two markers had a very pronounced all-or-none relationship with genes from the Macrophages-Monocytes and Cell Function class. Gene *FCGR1A* was found enriched with *GPI*, *CORO1C* and *PFKP* but depleted with genes like *PSMA7*, *ENO1* and P*GK1* among others. Similarly, gene *FGF7* was enriched with *DDIT4*, *PGAM1* but avoided *ACTB* and *SEC61G*.

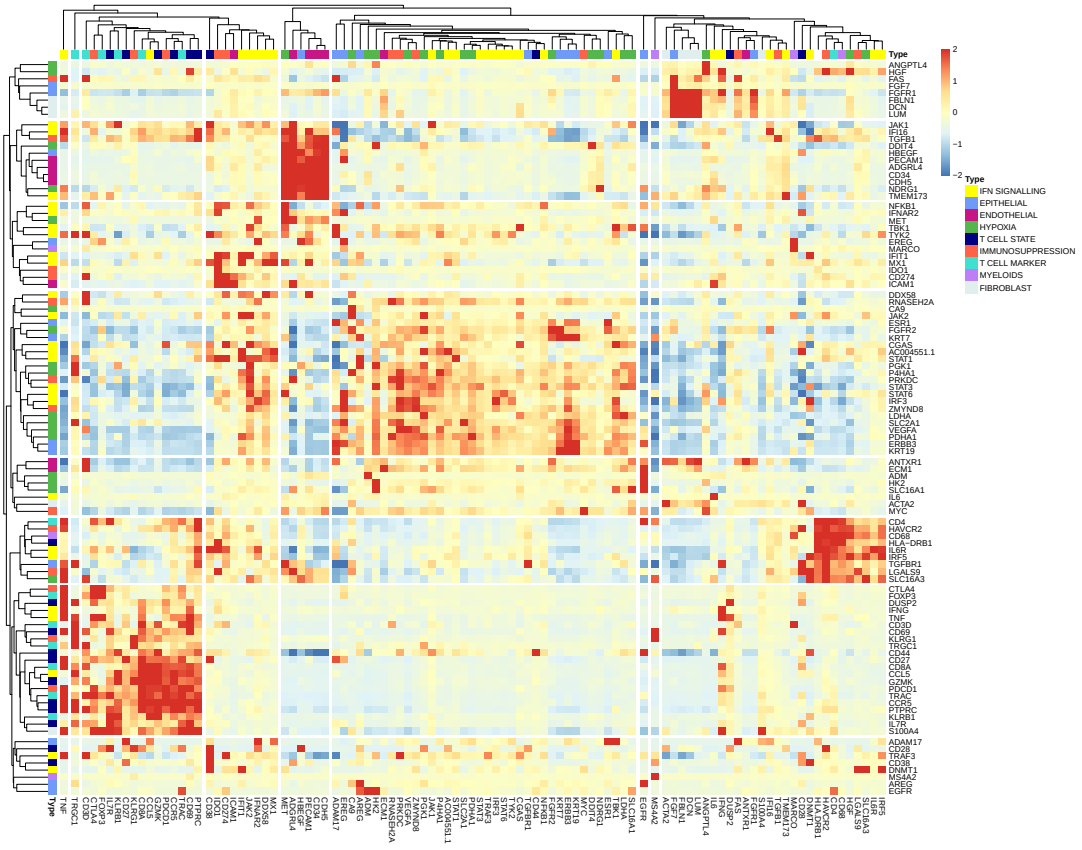


Figure 4.37: Tissue-Wide Colocalisation of Molecular Cartography Probes. Clustered heatmap of colocalisation scores for the Molecular Cartography gene probes averaged over 14 tissue samples. Search centre transcripts are given on the x-axis and linked partners are found on the y-axis. A high score is indicative of close spatial proximity and a low score suggests spatial avoidance. Probe class labels are colour-coded and dendrograms obtained using Euclidean distance measure are displayed.

Similar relationships were identified upon examining the MC matrix. Fibroblast markers clustered together and in the proximity of epithelial cell markers. Probes in the T

Cell Marker and T Cell State sets were found to colocalise together, along with markers from the Immunosuppression class (e.g. *CTLA4* and *PD-1*). Other markers from the Immunosuppression class like *ZMYND8*, *RNASEH2A* and *PRKDC* were found to primarily colocalise with themselves and hypoxia and IFN signalling markers. *KLRB1* expression strongly colocalises with *IL7R*, *CCL5* and *GZMK* but not *CD8* nor *CD4*. Genes from the IFN Signalling set predominantly clustered with hypoxia markers which in turn colocalised with epithelial and endothelial markers. Focusing on genes of the cGAS-STING signalling pathway, it was found that *cGAS* is enriched with *TBK1*, *RNASEH2A*, *JAK2* and *LDHA*. In turn, TBK1 strongly colocalised with epithelial marker *ADAM17*, hypoxia markers *SLC16A1* and *LDHA*, IFN-related markers *cGAS*, *IRF3* and *TYK2*, and *ZMYND8* from the Immunosuppression class. *ZMYND8* did not colocalise with T cell markers and mainly with epithelial markers (*ERBB3*, *KRT19*), hypoxia markers (*SLC2A1*, *VEGFA*, *PDHA1*) and IRF3. *DDX58* (RIG-I) was enriched with other IFN-related probes including *IFIT1*, *MX1*, *cGAS*, *STAT1*, *AC004551.1*, *STAT6*, *IRF3*. Endothelial markers (*CDH5*, *CD34*, *PECAM1*, *ADGRL4*) and the hypoxia marker *MET* strongly colocalised with *TMEM173* (STING). *IFNγ* transcripts closely associated with *CD8A* and the T cell activation marker *CD69* along with the hypoxia-related gene *HGF* and the cytokine *TNF*. *DNMT1* was found strongly enriched with IFN signalling genes *IRF3*, *IRF5* and *IFIT1*, with immune checkpoint *LGALS9* and hypoxia regulated gene *SLC16A3*. *MX1*, *DDX58* and *IFIT1* colocalised with an antisense lncRNA to *OAS1*, *AC004551.1*, a negative regulator of the IFN response. *EGFR* strongly colocalised with markers of hypoxia, endothelial cells and T cell markers *CD4* and *CD44*. Interestingly, *EGFR* and genes from the IFN Signalling class were strongly anticolocalizsed. *ICAM1* and *ANTXR1* which had been added to the panel of targets because of their potential as CAR T cell targets were found to colocalise with endothelial markers. Intriguingly, *CD274* (PD-L1) was primarily enriched with genes from the IFN Signalling class (*TRAF3*, *STAT1*, *cGAS*, *IFIT1*) and also the CAR target candidate *ICAM1*.

BBP1467

In order to visualise the colocalisation of T cell markers measured in the above section, transcript coordinates were overlayed onto the microscopy image as colour-coded spots (Fig. 4.38). Resident lymphocytes were found to be scattered across the tumour section based on the colocalisation of T cell markers *CD8A*, *CD8B*, *CD4*, *CD2D* and *CD3E*. A *CD4*:*CD8* ratio of 5.89 was calculated (3,000 *CD4* to 509 *CD8A*) and suggests a significantly greater level of CD4 tumour-infiltrating lymphocytes than CD8. Out of a total of 1,813,541 detected transcripts, the most abundant ones were *IGKC* (707,997), *ACTB* (149,972) *RPL11* (112,676), *KRT7* (108,714) and *S100A8* (92,019).

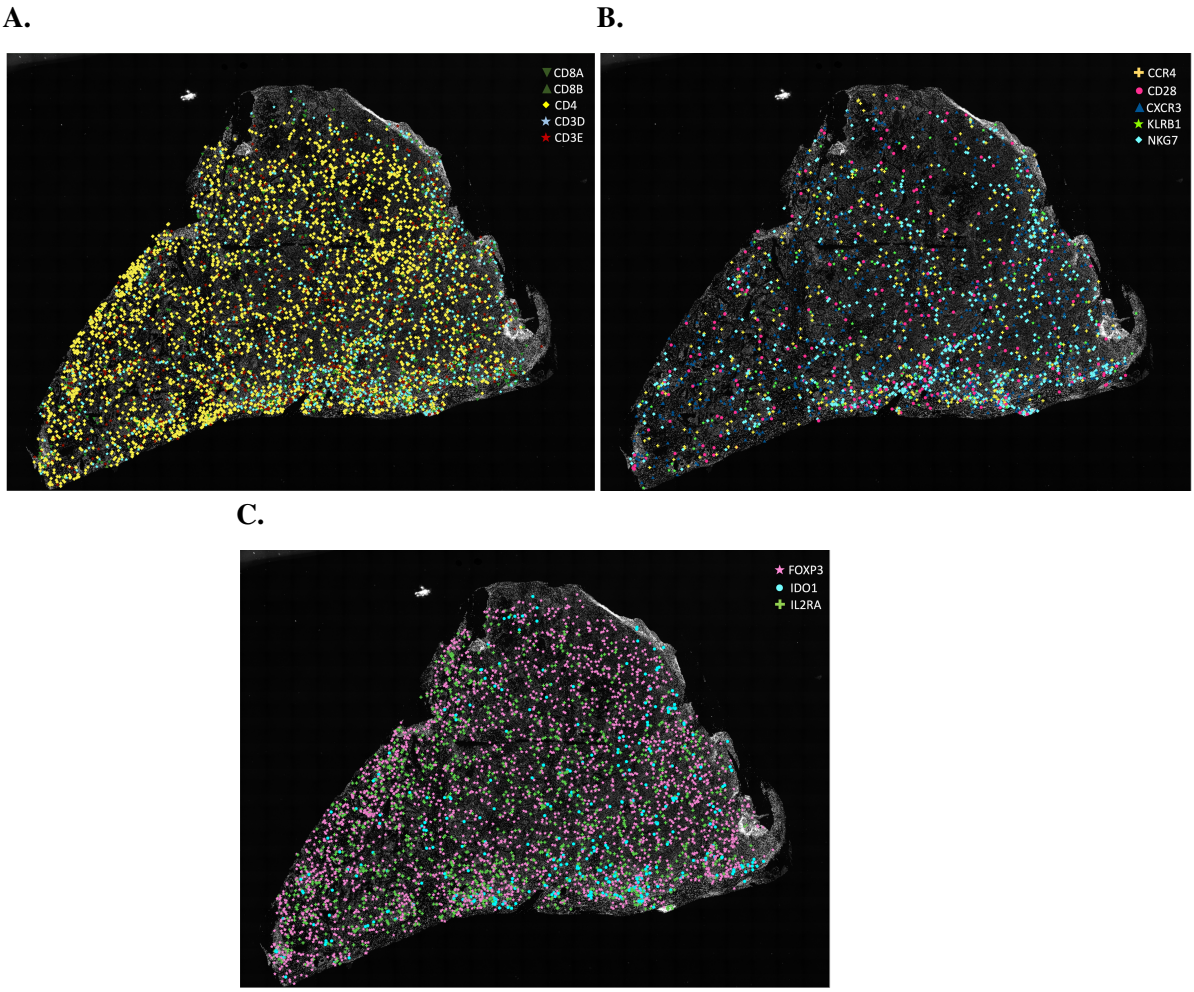


Figure 4.38: T Cell Markers Overlaying Tissue Sample BBP1467.
T cell markers targeted during the CARTANA ISS experiment were split into groups and projected onto the microscopy image of sample BBP1467. **A.** General T cell markers. **B.** Markers associated to a T cell state or subtype. **C.** Markers involved in suppressing T cell immunity.

BBP1557

The same process as above was repeated for BBP1557. Again, none of the markers were found to concentrate in a specific area, with the exception of *IDO1* which is found to be expressed in the tissue periphery and the *CD4* density being greatest in the central region. A *CD4*:*CD8* ratio of 5.82 was calculated (5,227 *CD4* to 898 *CD8A*). The total number of transcripts detected was 1,491,930 with the most expressed genes being *GAPDH* (365,652), *RPL11* (150,508), *ACTB* (137,729), *IGKC* (109,865), *KRT7* (88,392).

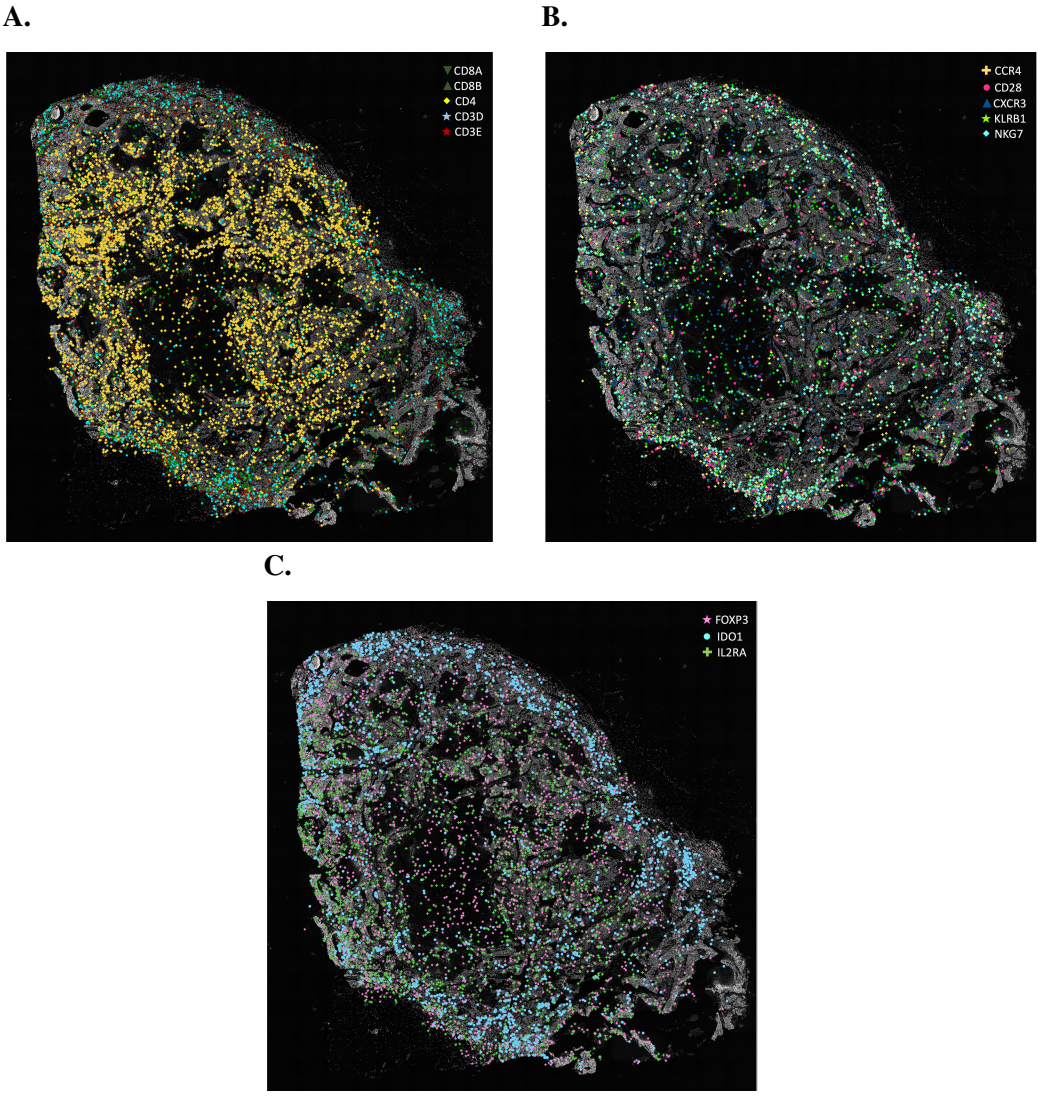


Figure 4.39: T Cell Markers Overlaying Tissue Sample BBP1557.
T cell markers targeted during the CARTANA ISS experiment were split into groups and projected onto the microscopy image of sample BBP1557. **A.** General T cell markers. **B.** Markers associated to a T cell state or subtype. **C.** Markers involved in suppressing T cell immunity.

BRO18

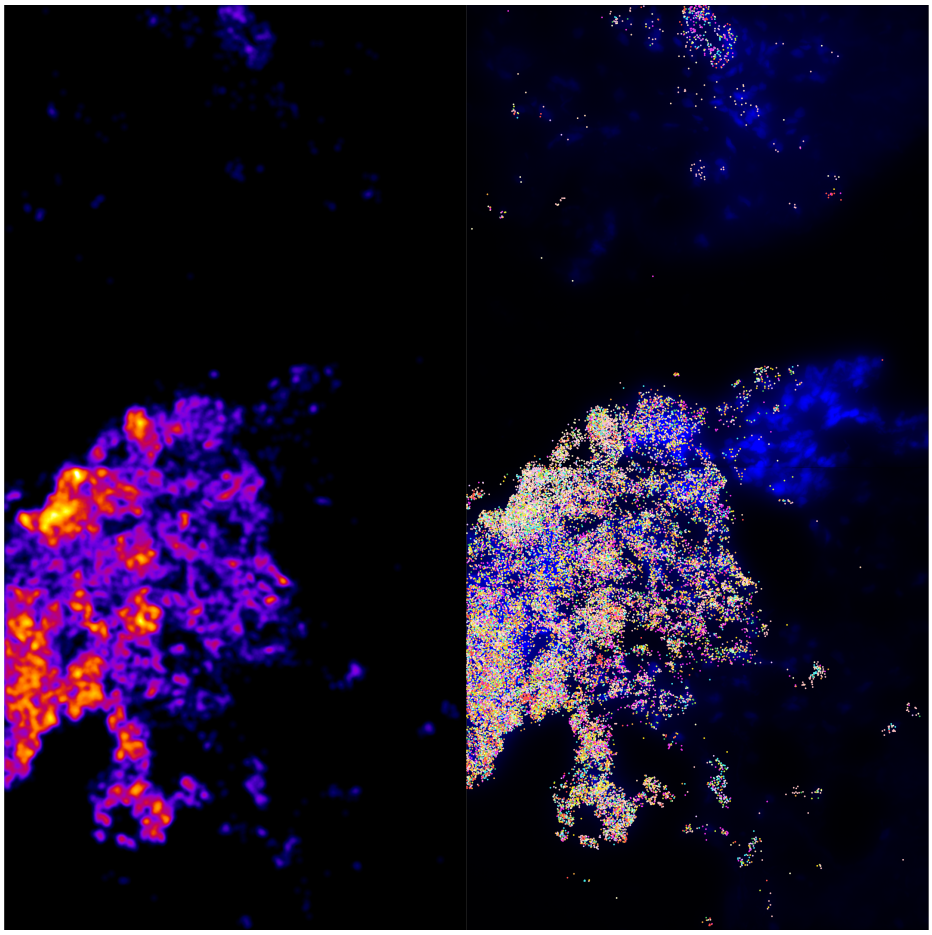
Patient sample BRO18 was split into four ROIs in well A1 (Table 4.4). During quality control, upon examining the resulting images it was found that A1-1 did not meet the minimum threshold of 90 genes detected. In addition, probes poorly attached to sample A1-4. As a result, the two ROIs were discarded from downstream analysis and only ROI A1-2 (Fig. 4.40A) and A1-3 (Fig. 4.40B) were used from this sample.

Table 4.4: Molecular Cartography Spatial Metrics for BRO18.

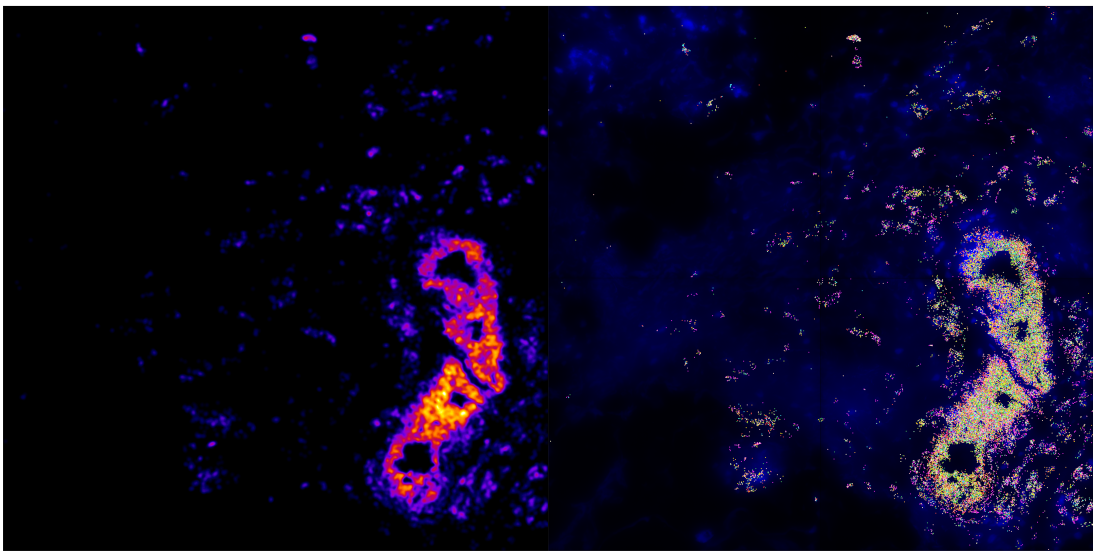
ROI	Detected Genes	Total Counts	Segmented Cells	Genes Per Cell	Signals Per Cell
A1-1	76	5,173	1,074	$\mu = 1.80$	$\mu = 4.52$
A1-2	93	32,675	623	$\mu = 13.23$	$\mu = 7.57$
A1-3	90	35,127	1,169	$\mu = 6.98$	$\mu = 27.74$
A1-4	26	179	148	$\mu = 0.39$	$\mu = 1.03$

From the combined images a total of 990 cells were matched to potential cell types. In A1-2, 388 ROIs were classified as T Cells (n=133), Macrophages (n=98), Fibroblasts (n=86), Epithelial (n=31), Endothelial (n=24) or MYC+ Endothelial (n=16), as shown in Figure 4.40C. Furthermore, there were signs of T cell infiltration in the area and interactions occurring between macrophages and T cells. In A1-3, the 602 ROIs were split into MYC+ VEGFA+ (n=221), Epithelial (n=116), Myoepithelial (n=86), Endothelial (n=75), Fibroblasts (n=62), T Cells (n=32) and Macrophages (n=10), as depicted in Figure 4.40D. The Myoepithelial cluster highly expressed the smooth muscle alpha-actin gene (ACTA2) which in the original probe panel was considered a marker of fibroblasts. However, upon examination of the spatial organisation of these clusters, it became apparent that this cluster represented myoepithelial cells, likely surrounding mammary gland ducts.

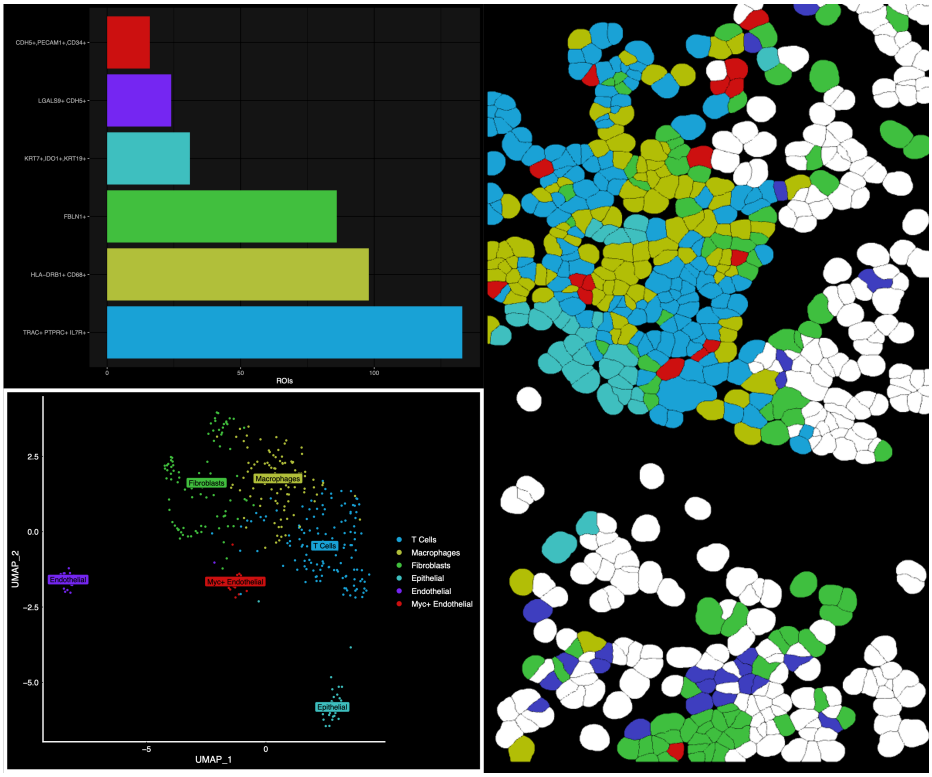
A.



B.



C.



D.

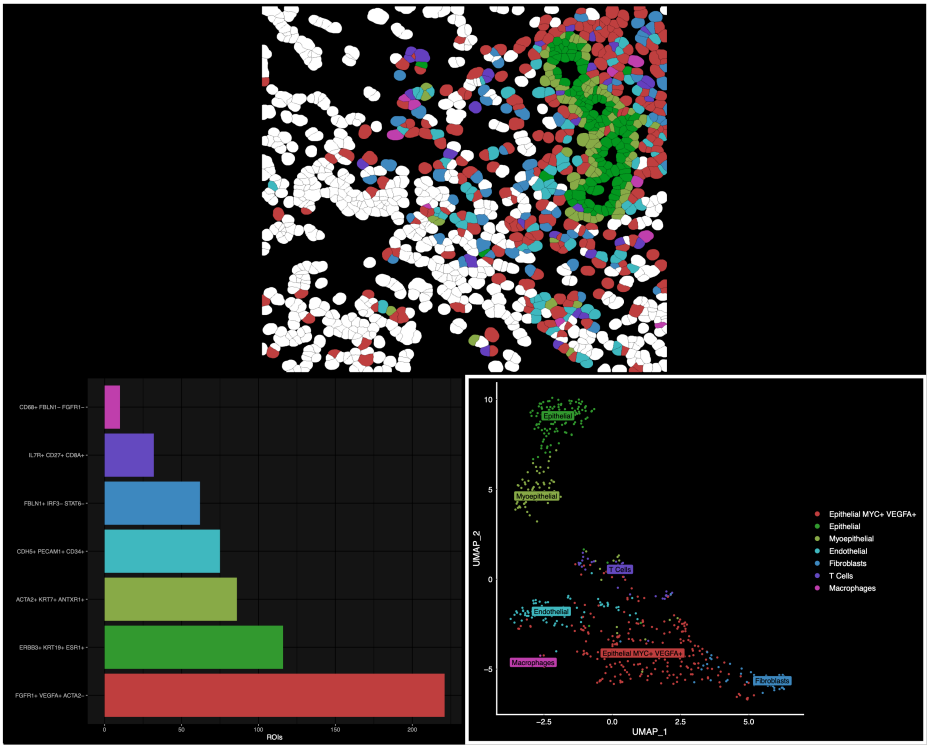


Figure 4.40: Spatial Organisation of Cell Populations for Sample BRO18. Density map of transcript abundance on the left and overlay of DAPI staining and RNA spots on the right for **A.** A1-2. **B.** A1-3. Tissue composition based on transcriptional profile projected onto segmented DAPI images for **C.** A1-2. **D.** A1-3.

BRO20

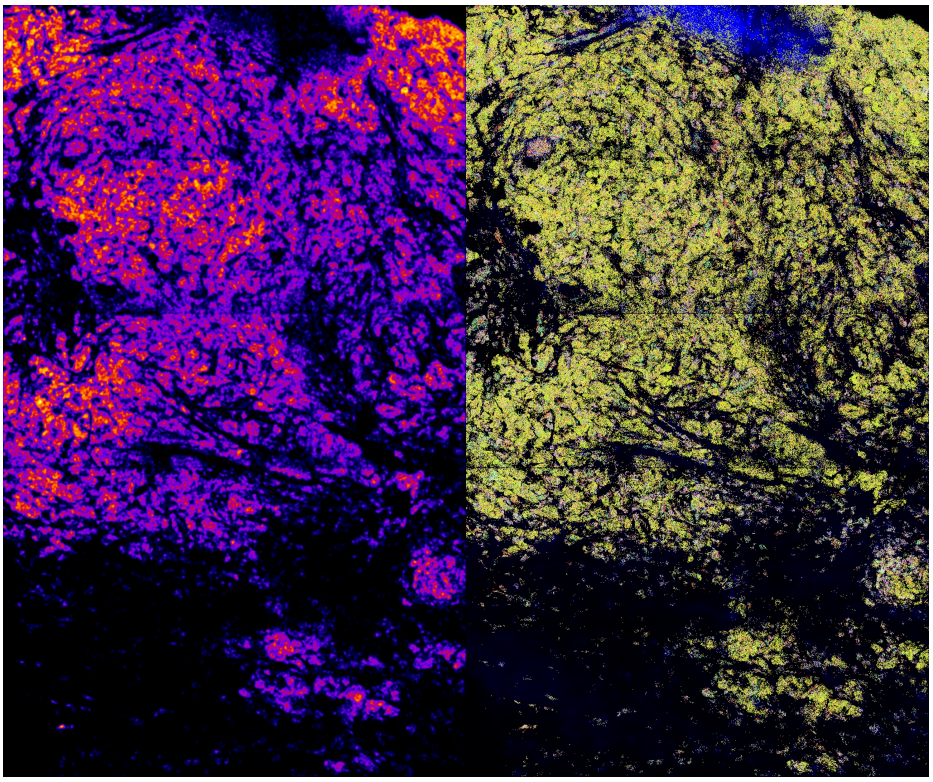
As described in Section 4.3.2, patient sample BRO20 consists of two biopsies taken from different areas of the tumour. One of the sample was split into three ROIs in well A2 and the other consists of one ROI in well B2 (Table 4.5).

Table 4.5: Molecular Cartography Spatial Metrics for BRO20.

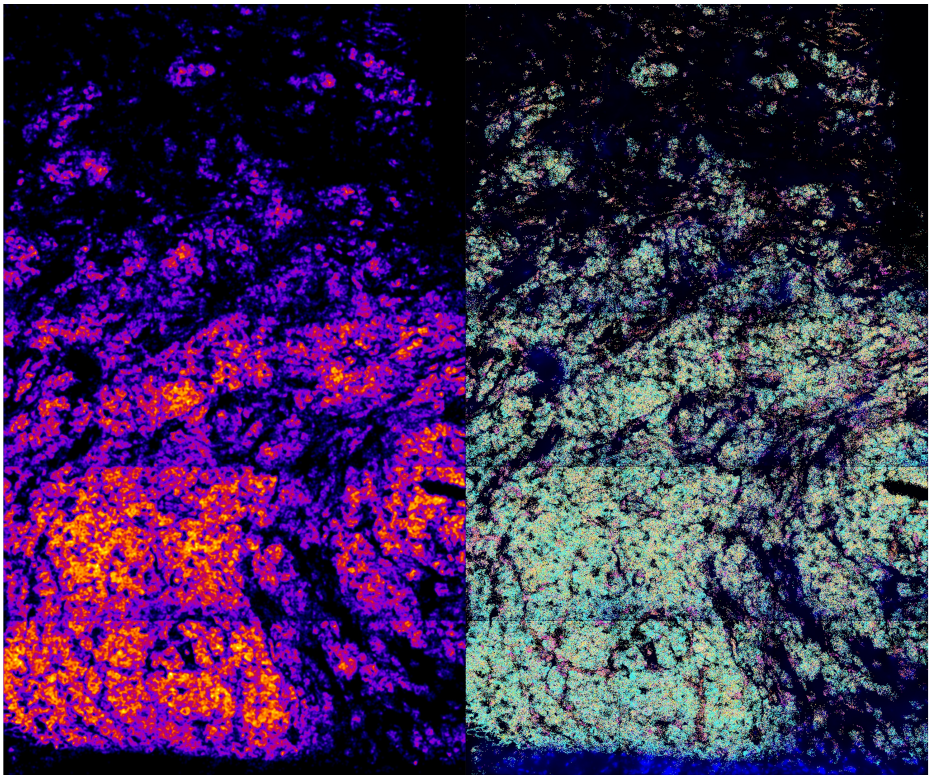
ROI	Detected Genes	Total Counts	Segmented Cells	Genes Per Cell	Signals Per Cell
A2-1	98	796,644	6,608	μ = 19.50	μ =114.01
A2-2	97	793,015	6,698	μ = 16.19	μ = 109.11
A2-3	92	88,948	2,807	μ = 7.74	μ = 29.67
B2-1	99	574,607	6,370	μ = 14.84	μ = 83.89

A total of 14,962 cells across the 3 ROIs in well A2 (Figs. 4.41A–4.41C) and 5,210 cells in ROI B2-1 were matched to potential cell types (Fig. 4.41D). As shown in Figure 4.41E, five populations were identified in A2-1: Hypoxic Epithelial (n=3963), Macrophages (n=1,262), Fibroblasts (n=740), Endothelial (n=237) and Epithelial (n=31). The population of macrophages was found to highly express *TGFB1*, a multi-functional cytokine associated with tumour development and progression whilst the majority of epithelial cells expressed high levels of hypoxia-related genes like *VEGFA*. Similarly, in Figure 4.41F it can be seen that cells in A2-2 were divided into: Hypoxic Epithelial (n=2,423), Hypoxic MYC+ IDO1+ HER3+ (n=1,587), Fibroblasts (n=666), Macrophages (n=663), Hypoxic SLC16A1+ Epithelial (n=550), Endothelial (n=229), T Cells (n=191) and Epithelial (n=52). Three cell groups differentially expressed hypoxic genes with one of them also expressing high levels of *IDO1* and oncogene *MYC*. ROI A2-3 (Fig. 4.41G) resulted in clusters labelled as Hypoxic Epithelial (n=1,081), Macrophages (n=712), Fibroblasts (n=411), Endothelial (n=128) and IFN Signature (n=36). The high expression of hypoxia genes in well A2 including *HK2*, *LDHA*, *P4HA1*, *PGK1*, *SLC16A1*, *SLC2A1*, *VEGFA*, *PDHA1* and *ADM* suggest that this biopsy was taken from an area with a low concentration of oxygen. Segmented cells in B2-1 (Fig. 4.41H) were classified into: Hypoxic Epithelial (n=1555), Epithelial (n=1,082), Endothelial (n=801), Fibroblasts (n=866), Macrophages (n=563), T Cells (n=325) and ANTXR1 (n=18) with two dense epithelial regions.

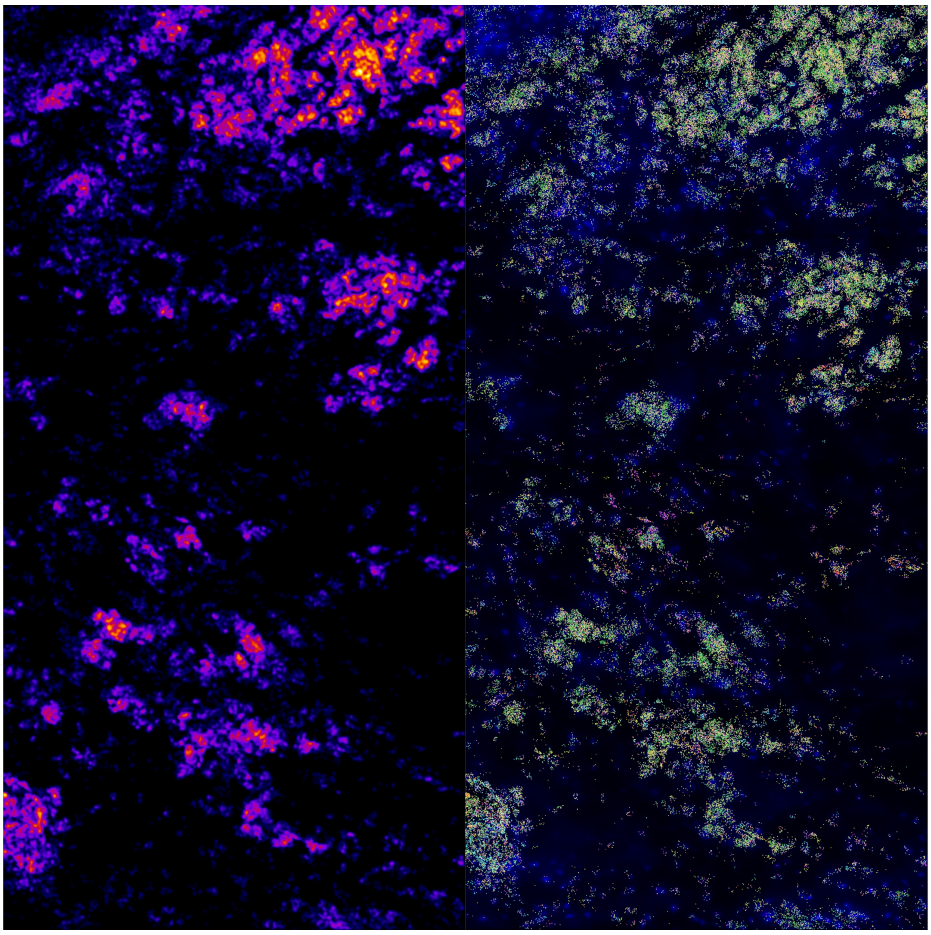
A.



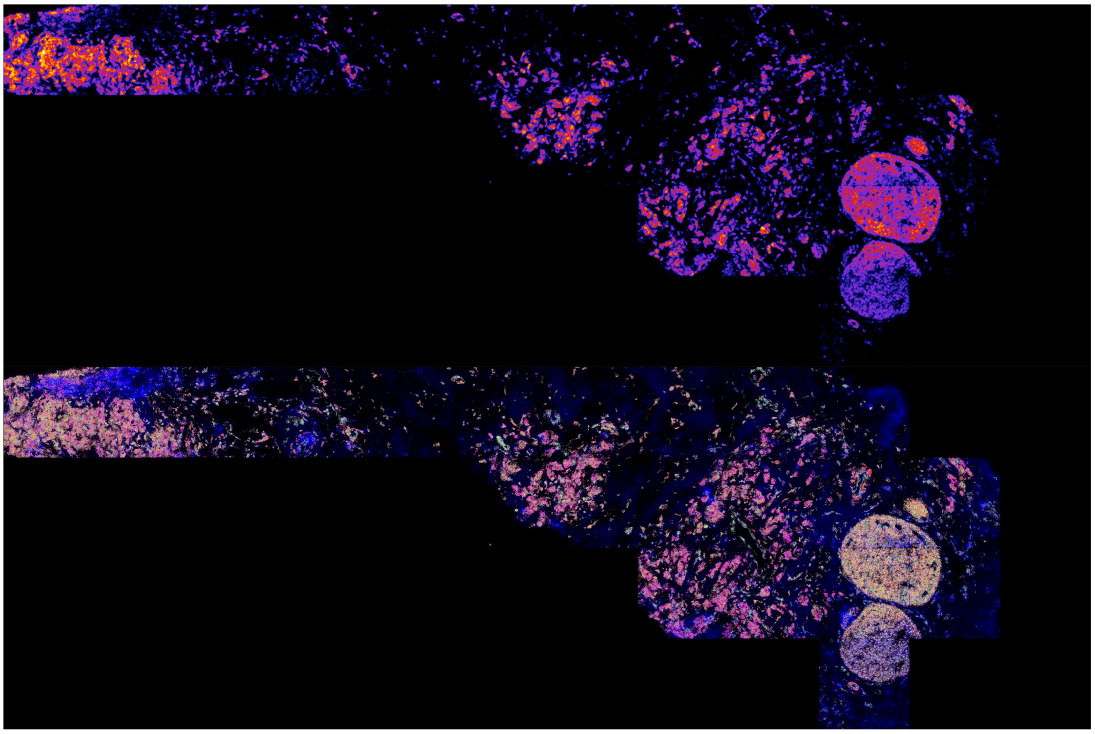
B.



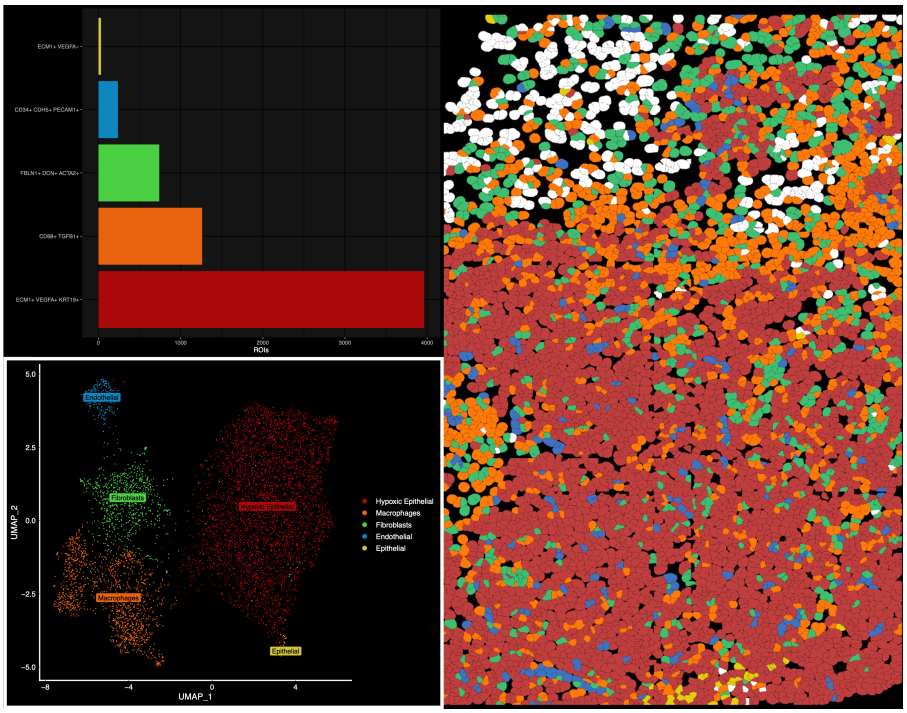
C.



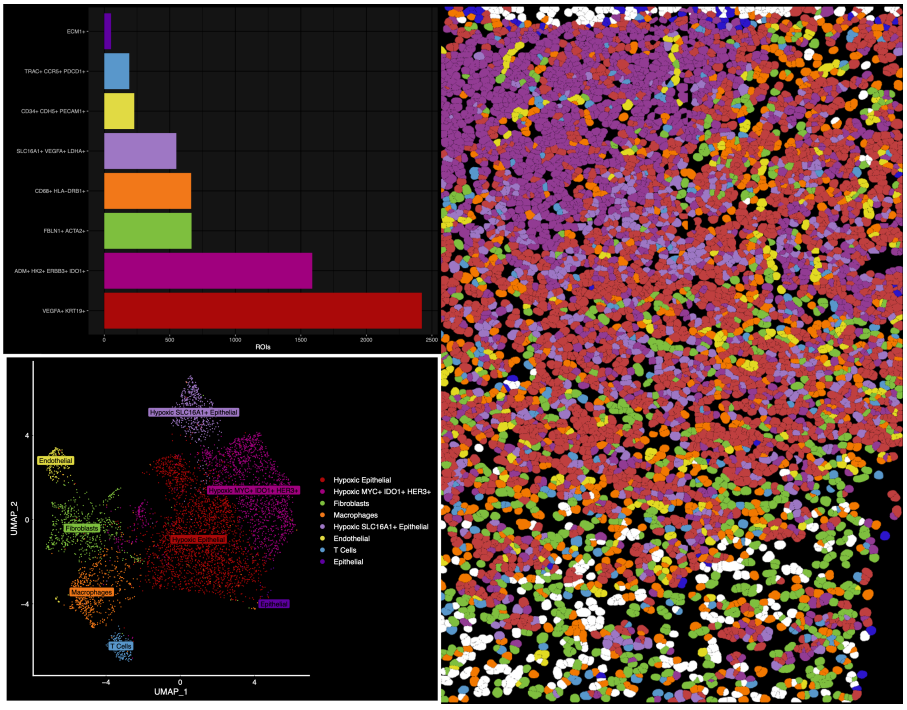
D.



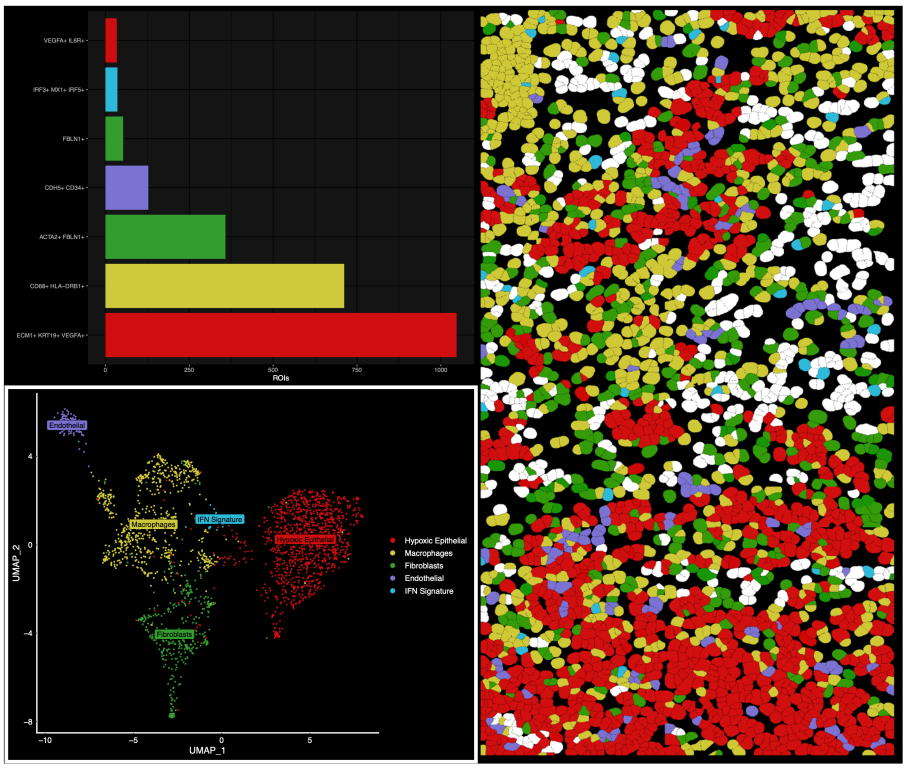
E.


ECM1+ VEGFA+
CD34+ CDH5+ PECAM1+
FBLN1+ DCN+ ACTA2+
CD68+ TGFB1+
ECM1+ VEGFA+ KRT19+
ROIs
Endothelial
Fibroblasts
Hypoxic Epithelial
Macrophages
Epithelial
UMAP_1
UMAP_2
Hypoxic Epithelial
Macrophages
Fibroblasts
Endothelial
Epithelial

F.


ECM1+
TRAC+ CCR6+ PDCD1+
CD34+ CDH5+ PECAM1+
SLC16A1+ VEGFA+ LDHA+
CD68+ HLA-DRB1+
FBLN1+ ACTA2+
ROIs
Hypoxic SLC16A1+ Epithelial
Endothelial
Hypoxic MYC+ IDO1+ HER3+
Fibroblasts
Macrophages
Hypoxic Epithelial
T Cells
Epithelial
UMAP_1
UMAP_2
Hypoxic Epithelial
Hypoxic MYC+ IDO1+ HER3+
Fibroblasts
Macrophages
Hypoxic SLC16A1+ Epithelial
Endothelial
T Cells
Epithelial

G.



H.

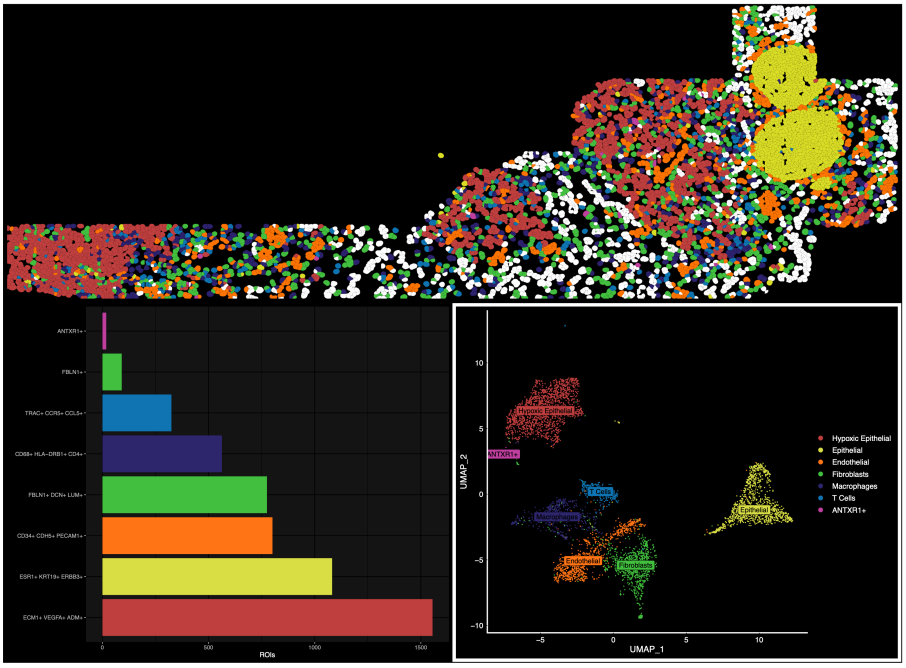


Figure 4.41: Spatial Organisation of Cell Populations for Sample BRO20. Density map of transcript abundance on the left/top and overlay of DAPI staining and RNA spots on the right/bottom for **A.** A2-1. **B.** A2-2. **C.** A2-3. **D.** B2-1. Tissue composition based on transcriptional profile projected onto segmented DAPI images for **E.** A2-1. **F.** A2-2. **G.** A2-3. **H.** B2-1.

BRO23

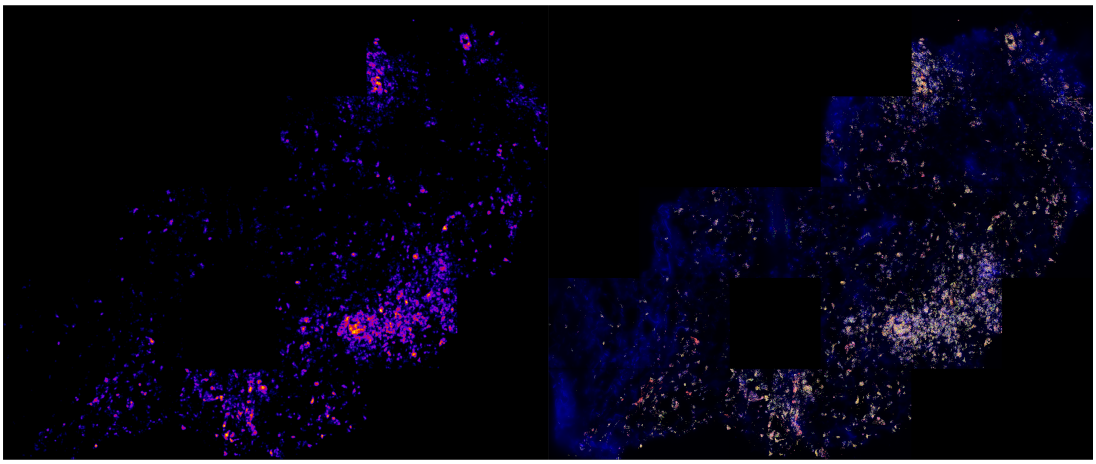
A single ROI was imaged for patient sample BRO23 placed in well B1 (Table 4.6).

Table 4.6: Molecular Cartography Spatial Metrics for BRO23.

ROI	Detected Genes	Total Counts	Segmented Cells	Genes Per Cell	Signals Per Cell
B1-1	97	88,821	4,835	μ = 6.82	μ = 17.07

A total of 2,753 cells were isolated from ROI B1-1 (Fig. 4.42A) and were labelled as follows: Mix (n=984), Endothelial (n=586), Fibroblasts (n=644), T Cells (n=367), Epithelial (n=100), TYK2+ (n=43) and Macrophages (n=29), as shown in Figure 4.42B. The cell cluster labelled Mix highly expressed genes characteristic of both IFN and hypoxic signatures, along with a mix of cell type markers for epithelial cells (*PECAM1*) and macrophages (*MARCO*). It could not be further separated into more meaningful cell populations. In the proximity of this group were a high number of cells classified as T cells.

A.



B.

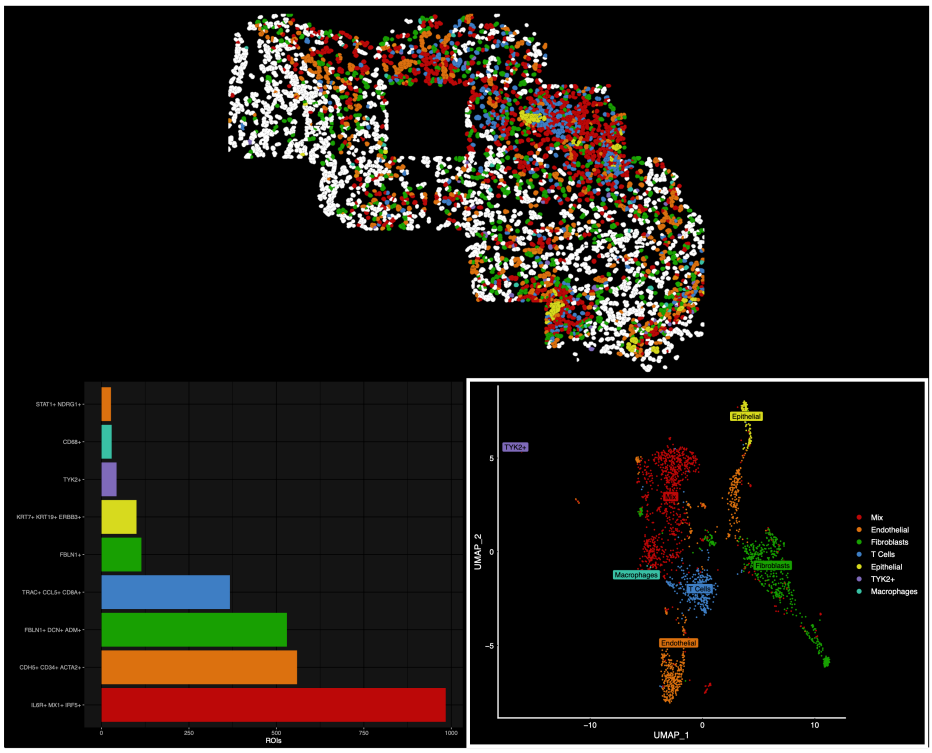


Figure 4.42: Spatial Organisation of Cell Populations for Sample BRO23. **A.** Density map of transcript abundance on the left and overlay of DAPI staining and RNA spots on the right for B1-1. **B.** Tissue composition based on transcriptional profile projected onto segmented DAPI images for B1-1.

TRN1

Two ROIs were selected for patient sample TRN1 in well C1 (Table 4.7). However, both ROIs failed QC and were omitted from downstream analysis. The majority of the transcripts in sample C1-1 were below the transcript quality score cutoff suggested by Resolve Biosciences. Furthermore, the number of detected genes in each ROI was lower than the decided QC threshold.

Table 4.7: Molecular Cartography Spatial Metrics for TRN1.

ROI	Detected Genes	Total Counts	Segmented Cells	Genes Per Cell	Signals Per Cell
C1-1	84	59,896	4,884	$\mu = 4.51$	$\mu = 11.58$
C1-2	78	15,024	2,088	$\mu = 3.19$	$\mu = 6.78$

TRN2

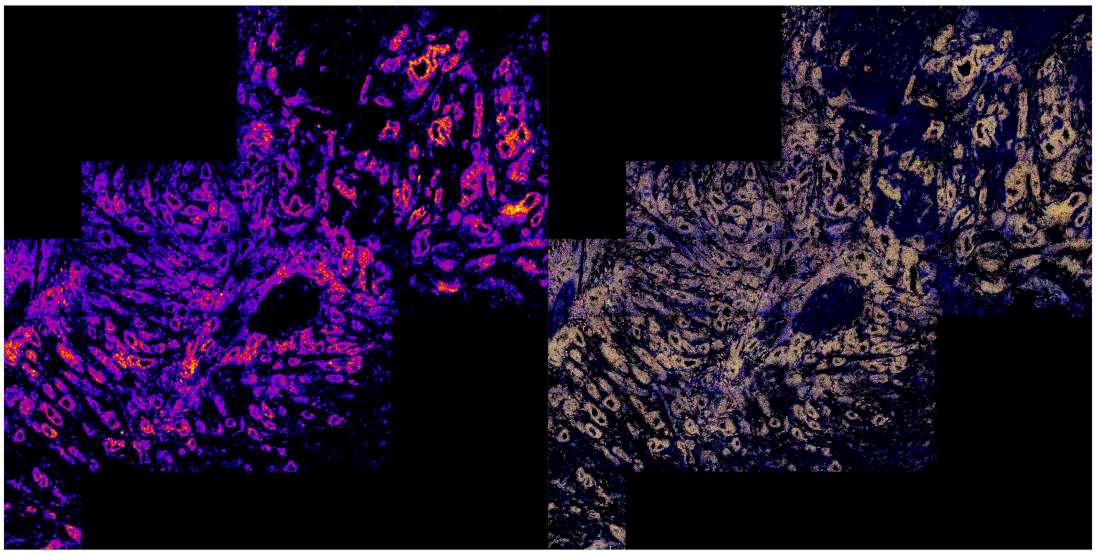
One ROI in well C2 was imaged for patient sample TRN2 (Table 4.8).

Table 4.8: Molecular Cartography Spatial Metrics for TRN2.

ROI	Detected Genes	Total Counts	Segmented Cells	Genes Per Cell	Signals Per Cell
C2-1	99	933,062	12,741	μ = 14.81	μ = 68.61

The cellular space was divided into Hypoxic HER3+ Epithelial (n=3,660), Fibroblasts (n=3123), HER+ Epithelial (n=2,111), Macrophages (n=1,165), Endothelial (n=737), P4HA+ RNASEH2A+ Epithelial (n=722), T Cells (n=251), Epithelial (n=166), STAT3- STAT6+ (n=156) and TYK2+ (n=62) giving a total of 12,153 cells. Four subpopulations of epithelial cells differed in the expression of hypoxic markers or in the expression level of markers like *RNASEH2A* and *P4HA*. In this tissue section, there was a high number of mammary ducts formed by epithelial cells from the Hypoxic HER3+ Epithelial and HER3+ Epithelial clusters. The lack of endothelial cells combined to the high expression of genes associated with epithelial-to-mesenchymal transition (EMT), mainly *KRT19*, *ERBB3* and *CD44* in cells lining some of these channels could also suggest the promotion of vasculogenic mimicry (VM).

A.



B.

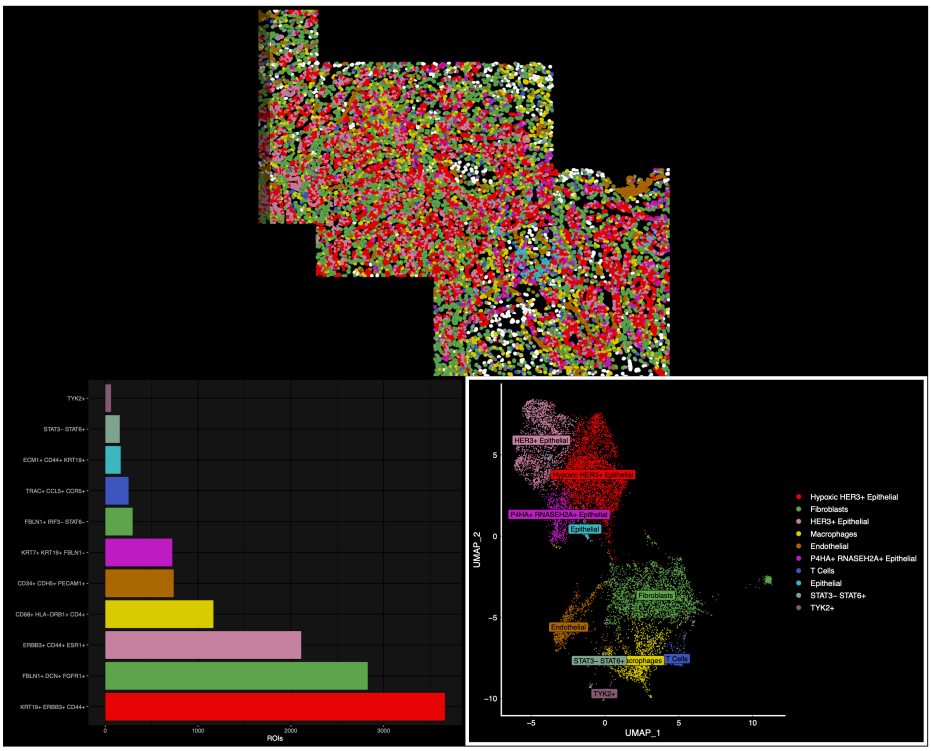


Figure 4.43: Spatial Organisation of Cell Populations for Sample TRN2.
A. Density map of transcript abundance on the left and overlay of DAPI staining and RNA spots on the right for C2-1. **B.** Tissue composition based on transcriptional profile projected onto segmented DAPI images for C2-1.

TRN3

Patient sample TRN3 was split into 2 ROIs (Table 4.9).

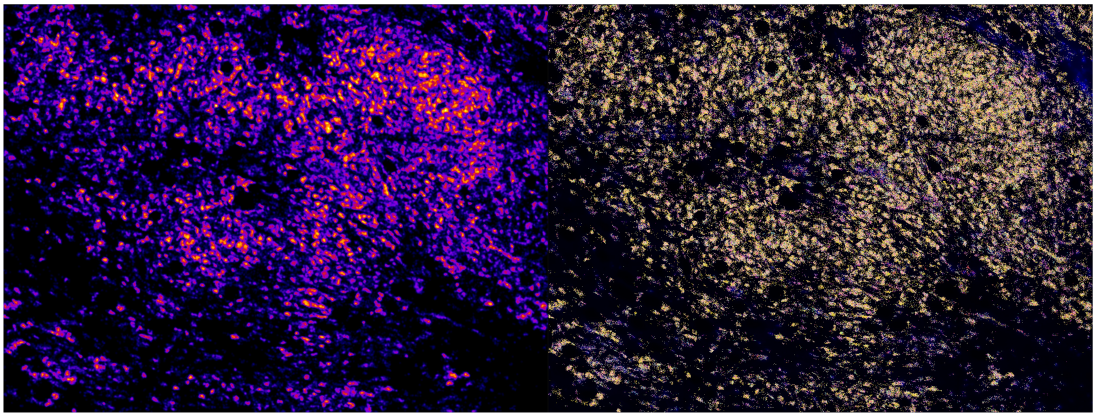
Table 4.9: Molecular Cartography Spatial Metrics for TRN3.

ROI	Detected Genes	Total Counts	Segmented Cells	Genes Per Cell	Signals Per Cell
D1-1	97	352,785	4,256	$\mu = 15.74$	$\mu = 77.84$
D1-2	99	794,275	11,920	$\mu = 14.17$	$\mu = 61.83$

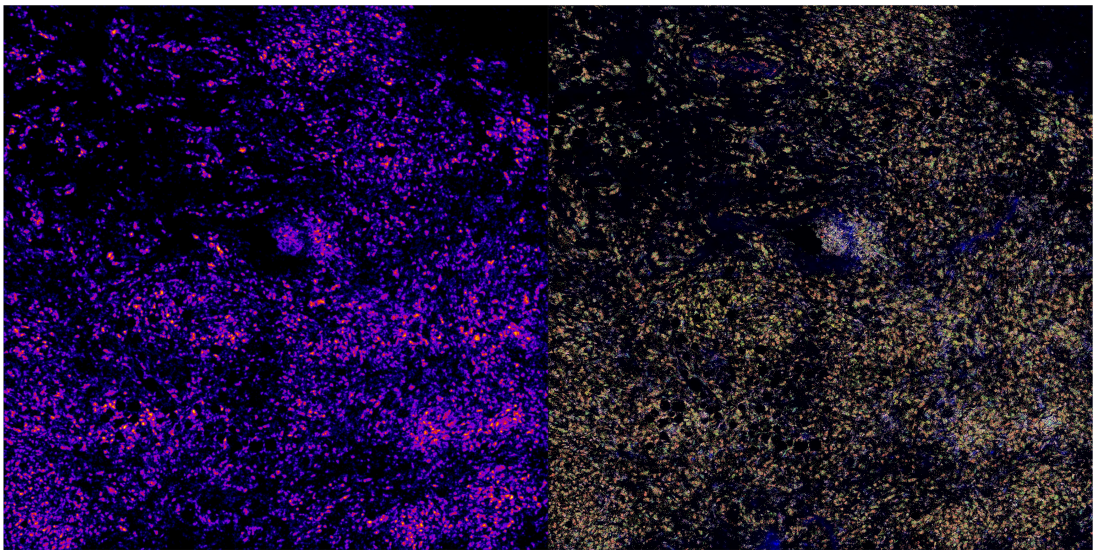
In D1-1 (Fig. 4.44A), 98% of the segmented cells were assigned to a cell label and the following groups were identified (Fig. 4.44C): Hypoxic Epithelial (n=2,169), Fibroblasts (n=836), Macrophages (n=383), Endothelial (n=322), T Cells (n=276) and CD27+ (n=198). ROI D1-2 (Fig. 4.44B) contained significantly more cells (n=11,490) split into

(Fig. 4.44D): Hypoxic Epithelial (n=4,477), Fibroblasts (n=1,869), Endothelial (n=1,284), CD27+ (n=1,229), HLA-DRB1+ CD68+ CD4+ (n=1,178), T Cells (n=911), HER3+ Epithelial (n=421), KRT19+ Epithelial (n=67), KRT7+ Epithelial (n=27) and MYC+ (n=27). Cluster CD27+ was scattered across both ROIs. It consists of cells highly expressing *CD27*, a co-stimulatory molecule constitutively expressed on T cell and B cell subsets. Both samples contained a large population of cells highly expressing hypoxia-related genes.

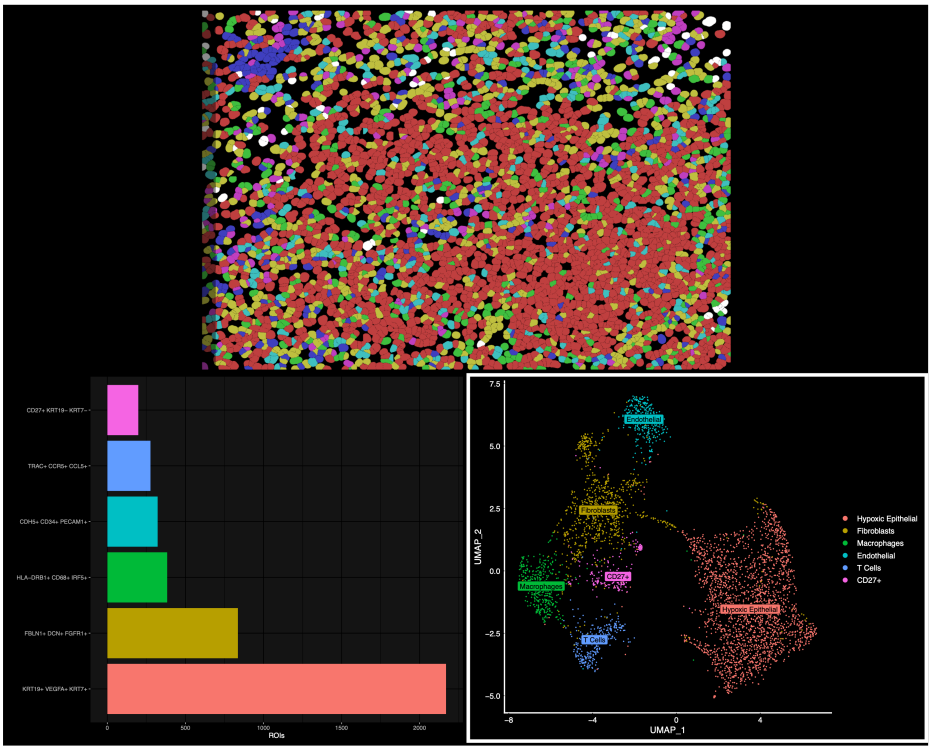
A.



B.



C.



D.

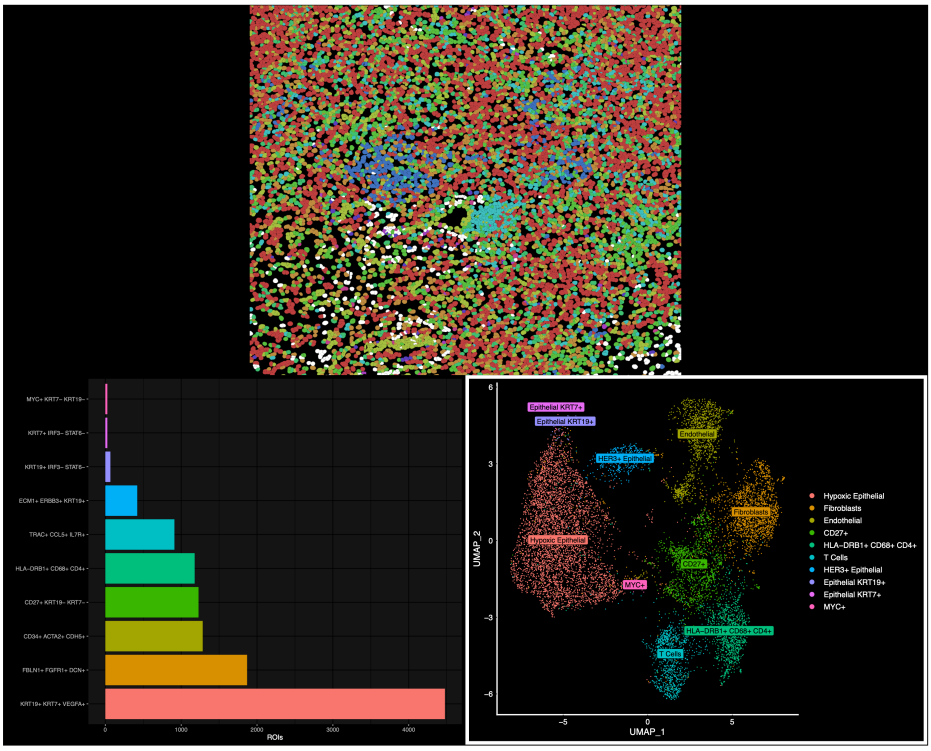


Figure 4.44: Spatial Organisation of Cell Populations for Sample TRN3. Density map of transcript abundance on the left/top and overlay of DAPI staining and RNA spots on the right/bottom for **A.** D1-1. **B.** D1-2. Tissue composition based on transcriptional profile projected onto segmented DAPI images for **C.** D1-1. **D.** D1-2.

TRN4

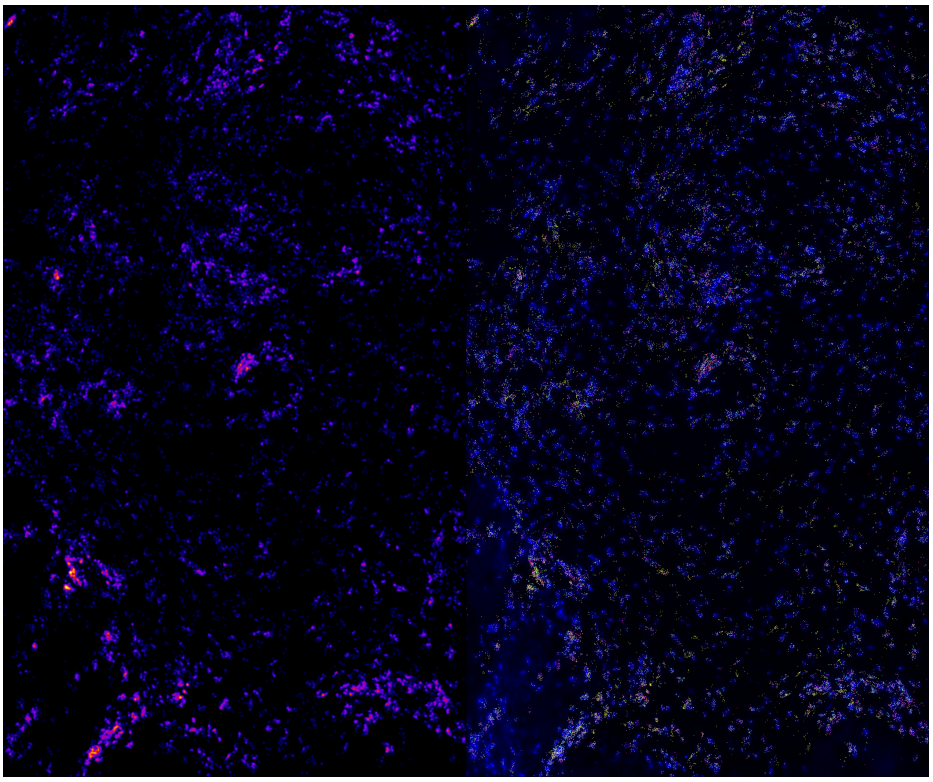
Patient sample TRN4 was split into four ROIs in well D2 (Table 4.10). Compared to other samples, a greater number of segmented cells were marked by only one differentially expressed gene. As a result, more stringent quality control thresholds were required leading to a decrease in the number of labelled cells (cf. Section 4.3.9).

Table 4.10: Molecular Cartography Spatial Metrics for TRN4.

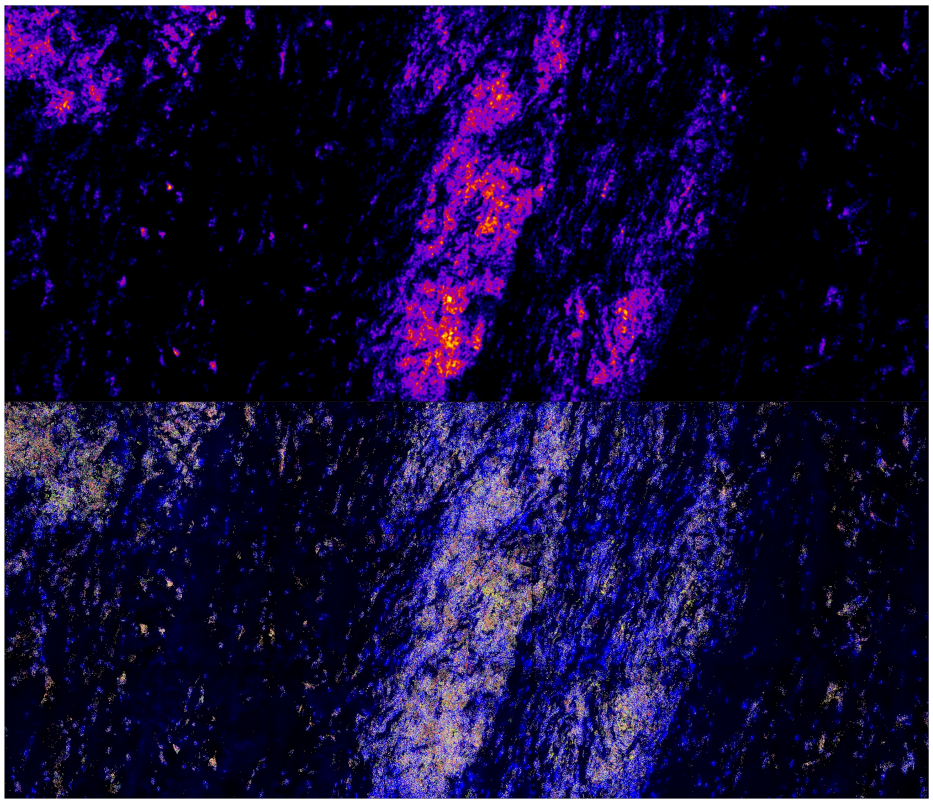
ROI	Detected Genes	Total Counts	Segmented Cells	Genes Per Cell	Signals Per Cell
D2-1	91	29,510	3,974	$\mu = 4.06$	$\mu = 6.86$
D2-2	99	267,190	8,973	$\mu = 10.28$	$\mu = 27.64$
D2-3	94	82,339	8,092	$\mu = 4.51$	$\mu = 9.42$
D2-4	91	28,533	4,754	$\mu = 3.08$	$\mu = 5.48$

Of the 2,250 cells labelled in ROI D2-1 (Fig. 4.45A) 984 were defined as FBLN1+ FGFR1+ CD68-, 628 CD68+ IRF5+ IL6R+, 562 TRAC+ CCL5+ CD27+ and 76 CD34+ CDH5+ IL6+ (Fig. 4.45E). In ROI D2-2 (Fig. 4.45B), 2,756 were marked as TRAC+ CCL5+ CD44+, 1,479 as CD68+ TRAC- HLA-DRB1+, 993 FBLN1+ DCN+ FGFR1+, 924 KRT7+ KRT19+ VEGFA+ and 318 CD34+ CDH5+ HBEGF+, for a total of 6,470 (Fig. 4.45F). As depicted in Figure 4.45G, 3,592 cells in ROI D2-3 (Fig. 4.45C) were categorised as follows: TRAC+ KRT7- VEGFA- (n=1,838), CD68+ HLA-DRB1+ CD4+ (n=702), FBLN1+ FGFR1+ DCN+ (n=492), KRT7+ KRT19+ ERBB3+ (n=466) and CD34+ CDH5+ NDRG1+ (n=94). In ROI D2-4 (Fig. 4.45D), 1,943 cells were divided between FBLN1+ FGFR1+ CD68- (n=780), CD68+ IRF5+ IL6R+ (n=571), CCL5+ TRAC+ CD27+ (n=474) CD34+ CDH5+ DDIT4+ (n=80) and ACTA2+ MYC+ FBLN1- (n=38).

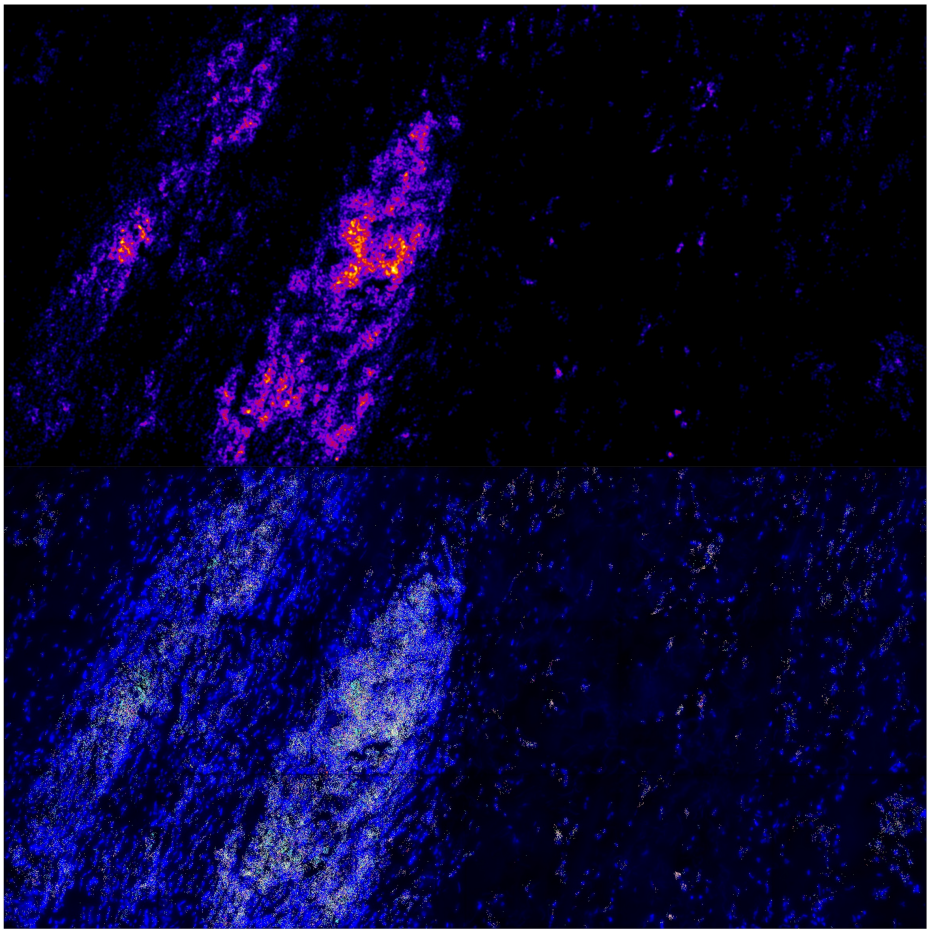
A.



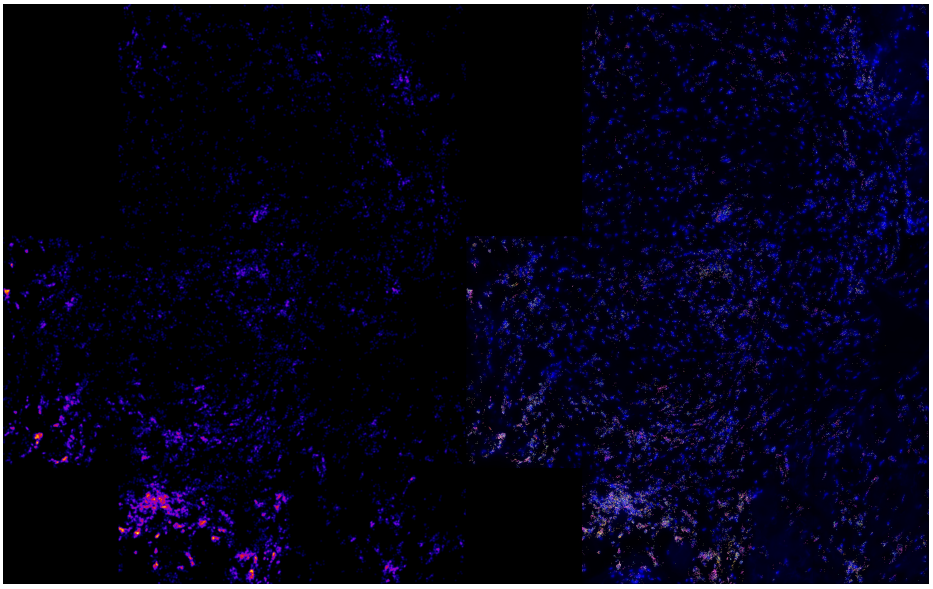
B.



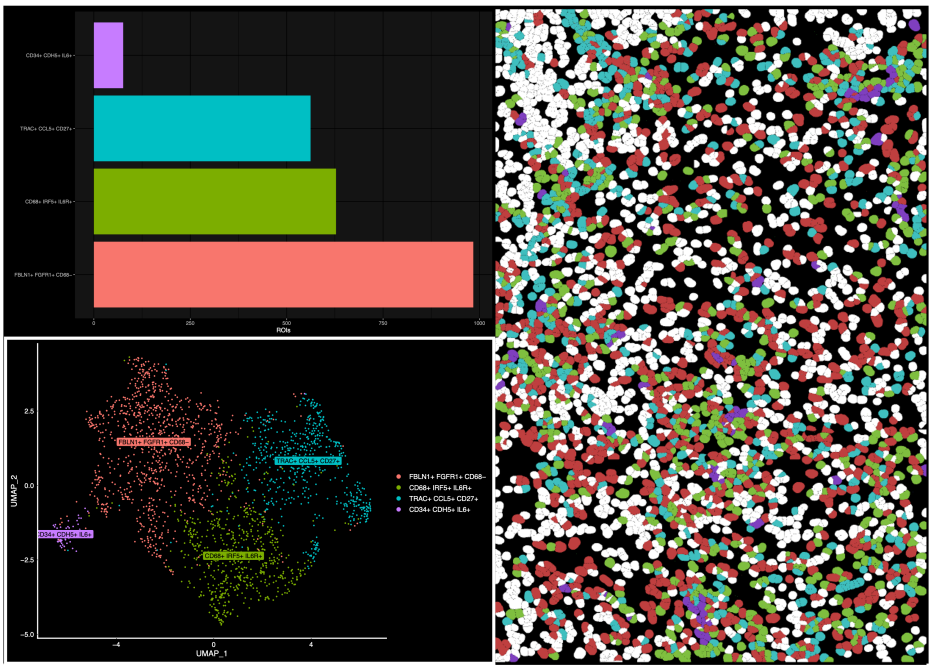
C.



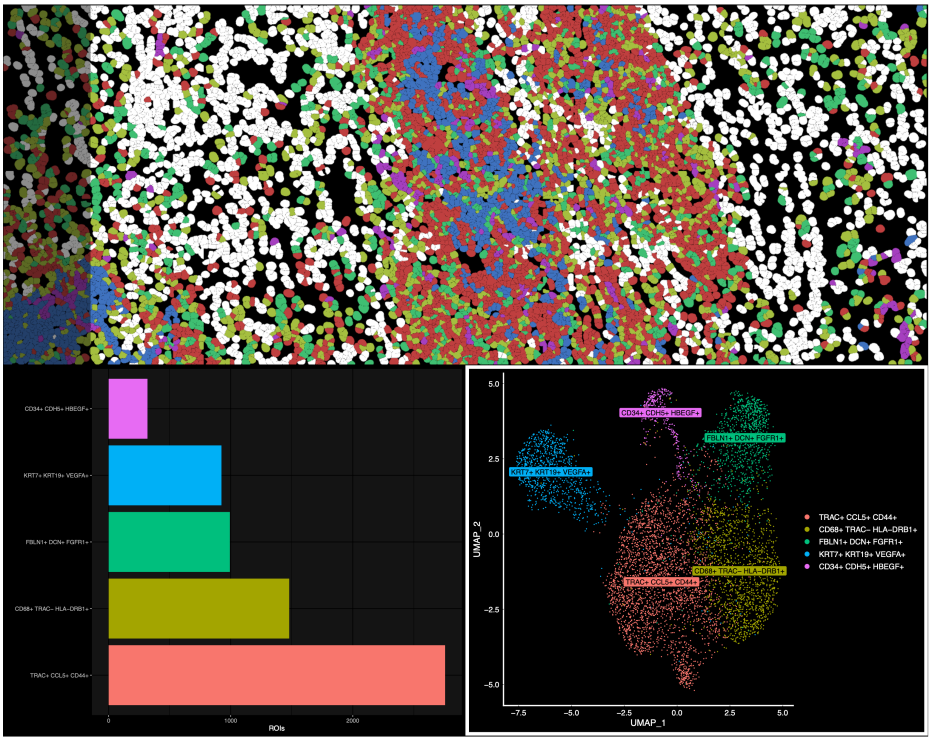
D.



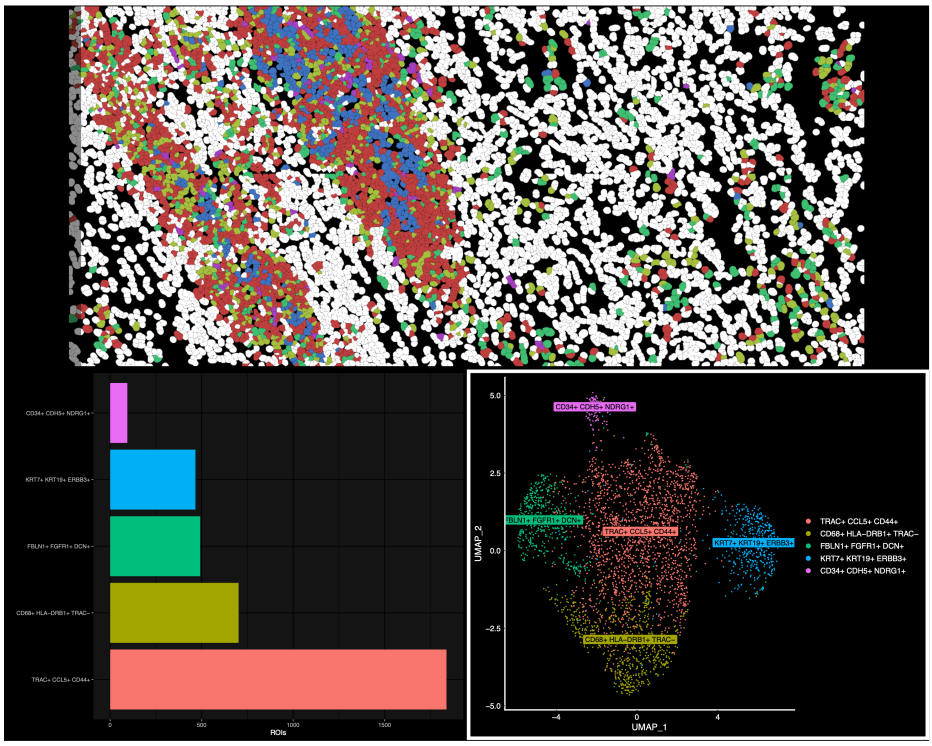
E.


CD34+ CDH5+ IL6+
TRAC+ CCL5+ CD27+
CD68+ IRF5+ IL6R+
FBLN1+ FGFR1+ CD68-
ROIs
UMAP_2
UMAP_1
FBLN1+ FGFR1+ CD68-
TRAC+ CCL5+ CD27+
CD34+ CDH5+ IL6+
CD68+ IRF5+ IL6R+
FBLN1+ FGFR1+ CD68-
CD68+ IRF5+ IL6R+
TRAC+ CCL5+ CD27+
CD34+ CDH5+ IL6+

F.


CD34+ CDH5+ HBEGF+
KRT7+ KRT19+ VEGFA+
FBLN1+ DCN+ FGFR1+
CD68+ TRAC- HLA-DRB1+
TRAC+ CCL5+ CD44+
ROIs
UMAP_2
UMAP_1
CD34+ CDH5+ HBEGF+
FBLN1+ DCN+ FGFR1+
KRT7+ KRT19+ VEGFA+
CD68+ TRAC- HLA-DRB1+
TRAC+ CCL5+ CD44+
TRAC+ CCL5+ CD44+
CD68+ TRAC- HLA-DRB1+
FBLN1+ DCN+ FGFR1+
KRT7+ KRT19+ VEGFA+
CD34+ CDH5+ HBEGF+

G.



H.

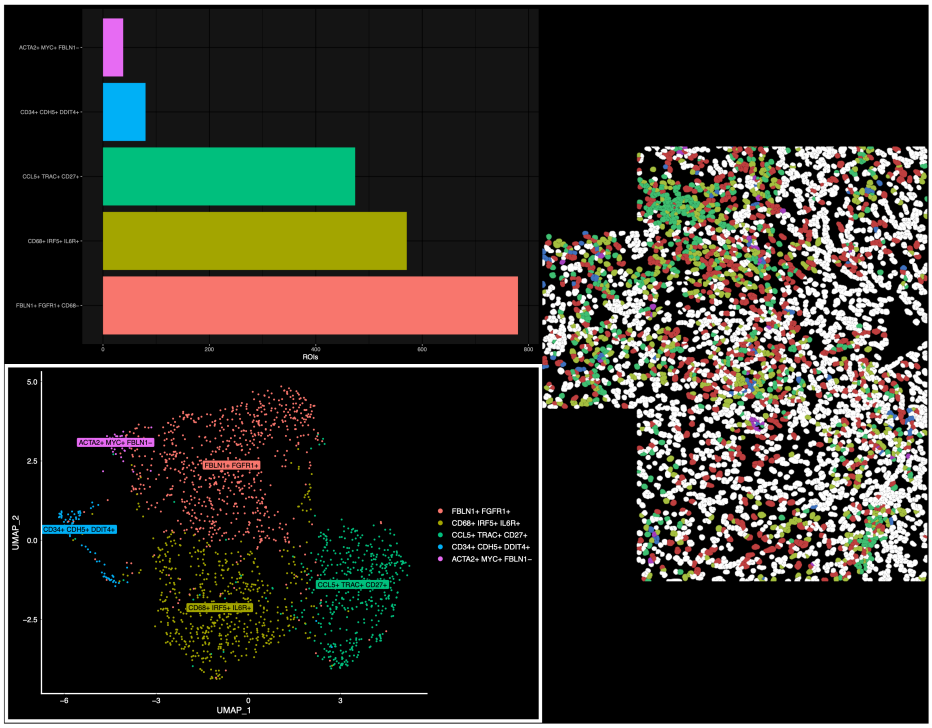


Figure 4.45: Spatial Organisation of Cell Populations for Sample TRN4.
Density map of transcript abundance on the left/top and overlay of DAPI staining and RNA spots on the right/bottom for **A.** D2-1. **B.** D2-2. **C.** D2-3. **D.** D2-4. Tissue composition based on transcriptional profile projected onto segmented DAPI images for **E.** D2-1. **F.** D2-2. **G.** D2-3. **H.** D2-4.

4.5 Discussion

A comprehensive analysis of tumour-infiltrating lymphocytes in triple-negative breast cancer was conducted in this chapter. The combination of scRNA-seq, scTCR-seq and ST enabled the dissection of gene expression and TCR sequences at single-cell resolution within the spatial context of the tumour microenvironment. In order to explore T cell heterogeneity and transcriptional variations, T cells were isolated from a larger population of cells that included fibroblasts (*ACTA2*, *COL1A1*, *PDGFRA*), endothelial (*CD34*, *PECAM1*, *ADGRL4*, *VWF*), epithelial (*KRT6*, *EPCAM*, *MUC1*) and other immune cells.

Enzymatic approaches used to dissociate tumour tissue prior to scRNA-seq can substantially influence the viability of the resulting cell suspension. Denisenko et al. (2020) identified protocol-specific biases that can have important effects on recovered cell type composition or induce transcriptional changes due to cellular stress. For instance, whilst cold temperatures were found to induce specific stress-response-related genes, major cell populations were absent from warm-dissociated samples. In the case of sample BRO18, treated with either of two available treatments, the proportion of recovered major cell populations and corresponding gene expression profiles did not differ significantly. Another paired comparison was made between 10X Genomics single-cell 3' and 5' assays. Both assays had similar results. However, the gene *ALDOA* was missing from the dataset that used the 3' sequencing library. Since the assays capture different ends of the polyadenylated transcript, it is hypothesised that this may be a result of alternative splicing occurring at the 3'-end of the *ALDOA* transcript. A more comprehensive study with additional paired samples was not possible because of the pandemic but would be required to better understand the effect of 3' and 5' kit selection on gene expression patterns.

Analysis of the immune repertoire of patients BRO20 and BRO23 revealed high diversity but only some clonal expansion. A recent study by Valpione et al. (2021) found high TIL TCR diversity at baseline to be associated with better prognosis across several cancers including breast cancer. On the other hand, their findings show that patients with higher clonality benefited from greater PD-1 blockade immunotherapy response. The literature

was also searched for sequences matching those of the most expanded TILs, defined as the proportion of T cells with identical CDR3 and VDJC genes greater than 1%. Though it was not possible to predict antigen specificity for those clones, the preferential usage of *TRBV19*, *TRBV20-1* and *TRAV12-1/2/3* had been reported in other disease conditions. A bias for *TRAV12-2*, *TRAV19*, *TRBV19* and *TRBV20-1* had also been identified in melanoma antigen-specific T cell repertoires (Simon et al., 2018). Skewed clonal proliferation can act as immunological signatures pointing to a potential parallel between both diseases that warrants further investigation, especially considering the growing evidence that breast cancer and melanoma share common molecular characteristics. For instance, overexpression of melanoma-associated chondroitin sulfate proteoglycan 4 (CSPG4) has been identified as a therapeutic target in TNBC (Ilieva et al., 2018).

Clustering of the isolated T cell population resulted in the detection of twelve T cell subgroups with characteristic transcriptional profiles. These were further examined using the reference-based methods from Chapter 3 which revealed that each group is likely to contain a mixture of T cells defined by shared transcriptional features rather than differentiation states. This result holds at varying levels of cluster resolution. Consistent with the findings from Chapter 3, boundaries between subsets get blurred when clustering scRNA-seq data. Nevertheless, one subgroup expressed a common core set of regulatory T cell genes including *FOXP3*, *TNFRSF18*, *TIGIT* and *CTLA4*. A literature search led to the finding of a recent paper by the Bowness Group (University of Oxford) in which they investigated FACS-sorted T regulatory cells from the PBMCs and synovial fluid of ankylosing spondylitis patients (Simone et al., 2021). Interestingly, they independently used a very similar approach to refine the T cell landscape and had identified overlapping populations based on the same top differentially expressed genes, mainly Cycling, IFN Signature, ANXA1+, IL7R KLRB1+ and FOXP3+ TReg. Since their dataset consists of experimentally isolated T regulatory cells, this is further evidence that lineage markers are not the primary defining feature. These results support the rationale behind the need to not solely rely on conventional marker-based clustering methods when refining T cell populations, and highlight the benefits of combining techniques and references to define closely related cell types. In the literature, T cell groups are broadly defined based

on population-averaged expression of canonical markers. In this thesis, T cell subgroups were labelled based on their transcriptional features instead of associating them to classic T cell subsets defined by a small set of cell surface markers. This unbiased approach is further supported by the findings outlined in X. Wang et al. (2022). When testing the relationship between scRNA-seq clustered T cell subgroups and cell surface marker-defined subtypes, they found a high level of overlap in the classification of T cell subsets. For instance, they show how the composition of a CD8 CTL group defined at the cluster level is complex and will commonly contain multiple T cell subsets, including MAIT and NKTs. This holds for CD4 T cells subtypes which exhibited similar expression patterns resulting in homogeneous cell populations with similar cell states.

The largest T cell group in the TNBC data was defined as IL7R+ KLRB1+ expressing high level of ribosomal content identified in Ricciardi et al. (2018) to be associated with CD4 naive and central memory, consistent with their expression of *CCR7*. This group had an overall pro-inflammatory $CD4^+$ Th17-like profile and interestingly also highly expressed *TPT1*, in line with the high level of ribosomal protein genes (Bommer & Telerman, 2020). A recent study found that high TPT1 in tumour cells correlates with poor prognosis, decreased T cell infiltration and resistance to T cell-mediated therapy (H.-J. Lee et al., 2022). In T cells specifically, *TPT1* was found to be one of the 10 top genes upregulated in HIV $CD4^+$ T cells in an article by Nunnari et al. (2015), HIV infection being characterised by T cell dysfunction. Searching through the literature it was found that knowledge on the implication of TPT1 overexpression in tumour-infiltrating T cells is largely absent and warrants further investigation. Canonical CTL was the second largest group of T cells, followed by the FOXP3+ TReg cluster. The latter consists mainly of Tregs which are known to present an anti-inflammatory barrier that correlates with poor prognosis in solid tumours. MTs High contained high levels of metallothioneins genes which encode for cysteine-rich proteins that sequester heavy metals and are increasingly believed to orchestrate inflammatory responses. They were found by Singer et al. (2016) to promote T cell dysfunction, however, conflicting studies suggest they promote T cell survival and hyperproliferation (Vignesh & Deepe, 2017). This group expressed granzyme genes and *RUNX3*, a regulator of cytotoxic $CD8^+$ effector T cells. Principal mediators of

type I IFN induction are interferon regulatory factors (IRFs), a family of transcription factors. IRF-binding sites in the promoters of IFN genes have varying affinities through which they control expression of interferon-stimulated genes (ISGs). Largely studied in the context of viral infections, IFNs have also displayed a wide range of anti-tumour properties prompting the use of IFN-based treatments like STING agonists. IFNs are produced by and act on a broad range of cells in the TME in response to the activation of pattern recognition receptors (PRRs), including toll-like receptors (TLRs), retinoic acid-inducible gene I (RIG-I)-like receptors (RLRs), NOD-like receptors (NLRs) and cyclic guanosine monophosphate-adenosine monophosphate synthase (cGAS). There is, however, accumulating evidence that implicates chronic and persistent activation of type I IFN signalling in tumourigenesis and treatment resistance, suggesting a dichotomous role. Since little is known about specific effects of inherent IFN signalling pathway activation on T cell states, survival and activity, the cluster identified as IFN Signature cluster was further analysed. Based on its transcription profile, this group is both responding to type I IFN exposure and promoting their production, as highlighted by the upregulation of genes from the JAK-STAT pathway (*STAT1*, *STAT2* and *MX1*), genes from the RIG-I pathway and interferon-sensitive response element containing genes like *OAS1* that respond to dsRNA. There was no sign of cGAS-STING pathway activation. On the other hand, the $CD4^+$ $CXCL13^{High}$ cluster which expressed the highest levels of *KLRB1* was found to differentially express *CTLA4*, *PDCD1* (PD-1), *TIGIT*, *DUSP4* and *TMEM173* (STING). *TMEM173* was also found to be one of the top upregulated gene in the FOXP3+ Treg group, along with *JUNB* and *CTLA4*. Overall, the cGAS-STING pathway was mostly upregulated in $CD4^+$ T cells with most significant *cGAS* expression in $Ki67^{high}$ T cells. A recent study found that activation of STING signalling in CD4 T cells enhances differentiation into Th1 and Th9 and anti-tumour functions (Benoit-Lizon et al., 2022). As mentioned previously, however, the effect of T cell-intrinsic STING activation has yet to be understood with studies pointing to an apoptosis-inducing effect in T cells (Gulen et al., 2017). An interesting finding that warrants further investigation was how different nucleic acid sensing pathways are upregulated in different T cell subsets. Whilst DNA sensing cGAS-STING was most active in $CD4^+$ T cell clusters, single-stranded RNA sensing RIG-I was upregulated

in IFN-responsive T cells and double-stranded RNA sensing TLR3 was most active in $CD8^+$ T cells. No strong correlation was identified between cGAS-STING pathway components and T cell exhaustion markers, albeit *TMEM173* having a slightly positive correlation with *TIGIT* and *CTLA4*. It is important to note that the correlation was weak with other genes expected to correlate together based on the underlying biology, and could be a reflection of the high drop-out rate of the method. T cells in the TME span a spectrum of functional states. Dysfunctional T cells that arise are characterised by reduced proliferative capacity and effector functions. They tend to upregulate expression of immune checkpoint molecules such as programmed cell death protein 1 (PD-1) and cytotoxic T lymphocyte antigen 4 (CTLA-4) which have been the focus of gene signatures defining this state, as seen when searching through the Molecular Signatures Database (MSigDB). However, PD-1 and CTLA-4 are also upregulated in response to T cell activation in order to limit inflammation and prevent autoimmune tissue damage. Furthermore, the transcriptional profile of T cell dysfunction, activation and effector function is highly intertwined, making it challenging to discern between them. As such, in order to accurately evaluate the level of T cell dysfunction and activation the literature was searched for a signature that specifically decouples these two states. Suitable mouse-derived activation and dysfunction gene modules tested on human melanoma $CD8^+$ TILs were found in Singer et al. (2016). The average expression levels of each program was calculated for both individual cells and T cell clusters. It was found that for the most part, T cells find themselves along a transcriptional gradient of activation and exhaustion. A mixture of scores was found across clusters. At the single cell level, signatures were only relevant at extreme values (highly exhausted vs. highly activated). Better gene signatures of dysfunction uncoupled from activation would be required to gain a deeper molecular understanding of T cell state transitions.

According to the literature, MIF is a pro-inflammatory cytokine that is abundantly released by T cells. It is commonly overexpressed in solid tumours including breast cancer and is associated with poor prognosis. A recent study by Charan et al. (2020) explored the role of MIF in TNBC progression and metastasis. They found that MIF/CD74 interaction promotes tumourigenesis through the activation of cell survival pathways such as

phosphoinositide-3-kinase (PI3K) signalling pathway and identified MIF as a promising target against aggressive TNBC. Interfering with MIF/CD74 signalling, could further decrease expression of immunosuppressive factors by macrophages (Figueiredo et al., 2018) and decrease the prevalence of myeloid-derived suppressor cells (MDSCs) suppressing immune responses (Simpson et al., 2012). Predicted interaction results also showed that a certain population of *CCR8* expressing *FOXP3*$^{+}$ T regulatory cells was present. Barsheshet et al. (2017) demonstrated that CCR8^{+} Treg are drivers of immunosuppression. CCR8 is upregulated in tumour-resident T cells and and has been suggested as a therapeutic target in breast cancer (Plitas et al., 2016). Similarly, TNFSF9 on cells from the Canonical CTL cluster was predicted to interact myeloid cells and neutrophils. A study by Y. Li et al. (2020) found that TILs expressing TNFSF9 share an effector and exhausted phenotype and were associated with poor clinical outcome in renal cell carcinoma. Their role in breast cancer warrants further investigation. Myeloid cells were found to promote an anti-inflammatory state through expression of *IL-10*. To further promote immunosuppression, the data clearly shows that myeloid cells are likely to be inhibiting T cell activation through MIF-CD74. Taken together, these findings support an important role for myeloid cells establishing an immunosuppressive TME.

Spatial transcriptomics enabled the generation of new observations through exploratory analysis. Combining transcriptome-wide and targeted technologies provided a unique opportunity to assess spatial patterns at different levels of resolution and explore the scRNA-seq data in the context of the TNBC microenvironment. In turn, the scRNA-seq data offered a powerful reference to infer and map cell types to the spatial data.

Analysis of the Slide-seq pucks revealed a qualitatively similar proportion of cell types to the H&E images taken prior to running the experiment. However, fibroblasts were not successfully detected. Sample BRO18 was found to have a high level of lymphocytic infiltration both through H&E staining and using the array-based method. The use of matched patient-specific scRNA-seq helped account for interpatient variability. A limitation of Slide-seq that influences the resulting cell populations is that single beads do not represent pure single cells. As a result of mRNA being captured at 10µm resolutions, the

gene expression profile associated to each bead is an average of the transcripts successfully captured in the area which can include multiple cells. Furthermore, though Slide-seq is said to have an mRNA capture efficiency that approaches that of scRNA-seq, this was not found to be the case. UMI counts per bead highly varied and the median was around 18 times lower on average than scRNA-seq. The increase noise and drop-out rate associated with the Slide-seq analysis made it challenging to accurately deconvolve cell types. A recent method to delineate spatial organisations was developed by Miller et al. (2022) and presents a possible solution to this problem. The elevated expression of *MALAT1* in puck two could indicate a high number of dead or dying cells. However, this long non-coding RNA is also commonly found to be expressed at high levels in breast cancer tissue and is known to promote inflammatory factors (Yue et al., 2021). Collagen-dense regions were identified in pucks two and three. Multiple studies have shown how increased deposition of collagen translates into high tumour tissue stiffness which is a driver of tumourigenesis and metastasis (Provenzano et al., 2008; S. Xu et al., 2019). In the puck which had the highest proportion of predicted T cells, expression of ISG genes, mainly *ISG15*, *MX1* and *IRF7* was specifically detected in beads labelled as Epithelial Cells. Beads were preferentially mapped to CD4 T cell subtypes, however, Slide-seq was not found to be the ideal platform to study T cell populations. This may in part be due to their small size, naive T cells have a diameter of around 5-7 µm (Tasnim et al., 2018) whilst epithelial cells range from 8-21 µm (Brown & Bron, 1987). Slide-seq beads are 10 µm in size, thus they are unlike to capture the transcriptome of pure T cells.

Molecular Cartography provided sub-cellular resolution of selected transcripts and enabled the computational reconstruction of cells using imaging-based methods. It provided an ideal way to examine IFN signalling pathways but the limited gene panel made it difficult to accurately identify cell types. Colocalisation analysis of the probes revealed biologically informative gene pairs which were further assessed both at the transcript and cellular level. Just as correlation does not imply causation, colocalisation does not imply biological relationship or same cell membership. However, it can be hypothesised that genes with a high colocalisation score are more likely to be expressed in the same cells or found in cells in close proximity to each other. Pair-wise scores were calculated for every

targeted transcript to identify genes significantly enriched or depleted around it.

Spatial colocalisation scores were found to recapitulate biological relationships with, for example, endothelial markers clustering together and the colocalisation of type I IFN pathway genes. Components of the cGAS-STING pathway including *TMEM173* and *cGAS* were found to be mainly expressed by cells expressing markers of endothelial cells including *CDH5*, *CD34* and *PECAM-1*, suggesting that the cGAS-STING pathway was the most active in this cell type. The ligand *PD-L1* (CD274) was found to highly colocalise with *TRAF3*, *STAT1*, *cGAS* and *IFIT1*, thereby providing further evidence of the relationship between IFN signalling and resistance to immune checkpoint blockade therapy. Experiments by Diaz et al. (2017) showed a strong correlation between PD-L1 and enrichment of interferon signatures, mainly the upregulation of STAT1 and IRF1. Whilst *IRF1* was not in the probe set, *STAT1* was indeed found to strongly colocalise with *PD-L1*. There was evidence linking *PD-L1* expression to the *JAK2*/*STAT1* axis. Similarly, *MX1* strongly colocalised with the immunosuppressive molecule *IDO1*. A strong link between immunosuppressive molecules *IDO1*, *PD-L1* and *LGALS9*, and type I IFN genes including *IFNAR2*, *MX1*, *STAT1* and *IFIT1* was identified. This finding provides further evidence of the relationship between type I IFN signalling and upregulation of the immunomodulatory molecule *IDO1*, as found in B. Chen et al. (2020). As described in Section 2.3, there is an ongoing debate whether cGAS-STING agonists can have immune suppressive functions, with Lemos et al. (2016) and M. Petrovic et al. (2021) showing that chronic stimulation of the pathway may lead to the upregulation of PD-L1 and IDO1 in cancer cells. The resulting colocalisation of *cGAS* and *PD-L1* supports this view. Moreover, these genes also strongly colocalised in cells expressing *ICAM1*, a CAR T cell target candidate for TNBC. For instance, H. Wei et al. (2020) generated ICAM1-specific CAR T cells and showed their high therapeutic potential against ICAM1-positive TNBC tumours. There was also evidence of *cGAS* expression being associated with *PD-1* transcripts in the proximity. The same spatial relationship was not detected with *TMEM173*. This is consistent with the results from a study by Ozawa et al. (2021) showing that tumour expression of cGAS, not STING, negatively effects PD-1/PD-L1 inhibitors efficacy. On the other hand, there was no sign of *cGAS* being associated to *CTLA4* nor *LGALS9* expression. As the synergistic combination

of cGAS-STING agonists with CAR T cells is actively being proposed by those in the field, these results represent an important finding which warrants further investigation. In the context of ICAM1-specific CAR T cells for TNBC, the suitability of a *cGAS* adjuvant should be considered in light of its effect on PD-L1 expression in the targeted cells or the use of CAR T cells designed to be resistant to PD-L1-mediated immunosuppression should be considered. Another CAR target candidate in the panel of probes was *ANTXR1* (*TEM8*). This gene was found to be predominantly expressed in fibroblasts and endothelial cells. A research article by K. Petrovic et al. (2019) highlights the potential on-target off-tumour risks with CARs targeting this marker, which will be the subject of Chapter 5. A recent study by Y. Wang et al. (2021) demonstrated how ZMYND8 helps breast cancer cells evade CTL surveillance through the inactivation of the cGAS-STING signalling pathway having identified a negative correlation with ISGs and T cell markers. The work presented therein further supports their findings with *ZMYND8* strongly colocalising with other immunosuppressive genes, *RNASEH2A* and *PRKDC*. In addition, *ZMYND8* was found to be enriched along with hypoxia and epithelial markers whilst being slightly anticolocalised with CTL markers. This suggests a potential link between low oxygen levels, *ZMYND8* upregulation in epithelial cells and subsequent suppression of anti-tumour responses. Zhao et al. (2022) identified a relationship between hypoxia-induced RNASEH2A and transcript inhibition of cGAS-STING inhibition. Though in the data there was no strong evidence of the link between hypoxia and *RNASEH2A*, the transcript was found to be associated with glucose metabolism gene *PDHA1* and not cGAS-STING genes. Moreover, in image A1-3, the epithelial cells lining the mammary gland ducts highly expressed *TMEM173*, *TBK1*, *IRF3* but not cGAS which could be due to the presence of *PRKDC* (DNA-PK) and *ZMYND8*. The crosstalk between cGAS and these two negative regulators needs to be further investigated. There is evidence of the potential presence of double-negative MAIT cells with the strong colocalisation in cells of *KLRB1*, *IL7R*, *CCL5*, *GXZK* and not *CD8A* nor *CD4*. However, cells expressing this set of genes could also represent natural killer cells since *KLRB1* is also associated with *CD69*, only slightly with *TRAC* and not with *CD3D*. *DNMT1* which colocalises with *MYC* was found to be highly anticolocalised with *cGAS*. The literature was searched for a potential underlying mechanism that could explain

this finding. In a recent study by S. Y. Wu et al. (2021), it was found that MYC induces immune evasion of TNBC cells through the MYC-DNMT1-STING axis. MYC positively regulates DNMT1 which inhibits the cGAS-STING pathways and thus suppresses type I IFN signalling. Another study by Beck et al. (2021) in keratinocytes reported how DNMT1 knockout activates signalling through cGAS-STING triggering inflammation in response to DNA release into the cytosol, further highlighting their negative correlation which is in line with the anticolocalisation scores.

The CARTANA experiments targeted a different panel of genes and transcripts, and were also found to highly colocalise based on their corresponding biological function. Understandably, the Cell Function and Hypoxia-Epithelial-Endothelial groups were found to cluster together. A hallmark of malignant cells is the expression of genes promoting cell cycle progression, genomic alterations and cell survival under hypoxic conditions. Nevertheless, the hypoxia gene *CA9* strongly colocalised with pro-apoptotic molecule *BNIP3*, glycolysis-promoting gene *PGK1*, lactate metabolism *LDHA*, and other hypoxia-linked genes like *ENO1*, *TPI1*, *MUC1*, *VEGFA*, *NDRG1* and *P4HA1*. There was a very strong positive spatial association between two markers found to correlate positively or negatively with immune infiltration, *FCGR1A* and *GPI*, respectively (li Xu & Guo, 2020; Han et al., 2021). *FCGR1A* was anticolocalised with *PSMA7*, a prognostic marker associated with poor prognosis at high levels (Xia et al., 2019). Visual inspection of the images in conjunction with the colocalisation scores confirmed the presence of T cells. Sets of T cell markers colocalised together but were found scattered across the tissue and a significantly greater level of CD4 T cells than CD8 T cells was identified. Interestingly, the spatial association between *CD4* and *CD14* could indicate the presence of T cell-macrophage complexes, as previously described in Burel et al. (2019), and consistent with the predicted interactions between myeloid cells and T cells in the scRNA-seq data. Furthermore, the high level of *FOXP3* transcripts was indicative of high infiltration by T regulatory cells which is strongly associated with poor survival as it suppresses anti-tumour immunity (C. Li et al., 2020).

A clear limitation in this study is the limited number of patient samples which was a di-

rect result of the COVID-19 pandemic. Around 15 TNBC patients were to take part in this study. All samples would have been sent for scRNA-seq, scTCR-seq, scBCR-seq and ST. Furthermore, the comparison of the 10x Genomics 3' and 5' assay experiment would have been done in triplicate. Separating signal from noise is critical before any firm conclusions can be drawn. More clinical samples would need to be collected to account for inter-patient variability and technical considerations. In addition, the loss of the patient PBMCs meant that T cells across the local and systemic environment could not be compared as originally intended. This led to a change in direction for the subsequent part of this project (Appendix B). Furthermore, normal healthy breast tissue was to be collected and ideally would have been used as a reference to properly investigate defects in T cells and IFN signalling arising within the TME. A large-scale scRNA-seq atlas of human breast spanning normal tissue was recently published by Y. Chen et al. (2022) and could be worth looking into. The severe delays in acquiring and receiving the data left very limited time to conduct a more comprehensive analysis of the spatial transcriptomics, given the nearing deadline for thesis submission. Technical limitations include the poor sensitivity and bias of transcript coverage intrinsic to scRNA-seq technology. Lowly expressed genes are more difficult to detect and dropout events are frequent. This likely explains why IFN genes were not detected at all. It would be worth sending the samples for bulk RNA-seq to get a more accurate measure of global gene expression at the cost of single-cell resolution. If IFN levels cannot be measured at the RNA transcript level, the use of proteomics methods may be warranted. As a first step, had the research not been interrupted, immunofluorescence staining using fluorescent-labelled antibodies specific to IFNs was an actively discussed option. The sensitivity of different cell populations to varying technologies should also be further investigated as it can influence the proportion of recovered cell types. For instance, a lower number than expected of epithelial cells was recovered from scRNA-seq, and other labs at Oxford had observed the same in their experiments. The different spatial transcriptomics platforms used in this chapters have their own technical limitations. Array-based methods like Slide-seq provide unbiased genome-scale coverage but suffer from a very low mRNA capture efficiency, as reported therein. In addition, cells often span multiple capture areas, and therefore single-cell resolution is

not achieved. Conversely, targeted approaches like Molecular Cartography (Resolve Sciences) and in situ sequencing (CARTANA) enabled the detection of RNA transcripts at a single-molecule resolution, revealing subcellular localisation at the cost of gene throughput. Moreover, Molecular Cartography has a high detection efficiency but restrictions are placed on the number and type of genes that can be targeted. Additional limitations in this study include the possibility of overcorrecting when using integration methods to combine different datasets. When addressing batch effects, there is always a risk of confounding the underlying biological signal. In addition, as presented in this chapter, appropriate gene signatures to accurately investigate and separate T cell activation from dysfunctional states are needed. These could be used in conjunction with trajectory inference approaches placing T cells on a continuum to better understand T cell transition towards dysfunctional states along with the dual role and influence of IFN signalling. Importantly, computational methods to process and analyse single-cell omics data are continuously improved and the lack of gold standards hinders the systematic evaluation of these methods, as was discussed in Chapter 3.

The primary aim of this chapter was to investigate the interplay between cGAS-STING signalling and T cell function. The integration of single-cell technologies including targeted and transcriptome-wide spatial transcriptomics to delve into the TNBC TME provided a powerful approach to study T cell immunity and IFN signalling pathways at varying levels of resolution. In addition, it enabled the evaluation and comparison of newly emerging spatial transcriptomics techniques. Despite the limitations listed above that made it challenging to elucidate the direct relationship between cGAS-STING signalling and T activity, results indicated that the cGAS-STING cascade is mostly activated in endothelial cells, and is associated with the expression of immune suppressive molecules known to impair T cell immunity. This warrants further examination as it has important implications for the development of therapies targeting this pathway. Due to the dichotomous role of the cGAS-STING pathway suggested by the conflicting evidence found in the literature, it would be worthwhile to investigate whether there is a bifurcation point between cGAS-STING-TBK1-IRF3 expression levels and pro-tumourigenic mechanisms. Finally, studying the possible effect of hypoxia on the spatial patterning of these transcripts

will be a crucial element of future work.

4.6 Conclusion

In this chapter, multi-omics methods were used to capture cellular profiles and their spatial organisation in breast cancer at different levels of resolution. Comprehensive single cell expression atlases for major cell types and T cell subtypes were constructed from tissue biopsies and subsequently used to navigate patient-matched spatial transcriptomics data. The integration of scRNA-seq, scTCR-seq, scBCR-seq, Slide-seq, in situ sequencing and Molecular Cartography provided a powerful approach to study immune infiltration in TNBC. Moreover, spatial transcriptomics is an emerging technique that had not yet been used to visualise type I IFN signalling pathways in tumour biopsies at scale. The cGAS-STING signalling pathway was found to be mostly transcribed in endothelial cells. In T cells, higher activity of the cGAS-STING cascade was observed in $CD4^{+}$ subpopulations with most significant cGAS expression identified in $Ki67^{high}$ T cells. Tissue-wide gene expression patterns revealed a strong link between *PD-L1*/*PD-1* and *cGAS* but not *TMEM173*. Similarly, markers of IFN signalling pathways were found to spatially associate with markers of hypoxia and immunosuppressive molecules. These findings have implications for understanding mediators of T cell dysfunction in the TME.

CHAPTER 5

Multi-Scale Modelling of Chimeric Antigen Receptor T Cell Therapy in Search for the Goldilocks T Cell Design

5.1 Overview

T cells can be genetically engineered to express a chimeric antigen receptor (CAR) that mediates MHC-unrestricted antigen recognition. CAR T cells have produced encouraging clinical outcomes in the context of liquid tumours but succumb to the numerous challenges posed by the immunosuppressive microenvironment of solid tumours explored in Chapter 4. With CAR T cells targeting tumour-associated antigens, self-antigens overexpressed on tumour cells but also found on healthy cells, a common side effect is on-target off-tumour toxicity. Advances in gene editing have made it possible to tweak quantitative parameters such as CAR expression level and could be leveraged to optimise therapy design. Further, compelling evidence suggests that IFN pathways could be exploited to tackle immune-inhibitory mechanisms. Models amenable to analysis can help predict the behaviour of a cellular system, thereby conveying valuable T cell engineering design information or suggesting potential drug targets and combinations. This chapter delves into CAR T cell dynamics at different levels of resolution using agent-based modelling and gene regulatory network modelling to demonstrate how the optimal level of CAR expression and strategic targeting of IFN pathways could result in reduced toxicity and higher therapy efficacy.

5.2 Introduction

Immune cells genetically engineered to express synthetic tumour-targeting receptors on their surface, called chimeric antigen receptors (CARs), have shown promising results in treating haematological malignancies but limited success against solid tumours (Hou et al., 2021). NY-ESO-1-specific CAR T cells are an attractive immunotherapy option for breast cancer patients (Stagg & Allard, 2013). A highly immunogenic target, NY-ESO-1 belongs to the cancer testis antigen (CTA) family which has restricted expression to immune privileged testicular and placental tissue in healthy individuals (Raza et al., 2020). Induction of CTA gene expression in tumour cells results from genome-wide epigenetic events affecting tightly controlled promoter-regions by altering DNA methylation and histone modifications. In the context of CAR T cell therapy, $CD4^+$ and $CD8^+$ T cells that are modified to target antigens abundantly expressed by tumour cells have complementary functions and are typically infused together. Downstream signalling cascades that are initiated by CAR ligation induce helper functions for immune activation and favourable cytotoxic effects in $CD4^+$ and $CD8^+$ T cells, respectively. The optimal CD4:CD8 ratio to synergise anti-tumour effects and their individual contributions in vivo are actively being researched. It has been well recognised that the level of $CD8^+$ tumour infiltrating lymphocytes (TILs) in many types of cancer, including breast cancer, correlates with better prognosis. Nevertheless, naturally occurring TILs and CAR-engineered T cells alike are exposed to numerous obstacles when on the tumour battleground. The complex tumour microenvironment (TME) supporting and surrounding solid tumours displays a large degree of antigen heterogeneity, is highly immunosuppressive, and acts itself as a physical barrier to cellular infiltration. Consequently, the TME hinders T cell trafficking, efficient tumour recognition and killing, and their ability to survive and proliferate. This has important implications when designing CAR T cells for their intended use in the clinic.

Quantitative parameters that correlate with in vivo safety and efficacy are being investigated to guide and optimise CAR T cell designs. High levels of CAR surface expression leading to spontaneous receptor clustering results in increased tonic signalling, a non-coordinated and sustained activation in the presence or absence of ligand. In non-

engineered endogenous T cells, transient interactions between the TCR and self peptide-MHC have been observed to drive tonic signalling. T cells exhibit low basal levels of this constitutive signal that has been found to influence the phosphorylation of proximal signalling molecules, TCRζ and Lck. Tonic signals are believed to be important for T cell survival and regulation of T cell metabolism and effector functions in both wild-type and CAR T cells. However, tonic signalling has also been found to induce early CAR T cell exhaustion, thereby reducing tumour killing and causing a disparity between in vitro cytolytic capacity and in vivo anti-tumour efficacy. Similar to chronic antigen exposure, CAR-mediated tonic signalling promotes T cell dysfunctional states leading to premature activation-induced cell death. Furthermore, target antigen expression level also strongly influences CAR-mediated response, both the release of cytolytic granules and the lysis of target cells (Anikeeva et al., 2021). CAR T cells have been found to be less sensitive than TCR T cells in the detection of target cells, and thus require a higher level of tumour associated antigens (TAAs) for the number of engaged CARs to be sufficient. Nevertheless, since TAAs are not tumour-specific and are found expressed on normal tissue, albeit at low to intermediate levels, CAR T cells raise safety concerns as they do not discriminate between normal and malignant cells, resulting in serious adverse effects known as on-target off-tumour toxicities. The right balance between the level of TAAs and CARs has not been elucidated yet but could be exploited to optimise therapy for tumours presenting with varying levels of targeted molecules and to minimise off-tumour effects.

As described in Chapter 2, CAR constructs conventionally consist of a single chain variable fragment (scFv), a transmembrane domain (TM), a spacer or hinge region linking the scFv and TM, and endodomains responsible for transmitting activation and costimulatory signals. Molecular refinements to any of these components will influence CAR T cell function. Recently developed molecular tools enabling the precise control of gene expression levels, such as microRNA silencing-mediated fine-tuners (miSFITs) developed by Dr. Fulga's group at the University of Oxford (Michaels et al., 2019), make it possible to tune endogenous or therapeutic transgenes. The miSFIT system comprises of engineered synthetic microRNA (miRNA) response elements (MREs) with varying complementarity to the miRNA associated with the transcript of interest, thus regulating the magnitude of

miRNA-mediated repression. The need to further optimise CAR constructs to enhance anti-tumour potency prompted the Fulga Lab to use miSFIT to investigate how CAR expression level influences CAR T cell function and off-target toxicity. A CAR T cell design with just the right CAR expression level could prove to be a more effective therapy.

Mathematical and computational models incorporating knowledge extracted from experimental data can help improve our understanding of biological systems and influence engineering design decisions. In oncology, both types of models are frequently used to investigate how cancer evolves and spreads, thereby providing clues on how to thwart drug resistance and optimise treatments. They are valuable tools to test hypothesis and predict dynamics of interacting cellular systems. CAR T cell therapy is a relatively recent type of cancer treatment which has shown minimal efficacy against solid tumours. Models of CAR T cell dynamics in the context of solid tumours are scarce, and a thorough search of the relevant literature yielded no related article that considers different CAR T cell designs. In an attempt to model CAR T cell killing in glioma, Sahoo et al. (2020) built upon the classical predator-prey Lotka-Volterra equations. However, their model assumes that populations are well mixed and that they interact via the law of mass action, which is not representative of the biological context of solid tumours. Moreover, their model alone cannot be used to analyse CAR T cell dynamics at multiple scales of resolutions and how these interact with each other. In this chapter, different modelling paradigms are used to study CAR T cell dynamics in the context of solid tumours. An agent-based model is created to explore the effect of CAR tuning on therapy efficacy and safety. Unknown model parameters are estimated from functional assays and scRNA-seq data that was generated in the laboratory (Fulga Lab) from Jurkat cells engineered with varying CAR expression levels and human primary $CD8^+$ T cells. The expansion of previously published gene regulatory networks (GRNs) enables the modelling of intracellular interactions to test how inhibition or enhancement of cell-intrinsic cGAS-STING signalling pathway components influences the resulting phenotype. The work presented therein will pave the way for in silico methods tailored to optimising CAR T cell designs and predicting synergistic combinations. Moreover, the research conducted in this chapter feeds into a previously published modelling framework used to test and generate new hypotheses.

5.3 Materials and Methods

5.3.1 Experimental Data

The experimental data presented in this section was generated by Andrew Ramos in the Fulga Lab (University of Oxford) for his DPhil thesis with the help of a former member, Dr. Yale Michaels. It stems from a collaborative project between the Dushek Lab and the Fulga Lab at the Sir William Dunn School of Pathology and MRC Weatherall Institute of Molecular Medicine, respectively. The experiments were severely delayed with the COVID-19 pandemic, and as a result, the data was not received until late in the DPhil.

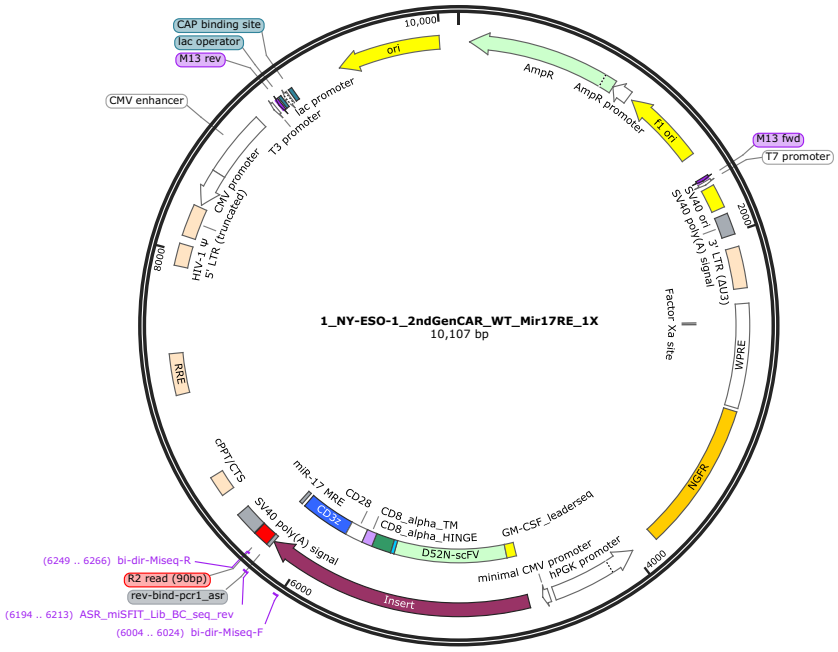


Figure 5.1: Lentiviral Vector Encoding Tunable Anti-NY-ESO-1 CAR
Circular plasmid map of NY-ESO-1_2ndGenCAR_WT_Mir17RE_1X generated using SnapGene. DNA sequence provided by Andrew Ramos (Fulga Lab).

Briefly, an MRE variant library was constructed with varying complementarity to hsa-miR-17, as done previously by Michaels et al. (2019). A NY-ESO-1 2nd generation DN-28z CAR (cf. Section 2.4) with downstream miR-17-MREs lentiviral expression vector (NY-ESO-1_2ndGenCAR_WT_Mir17RE_1X) was generated and transduced into Jurkat cells, an immortalised line of human T lymphocyte cells (Fig. 5.1). The CAR is designed

by fusing the D53N of an antibody-derived single-chain variable fragment (scFv) to the CD8α hinge and transmembrane domains coupled to the CD28 signalling chain (Maus et al., 2016). CAR-engineered Jurkat cells were co-cultured with a synthetic antigen presentation system (sAPCs) pulsed with varying concentrations of NY-ESO-1 9V. The early activation marker, NF-κB:EGFP, was measured. Primary CD8$^+$ T cells were similarly transduced and sent for scRNA-seq.

A positive correlation between average CAR expression and the following genes was identified in the scRNA-seq (Fig. 5.2): CD69, IFNγ, IL2, PRF1, GNLY, LAG3, CD28, CTLA4. As a proxy for expression level in the Jurkat cells, the median fluorescence intensity (MFI) values of CAR and NF-κB:EGFP were measured using flow cytometry. The different miSFIT variants showed varying levels of CAR repression ranging from maximum repression with 4X Perfect, resulting in lowest expression, to lowest repression, thus highest expression, with Scramble. The results from the functional assays suggest a sigmoidal relationship between activation marker expression, NF-κB:EGFP, and tumour antigen concentration, NY-ESO-1 (Fig. 5.3A). Higher CAR expression causes an upward shift of the plateau phase which translates in a higher NF-κB:EGFP expression level (EMax) being reached and a lower half maximal effective concentration (EC50) value (Fig. 5.3B).

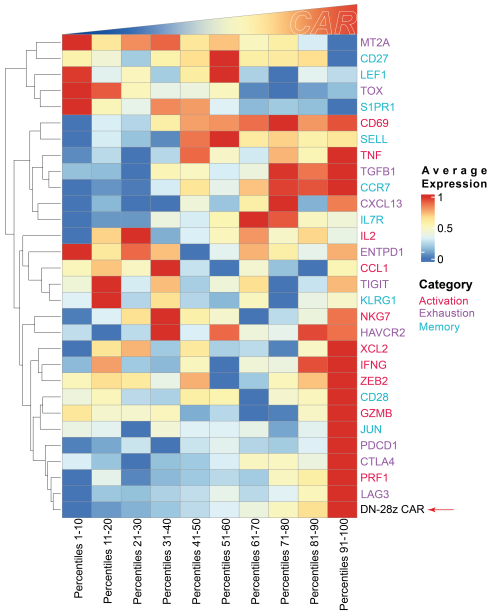
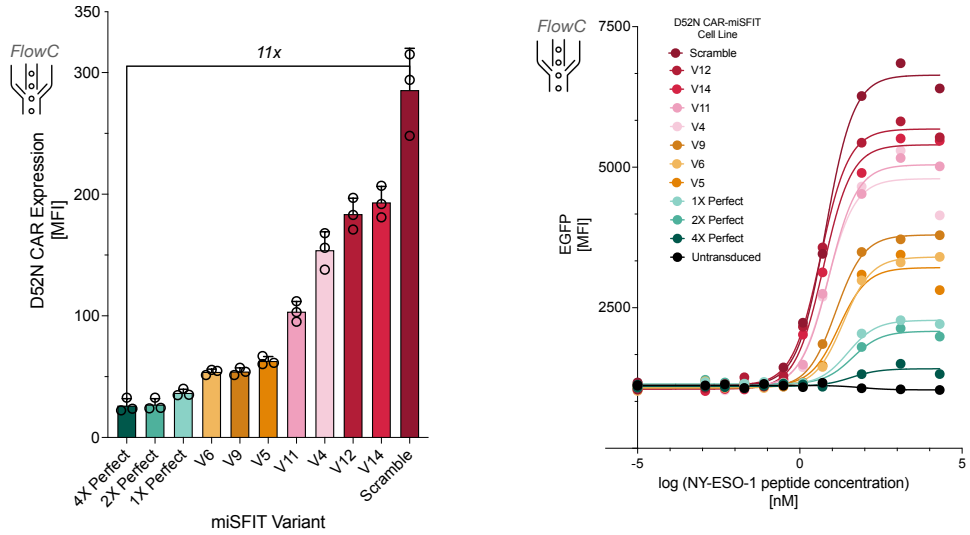


Figure 5.2: Gene Expression Heatmap of CAR Tuning.
Measured mRNA levels by scRNA-seq of primary CD8$^+$ T cells expressing varying levels of CARs. Courtesy of Andrew Ramos (Fulga Lab).

A.



B.

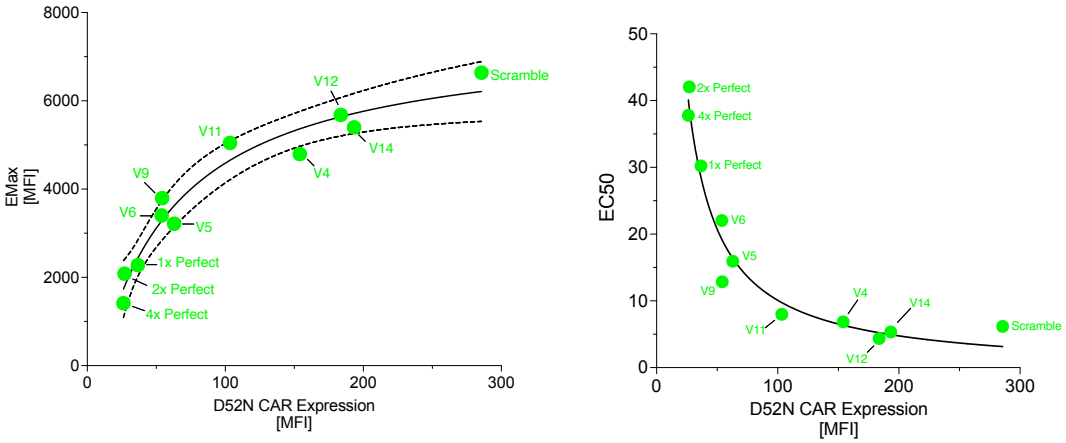


Figure 5.3: CAR Expression-Activation Relationship
Functional assay results for anti-NY-ESO-1 CAR-engineered Jurkat cells with miSFIT-tunable expression level exposed to varying concentrations of NY-ESO-1. **A.** Bar chart on the left-hand side gives CAR MFI for each miSFIT variant. The sigmoidal relationship between NFκB-EGFP reporter MFI and NY-ESO-1 peptide concentration exposure is shown on the right. EGFP was measured after 24h of co-culture. **B.** Maximum MFI and EC50 as a function of CAR expression. Courtesy of Andrew Ramos (Fulga Lab).

5.3.2 Modelling CAR T Cell-Tumour Interactions

Agent-Based Modelling

Agent-based models (ABMs) have become increasingly popular in computational biology for modelling and simulating complex systems, in part due to their unique ability to integrate heterogeneity into exploring dynamic processes over spatial and temporal domains. ABMs model systems composed of autonomous, interacting agents (Fig. 5.4) and consist of three core components:

- **Agents** Computational entities with encapsulated attributes capable of altering their behaviour or state in response to other agents and/or the environment. Agents can be used to represent individuals (e.g. a cell) or a group of individuals (e.g. a tumour).
- **Environment** The simulation space (i.e. virtual world) in which agents exist.
- **Logic** An underlying topology of connectedness, rules that govern how agents learn, adapt and interact with each other and/or their environment.

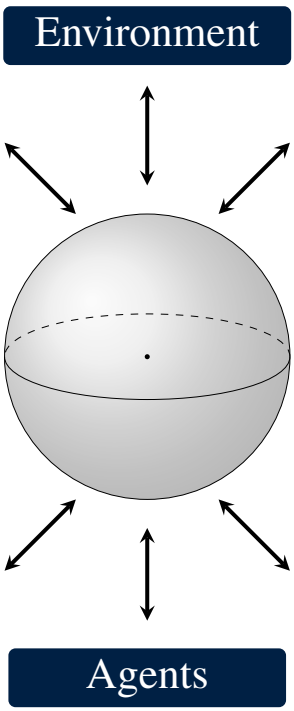


Figure 5.4: Visual Representation of an Agent
Schematic representation of a self-contained agent represented as a sphere. Double-headed arrows (↔) depict interactions with other agents or the environment.

Predefined logical rules determine how agents interact and respond to features and stimuli within their environment (Fig. 5.5). Their state varies over time and the state of the system consists of the collection of states of all agents. For an ABM with discrete time $t = 1, 2, 3, ...$, the macroscopic state, $X(t)$, of the modelled system is given by the set of microscopic states of n agents:

$$X(t) = (x_i(t))_{i=0,...,n}.$$

Spatial ABMs can be distinguished between those operating on fixed lattices (e.g. cellular automata, cellular Potts models) or off-lattice. The former consists of agents represented on a grid in $\mathbb{R}^2$ (e.g. hexagonal or squared) or a grid in $\mathbb{R}^3$ (e.g. dodecahedron or cubic). The lattice site can either be occupied by one agent, multiple agents, or an agent can be occupying multiple lattice sites. Movement and interactions are generally constrained to adjacent regions. For instance, on a two-dimensional square lattice the Moore neighbourhood can be used to define a set of cells that are orthogonally or diagonally-adjacent to a given cell (Tzedakis et al., 2015). Given a range r, the Moore neighbourhood consists of $(2r + 1)^2$ cells and is defined by (Weisstein, 2022):

$$N^M_{(x_0,y_0)} = (x, y) : |x - x_0| \leq r, |y - y_0| \leq r.$$

Conversely, lattice-free ABMs provide unconstrained movement, whereby agents are represented individually and can move continuously in space. Lattice-free models include centre-based models, in which agent interactions are based on forces or energies between their centres and do not resolve agent shape, and deformable cell models, which take into account the agent's shape and the area/volume they occupy in space (Kennedy et al., 2016).

ABMs are generally much more coding-intensive than other types of modelling techniques. Although any programming language can be used to build these models, object-oriented programming (OOP) languages, such as Java, Python, C++, are best suited for the task. They support encapsulation of data and programming methods or procedures as objects which can manipulate their own data and interact with other programming objects.

In addition, OOP supports a concept known as inheritance, in which a new object can acquires the property of another object. Indisputably, the OOP paradigm can easily be leveraged to design and build software agent systems.

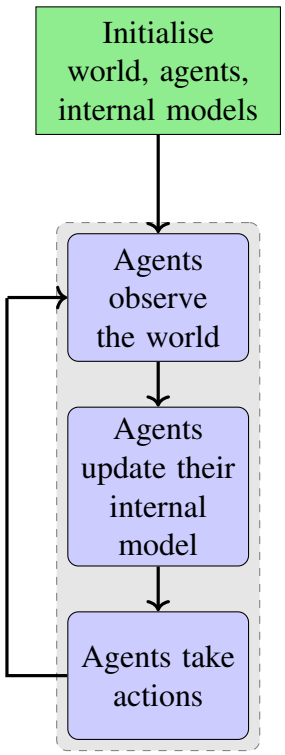


Figure 5.5: Overview of the Flow of Information in an ABM
High-level description of the flow of information as an agent interacts with other agents and the environment. Adapted from Rand (2006).

In addition to benefiting from advances in computer performance, the availability of data at increasing levels of granularity has led to increasingly sophisticated agents, larger models, and more complex interactions. This myriad of details at the agent scale is used to explore and interpret system behaviour and predict subsequent emergent phenomena. However, the high flexibility of this approach comes at a computational cost — the more details embedded into the model, the higher computational complexity. Furthermore, due to their stochastic nature, ABM simulations are run multiple times for the result to be summarised across a distribution of outcomes (Auchincloss & Garcia, 2015). It is thus crucial to find the right balance between simplicity, validity and robustness.

Model Design

Agent-based modelling was used to investigate the effect of CAR tuning on biological activity and on-target off-tumour toxicity. The objective was to identify the optimal CAR expression level that offers the right balance between safety and therapy efficacy. The model was designed to evaluate treatment response both at the primary tumour site and in the metastatic setting of breast cancer but could be generalised for many tumour types as it does not take into account tissue morphology nor molecular features.

The dynamics of tumour cells and CAR T cells were modelled as a close system in a finite lattice-based two-dimensional Euclidean space with reflective boundaries. Each lattice site can be occupied by at most one agent at a time. Three classes of agents were created, each one corresponding to a different type of single cell: CAR T agent, tumour agent and wild-type agent. Individual agents receive and respond to varying stimuli as they interact with one another or their environment, and movement and interactions are constrained to adjacent regions in a Moore neighbourhood of range 1.

The tumour is defined by the collection of tumour agents and the total population size is limited by the space. The tumour grows as a dense packing of congruent circles residing in square-shaped lattice sites at a rate which decreases as it approaches carrying capacity, akin to the classical Gompertz function (Gerlee, 2013) used to describe tumour dynamics. Tumour growth results from the proliferation of dividing cell agents at the periphery of the tumour. A new daughter cell can only be created if there is a free lattice point to place it within the surrounding neighbourhood. Tumour agents in the primary tumour site setting do not migrate and stay in their assigned square-shaped grid cell. In the metastatic context tumour cells are randomly scattered and assumed to perform a Brownian-type (diffusive) random walk on the grid, taking up to one step on each iteration if space allows. The level of tumour-associated antigens (TAA) expressed on the surface of tumour cells can be predefined but is assumed to be homogeneously expressed by all tumour cells, and thus the model does not account for tumour heterogeneity.

Wild-type cells representing healthy tissue can be included and set to express a specific level of TAA. Wild-type cells do not undergo cell division and can be organised into a

compact structure, randomly scattered or placed along the lattice boundary. Since these agents correspond to structural or tissue-resident cells, they remain stationary and cannot migrate to new locations. Wild-type cells are sensitive to nutrient deprivation, therefore the probability of cell death was designed to increase if surrounded by tumour cells, akin to the Chinese strategy game of Go, for an extended period.

A predefined starting number of TAA-specific CAR T cells can be added with a fix number of CARs on their surface. CAR T cell-tumour interaction primarily occurs at the tumour periphery for solid tumours. Interaction with their cognate antigen can stimulate them to become activated and proliferate at a probability proportional to CAR expression level and TAA exposure. The model does not consider varying levels of CAR-TAA affinity. Activated CAR T cells upon re-encountering their cognate antigen will mediate their cytolytic effector functions, achieving a level of potency proportional to the number of engaged CARs. It is further assumed that CAR T cell migration patterns are similar to those of wild-type T cells. T cells searching for their cognate antigen have previously been described to use a Brownian-type (diffusive) or Lévy-type (subdiffusive) random walk depending on their state and the microenvironmental context. As reported in Krummel et al. (2016), an adaptive switch of migration mode, from Lévy to Brownian walk, has been observed in population dynamic studies in response to different target densities. Despite the lack of clarity in empirical studies which has led to these patterns of T cell movement being disputed, as highlighted in Fricke et al. (2016), these concepts were applied to the migration of CAR T cells. CAR T cells move subdiffusively on the two-dimensional square lattice, and switch to a Brownian-type pattern once activated in the tumour periphery region which represents a high target density area. The model is further simplified by assuming that the interactions take place over a short time period, and thus exhaustion is not expected to play a meaningful role on CAR T cell activity.

Model Development

The ABM was built in Python (v3.8.13) because of its flexible OOP paradigm and rich libraries like Mesa (Kazil et al., 2020) to facilitate model development. The Mesa Wolf-Sheep Predation Model which is a re-implementation of the NetLogo Wolf-Sheep Preda-

tion Model (Wilensky, 1997) was used as a starting point to leverage the graphical user interface the framework offers to help with model design and troubleshooting. Once the ABM was completed, all simulations were run in batch mode from the command line.

A class called `TumourCARTCell` was created to hold all the general parameters, the environment, the scheduler and handle the agents and data collectors. The latter consists of Python dictionaries that will record model-level and agent-level data during a simulation run. The scheduler is an object that keeps track of time (Δt) in discrete steps and manages agent activation order. Separate classes were created for each of the three agent types, each with their own attributes and methods: `TumourCell`, `WTCell` and `CARTCell`. The environment consists of a regular 40x40 square grid, limiting the maximum total agent population size to 1,600 since each grid cell was designed to hold up to one agent.

The model requires the following input parameters:

Tumour Type	Solid | Liquid
Tissue Type	TME | Compact | Sparse | Lining
Tumour Cell Proliferation Rate	$\{x \in \mathbb{R} : 0 \leq x \leq 1\}$
Tumour Cell TAA Level	[NY-ESO-1]
WT Cell TAA Level	[NY-ESO-1]
T Cell CAR Level	miSFIT Variant
Initial Tumour Size	$\{x \in \mathbb{Z} : 0 \leq x \leq 1600\}$
Initial Healthy Tissue Size	$\{x \in \mathbb{Z} : 0 \leq x \leq 1600\}$
Initial CAR T Cells	$\{x \in \mathbb{Z} : 0 \leq x \leq 1600\}$

At the start of each model run, all classes are instantiated with the inputted parameter values, starting with the class `TumourCARTCell`. In turn, the 2D discrete space is created and the individual agents are randomly positioned on the grid and added to the scheduler. Grid item placement algorithms were developed to implement the different tumour and tissue types shown in Figure 5.6.

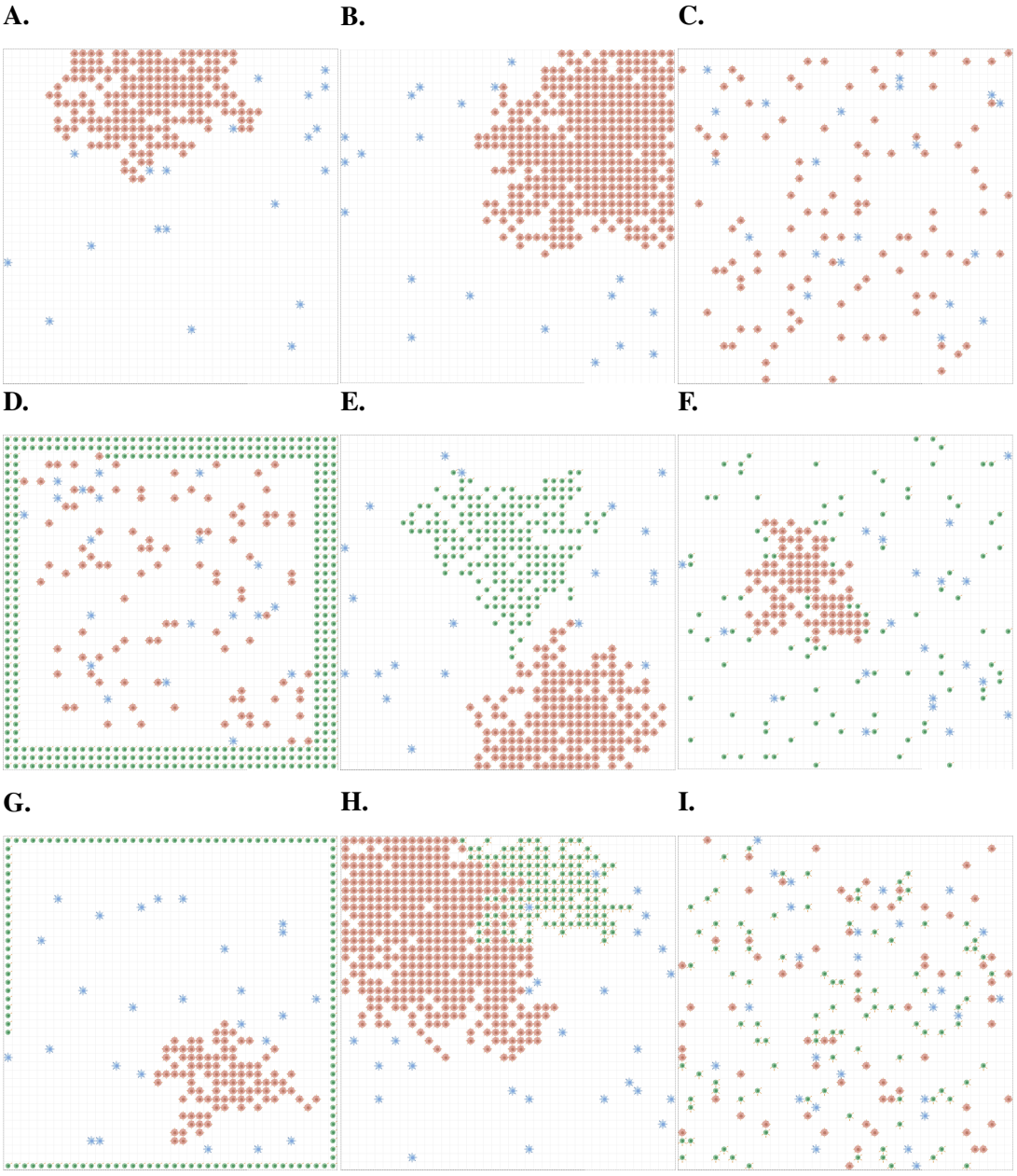


Figure 5.6: Mosaic of Model Topologies.
Collection of model topologies representing different tumour and tissue parameters with `CARTCell` agents shown in blue, `TumourCell` in pink and `WTCell` in green: **A.** Tumour Solid & Tissue TME. **B.** Tumour Solid & Tissue TME. **C.** Tumour Liquid & Tissue TME. **D.** Tumour Liquid & Tissue Lining. **E.** Tumour Solid & Tissue Compact. **F.** Tumour Solid & Tissue Sparse. **G.** Tumour Solid & Tissue Lining. **H.** Tumour Solid & Tissue Compact. **I.** Tumour Liquid & Tissue Sparse.

In order to model different scenarios, various structural contexts were implemented. These can be set using the Tumour Type and Tissue Type parameters. In the first instance, a tumour of type Solid or Liquid can be constructed using Algorithm 1 or 2, respectively. The aim of Algorithm 1 is to densely pack the agents together on the grid, thereby forming a compact structure. In the context of the tumour agents, this is meant to represent the formation of a solid tumour at the primary site. The same structure can be created using WT agents with the Tissue Type parameter set to Compact. Conversely, Algorithm 2 randomly scatters the agents across the grid to represent cells either sparsely organised or travelling through the circulatory system. This can be implemented for the tumour agents using the Liquid parameter and for the WT agents by selecting Sparse for Tissue Type. In addition, the Tissue Type can be set to Lining in order to place WT agents along the grid border, akin to healthy tissue lining the surfaces of organs or blood vessels. If the Tissue Type is set to TME, only tumour agents and CAR T agents will appear on the grid. CAR T agents are subsequently randomly allocated to grid cells.

The level of TAAs expressed on the surface of `WTCell` and `TumourCell` agents, along with the number of CARs on `CARTCell` agents can be defined before each experiment (Fig. 5.7). Once the simulation starts, at each time step Δt the agents' step functions will be executed in random order (Fig. 5.8). This process will be repeated until the end of the simulation is reached. The user can choose to stop the simulation at any time but it is set by default to stop once either the grid reaches capacity, or there are no more agents of class `TumourCell` or a maximum of 500 steps have been reached. Using the browser-based interface, values stored in the data collectors can be visualised in real-time. At the end of the simulation, the data is exported to a CSV file for downstream analysis.

During the simulation, `WTCell` agents stay in their assigned grid cell (Fig. 5.9) whilst `TumourCell` agents either perform a Brownian-type random walk, if the Tumour Type was set to Liquid, or stay in their assigned grid cell, if set to Solid (Fig. 5.10). Naive `CARTCell` agents perform a Lévy-type random walk (Fig. 5.11). The probability of CAR T cell engagement with its cognate antigen is proportional to the number of CARs and TAAs in the neighbourhood. Upon its first encounter with a TAA, the state of a

`CARTCell` agent changes to activated. At that instance, if space allows, the CAR T agent makes a copy of itself into a neighbouring location. Once activated, agents switch to a Brownian-type walk. An activated `CARTCell` agent's Activation Marker score increases upon re-engagement with TAA at a rate that diminishes with its age. This Activation Marker is used to measure the resulting cytolytic activity towards cognate target cells.

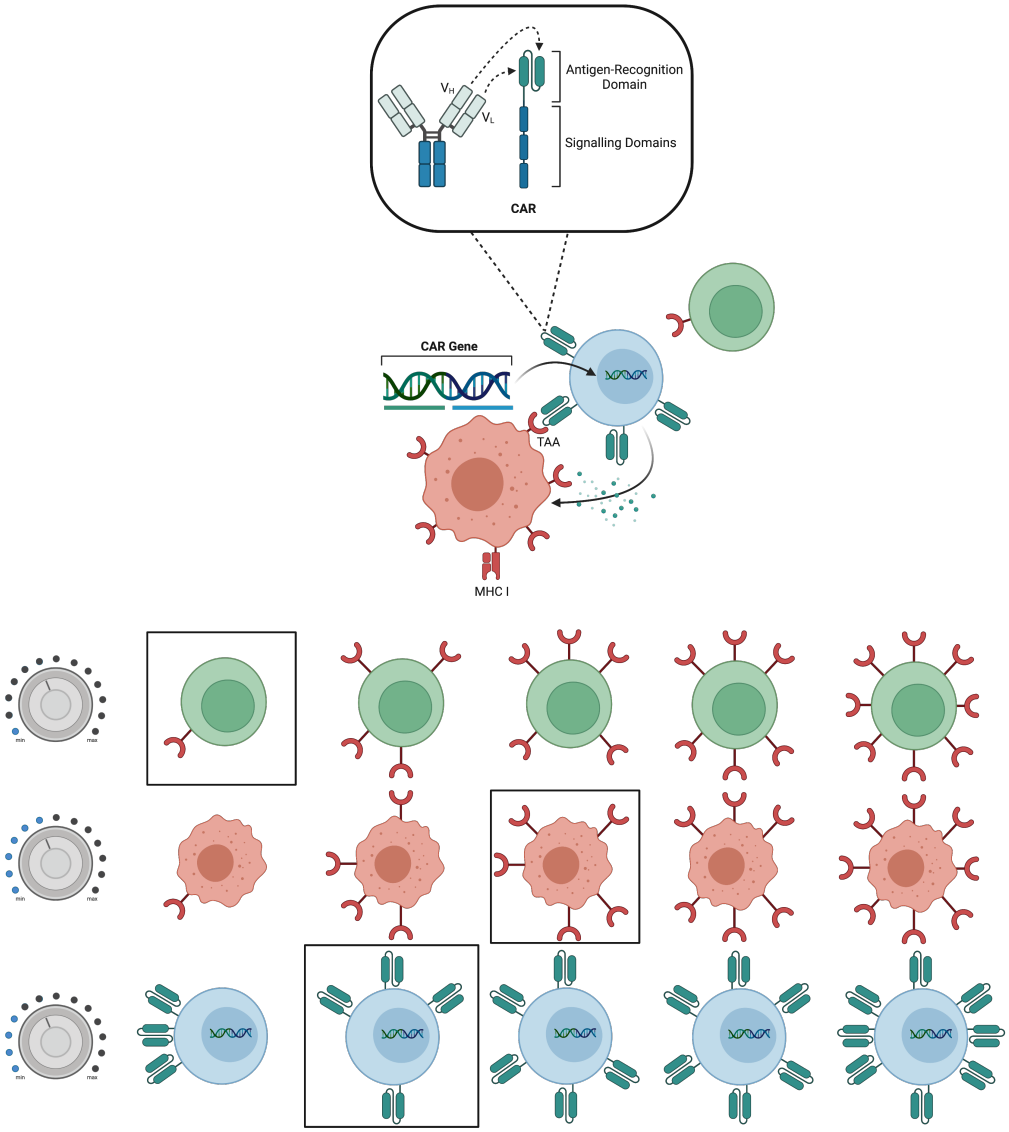


Figure 5.7: Schematic Representation of the ABM.
The number of TAAs on `WTCell` and `TumourCell` agents, and the number of CARs on `CARTCell` agents are selected prior to running the simulation. The resulting cytolytic activity towards TAA-expressing agents is a function of the level of CAR-TAA engagement. Created with BioRender.com and icon by Freepik from Flaticon.com.

Algorithm 1: Dense Packing

```
function DrawCompact (a);
  Input  : Number of agents a
  Output: Dense packing of agents on grid
for i in range [0,a) do
    if i = 0 then
        x = random integer in range [0, X_grid]
        y = random integer in range [0, Y_grid]
        while (x,y) not empty do
            x = random integer in range [0, X_grid]
            y = random integer in range [0, Y_grid]
        end
        create new agent
        place agent on grid at (x,y)
        add agent to scheduler
        get Moore neighbours of (x,y)
        save to set of coordinates
    else
        select empty coordinate from set
        remove selected coordinate from set
        create new agent at new position
        add agent to scheduler
        iterate neighbourhood for free positions
        save to set of coordinates
    end
end
```

Algorithm 2: Scattered

```
function DrawScattered (a);
  Input  : Number of agents a
  Output: Agents scattered on grid
for i in range [0,a) do
    x = random integer in range [0, X_grid]
    y = random integer in range [0, Y_grid]
    while (x,y) not empty do
        x = random integer in range [0, X_grid]
        y = random integer in range [0, Y_grid]
    end
    create new agent
    place agent on grid at (x,y)
    add agent to scheduler
end
```

Algorithm 3: Lining Environment

```
function DrawLining (a);
  Input  : Number of agents a
  Output: Line agents along grid border
create iterator ['right', 'up', 'left', 'down']
stop = (0,1)
x_max, y_max, x_min, y_min = X_grid, Y_grid, 0, 0
for i in range [0,a) do
    if i = 0 then
        x = 0
        y = 0
        create new agent
        place agent on grid at (x,y)
        add agent to scheduler
    else
        if iterator on 'right' and x < x_max − 1 then
            x += 1
            create agent at (x,y)
            if x = x_max − 1 then
                next iterator element
        else if iterator on 'up' and y < y_max − 1 then
            y += 1
            create agent at (x,y)
            if y = y_max − 1 then
                next iterator element
        else if iterator on 'left' and x > x_min then
            x -= 1
            create agent at (x,y)
            if x = x_min then
                next iterator element
        else if iterator on 'down' and y > y_min then
            y -= 1
            create agent at (x,y)
            if y = y_min then
                next iterator element
        else
            pass
    if (x,y) = stop then
        next iterator element
        stop = (x+1, y+1)
        x_max, y_max= x_max − 1, y_max − 1
        x_min, y_min = x_min + 1, y_min + 1
end
```

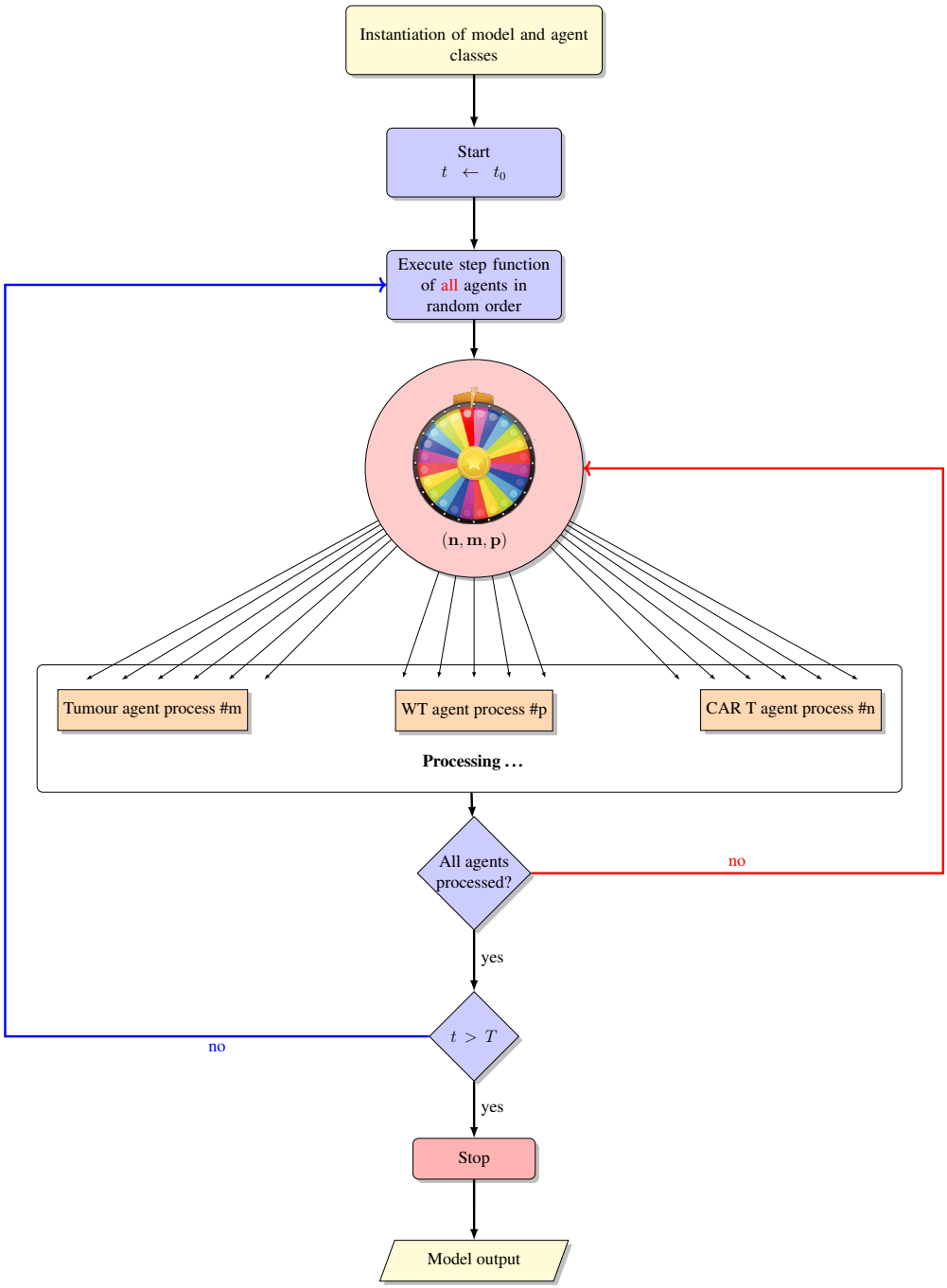


Figure 5.8: Flow Chart of the ABM Simulation Algorithm.
The ABM starts once all objects have been instantiated. At each time step Δt, the agents are activated in random order. The simulation runs until t reaches T.

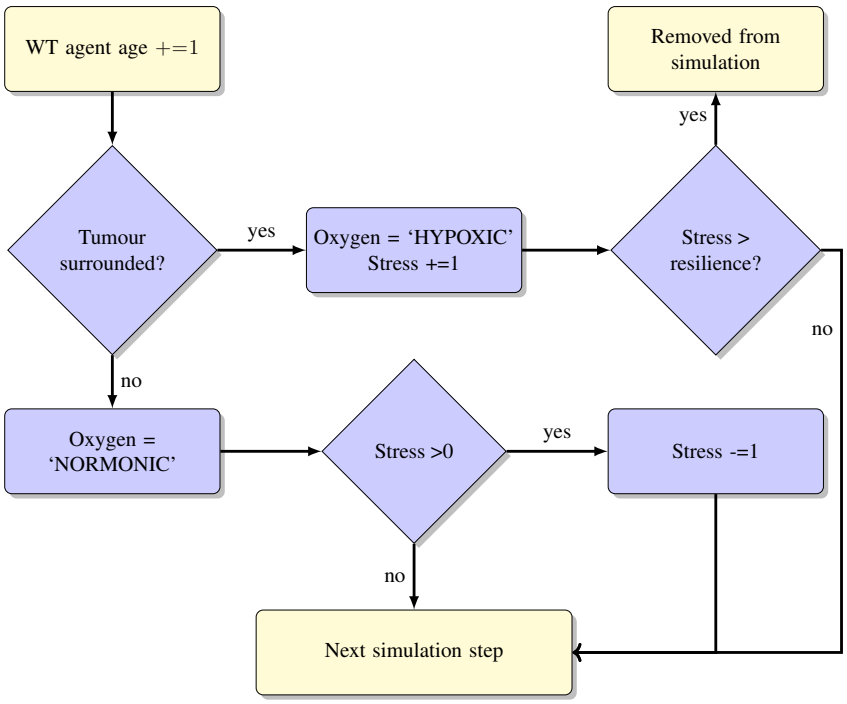


Figure 5.9: Wild-Type Agent Step Function.

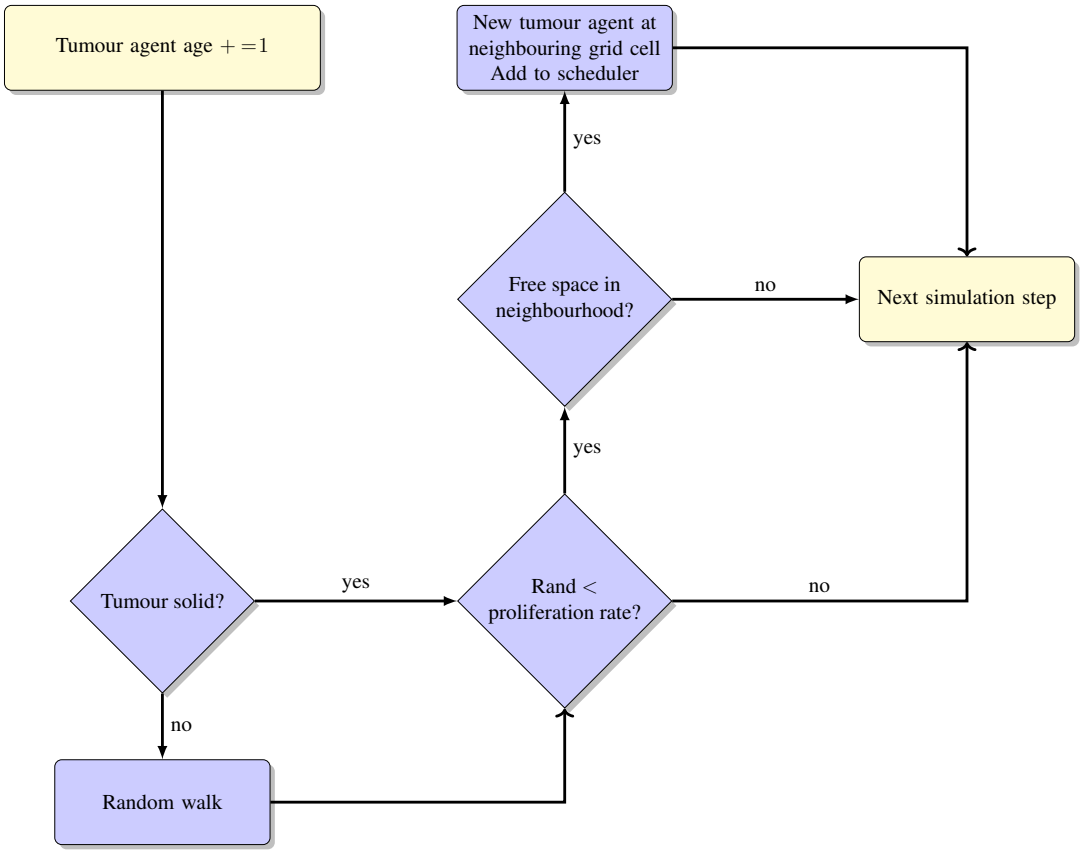


Figure 5.10: Tumour Agent Step Function.

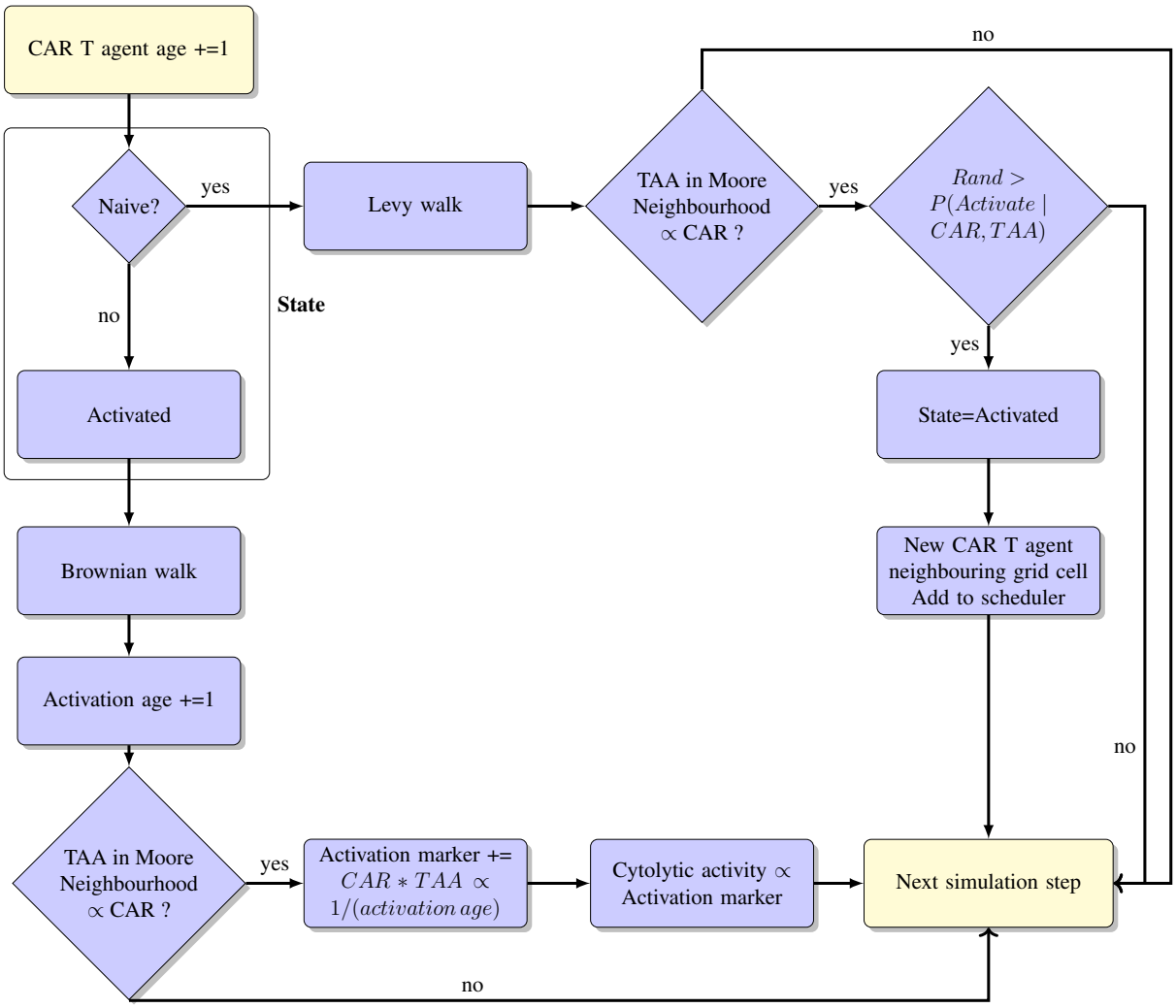


Figure 5.11: CAR T Agent Step Function.

Model Analysis

The ABM was calibrated by testing different parameter sets, and comparing the output to the experimental data presented in Section 5.3.1 as applicable. Simulation experiments were run in batch mode with 50 replicates per scenario, and the average for all the runs was calculated. All corresponding files can be found in the repository for this chapter (Appendix F). Results were analysed in Python (v3.8.13) and Prism (v9.4.1). In the first experiment, different tumour growth rates were tested without the addition of CAR T cell agents to see how the behaviour compares across the two tumour types, namely solid and liquid. In a second experiment, varying combinations of CAR expression were tested with

fixed numbers of TAAs on agents representing wild-type cells and tumour cells. The objective was to test whether there exists an optimal number of CAR that maximises tumour killing whilst sparing healthy tissue. Numerical analysis of ordinary differential equations was conducted in MATLAB.

5.3.3 Modelling CAR T Cell Gene Regulatory Network

Boolean Network Modelling

A network, also known as a graph, is formally defined as a set of objects (i.e. nodes) that are connected together by edges. Molecular systems orchestrating the biology of the cell involve a complex network of interactions among various components (i.e. genes, proteins, molecules), which can be studied using a class of discrete models known as logical networks (Xiao, 2009). This approach is well-adapted to study the switch-like behaviour exhibited by cellular systems when regulating their functional states (Leshi et al., 2018). Unlike continuous models which couple differential equations, modelling logical networks does not require knowledge of each interaction's kinetic rate constant. Instead, it relies on Boolean algebra performing the logical operations AND, OR and NOT on binary numbers. The Gene Interaction Network Simulation (GINSim) Environment (Naldi et al., 2018) is a well-established framework in systems biology to model and analyse gene regulatory networks (GRNs). To facilitate model development, python bindings are available through the package ginsim. Simple geometric vector shapes such as rectangles or ellipses define the nodes (i.e. gene, proteins) whilst line segments drawn between nodes define the interaction type (i.e. activating or inhibitory). Beyond this visual structure, logical parameters governing the activation of individual nodes in the network can be defined using the Boolean expressions: & = AND, ! = NOT, | = OR. Each node can take a value of 1 or 0, ON or OFF. Of note, in the non-Boolean case GINSim enables the generation of multi-valued nodes with expression greater than 1. Single or multiple perturbations corresponding to mutant conditions such as loss-of-function or overexpression can be tested by forcing targeted node values to 0 or 1, respectively. The GINSim architecture offers an embedded simulation framework which results in the creation of a state transition graph

(STG) that represents the dynamical behaviour of the system. The size of a STG grows exponentially, with up to 2^N states possible for a model of N components. The updating policy can be specified as asynchronous or synchronous, which considers all updating tasks either independently or simultaneously. To reduce the computational cost associated with large models, GINSim implements an algorithm described in Naldi et al. (2011) that bypasses the need of running the simulation to compute the stable states, the terminal nodes of the STG. These are states where if reached, the dynamics of the network does not change over time. When modelling a network of interacting intracellular elements this can translate into the long-term behaviour of a signalling network that is related to a certain cellular phenotype. At an abstract level, a Boolean network, $G = (N, A)$, is composed of a collection of nodes N and directed arcs A. These networks are binary rule-based dynamic systems where each node, $x_1, x_2, \ldots, x_n$, takes a value of 1 or 0 respectively. State transitions, from 0 to 1 or 1 to 0, are influenced by the node's parent interactions, where relationships are either positive (up-regulating) or negative (down-regulating). If x_i has k_i parent nodes $x_{i_1}, x_{i_2}, \ldots, x_{k_i}$, then the value of x_i at time $t+1$ is determined by its parent nodes at time t and the Boolean function (Xiao & Dougherty, 2007):

$$x_i(t+1) = f_i\left((x_{i_1}(t), x_{i_2}(t), \ldots, x_{k_i}(t)\right)$$

Assuming that the Boolean functions are time invariant, this follows that:

$$\begin{cases} x_1 &= f_1\left((x_{1_1}(t), x_{1_2}(t), \ldots, x_{k_1}(t)\right) \\ x_2 &= f_2\left((x_{2_1}(t), x_{2_2}(t), \ldots, x_{k_2}(t)\right) \\ &\vdots \\ x_n &= f_n\left((x_{n_1}(t), x_{n_2}(t), \ldots, x_{k_n}(t)\right) \end{cases}$$

The global state of the network at any time point is the collection of states, as defined by 0 or 1, of each of the node:

Vector state:

$$x(t) = ((x_1(t), x_2(t), \ldots, x_n(t))$$

State space:

$$\{0,1\}^n = \{(0,0,\ldots,0),(0,0,\ldots,1),\ldots,(1,1,\ldots,1)\}$$

The transitions of a three node network, a, b, c, can be defined by three functions:

$$\begin{cases} a &= f_a\left((x_a(t), x_b(t), x_c(t)\right) \\ b &= f_b\left((x_a(t), x_b(t), x_c(t)\right) \\ c &= f_c\left((x_a(t), x_b(t), x_c(t)\right) \end{cases}$$

The truth table (Table 5.1) indicates the 2^3 possible states the network can undertake. A schematic representation of these results is shown in Figure 5.12.

Table 5.1: Truth Table of the Transition Functions.

State	$f_a(x_a, x_b, x_c)$	$f_b(x_a, x_b, x_c)$	$f_c(x_a, x_b, x_c)$
$(0,0,0)$	0	0	0
$(0,0,1)$	1	0	1
$(0,1,0)$	0	1	1
$(0,1,1)$	1	1	1
$(1,0,0)$	0	0	0
$(1,0,1)$	1	0	0
$(1,1,0)$	0	1	0
$(1,1,1)$	1	1	0

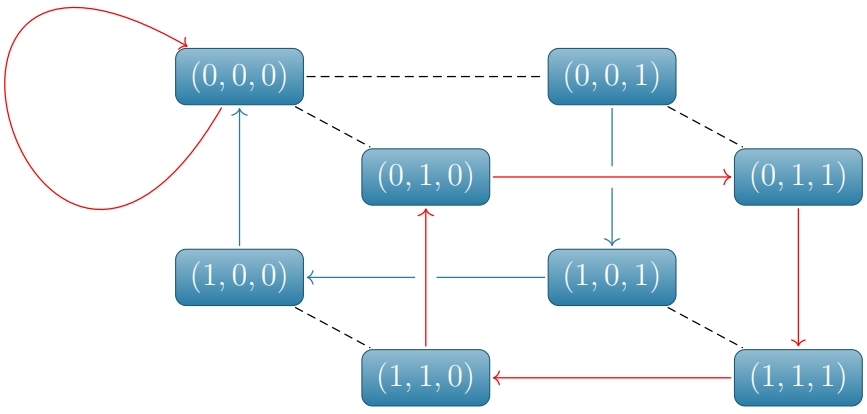


Figure 5.12: Schematic Diagram of Boolean State Space.
State space diagram of a three-node Boolean model with stable-states depicted in red.

Model Expansion

The BioModels database (Malik-Sheriff et al., 2020) and the GINSim repository were searched for GRN models of T cells. A suitable model had been published Rodriguez-Jorge et al. (2019). It consists of a validated model encompassing 110 nodes of antigen-driven signalling through the T cell receptor of CD4 T cells and costimulatory receptors. The two externally controlled input nodes, TCR and CD28 were merged to create a CAR node. Steady-state analysis confirmed that this modification did not modify downstream dynamics. A query was performed using the core components of the cGAS-STING pathway, mainly cGAS, STING, TBK1, and IRF3 to retrieve known and predicted interactions from the STRING database (Fig. 5.13). The cytosolic DNA-sensing pathway (hsa04623) from the KEGG database (Kanehisa et al., 2021) presented in Section 4.4.6 was used in conjunction. The literature was searched for evidence supporting the underlying Boolean rules governing each interaction.

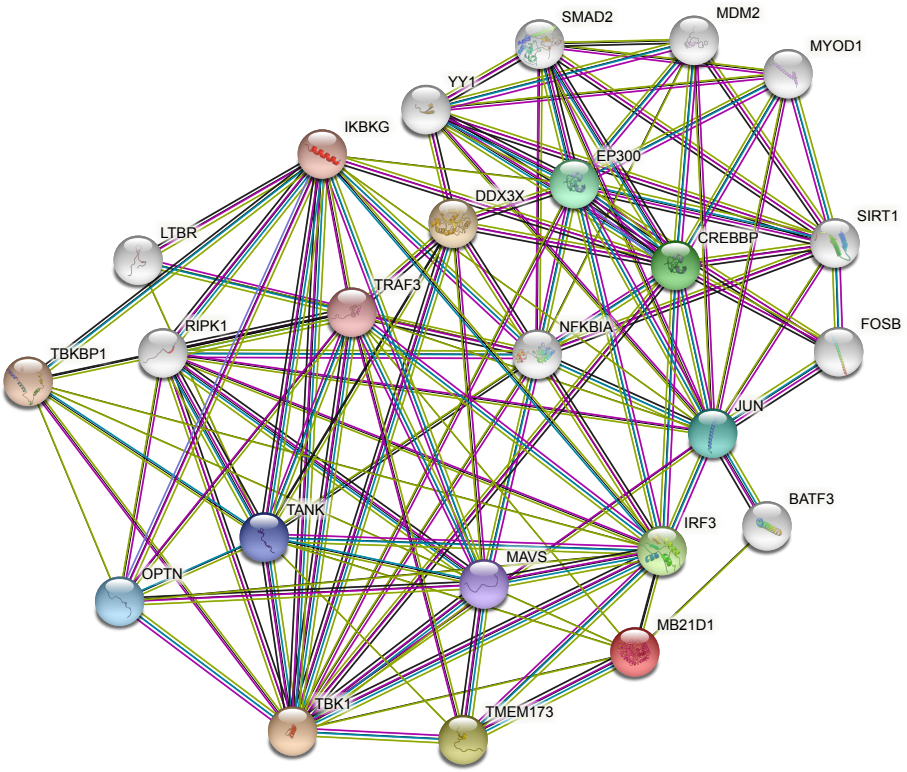


Figure 5.13: Protein-Protein Interaction Network of cGAS-STING Components. Known (pink or blue) and/or predicted physical and/or functional associations between core members of the cGAS-STING pathway with nodes representing proteins. Generated using the STRING database (Szklarczyk et al., 2015).

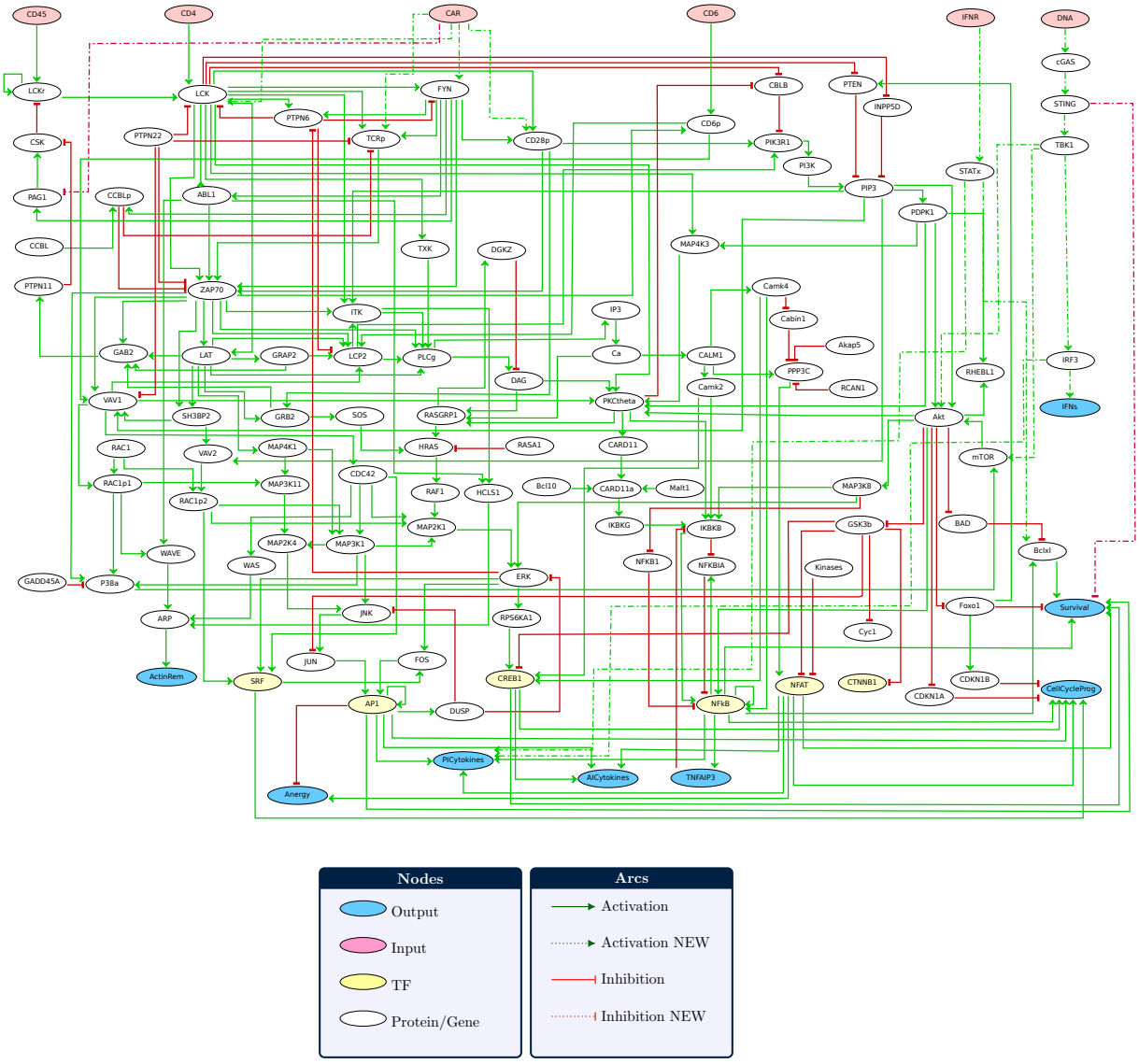


Figure 5.14: CAR T Agent Gene Regulatory Network.
A validated T cell receptor signalling GRN model published by Rodriguez-Jorge et al. (2019) was adapted for CAR T cell modelling. The cGAS-STING signalling pathway was incorporated.

5.3.4 Combining Modelling Paradigms

Modelling paradigms described in Sections 5.3.2 and 5.3.3 can be combined to create multi-scale models that bridge intracellular pathways and cell population dynamics. For instance, to model the complexity of the TME, a three-dimensional virtual microenvironment, MicroC, was created by the Buffa Lab at the University of Oxford (Voukantsis et al., 2019). Using cancer as a model system, the aim of MicroC is to provide an in silico method to study evolution and growth of heterogeneous cell populations across diverse spatial, temporal and functional domains. Moreover, this modelling framework can be used to investigate the effect of mutations on cell to cell interactions and cell growth dynamics. Since its original publication, MicroC has undergone numerous changes. The most recent version, developed by Pedro Victori (Buffa Lab), is described below.

Current Framework

MicroC is a modular framework implemented and developed in Java. Every cell in the model is represented as a computational agent (cf. Section 5.3.2), capable of receiving varying stimuli as it interacts with other entities, outputting a behaviour or change of state in response. The simulation space is lattice-free and cellular agents are approximated as spheres in space (Fig. 5.15):

$$\psi = \frac{\pi^{\frac{1}{3}}(6V)^{\frac{2}{3}}}{A}$$

Where:

- ψ Sphericity
- V Volume of the agent
- A Surface area

Cellular agents are equipped with Boolean network models consisting of abstract representations of genes, proteins or molecules interacting with one another, as described in Section 5.3.3. In such manner, the underlying network model defines the agent's particular

behaviour, akin to how cellular genotype influences cellular phenotype. The networks can be perturbed in response to stimuli, to test how differences in agent genotype influence the system's behaviour. Similarly, mutations which are permanent changes in the activation status of a node may be randomly generated as the agents multiply. Probabilistic Boolean network (PBN) models have recently been introduced through the implementation of a novel algorithm called Ingres (Inferring Probabilistic Boolean Networks of Gene Regulation Using Protein Activity Enrichment Scores) that maps scRNA-seq data to Boolean networks (Victori & Buffa, 2022). Furthermore, MicroC has been adapted to integrate spatial transcriptomics data to inform the positioning of cellular agents (Victori personal communication). The incorporation of scRNA-seq and spatial transcriptomics enables the in silico reconstruction of patient-specific transcriptional profiles and tissue organisation.

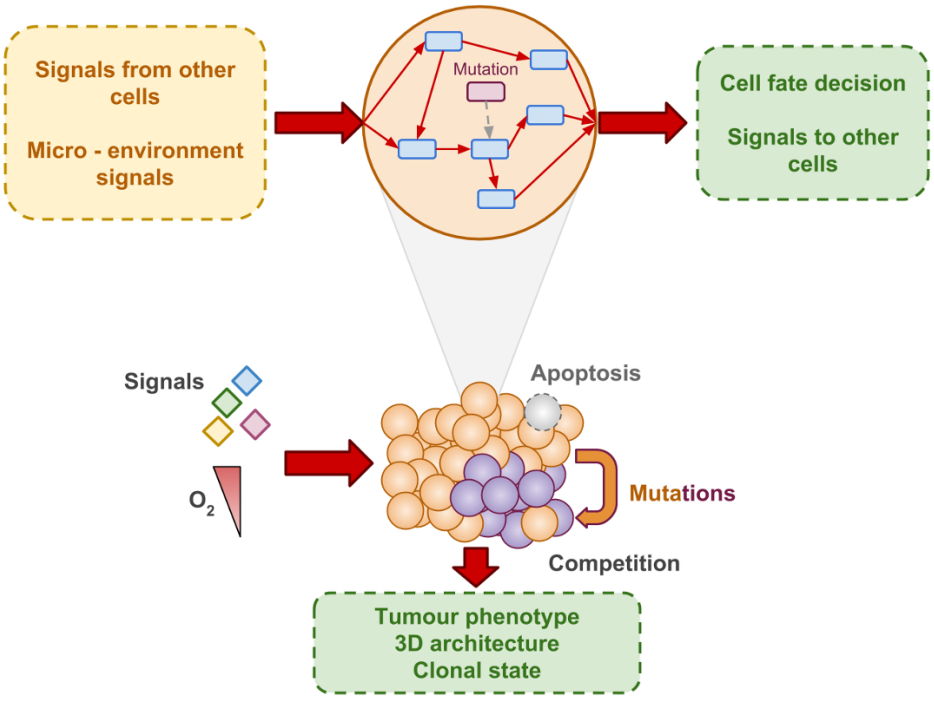


Figure 5.15: MicroC Environment.
Conceptual representation of the MicroC environment and simulation components linking genotype to phenotype. Reprinted from Victori and Buffa (2019).

The simulations can be run in the presence of diffusible substance agents representing, for example, drugs or molecules acting upon input nodes of the GRNs. Varying levels of oxygen can be simulated to test how oxygen influences population evolution. Although

cellular agents themselves are modelled in lattice-free space, diffusion is modelled through a voxel cube, as depicted in Figure 5.16. Diffusion simulations are based on Fick's law which describes the spread of particles through random motion in response to a concentration gradient, moving from an area of high concentration to an area of low concentration. For a substance like oxygen, the boundary conditions satisfy the Dirichlet condition which fixes substrate concentrations at the boundaries, akin to oxygen diffusing out from blood vessels.

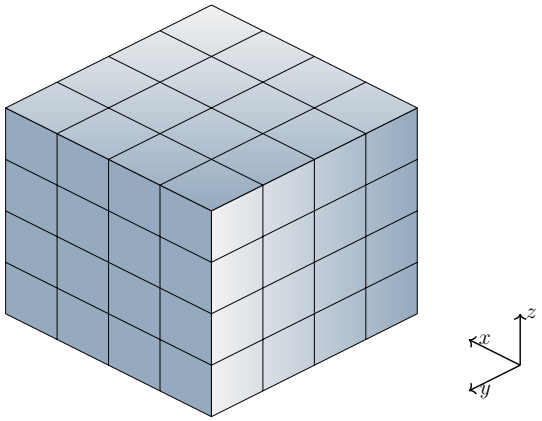


Figure 5.16: Voxel Cube Representation.
Space subdivided into equally sized cubic voxels used to simulate the diffusion of substances in MicroC.

Cellular agents are sinks that consume oxygen at a rate proportional to the oxygen concentration in the voxel:

$$S_{O_2} = R_0 \frac{C - C_{N,f}}{C_{O_2,opt} - C_{N,f}}$$

Where:

- S_{O_2} Oxygen consumption
- R_0 Initial consumption rate
- C Concentration of oxygen in the specific voxel
- $C_{N,f}$ Threshold value for lowest possible oxygen concentration; 80%
- $C_{O_2,opt}$ Optimal oxygen concentration; 0.28 mM

Cellular interaction through the production and consumption of cytokines is described by:

$$S_u = R_0 \cdot I_{i,j,k}^{receptor} + P_0 \cdot I_{i,j,k}^{output}$$

Whereby i, j, k are the position coordinates of the cell in space, either consuming the diffusible substance at a fixed rate of R_0 or producing it at a fixed rate of P_0, a behaviour conditional to the activation status of its corresponding receptor or output node, $I_{i,j,k}$ taking on a value of 0 or 1.

Expanding Framework

MicroC is still in its infancy and currently only supports the modelling of tumour cells without taking into account components of the immune system. However, it provides the ideal framework to model cell-based therapy and patient-specific responses. The T cell GRN from Section 5.3.3 will be used in the creation of a CAR T agent (Fig. 5.17), integrating knowledge about the effect of CAR tuning on T cell activity from Section 5.3.2.

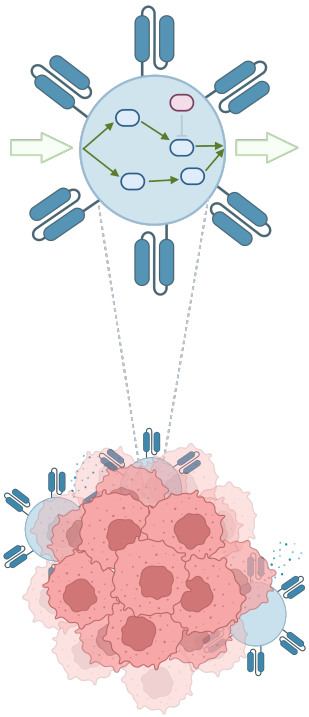


Figure 5.17: Modelling T Cell-Based Immunotherapy.
Schematic of a CAR T cell agent equipped with a GRN model interacting with tumour agents in its environment. Created with BioRender.com.

The tumour agent in MicroC uses an existing mitogen-activated protein kinase (MAPK) GRN model published by Grieco et al. (2013). It consists of highly interconnected signalling pathways involved in various cellular processes, including cell proliferation, growth arrest and cell death. The published model had been adapted for MicroC to include oxygen sensing. To enable the future modelling of tumour-T cell interactions and combination therapy in MicroC, the MAPK GRN was further expanded. A similar approach to the one in Section 5.3.3 was used to include the cGAS-STING pathway, expression of PD-L1 and response to oncolytic virus (OV) therapy. Results from Chapter 4 were used in the process, along with the manual curation of gene pairs from the literature and the KEGG database (Kanehisa et al., 2021). Both IFNs and OVs will be implemented as diffusible agents in the future.

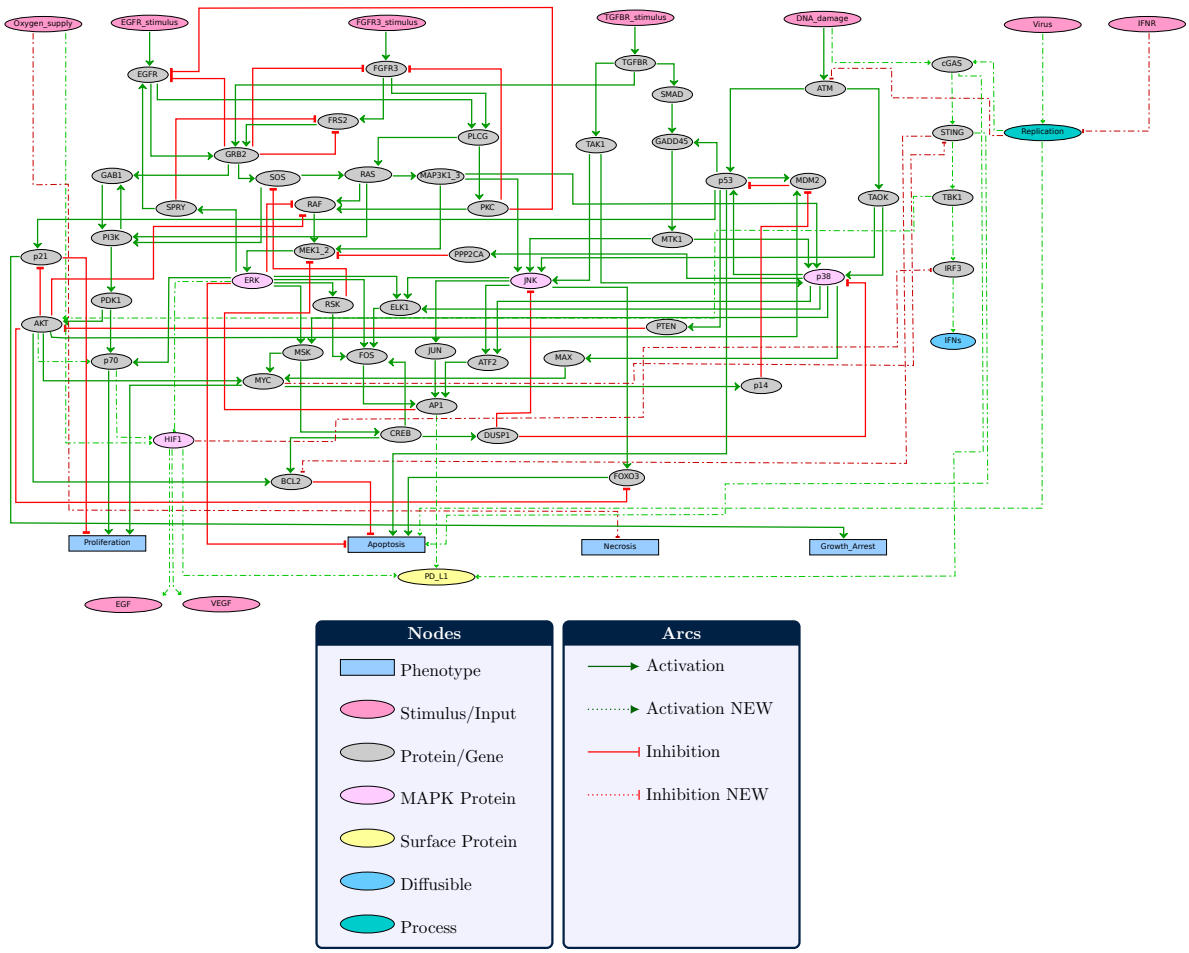


Figure 5.18: Tumour Agent Gene Regulatory Network.
The GRN model of the MicroC tumour agents was adapted to include PD-L1 expression, cGAS-STING signalling and the effect of oncolytic virotherapy.

5.4 Results

5.4.1 Comparison of Tumour Growth Rates

Sigmoid functions including Gompertzian and generalised logistic curves are commonly used to fit tumour growth data. As described in Section 5.3.2, carrying capacity, K, is defined in the ABM by the size of the grid environment. To test the implementation of tumour growth dynamics, simulations were run at varying tumours growth rates starting with one `TumourCell` agent. No other agent type was added. The resulting curves could then be compared to the Gompertz equation which was numerically solved using the same growth rates, λ, and carrying capacity, K (Fig. 5.19A). At the same growth rates, it was found that the tumour reaches carrying capacity faster in the liquid scenario (Fig. 5.19B) than in the solid (Fig. 5.19C).

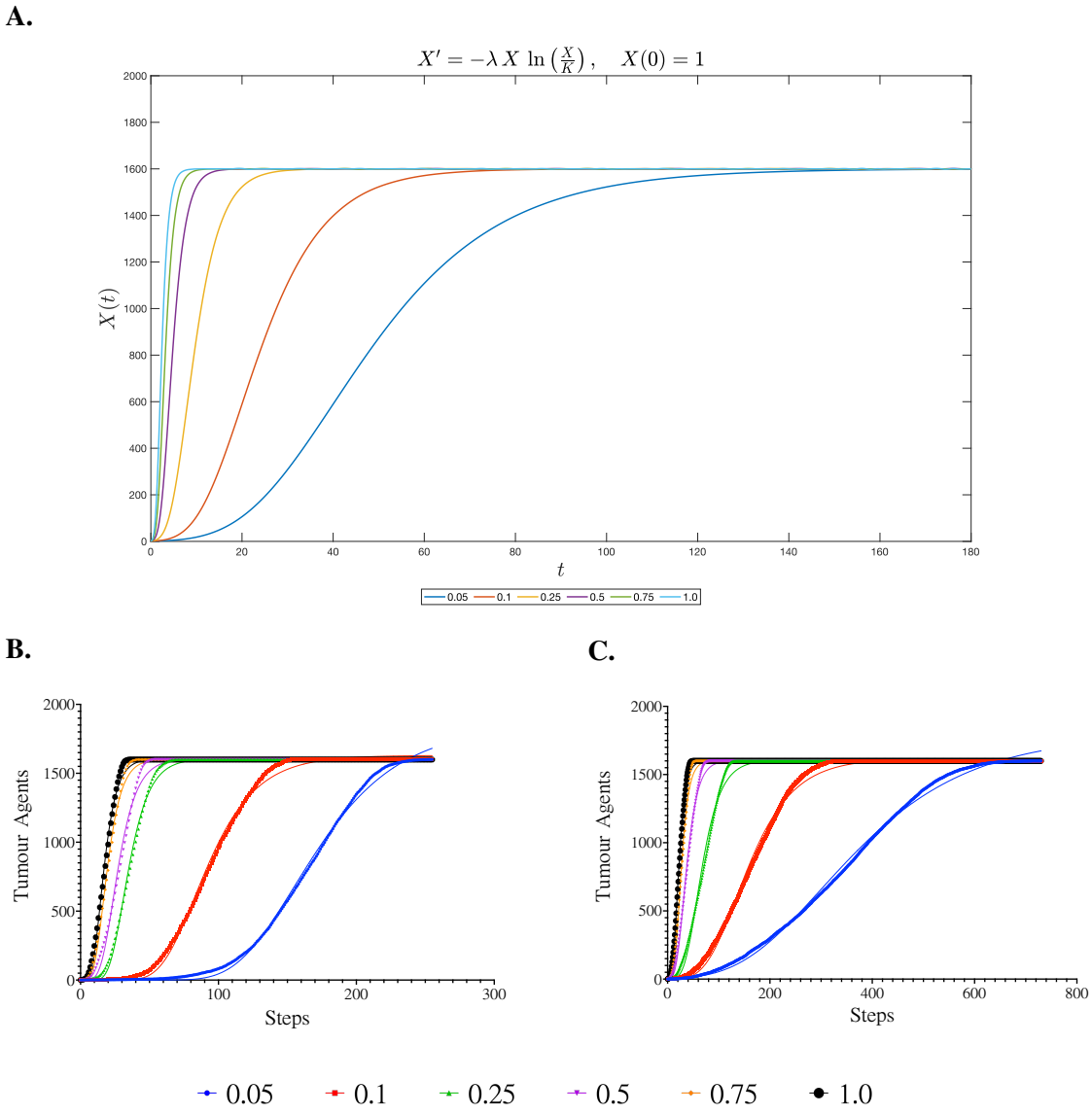
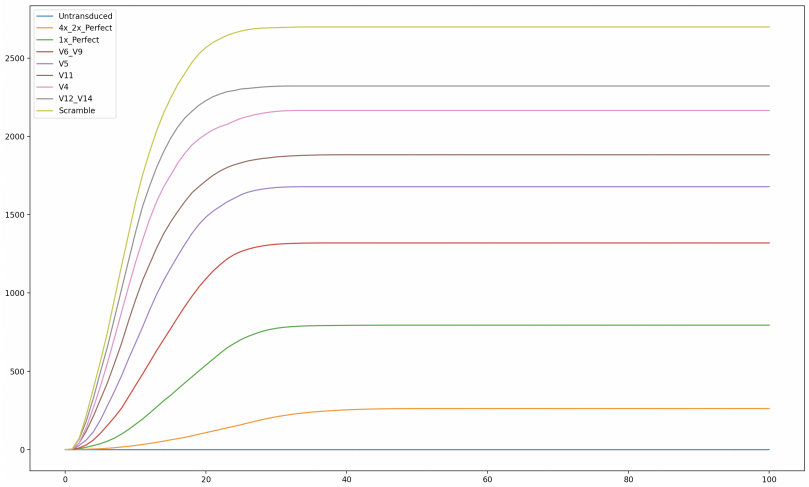


Figure 5.19: Tumour Growth Curves.
A. Numerical analysis of the Gompertz differential equation to describe tumour growth at varying growth rates for $K = 1600$. **B.** Tumour growth at varying growth rates in the ABM liquid tumour context. **C.** Tumour growth at varying growth rates in the ABM solid tumour context.

5.4.2 CAR T Agent Activation Tuning

Simulations were run using `CARTCell` agents with varying levels of CARs corresponding to the miSFIT variants used in the laboratory experiments. The activation marker was set to grow as a function of the number of CAR:TAA engagements. The growth of the total activation marker score across 10 CAR T agents was tracked for 100 time steps (Fig. 5.20). The CAR T cell agents were exposed to a growing number of `TumourCell` agents, growing at a rate of 0.1 and expressing a fixed level of TAA. In these simulation experiments, `CARTCell` agents were set not to exert their cytolytic functions.

A.



B.

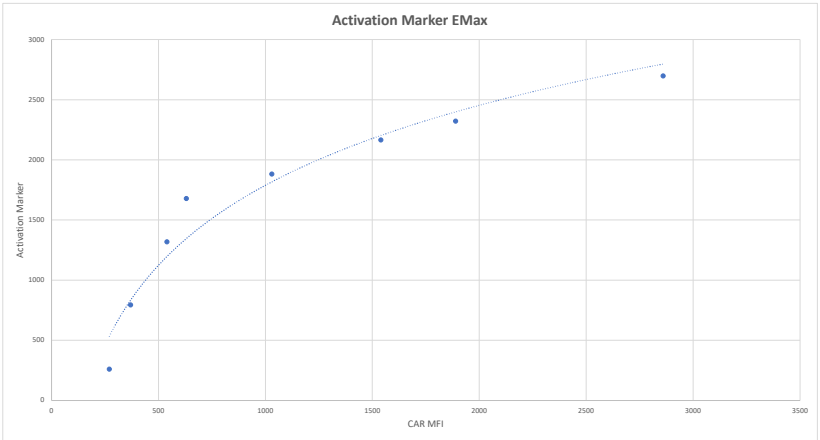
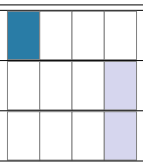
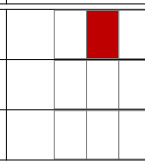
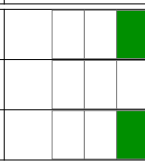
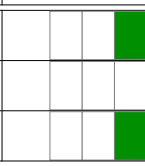


Figure 5.20: Activation Marker in Response to CAR Tuning.
A. ABM simulation results showing the activation marker score as a function of time for each CAR expression level. **B.** Maximum activation marker score reached for each CAR expression level.

5.4.3 Influence of cGAS-STING Activation on T Cell Stable States

Results of the CAR T cell GRN steady-state analysis under different perturbations and node stimulation are shown in Table 5.2. CAR stimulation is equivalent to the combined TCR and CD28. Perturbation with a STING agonist inhibits survival, contrary to the TBK1 agonist that further promotes it.

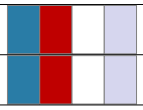
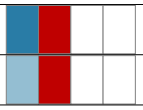
Table 5.2: Stable States of CAR T Cell Gene Regulatory Network.

Perturbations	None	TCR	TCR + CD28	CAR
None				
STING E1				
TBK1 E1				
Legend:	*ActinRemod*=0 *Cytokines*=0 *CellCycleProg*=0 *Anergy*=0 *Survival*=0	*ActinRemod*=0 *Cytokines*=0 *CellCycleProg*=0 *Anergy*=1 *Survival*=0	*ActinRemod*=1 *Cytokines*=1 *CellCycleProg*=1 *Anergy*=0 *Survival*=1	*ActinRemod*=0|1 *Cytokines*=1 *CellCycleProg*=0|1 *Anergy*=0 *Survival*=0

5.4.4 Tumour Cell Therapeutic Stable State Responses

End nodes reached by the stable states of the tumour cell GRN with a p53 KO were computed under different conditions (Table 5.3). The cGAS agonist promotes the expression of both IFN and PD-L1 under normoxic conditions but cannot stimulate the production of IFN under hypoxic conditions. The STING agonist results in similar stable states with less states expressing PD-L1 because expression in that scenario is only controlled by hypoxia but not cGAS-STING signalling. Perturbation induced by an oncolytic virus results in an apoptotic state with IFN production but leads to upregulation of PD-L1.

Table 5.3: Stable States of Tumour Cell △p53 Gene Regulatory Network.

Perturbations	Oxygen E1	Oxygen KO	EGFR E1	Virus E1
cGAS E1				
STING E1				
Legend:	 $PD{-}L1{=}1$	 $Apoptosis/Necrosis{=}1$ $GrowthArrest{=}0\|1$	 $Proliferation{=}1$	 $IFN{=}1$

5.5 Discussion

Motivated to delve into CAR T cell dynamics from a systems biology perspective, this chapter started with the development of an agent-based model (ABM) to explore the effect of CAR expression level on on-target off-tumour toxicity. Tumour growth was implemented as a compact structure of congruent circles or scattered across the grid environment in Brownian motion. Carrying capacity was set by the size of the grid and it was found that the scattering of the tumour agents resulted in faster tumour growth than the compact scenario at the same growth rate. This is likely due to the tumour agents moving around, creating space between them, thereby liberating nearby grid cells for them to replicate into. On the other hand, in the context of the compact structure, analogous to a solid tumour, growth was only occurring at the tumour periphery. An iterative process was used to calibrate the ABM to the experimental data. A similar step function to the experimentally-measured one describing the relationship between activation marker and TAA exposure at varying CAR levels was achieved with the ABM. However, the data was received too close to the thesis submission deadline for the calibration section to be completed. Preliminary simulation results show that on-target off-tumour toxicities cannot be avoided whilst maintaining efficacy but can be reduced by decreasing the level of CAR (Appendix F). Once properly calibrated, the ABM will be used to evaluate on-target off-tumour effects of CAR tuning under different spatial scenarios. The model was designed to predict which miSFIT variant may offer the optimal trade-off between safety and effi-

cacy in different tumour and healthy tissue contexts. Importantly, the ABM will be useful to test whether areas with greater antigen density are more at risk of on-target off-tumour toxicities. The sensitivity of the model to different doses of CAR T agents will also need to be evaluated. Furthermore, the ABM does not consider varying levels of CAR affinity which are sensitive to antigen density and could influence target cell discrimination. It would be important to study the influence of the ABM scheduling pattern and how different activation schemes impact the model. Since CAR tuning alone is expected based on preliminary results to be insufficient to prevent normal tissue toxicity, the model could be further extended to consider other strategies, including ones that restrict CAR T cell activation, limit their persistence or target them to the tumour site (Caruso et al., 2018). In addition, chronic antigen exposure and CAR-mediated tonic signalling have been recognised as drivers of CAR T cell exhaustion. Another potential application of this ABM would be to consider the effect of CAR tuning on balancing reduced exhaustion rate and toxicity.

Gene regulatory networks (GRNs) are complementary modelling paradigms to ABMs. In order to study CAR T cell dynamics at the level of intracellular interactions and explore the effect of cGAS-STING targeting therapies, a GRN published by Rodriguez-Jorge et al. (2019) of T cell receptor signalling was modified. As described in Section 2.4, CARs consist of an antigen-binding moiety and a costimulatory domain. As such, the CAR node was created by merging the TCR and CD28 nodes. The cGAS-STING cascade was introduced and linked to the main GRN, where applicable. For instance, TBK1 has been reported to promote survival through the activation of Akt and mTOR (Ou et al., 2011). In addition, IRF3 is a transcriptional regulator of type I IFN gene and proinflammatory cytokines (Yanai et al., 2018). STING agonists like koAA have been proposed to promote T cell activation and survival in mouse models. Indeed, the production of type I IFNs downstream of the cGAS-STING cascade can act upon T cells through a positive feedback loop that results in enhanced effector functions and survival (Zhu et al., 2019). However, as previously described increasing evidence points to STING-triggered cell apoptosis in T cells. Though the mechanism through which STING regulates anti-tumour immunity has not been elucidated, evidence suggests that its tumour-promoting activity occurs in a

type I IFN-independent manner. Having incorporated these assumptions into the GRN, stable-state analysis in the presence of a STING agonist shows direct impairment of T cell survival. On the other hand, simulation of a TBK1 agonist resulted in improved survival and the promotion of inflammatory cytokines. The development of small-molecule agonists of TBK1 could present a potential solution to bypass the apoptotic feature of STING whilst stimulating type I IFN production to further promote anti-tumour immunity.

The cGAS-STING cascade was added to the MAPK GRN used in MicroC to model tumour cells. Relationships between nodes were inferred from literature search results, findings from Chapter 4, and predicted or known genetic interactions found using the STRING and KEGG databases. The cGAS-STING cascade is triggered by either DNA damage or viral replication. As reported by S. Y. Wu et al. (2021), MYC inhibits STING whilst hypoxia-induced HIF-1 negatively regulates IRF3 (T. Peng et al., 2021). As stated above, there is also evidence in the literature that TBK1 engages with Akt to promote survival (Ou et al., 2011). Similarly, STING has been found to downregulate anti-apoptotic protein BCL-2. PD-L1 was added downstream of AP1 which has been reported to bind and transactivate its expression (Y. Wang et al., 2018). Activated cGAS was also set to promote PD-L1 expression based on the spatial transcriptomics results. Oncolytic therapy was added as an input node promoting intracellular replication which can be inhibited by the IFN receptor input node. This was meant to reflect the signalling cascade triggered by type I IFN binding to IFNAR that functions to inhibit viral replication (Yang & Li, 2020). Viral replication on the other hand is detected by cGAS, inhibits ATM from the DNA damage response pathway and promotes cellular apoptosis (Kleinberger, 2020). Though tumour cells with STING deficiency have been identified to be more susceptible to oncolytic therapy, viral mechanisms are known to inhibit STING activation. For simplicity, this GRN does not consider potential inhibitory mechanisms of the cGAS-STING cascade by the virus. Interestingly, a recent study by Froechlich et al. (2020) suggested that STING-knockout favours viral replication but results in impaired immunogenic cell death. Their results suggest that STING contributes to the immunotherapeutic efficacy of oncolytic viruses. As such, when considering the combination of oncolytic virotherapy and T cells, STING integrity may be key to bridge the innate anti-viral and adaptive anti-

tumour immunity. With all these encoded interacting relationships, the stable-states of the network were computed under different conditions for p53 KO cells. Under hypoxic conditions, cGAS agonists did not lead to the expression of IFNs due to the inhibition of IRF3 by HIF1. It resulted in two cyclic attractors both undergoing necrosis and expressing PD-L1 with one attractor inhibiting STING through MYC. A switch to a normoxic state resulted in IFNs being produced. Promoting STING activation instead of cGAS helped bypass PD-L1 expression in one of the cyclic attractor. Importantly, the combination of oncolytic viruses, in conjunction with the inhibition of IFNR and PD-L1 was shown to strongly induce cell apoptosis and IFN production. Viral infection promotes the transcription of IFN genes through stimulation of the cGAS-STING cascade. The release of IFNs into the environment could further promote T cell infiltration. The simultaneous inhibition of IFNR using an antibody like anifrolumab (AstraZeneca) would prevent these IFNs from inhibiting oncolytic virotherapy. In addition, PD-L1 blockade using an antibody like avelumab (Merck KGaA and Pfizer) would prevent the inhibition of infiltrating T cells through the PD-1/PDL-1 axis. This is consistent with findings from H. Wang et al. (2017) and El-Sayes et al. (2022) who show that IFNAR and PD-L1 blockade could synergise with oncolytic virotherapy and T cell immunotherapy.

Boolean networks do not consider varying expression levels and are confined to binary states. To overcome this limitation and accommodate the stochastic nature of biological processes, the GRNs could be converted to probabilistic Boolean networks using the scRNA-seq data presented in Chapter 4. This approach could help capture the effect of tuning cGAS-STING cascade elements on cellular phenotypes. Of note, due to the conflicting evidence reported in the literature on the cGAS-STING pathway, important assumptions had to be made and future work should revisit interaction rules in the light of new evidence as it emerges.

All the work presented in this thesis will ultimately feed into the novel multiscale modelling framework for perturbation biology, MicroC, developed by the Buffa Lab at the University of Oxford. First, the GRN in tumour agents will be replaced by the new version presented in this chapter. Secondly, the GRN adapted for CAR T cell modelling will be

used to create CAR T cell agents which will incorporate CAR design considerations that were investigated using the ABM in this chapter. Furthermore, the T cell expression atlas constructed in Chapter 4 will be mapped to the GRN to accommodate probabilistic interactions. It would also be interesting to map the scRNA-seq data generated using the miSFIT variants to study the effect of CAR tuning at the intracellular level. The knowledge gained on T cell identification methods in Chapter 3 will help further isolate T cell subtypes from scRNA-seq experiments to select data for GRN fitting. Similarly, the epithelial cluster identified in the scRNA-seq data will be mapped to the tumour cell GRN. The spatial transcriptomics data could also be used to position cell populations. Both IFN expression and oncolytic virus will be implemented as diffusible agents. These new additions will enable the modelling of cGAS-STING signalling and immunotherapeutic strategies in the context of TNBC. More specifically, combinations of immune checkpoint inhibitors, CAR T cell therapy, oncolytic virotherapy and the addition of cGAS-STING agonists or antagonists will be testable in the MicroC environment. MicroC is an ongoing project that now aims to incorporate machine learning algorithms to create truly adaptive agents which can not only modify their actions but the strategies they use to determine them. The marriage between the two will enable agents to refine their internal models using algorithms like reinforcement learning, and they, in turn, will use the ABM as an environment and reward generator (Fig. E.1). This will transform MicroC into a multi-agent system that will be used to model tumour cell evolution and study cancer drug resistance, making it the ideal platform to test therapeutic strategies in silico. Diverse tools have emerged from the field of synthetic biology to program immune cells and deploy them against a broad range of cancers. Engineering designs must primarily ensure their safety and efficacy. By guiding the design of novel technologies and artificial biological systems, bridging the gap between systems biology and synthetic biology has the power to drive significant advances in biomedicine.

5.6 Conclusion

In this chapter, modelling paradigms were exploited to explore the effect of CAR tuning on on-target off-tumour toxicity and how varying strategies targeting T cell-intrinsic type I IFN pathways influence CAR T cell activity. Specifically, an ABM that can be calibrated to experimental data was developed to test whether an optimal level of CAR expression can reduce recognition of TAA-expressing wild-type cells whilst maintaining cytolytic activity directed towards TAA-overexpressing tumour cells. Finally, steady-state analysis of a GRN model suggests that careful targeting of downstream components of the cGAS-STING pathway holds promise to further combat T cell dysfunctional states. Models tailored to CAR T cell therapy in solid tumours are scarce, thus the work presented therein represents a starting point for the development of novel approaches to generate or test new hypotheses relevant to CAR T cell design. Importantly, the research conducted in this chapter will feed into a previously published in silico framework to enable the modelling of patient-specific responses to T cell-based immunotherapy and synergistic combinations.

CHAPTER 6

Conclusions

The main focus of this thesis was to investigate the interplay between type I interferon (IFN) signalling pathways and T cell immunity using a multi-omics approach. More specifically, the GMP-AMP synthase (cGAS)-stimulator of IFN genes (STING) pathway was studied in the context of the tumour microenvironment (TME) of triple-negative breast cancer (TNBC), an aggressive cancer that lacks effective therapies. Genetically modifying T cells with chimeric antigen receptors (CARs) offers a promising immunotherapeutic strategy. However, once at the tumour site CAR T cells face several challenges that dampen their therapeutic efficacy. Engineering strategies or drug combinations need to be carefully orchestrated to counteract the immunosuppressive milieu and enhance anti-tumour immunity.

6.1 Identifying T Cells with Single-Cell RNA Sequencing

Motivated to dive into the diverse T cell landscape, the thesis began with the analysis of existing single-cell RNA sequencing (scRNA-seq) datasets to dissect gene expression programs underlying T cell subtypes. It was shown how the traditional cell-type identification method based on the partitioning and clustering of single cells risks blurring the distinction between T cell populations. Gene signatures of both conventional and unconventional T cells were identified and different cell type prediction methods were evaluated. Performance was found to highly vary for different T cell classes, each method having its own advantages and limitations. These findings led to the recommendation of applying multi-

ple complementary methods for T cell subtype identification in scRNA-seq data. Future work should consider developing a framework to facilitate the systematic assessment of cell type identifiers being used to describe T cell heterogeneity.

6.2 Multi-Omics Analysis of T Cells and cGAS-STING

In order to examine T cell populations and IFN pathways in the context of breast cancer, the next chapter combined individual omics data generated from tissue biopsies. More specifically, scRNA-seq, scTCR-seq, scBCR-seq, Slide-seq, in situ sequencing and Molecular Cartography were used to capture cellular profiles and gene expression patterns at varying levels of resolution. The comprehensive analysis of T cells revealed the relative contribution of distinct transcriptomic programmes to the immune infiltrate. Furthermore, this study enabled the evaluation and comparison of spatial transcriptomics platforms for exploring the spatial organisation of T cells in the breast TME. Genes of the cGAS-STING signalling pathway were found to be most actively transcribed in endothelial cells. In T cells, *cGAS* was predominantly expressed in $Ki67^{High}$ T cells. The spatial distribution of RNA transcripts in the breast TME indicated that *PD-L1* and *PD-1* transcripts are strongly enriched with *cGAS*. Similarly, transcripts encoding the immunosuppressive proteins IDO1, PD-L1 and LGALS9 were found to highly colocalise with *IFNAR2*, *MX1*, *STAT1* and *IFIT1*. In addition, *TBK1* transcripts associated with hypoxia markers. Further studies should be conducted to confirm these relationships and investigate their potential role in immune escape. Moreover, it would be interesting to use bifurcation analysis to further explore the dichotomous nature of cGAS-STING signalling in cancer. Fine-tuning the expression of cGAS-STING components may be key to exploiting this pathway for therapeutic benefit and to combat T cell exhaustion.

6.3 Modelling T Cell Immunotherapy

Finally, modelling paradigms were implemented to study cancer immunotherapy at the molecular and cellular level. An agent-based model (ABM) was designed to simulate the

effect of CAR tuning on on-target off-tumour toxicity using properties derived from both theory and data generated in the laboratory. Multiple hypothetical scenarios were set up in order to compare the outcomes of different parameter sets. Whilst the model needs further refinement and validation, initial results suggest how optimising the level of CAR can help minimise interaction with normal tissue minimally expressing the target antigen but it cannot prevent it. Future work should also consider other aspects of CAR design including CAR affinity and the choice of costimulatory domain. Next, two gene regulatory network (GRN) models were expanded to simulate the effect of cGAS-STING agonists and inhibitors on CAR T cells and tumour cells. Steady-state analysis results reflect the dichotomous functions of this pathway which warrants further investigation. Elucidating the T cell-intrinsic role of the cGAS-STING cascade will be particularly important. An interesting extension of this work will be the conversion of these GRNs into probabilistic Boolean networks that can be used to fine-tune cGAS-STING interactions. Expression of PD-L1 and susceptibility to oncolytic virotherapy were also implemented into the tumour GRN. These additions will further enable the modelling of synergistic combinations once the GRNs are incorporated into the novel multiscale virtual microenvironment, MicroC.

6.4 Concluding Remarks

Integrating omics data from multiple modalities plays a central role in cataloguing and understanding the relationship between genotypes and phenotypes. Moreover, mathematical and computational modelling provide powerful means to widely explore parameter space and test hypotheses. Combining multi-omics and modelling paradigms could facilitate the discovery of more effective therapeutic strategies to maximise the benefits of T cell immunotherapy. Notably, it could help weigh the benefits and risks of different CAR T cell designs and suggest ways to harness the cGAS-STING pathway to drive anti-tumour responses whilst reducing immune-suppressive effects. The work presented in this thesis will hopefully represent a small scientific leap forward towards this endeavour and pave the way for the development of modelling approaches to understand variability in patient-response to immunotherapy.

Analysis of the data presented in Chapter 4 continued following submission of the thesis for examination: Pavillet et al. (2023). Spatial Multi-Omics Characterisation of cGAS-STING Signalling and Tumour-Infiltrating T Cells in Breast Cancer. Manuscript in preparation.

Bibliography

Abdelaal, T., Michielsen, L., Cats, D., Hoogduin, D., Mei, H., Reinders, M. J. T., & Mahfouz, A. (2019). A comparison of automatic cell identification methods for single-cell RNA sequencing data. *Genome Biology*, *20*. doi: 10.1186/s13059-019-1795-z

Ahlmann-Eltze, C., & Huber, W. (2021). Transformation and Preprocessing of Single-Cell RNA-Seq Data. *bioRxiv*. doi: 10.1101/2021.06.24.449781

Allen, B. M., Hiam, K. J., Burnett, C. E., Venida, A., DeBarge, R., Tenvooren, I., ... Spitzer, M. H. (2020). Systemic dysfunction and plasticity of the immune macroenvironment in cancer models. *Nat Med*, *26*, 125–1134. doi: 10.1038/s41591-020-0892-6

Alquicira-Hernandez, J., Sathe, A., Ji, H. P., Nguyen, Q., & Powell, J. E. (2019). scPred: accurate supervised method for cell-type classification from single-cell RNA-seq data. *Genome Biol*, *20*. doi: 10.1186/s13059-019-1862-5

Andreatta, M., Corria-Osorio, J., Muller, S., Cubas, R., Coukos, G., & Carmona, S. J. (2021). Interpretation of T cell states from single-cell transcriptomics data using reference atlases. *Nat Commun*, *12*. doi: 10.1038/s41467-021-23324-4

Andreattaa, M., & Carmonaa, S. J. (2021). UCell: Robust and scalable single-cell gene signature scoring. *Comput Struct Biotechnol J*, *19*, 3796–3798. doi: 10.1016/j.csbj.2021.06.043

Anikeeva, N., Panteleev, S., Mazzanti, N. W., Terai, M., Sato, T., & Sykulev, Y. (2021). Efficient killing of tumor cells by CAR-T cells requires greater number of engaged CARs than TCRs. *GigaScience*, *297*. doi: 10.1016/j.jbc.2021.101033

Aran, D., Looney, A. P., Liu, L., Wu, E., Fong, V., Hsu, A., ... Bhattacharya,

M. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature Immunology*, *20*, 163–172. doi: 10.1038/s41590-018-0276-y

Arnaiz, E., & Harris, A. L. (2022). Role of Hypoxia in the Interferon Response. *Front. Immunol.*, *13*. doi: 10.3389/fimmu.2022.821816

Arunachalam, P. S., Scott, M. K. D., Hagan, T., Li, C., Feng, Y., Wimmers, F., … Pulendran, B. (2021). Systems vaccinology of the BNT162b2 mRNA vaccine in humans. *Nature*, *596*, 410–416. doi: 10.1038/s41586-021-03791-x

Aslam, B., Basit, M., Nisar, M. A., Khurshid, M., & Rasool, M. H. (2017). Proteomics: Technologies and Their Applications. *Journal of Chromatographic Science*, *55*, 182–196. doi: 10.1093/chromsci/bmw167

Auchincloss, A. H., & Garcia, L. M. T. (2015). Brief introductory guide to agent-based modeling and an illustration from urban health research. *Cad Saude Publica*, *1*, 65–78. doi: 10.1590/0102-311X00051615

Baghban, R., Roshangar, L., Jahanban-Esfahlan, R., Seidi, K., Ebrahimi-Kalan, A., Jaymand, M., … Zare, P. (2020). Tumor microenvironment complexity and therapeutic implications at a glance. *Cell Communication and Signaling*, *18*. doi: 10.1186/s12964-020-0530-4

Bankhead, P., Loughrey, M. B., Fernandez, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., … Hamilton, P. W. (2017). QuPath: Open source software for digital pathology image analysis. *Scientific Reports*, *7*. doi: 10.1038/s41598-017-17204-5

Baranova, A., Cao, H., & Zhang, F. (2021). Unraveling Risk Genes of COVID-19 by Multi-Omics Integrative Analyses. *Front. Med.*. doi: 10.3389/fmed.2021.738687

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., … Soboleva, A. (2013). NCBI GEO: Archive for functional genomics data sets – update. *Nucleic Acids Res.*, *41*. doi: 10.1093/nar/gks1193

Barsheshet, Y., Wildbaum, G., Levy, E., Vitenshtein, A., Akinseye, C., Griggs, J., … Karin, N. (2017). CCR8+ FOXp3+ Treg cells as master drivers of immune

regulation. *Proceedings of the National Academy of Sciences*, *114*(23), 6086–6091. doi: 10.1073/pnas.1621280114

Beck, M. A., Fischer, H., Grabner, L. M., Groffics, T., Winter, M., Tangermann, S., ... Seiser, C. (2021). DNA hypomethylation leads to cGAS-induced autoinflammation in the epidermis. *The EMBO Journal*, *40*(22). doi: https://doi.org/10.15252/embj.2021108234

Bellone, M., Calcinotto, A., Filipazzi, P., Milito, A. D., Fais, S., & Rivoltini, L. (2013). The acidity of the tumor microenvironment is a mechanism of immune escape that can be overcome by proton pump inhibitors. *Oncoimmunology*, *2*. doi: 10.4161/onci.22058

Benoit-Lizon, I., Jacquin, E., Vargas, T. R., Richard, C., Roussey, A., Zuffo, L. D., ... Apetoh, L. (2022). CD4 T cell-intrinsic STING signaling controls the differentiation and effector functions of TH1 and TH9 cells. *Journal for ImmunoTherapy*, *10*. doi: 10.1136/jitc-2021-003459

Bommer, U.-A., & Telerman, A. (2020). Dysregulation of TCTP in Biological Processes and Diseases. *Cells*, *9*. doi: 10.3390/cells9071632

Borcherding, N., Bormann, N. L., & Kraus, G. (2020). scRepertoire: An R-based toolkit for single-cell immune receptor analysis. *F1000Research*, *9*([version 2; peer review: 2 approved]). doi: 10.12688/f1000research.22139.2

Boukhaled, G. M., Harding, S., & Brooks, D. G. (2021). Opposing Roles of Type I Interferons in Cancer Immunity. *Annu Rev Pathol*, *24*, 167–198. doi: 10.1146/annurev-pathol-031920-093932

Brandstadter, J. D., & Maillard, I. (2019). Notch signalling in T cell homeostasis and differentiation. *Open biology*, *9*. doi: 10.1098/rsob.190187

Brown, N. A., & Bron, A. J. (1987). An estimate of the human lens epithelial cell size in vivo. *Exp Eye Res*, *44*, 899–906. doi: 10.1016/s0014-4835(87)80052-0

Buenrostro, J., Wu, B., Chang, H., & Greenleaf, W. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current protocols in molecular biology*, *109*, 669–680. doi: 10.1002/0471142727.mb2129s109

Burel, J. G., Pomaznoy, M., Lindestam Arlehamn, C. S., Weiskopf, D., da Silva An-

tunes, R., Jung, Y., ... Peters, B. (2019). Circulating T cell-monocyte complexes are markers of immune perturbations. *eLife*, *8*, e46045. doi: 10.7554/eLife.46045

Butler, A., Hoffman, P., Hafemeister, C., Papalexi1, E., Mauck, W. M., Hao, Y., ... Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, *177*, 1888–1902. doi: 10.1016/j.cell.2019.05.031

Cano-Gamez, E., Soskic, B., Roumeliotis, T. I., Ernest So, D. J. S., Baldrighi, M., Wille, D., ... Trynka, G. (2020). Single-cell transcriptomics identifies an effectorness gradient shaping the response of CD4+ T cells to cytokines. *Nature Communications*, *11*. doi: 10.1038/s41467-020-15543-y

Caruso, H. G., Heimberger, A. B., & Cooper, L. J. N. (2018). Steering CAR T cells to distinguish friend from foe. *Oncoimmunology*, *8*. doi: 10.1080/2162402X.2016.1271857

Cerboni, S., Jeremiah, N., Gentili, M., Gehrmann, U., Conrad, C., Stolzenberg, M.-C., ... Manel, N. (2017). Intrinsic antiproliferative activity of the innate sensor STING in T lymphocytes. *J Exp Med*, *214*, 1769–1785. doi: 10.1084/jem.20161674

Charan, M., Das, S., Mishra, S., Chatterjee, N., Varikuti, S., Kaul, K., ... Ganju, R. K. (2020). Macrophage migration inhibitory factor inhibition as a novel therapeutic approach against triple-negative breast cancer. *Cell Death & Disease*, *11*. doi: 10.1038/s41419-020-02992-y

Chari, T., Banerjee, J., & Pachter, L. (2021). The Specious Art of Single-Cell Genomics. *bioRxiv*. doi: 10.1101/2021.08.25.457696

Chechlinska, M., Kowalewska, M., & Nowak, R. (2010). Systemic inflammation as a confounding factor in cancer biomarker discovery and validation. *Nature Reviews Cancer*, *10*. doi: 10.1038/nrc2782

Chen, B., Alvarado, D. M., Iticovici, M., Kau, N. S., Park, H., Parikh, P. J., ... Ciorba, M. A. (2020). Interferon-Induced IDO1 Mediates Radiation Resistance and Is a Therapeutic Target in Colorectal Cancer. *Cancer Immunol*, *8*, 451–464. doi: 10.1158/2326-6066.CIR-19-0282

Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S., & Zhuang, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. *Science*, *348*. doi: 10.1126/science.aaa6090

Chen, Y., Pal, B., Lindeman, G. J., Visvader, J. E., & Smyth, G. K. (2022). R code and downstream analysis objects for the scRNA-seq atlas of normal and tumorigenic human breast tissue. *Scientific Data*, *9*. doi: 10.1038/s41597-022-01236-2

Cock, P., Fields, C., Goto, N., Heuer, M., & Rice, P. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, *6*, 1767–1771.

COMBAT. (2022). A blood atlas of COVID-19 defines hallmarks of disease severity and specificity. *Cell*, *185*, 916–938. doi: 10.1016/j.cell.2022.01.012

Crespo, J., Sun, H., Welling, T. H., Tian, Z., & Zou, W. (2013). T cell anergy, exhaustion, senescence, and stemness in the tumor microenvironment. *Curr Opin Immunol*, *25*, 214–221. doi: 10.1016/j.coi.2012.12.003

Critchley-Thorne, R. J., Simons, D. L., Yan, N., Miyahira, A. K., Dirbas, F. M., Johnson, D. L., … Lee, P. P. (2009). Impaired interferon signaling is a common immune defect in human cancer. *Proc. Natl. Acad. Sci. U. S. A.*, *106*, 9010–9015. doi: https://doi.org/10.1073/pnas.0901329106

David, M., Dursi, L., Yao, D., Boutros, P., & Simpson, J. (2017). Nanocall: an open source basecaller for Oxford Nanopore sequencing data. *Bioinformatics*, *1*, 49–55.

Davison, A. C., & Hinkley, D. V. (1997). Bootstrap Methods and Their Application. *Cambridge university press*.

Decout, A., Katz, J. D., Venkatraman, S., & Ablasser, A. (2021). The cGAS-STING pathway as a therapeutic target in inflammatory diseases. *Nat Rev Immunol*, *21*, 548–569. doi: 10.1038/s41577-021-00524-z

Denisenko, E., Guo, B. B., Jones, M., Hou, R., de Kock, L., Lassmann, T., … Forrest, A. R. R. (2020). Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol*, *21*. doi: 10.1186/s13059-020-02048-6

Derosa, L., Melenotte, C., Griscelli, F., Gachot, B., Marabelle, A., Kroemer, G., & Zitvogel, L. (2020). The immuno-oncological challenge of COVID-19. *Nat Cancer*, *1*, 946–964. doi: 10.1038/s43018-020-00122-3

Dhatchinamoorthy, K., Colbert, J. D., & Rock, K. L. (2021). Cancer Immune Evasion Through Loss of MHC Class I Antigen Presentation. *Front Immunol*, *12*. doi: 10.3389/fimmu.2021.636568

Diaz, A. G., Shin, D. S., Moreno, B. H., Saco, J., Escuin-Ordinas, H., Rodriguez, G. A., ... Ribas, A. (2017). Interferon Receptor Signaling Pathways Regulating PD-L1 and PD-L2 Expression. *Cell Rep.*, *19*, 1189–1201. doi: 10.1016/j.celrep.2017.04.031

Dirks, R. M., Bois, J. S., Schaeffer, J. M., Winfree, E., & Pierce, N. A. (2007). Thermodynamic Analysis of Interacting Nucleic Acid Strands. *SIAM Review*, *49*, 65–88. doi: 10.1137/060651100

Duraiswamy, J., Ibegbu, C. C., Masopust, D., Miller, J. D., Araki, K., Doho, G. H., ... Ahmed, R. (2011). Phenotype, function, and gene expression profiles of programmed death-1(hi) CD8 T cells in healthy human adults. *J Immunol*, *186*, 4200–12. doi: 10.4049/jimmunol.1001783

Eddy, S. (2004). What is a hidden Markov model? *Nature Biotechnology*, *22*, 1315–1316.

Efremova, M., Vento-Tormo, M., Teichmann, S. A., & Vento-Tormo, R. (2020). CellPhoneDB: Inferring cell-cell communication from combined expression of multi-subunit receptor-ligand complexes. *Nature Protocols*, *15*, 1484–1506. doi: 10.1038/s41596-020-0292-x

El-Sayes, N., Walsh, S., Vito, A., Reihani, A., Ask, K., Wan, Y., & Mossman, K. (2022). IFNAR blockade synergizes with oncolytic VSV to prevent virus-mediated PD-L1 expression and promote antitumor T cell activity. *Molecular Therapy - Oncolytics*, *25*, 16-30. doi: https://doi.org/10.1016/j.omto.2022.03.006

Evgin, L., Kottke, T., Tonne, J., Thompson, J., Huff, A. L., van Vloten, J., ... Vile, R. G. (2022). Oncolytic virus-mediated expansion of dual-specific CAR T cells

improves efficacy against solid tumors in mice. *Science Translational Medicine*, *14*. doi: 10.1126/scitranslmed.abn22

Eyre, T. A., Wright, M. W., Lush, M. J., & Bruford, E. A. (2007). HCOP: a searchable database of human orthology predictions. *Briefings in Bioinformatics*, *8*, 2–5. doi: 10.1093/bib/bbl030

Figueiredo, C. R., Azevedo, R. A., Mousdell, S., Lara, P. T. R., Ireland, L., Santos, A., ... Mielgo, A. (2018). Blockade of MIF CD74 Signalling on Macrophages and Dendritic Cells Restores the Antitumour Immune Response Against Metastatic Melanoma. *Front. Immunol.*, *9*. doi: 10.3389/fimmu.2018.01132

Fischer, J., & Ayers, T. (2021). Single nucleus RNA-sequencing: how it's done, applications and limitations. *Emerg Top Life Sci*, *5*, 687–690. doi: 10.1042/ETLS20210074

Flament, H., Rouland, M., Beaudoin, L., Toubal, A., Bertrand, L., Lebourgeois, S., ... Lehue, A. (2021). Outcome of SARS-CoV-2 infection is linked to MAIT cell activation and cytotoxicity. *Nat Immunol*, *22*, 322–335. doi: 10.1038/s41590-021-00870-z

Fricke, G. M., Letendre, K. A., Moses, M. E., & Cannon, J. L. (2016). Persistence and Adaptation in Immunity: T Cells Balance the Extent and Thoroughness of Search. *PLoS Comput Biol*, *12*. doi: 10.1371/journal.pcbi.1004818

Fridman, W. H. (2018). From Cancer Immune Surveillance to Cancer Immunoediting: Birth of Modern Immuno-Oncology. *The Journal of Immunology*, *201*(3), 825–826. doi: 10.4049/jimmunol.1800827

Froechlich, G., Caiazza, C., Gentile, C., D'Alise, A. M., Lucia, M. D., Langone, F., ... Zambrano, N. (2020). Integrity of the Antiviral STING-mediated DNA Sensing in Tumor Cells Is Required to Sustain the Immunotherapeutic Efficacy of Herpes Simplex Oncolytic Virus. *Cancers (Basel)*, *12*, 3407. doi: 10.3390/cancers1211340

Furey, T. S. (2012). ChIP-seq: and beyond new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet*, *13*, 840–852. doi: 10.1002/nrg3306

Gerlee, P. (2013). The Model Muddle: In Search of Tumor Growth Laws. *Cancer Research*, *73*(8), 2407–2411. doi: 10.1158/0008-5472.CAN-12-4355

Gerstung, M., Jolly, C., Leshchiner, I., Dentro, S. C., Gonzalez, S., Rosebrock, D., ... Consortium, P. (2020). The evolutionary history of 2,658 cancers. *Nature*, *578*, 122–128. doi: 10.1038/s41586-019-1907-7

Golubovskaya, V., & Wu, L. (2016). Different Subsets of T Cells, Memory, Effector Functions, and CAR-T Immunotherapy. *Cancers*, *8*(36). doi: 10.3390/cancers8030036

Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for Multi-Class Classification: An Overview. *arXiv*. doi: 10.48550/arXiv.2008.05756

Green, D. R., Droin, N., & Pinkoski, M. (2003). Activation-induced cell death in T cells. *Immunological reviews*, *193*, 70–81. doi: 10.1034/j.1600-065x.2003.00051.x

Grieco, L., Calzone, L., Bernard-Pierrot, I., Radvanyi, F., Kahn-Perles, B., & Thieffry, D. (2013). Integrative Modelling of the Influence of MAPK Network on Cancer Cell Fate Decision. *PLoS Comput Biol*, *9*(10). doi: 10.1371/journal.pcbi.1003286

Guedan, S., & Alemany, R. (2018). CAR-T Cells and Oncolytic Viruses: Joining Forces to Overcome the Solid Tumor Challenge. *Front Immunol*, *9*. doi: 10.3389/fimmu.2018.02460

Gulen, M. F., Ute Koch, S. M. H., Schuler, F., Apetoh, L., Villunger, A., Radtke, F., & Ablasser, A. (2017). Signalling strength determines proapoptotic functions of STING. *Nat Commun*, *8*. doi: 10.1038/s41467-017-00573-w

Habiger, J. D., & Pena, E. A. (2011). Randomised P-values and nonparametric procedures in multiple testing. *Nonparametr Stat*, *23*, 583–604. doi: 10.1080/10485252.2010.482154

Hadjadj, J., Yatim, N., Barnabei, L., Corneau, A., Boussier, J., Smith, N., ... Terrier, B. (2020). Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science*, *369*, 718–724. doi: 10.1126/science.abc6027

Hafemeister, C., & Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol*, *23*(296). doi: 10.1186/s13059-019-1874-1

Han, J., Deng, X., Sun, R., Luo, M., Liang, M., Gu, B., ... Luo, Z. (2021). GPI Is a Prognostic Biomarker and Correlates With Immune Infiltrates in Lung Adenocarcinoma. *Front. Oncol.*, *11*. doi: 10.3389/fonc.2021.752642

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., ... Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, *184*, 3573–3587.e29. doi: 10.1016/j.cell.2021.04.048

Haque, A., Engel, J., Teichmann, S. A., & Lonnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Medicine*, *9*. doi: 10.1186/s13073-017-0467-4

Herck, S. V., Feng, B., & Tang, L. (2021). Delivery of STING agonists for adjuvanting subunit vaccines. *Advanced Drug Delivery Reviews*, *179*. doi: 10.1016/j.addr.2021.114020

Hesterberg, T. (2011). Bootstrap. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*, 497–526.

Hinks, T. S. C., & Zhang, X. W. (2020). MAIT Cell Activation and Functions. *Frontiers in Immunology*, *11*. doi: 10.3389/fimmu.2020.01014

Hosseinkhani, N., Derakhshani, A., Kooshkaki, O., Shadbad, M. A., Hajiasgharzadeh, K., Baghbanzadeh, A., ... Baradaran, B. (2020). Immune Checkpoints and CAR-T Cells: The Pioneers in Future Cancer Therapies? *Int J Mol Sci*, *21*. doi: 10.3390/ijms21218305

Hou, A. J., Chen, L. C., & Chen, Y. Y. (2021). Navigating CAR-T cells through the solid-tumour microenvironment. *Nature Reviews Drug Discovery*, *20*, 531–550. doi: 10.1038/s41573-021-00189-2

Huang, Q., Liu, Y., Du, Y., & Garmire, L. X. (2021). Evaluation of Cell Type Annotation R Packages on Single-cell RNA-seq Data. *Genomics Proteomics Bioinformatics*, *19*, 267–281. doi: 10.1016/j.gpb.2020.07.004

Ianevski, A., Giri, A. K., & Aittokallio, T. (2022). Fully-automated and ultra-fast cell-

type identification using specific marker combinations from single-cell transcriptomic data. *Nat Commun*, *13*. doi: 10.1038/s41467-022-28803-w

Ilieva, K. M., Cheung, A., Mele, S., Chiaruttini, G., Crescioli, S., Griffin, M., ... Karagiannis, S. N. (2018). Chondroitin Sulfate Proteoglycan 4 and Its Potential As an Antibody Immunotherapy Target across Different Tumor Types. *Front. Immunol.*, *8*. doi: 10.3389/fimmu.2017.01911

ImmunoMind. (2019). immunarch: An R Package for Painless Analysis of Large-Scale Immune Repertoire Data. *Online*. doi: 10.5281/zenodo.3367200

Interlandi, M., Kerl, K., & Dugas, M. (2022). InterCellar enables interactive analysis and exploration of cell-cell communication in single-cell transcriptomic data. *Communications Biology*, *5*. doi: 10.1038/s42003-021-02986-2

Ivetic, A., Green, H. L. H., & Hart, S. J. (2019). L-selectin: A Major Regulator of Leukocyte Adhesion, Migration and Signaling. *Front. Immunol.*. doi: 10.3389/fimmu.2019.01068

Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology*, *17*. doi: 10.1186/s13059-016-1103-0

Jiang, M., Chen, P., Wang, L., Li, W., Chen, B., Liu, Y., ... Zhou, C. (2020). cGAS-STING, an important pathway in cancer immunotherapy. *J Hematol Oncol*, *13*. doi: 10.1186/s13045-020-00916-z

Joyce, M., & James, M. (2011). Kullback-Leibler Divergence. *International Encyclopedia of Statistical Science*, *1*, 720–722.

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., & Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res*, *49*, D545–D551. doi: 10.1093/nar/gkaa970

Karlsson, M., Zhang, C., Mear, L., Zhong, W., Digre, A., Katona, B., ... Lindskog, C. (2021). A single-cell type transcriptomics map of human tissues. *Sci Adv*, *7*. doi: 10.1126/sciadv.abh2169

Katlinski, K., Gui, J., Katlinskaya, Y., Ortiz, A., Chakraborty, R., Bhattacharya, S., ... Fuchs, S. (2017). Inactivation of Interferon Receptor Promotes the Es-

tablishment of Immune Privileged Tumor Microenvironment. *Cancer Cell*, *31*, 194–207. doi: https://doi.org/10.1016/j.ccell.2017.01.004

Kazil, J., Masad, D., & Crooks, A. (2020). Utilizing Python for Agent-Based Modeling: The Mesa Framework. *Social, Cultural, and Behavioral Modeling*, 308–317.

Kennedy, R. C., Ropella, G. E., & Hunt, A. (2016). A cell-centered, agent-based framework that enables flexible environment granularities. *Theor Biol Med Model*, *13*(4). doi: 10.1186/s12976-016-0030-9

Kieran, J., William, J., & Graham, A. (2021). Non-Epithelial Stromal Cells in Thymus Development and Function. *Frontiers in Immunology*, *12*. doi: 10.3389/fimmu.2021.634367

Kim, J.-H., Kim, B. S., & Lee, S.-K. (2020). Regulatory T Cells in Tumor Microenvironment and Approach for Anticancer Immunotherapy. *Immune Netw*, *20*. doi: doi:10.4110/in.2020.20.e4

Kleinberger, T. (2020). En Guard! The Interactions between Adenoviruses and the DNA Damage Response. *Viruses*, *12*(996). doi: 10.3390/v12090996

Knapp, D. J. H. F., Michaels, Y. S., Jamilly, M., Ferry, Q. R. V., Barbosa, H., Milne, T. A., & Fulga, T. A. (2019). Decoupling tRNA promoter and processing activities enables specific Pol-II Cas9 guide RNA expression. *Nat Commun*, *10*. doi: 10.1038/s41467-019-09148-3

Krijgsman, D., Hokland, M., & Kuppen, P. J. K. (2018). The Role of Natural Killer T Cells in Cancer - A Phenotypical and Functional Approach. *Frontiers in Immunology*, *9*. doi: 10.3389/fimmu.2018.00367

Krummel, M. F., Bartumeus, F., & Gerard, A. (2016). T cell migration, search strategies and mechanism. *Nat Rev Immunol*, *16*, 193–201. doi: 10.1038/nri.2015.16

Kumar, B. V., Connors, T., & Farber, D. L. (2018). Human T Cell Development, Localization, and Function throughout Life. *Immunity*, *48*, 202–213. doi: 10.1016/j.immuni.2018.01.007

Kwon, J., & Bakhoum, S. F. (2020). The Cytosolic DNA-Sensing cGAS-STING Pathway in Cancer. *Cancer Discov*, *10*, 26–39. doi: 10.1158/2159-8290.CD-19

-0761

Labani-Motlagh, A., Ashja-Mahdavi, M., & Loskog, A. (2020). The Tumor Microenvironment: A Milieu Hindering and Obstructing Antitumor Immune Responses. *Frontiers in Immunology*, *11*. doi: 10.3389/fimmu.2020.00940

Lahnemann, D., Koster, J., Szczurek, E., McCarthy, D. J., Hicks, S. C., Robinson, M. D., ... Schonhuth, A. (2020). Eleven grand challenges in single-cell data science. *Genome Biology*, *21*. doi: 10.1186/s13059-020-1926-6

Lai, Y., Weng, J., Wei, X., Qin, L., Lai, P., Zhao, R., ... Li, P. (2018). Toll-like receptor 2 costimulation potentiates the antitumor efficacy of CAR T Cells. *Leukemia*, *32*, 801–808. doi: 10.1038/leu.2017.249

Larkin, B., Ilyukha, V., Sorokin, M., Buzdin, A., Vannier, E., & Poltorak, A. (2017). Cutting Edge: Activation of STING in T Cells Induces Type I IFN Responses and Cell Death. *J Immunol*, *199*, 397–402. doi: 10.4049/jimmunol.1601999

Lee, H., Salas, S. M., Gyllborg, D., & Nilsson, M. (2022). Direct RNA targeted in situ sequencing for transcriptomic profiling in tissue. *Scientific Reports*, *12*. doi: 10.1038/s41598-022-11534-9

Lee, H.-J., Song, K.-H., Oh, S. J., Kim, S., Eunho Cho, J. K., gyu Park, Y., ... Kim, T. W. (2022). Targeting TCTP sensitizes tumor to T cell-mediated therapy by reversing immune-refractory phenotypes. *Nat Commun*, *13*. doi: 10.1038/s41467-022-29611-y

Lee, J. H., Daugharthy, E. R., Scheiman, J., Kalhor, R., Ferrante, T. C., Yang, J. L., ... Church, G. M. (2014). Highly multiplexed subcellular RNA sequencing in situ. *Science*, *343*, 1360–1363. doi: 10.1126/science.1250212

Lee, J. S., Park, S., Jeong, H. W., Ahn, J. Y., Choi, S. J., Lee, H., ... Shin, E.-C. (2020). Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. *Science Immunology*, *5*. doi: 10.1126/sciimmunol.abd1554

Lemos, H., Mohamed, E., Huang, L., Ou, R., Pacholczyk, G., Arbab, A. S., ... Mellor, A. L. (2016). STING Promotes the Growth of Tumors Characterized by Low Antigenicity via IDO Activation. *Cancer Res*, *76*, 2076–2081. doi: 10.1158/

0008-5472.CAN-15-1456

Leshi, C., Don, K., & Sandhya, S. (2018). A Novel Data-Driven Boolean Model for Genetic Regulatory Networks. *Frontiers in Physiology*, *9*, 1328.

Lewis, S. A., Sureshchandra, S., Zulu, M. Z., Doratt, B., Jankeel, A., Ibraim, I. C., ... Messaoudi, I. (2021). Differential dynamics of peripheral immune responses to acute SARS-CoV-2 infection in older adults. *Nature Aging*, *1*, 1038–1052. doi: 10.1038/s43587-021-00127-2

Li, C., Jiang, P., Wei, S., Xu, X., & Wang, J. (2020). Regulatory T cells in tumor microenvironment: new mechanisms, potential therapeutic strategies and future prospects. *Mol Cancer*, *19*. doi: 10.1186/s12943-020-01234-1

Li, H., Handsaker, B., & Wysoker, A. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *16*, 2078–2079.

Li, W., Lu, L., Lu, J., Wang, X., Yang, C., Jin, J., ... Deng, L. (2020). cGAS-STING-mediated DNA sensing maintains CD8+ T cell stemness and promotes antitumor T cell therapy. *Science Translational Medicine*, *12*(549). doi: https://doi.org/10.1126/scitranslmed.aay9013

Li, X., Luo, O. J., Wang, P., Zheng, M., Wang, D., Piecuch, E., ... Ruan, Y. (2017). Long-read ChIA-PET for base-pair-resolution mapping of haplotype-specific chromatin interactions. *Nature Protocols*, *12*, 899–915. doi: 10.1038/nprot.2017.012

Li, Y., Wang, Z., Jiang, W., Zeng, H., Liu, Z., Lin, Z., ... Xu, J. (2020). Tumor-infiltrating TNFRSF9 CD8 T cells define different subsets of clear cell renal cell carcinoma with prognosis and immunotherapeutic response. *Oncoimmunology*, *9*. doi: 10.1080/2162402X.2020.1838141

Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*, *1*, 417–425. doi: 10.1016/j.cels.2015.12.004

li Xu, J., & Guo, Y. (2020). FCGR1A Serves as a Novel Biomarker and Correlates With Immune Infiltration in Four Cancer Types. *Front. Mol. Biosci.*, *7*. doi: 10.3389/fmolb.2020.581615

Lovely, G. A., Brewster, R. C., Schatz, D. G., Baltimore, D., & Phillips, R. (2015). Single-molecule analysis of RAG-mediated V(D)J DNA cleavage. *Proceedings of the National Academy of Sciences*, *112*(14), E1715-E1723. doi: 10.1073/pnas.1503477112

Lu, C., Klement, J. D., Smith, A. D., Yang, D., Waller, J. L., Browning, D. D., ... Liu, K. (2020). p50 suppresses cytotoxic T lymphocyte effector function to regulate tumor immune escape and response to immunotherapy. *J Immunother Cancer*, *8*. doi: 10.1136/jitc-2020-001365

Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics*, *14*(5), 265–279. doi: 10.1016/j.gpb.2016.05.004

Maekiniemi, A., Singer, R. H., & Tutucci, E. (2020). Single molecule mRNA fluorescent in situ hybridization combined with immunofluorescence in S. cerevisiae: Dataset and quantification. *Data in Brief*, *30*, 105511. doi: 10.1016/j.dib.2020.105511

Malekzadeh, P., Pasetto, A., Robbins, P. F., Parkhurst, M. R., Paria, B. C., Jia, L., ... Deniger, D. C. (2019). Neoantigen screening identifies broad TP53 mutant immunogenicity in patients with epithelial cancers. *J Clin Invest*, *129*, 1109–1114. doi: 10.1172/JCI123791

Malik-Sheriff, R. S., Glont, M., Nguyen, T. V. N., Tiwari, K., Roberts, M. G., Xavier, A., ... Hermjakob, H. (2020, 1). BioModels - 15 years of sharing computational models in life science. *Nucleic Acids Research*, *48*(D1), D407-D415. doi: 10.1093/nar/gkz1055

Manna, M. P. L., Orlando, V., Tamburini, B., Badami, G. D., Dieli, F., & Caccamo, N. (2020). Harnessing Unconventional T Cells for Immunotherapy of Tuberculosis. *Frontiers in Immunology*, *11*. doi: 10.3389/fimmu.2020.02107

Mao, S., Zhang, Y., Seelig, G., & Kannan, S. (2022). CellMeSH: probabilistic cell-type identification using indexed literature. *Bioinformatics*, *38*, 1393–1402. doi: 10.1093/bioinformatics/btab834

Martin, M. D., & Badovinac, V. P. (2018). Defining Memory CD8 T Cell. *Frontiers*

in Immunology, *9*. doi: 10.3389/fimmu.2018.02692

Marx, V. (2021). Method of the Year: spatially resolved transcriptomics. *Nat Methods*, *18*, 9–14. doi: 10.1038/s41592-020-01033-y

Maude, S., Laetsch, T., J. Buechner, S. R., M. Boyer, H. B., P. Bader, M. V., Stefanski, H., ... Grupp, S. (2018). Tisagenlecleucel in Children and Young Adults with B-Cell Lymphoblastic Leukemia. *The New England journal of medicine*, *378*, 439–448. doi: 10.1056/NEJMoa1709866

Maus, M. V., Plotkin, J., Jakka, G., Wolchok, J., Renner, C., & Sadelain, M. (2016). An MHC-restricted antibody-based chimeric antigen receptor requires TCR-like affinity to maintain antigen specificity. *Molecular Therapy Oncolytics*, *3*. doi: 10.1038/mto.2016.23

Michaels, Y. S., Barnkob, M. B., Barbosa, H., Baeumler, T. A., Thompson, M. K., Andre, V., ... Fulga, T. A. (2019). Precise tuning of gene expression levels in mammalian cells. *Nature Communications*, *10*. doi: 10.1038/s41467-019-08777-y

Miller, B. F., Huang, F., Atta, L., Sahoo, A., & Fan, J. (2022). Reference-free cell type deconvolution of multi-cellular pixel-resolution spatially resolved transcriptomics data. *Nat Commun*, *13*. doi: 10.1038/s41467-022-30033-z

Mishra, S., Srinivasan, S., Ma, C., & Zhang, N. (2021). CD8+ Regulatory T Cell - A Mystery to Be Revealed. *Front. Immunol.*, *12*. doi: 10.3389/fimmu.2021.708874

Mittrucker, H.-W., Visekruna, A., & Huber, M. (2014). Heterogeneity in the Differentiation and Function of CD8+ T Cells. *Arch. Immunol. Ther. Exp.*, *62*, 449–458. doi: 10.1007/s00005-014-0293-y

Monaco, G., Lee, B., Xu, W., Mustafah, S., Hwang, Y. Y., Carre, C., ... Larbi, A. (2019). RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Rep*, *26*, 1627–1640. doi: 10.1016/j.celrep.2019.01.041

Muller, C. A., Boemo, M. A., Spingardi, P., Kessler, B. M., Kriaucionis, S., Simpson, J. T., & Nieduszynski, C. A. (2019). Capturing the dynamics of genome repli-

cation on individual ultra-long Nanopore sequence reads. *Nature Methods*, *16*, 429–436. doi: 10.1038/s41592-019-0394-y

Muus, C., Luecken, M. D., Eraslan, G., Sikkema, L., Waghray, A., Heimberg, G., ... Network, T. H. C. A. L. B. (2021). Single-cell meta-analysis of SARS-CoV-2 entry genes across tissues and demographics. *Nature Medicine*, *27*, 546–559. doi: 10.1038/s41591-020-01227-z

Naldi, A., Hernandez, C., Abou-Jaoude, W., Monteiro, P. T., Chaouiya, C., & Thieffry, D. (2018). Logical Modeling and Analysis of Cellular Regulatory Networks With GINsim 3.0. *Front. Physiol.*. doi: 10.3389/fphys.2018.00646

Naldi, A., Remy, E., Thieffry, D., & Chaouiya, C. (2011). Dynamically consistent reduction of logical regulatory graphs. *Theoretical Computer Science*, *412*, 2207–2218. doi: 10.1016/j.tcs.2010.10.021

Naour, J. L., Zitvogel, L., Galluzzi, L., Vacchelli, E., & Kroemer, G. (2020). Trial watch: STING agonists in cancer therapy. *Oncoimmunology*, *9*. doi: 10.1080/2162402X.2020.1777624

Nguyen, V., & Griss, J. (2022). scAnnotatR: framework to accurately classify cell types in single-cell RNA-sequencing data. *BMC Bioinformatics*, *23*. doi: 10.1186/s12859-022-04574-5

Nunnari, G., Malaguarnera, L., Lazzara, F., Cambria, D., Distefano, G., Longo, A., ... Rosa, M. D. (2015). Th17 common genes in CD4 T-cells of HIV-1-infected naive patients and elite controllers. *Infectious Diseases and Tropical Medicine*, *1*.

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., ... Phillippy, A. M. (2022). The complete sequence of a human genome. *Science*, *376*(6588), 44–53. doi: 10.1126/science.abj6987

Ou, Y.-H., Torres, M., Ram, R., Formstecher, E., Roland, C., Cheng, T., ... White, M. A. (2011). TBK1 directly engages Akt/PKB survival signaling to support oncogenic transformation. *Mol Cell*, *41*, 458–470. doi: 10.1016/j.molcel.2011.01.019

Oxford COVID-19 Literature Initiative. (2020). *COVID-19 Literature Reviews.*

Retrieved 01-06-2022, from https://www.immunology.ox.ac.uk/covid-19/covid-19-immunology-literature-reviews

Ozawa, Y., Harutani, Y., Oyanagi, J., Akamatsu, H., Murakami, E., Shibaki, R., ... Yamamoto, N. (2021). Tumor expression of cGAS, not STING, negatively impacts on the efficacy of PD-1/L1 inhibitors in non-small cell lung cancer with PD-L1 TPS $\geq$ 50. *Journal of Clinical Oncology*, *39*, e21048-e21048. doi: 10.1200/JCO.2021.39.15_suppl.e21048

Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A. D., Rawlik, K., Pasko, D., ... Baillie, J. K. (2021). Genetic mechanisms of critical illness in COVID-19. *Nature*, *591*, 92–98. doi: 10.1038/s41586-020-03065-y

Pasquini, G., Arias, J. E. R., Schafer, P., & Busskamp, V. (2021). Automated methods for cell type annotation on scRNA-seq data. *Computational and Structural Biotechnology Journal*, *19*, 961–969. doi: 10.1016/j.csbj.2021.01.015

Patil, V. S., Schmiedel, M. B. J., Clarke, J., O'Rourke, P., de Silva, A. D., Harris, E., ... Vijayanand, P. (2018). Precursors of human CD4+ cytotoxic T lymphocytes identified by single-cell transcriptome analysis. *Science Immunology*, *3*. doi: 10.1126/sciimmunol.aan8664

Peng, T., Du, S.-Y., Son, M., & Diamond, B. (2021). HIF-1α is a negative regulator of interferon regulatory factors: Implications for interferon production by hypoxic monocytes. *Proceedings of the National Academy of Sciences*, *118*(26), e2106017118. doi: 10.1073/pnas.2106017118

Peng, Y., Mentzer, A. J., Liu, G., Yao, X., Yin, Z., Dong, D., ... Dong, T. (2020). Broad and strong memory CD4 and CD8 T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. *Nat Immunol*, *21*, 1336–1345. doi: 10.1038/s41590-020-0782-6

Pennock, N. D., White, J. T., Cross, E. W., Cheney, E. E., Tamburini, B. A., & Kedl, R. M. (2013). T cell responses: naive to memory and everything in between. *Adv Physiol Educ*, *37*, 273–283. doi: 10.1152/advan.00066.2013

Petrovic, K., Robinson, J., Whitworth, K., Jinks, E., Shaaban, A., & Lee, S. P. (2019). TEM8/ANTXR1-specific CAR T cells mediate toxicity in vivo. *PLOS*. doi:

10.1371/journal.pone.0224015

Petrovic, M., Borchard, G., & Jordan, O. (2021). Considerations for the delivery of STING ligands in cancer immunotherapy. *Journal of Controlled Release*, *339*, 235–247. doi: 10.1016/j.jconrel.2021.09.033

Pielawski, N., Andersson, A., Avenel, C., Behanova, A., Chelebian, E., Anna Klemm, F. N., ... Wahlby, C. (2022). TissUUmaps 3: Interactive visualization and quality assessment of large-scale spatial omics data. *bioRxiv*. doi: 10.1101/2022.01.28.478131

Pizzolato, G., Kaminski, H., Tosolini, M., Franchini, D.-M., Pont, F., Martins, F., ... Fourniea, J.-J. (2019). Single-cell RNA sequencing unveils the shared and the distinct cytotoxic hallmarks of human TCRVδ1 and TCRVδ2 $\gamma\delta$ T lymphocytes. *Proc Natl Acad Sci U S A*, *116*, 11906–11915. doi: 10.1073/pnas.1818488116

Plitas, G., Konopacki, C., Wu, K., Bos, P., Morrow, M., Putintseva, E. V., ... Rudensky, A. Y. (2016). Regulatory T Cells Exhibit Distinct Features in Human Breast Cancer. *Immunity*, *45*, 1122–1134. doi: 10.1073/pnas.162128011410.1016/j.immuni.2016.10.032

Prosser, M. E., Brown, C. E., Shami, A. F., Forman, S. J., & Jensen, M. C. (2012). Tumor PD-L1 co-stimulates primary human CD8+ cytotoxic T cells modified to express a PD1:CD28 chimeric receptor. *Molecular Immunology*, *51*(3), 263–272. doi: 10.1016/j.molimm.2012.03.023

Provenzano, P. P., Inman, D. R., Eliceiri, K. W., Knittel, J. G., Yan, L., Rueden, C. T., ... Keely, P. J. (2008). Collagen density promotes mammary tumor initiation and progression. *BMC Med*, *6*. doi: 10.1186/1741-7015-6-11

Qiu, P. (2020). Embracing the dropouts in single-cell RNA-seq analysis. *Nature Communications*, *1169*. doi: 10.1038/s41467-020-14976-9

Qiu, X., Zhao, T., Luo, R., Qiu, R., & Li, Z. (2022). Tumor-Associated Macrophages: Key Players in Triple-Negative Breast Cancer. *Front. Oncol.*, *12*. doi: 10.3389/fonc.2022.772615

Rand, W. (2006). Machine learning meets agent-based modeling : when not to go to a bar. [PDF file].

Rang, F. J., Kloosterman, W. P., & de Ridder, J. (2018). From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biology*, *19*. doi: 10.1186/s13059-018-1462-9

Rao, A., Barkley, D., Franca, G. S., & Yanai, I. (2021). Exploring tissue architecture using spatial transcriptomics. *Nature*, *596*, 211–220. doi: 10.1038/s41586-021-03634-9

Raza, A., Merhi, M., Inchakalody, V. P., Krishnankutty, R., Relecom, A., Uddin, S., & Dermime, S. (2020). Unleashing the immune response to NY-ESO-1 cancer testis antigen as a potential target for cancer immunotherapy. *J Transl Med*, *18*. doi: 10.1186/s12967-020-02306-y

Rezaei, R., Ghaleh, H. E. G., Farzanehpour, M., Dorostkar, R., Ranjbar, R., Bolandian, M., ... Alvanegh, A. G. (2022). Combination therapy with CAR T cells and oncolytic viruses: a new era in cancer immunotherapy. *Cancer Gene Ther*, *29*, 647–660. doi: 10.1038/s41417-021-00359-9

Ricciardi, S., Manfrini, N., Alfieri, R., Calamita, P., Crosti, M. C., Gallo, S., ... Biffo, S. (2018). The Translational Machinery of Human CD4 T Cells Is Poised for Activation and Controls the Switch from Quiescence to Metabolic Remodeling. *Cell*, *28*, 895–906.

Rivadeneira, D. B., DePeaux, K., Wang, Y., Kulkarni, A., Tabib, T., Menk, A. V., ... Delgoffe, G. M. (2019). Oncolytic Viruses Engineered to Enforce Leptin Expression Reprogram Tumor-Infiltrating T Cell Metabolism and Promote Tumor Clearance. *Immunity*, *51*(3), 548-560.e4. doi: 10.1016/j.immuni.2019.07.003

Robinson, J., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E., Getz, G., & Mesirov, J. (2011). Integrative Genomics Viewer. *Nature Biotechnology*, *29*, 24–26.

Rodriguez-Jorge, O., Kempis-Calanis, L. A., Abou-Jaoude, W., Gutierrez-Reyna, D. Y., Hernandez, C., Ramirez-Pliego, O., ... Thieffry, D. (2019). Cooperation between T cell receptor and Toll-like receptor 5 signaling for CD4+ T cell activation. *Sci Signal*, *12*. doi: 10.1126/scisignal.aar3641

Roth, D. B. (2014). V(D)J Recombination: Mechanism, Errors, and Fidelity. *Micro-*

biology spectrum, *2*, 70–81. doi: 10.1128/microbiolspec.MDNA3-0041-2014

Roybal, K. T., Rupp, L. J., Morsut, L., Walker, W. J., McNally, K. A., Park, J. S., & Lim, W. A. (2016). Precision Tumor Recognition by T Cells With Combinatorial Antigen-Sensing Circuits. *Cell*, *164*, 770–779. doi: 10.1016/j.cell.2016.01.011

Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T., & Eliceiri, K. W. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*, *18*. doi: 10.1186/s12859-017-1934-z

Sahoo, P., Yang, X., Abler, D., Maestrini, D., Adhikarla, V., Frankhouser, D., ... Rockne, R. C. (2020). Mathematical deconvolution of CAR T-cell proliferation and exhaustion from real-time killing assay data. *Journal of the Royal Society Interface*, *17*, 20190734. doi: 10.1098/rsif.2019.0734

Salic, A., & Mitchison, T. (2008). A chemical method for fast and sensitive detection of DNA synthesis in vivo. *PNAS*, *7*, 2415–2420.

Satija Lab. (2022). *Tips for Integrating Large Datasets.* Retrieved 01-06-2022, from https://satijalab.org/seurat/articles/integration_large_datasets.html

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., ... Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, *9*, 676–682. doi: d10.1038/nmeth.2019

Seaby, E. G., Pengelly, R. J., & Ennis, S. (2016). Exome sequencing explained: a practical guide to its clinical application. *Briefings in Functional Genomics*, *15*, 374–384. doi: 10.1093/bfgp/elv054

Senabouth, A., Andersen, S., Shi, Q., Shi, L., Jiang, F., Zhang, W., ... Powell, J. E. (2020). Comparative performance of the BGI and Illumina sequencing technology for single-cell RNA-sequencing. *NAR Genomics and Bioinformatics*, *2*. doi: 10.1093/nargab/lqaa034

Seo, H., Gonzalez-Avalos, E., Zhang, W., Ramchandani, P., Yang, C., Lio, C.-W. J., ... Hogan, P. G. (2021). BATF and IRF4 cooperate to counter exhaustion in tumor-infiltrating CAR T cells. *Nature Immunology*, *22*, 983–995. doi: 10.1038/

s41590-021-00964-8

Sharpe, A. H., & Pauken, K. E. (2018). The diverse functions of the PD1 inhibitory pathway. *Nat Rev Immunol*, *18*, 153–167. doi: 10.1038/nri.2017.108

Shimizu, S., Hiratsuka, H., Koike, K., Tsuchihashi, K., Sonoda, T., Ogi, K., ... Miyazaki, A. (2019). Tumor-infiltrating CD8+ T-cell density is an independent prognostic marker for oral squamous cell carcinoma. *Cancer medicine*, *8*, 80–93. doi: 10.1002/cam4.1889

Si, W., Shen, J., Zheng, H., & Fan, W. (2019). The role and mechanisms of action of microRNAs in cancer drug resistance. *Clinical Epigenetics*, *11*. doi: 10.1186/s13148-018-0587-8

Simon, S., Wu, Z., Cruard, J., Vignard, V., Fortun, A., Khammari, A., ... Labarriere, N. (2018). TCR Analyses of Two Vast and Shared Melanoma Antigen-Specific T Cell Repertoires: Common and Specific Features. *Front. Immunol.*, *9*. doi: 10.3389/fimmu.2018.01962

Simone, D., Penkava, F., Ridley, A., Sansom, S., Al-Mossawi, M. H., & Bowness, P. (2021). Single cell analysis of spondyloarthritis regulatory T cells identifies distinct synovial gene expression patterns and clonal fates. *Commun Biol*, *4*. doi: 10.1038/s42003-021-02931-3

Simpson, K. D., Templeton, D. J., & Cross, J. V. (2012). Macrophage Migration Inhibitory Factor promotes tumor growth and metastasis by inducing Myeloid Derived Suppressor Cells in the tumor microenvironment. *J Immunol.*, *189*, 5533–5540. doi: 10.4049/jimmunol.1201161

Singer, M., Wang, C., Cong, L., Marjanovic, N. D., Kowalczyk, M. S., Zhang, H., ... Anderson, A. C. (2016). A Distinct Gene Module for Dysfunction Uncoupled from Activation in Tumor-Infiltrating T Cells. *Cell*, *166*, 1500–1511. doi: 10.1016/j.cell.2016.08.052

Single Cell Gene Expression Dataset by Cell Ranger 1.1.0, 10x Genomics. (2017). *T Cells from PBMCs (v1 Chemistry).* Retrieved 15-12-2019, from https://www.10xgenomics.com/resources/datasets?query=zheng&page=1&configure%5Bfacets%

5D%5B0%5D=chemistryVersionAndThroughput&configure%5Bfacets%5D%5B1%5D=pipeline.version&configure%5BhitsPerPage%5D=500&menu%5Bproducts.name%5D=

Stagg, J., & Allard, B. (2013). Immunotherapeutic approaches in triple-negative breast cancer: latest research and clinical prospects. *Ther Adv Med Oncol*, *5*, 169–181. doi: 10.1177/1758834012475152

Stephenson, E., Reynolds, G., Botting, R. A., Calero-Nieto, F. J., Morgan, M. D., Tuong, Z. K., ... Haniffa, M. (2021). Single-cell multi-omics analysis of the immune response in COVID-19. *Nat Med*, *27*, 904–916. doi: 10.1038/s41591-021-01329-2

Stickels, R. R., Murray, E., Kumar, P., Li, J., Marshall, J. L., Bella, D. J. D., ... Chen, F. (2021). Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat Biotechnol*, *39*, 313–319. doi: 10.1038/s41587-020-0739-1

Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., ... Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, *19*. doi: 10.1186/s12864-018-4772-0

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., ... Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, *177*. doi: 10.1016/j.cell.2019.05.031

Sureshchandra, S., Lewis, S. A., Doratt, B. M., Jankeel, A., Ibraim, I. C., & Messaoudi, I. (2021). Single-cell profiling of T and B cell repertoires following SARS-CoV-2 mRNA vaccine. *JCI Insight*, *6*. doi: 10.1172/jci.insight.153201

Szabo, P. A., Levitin, H. M., Miron, M., Snyder, M. E., Senda, T., Yuan, J., ... Sims, P. A. (2019). Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nature Communications*, *10*. doi: 10.1038/s41467-019-12464-3

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., ... von Mering, C. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*, *43*. doi: 10.1093/nar/

gku1003

Tan, Y., & Cahan, P. (2019). SingleCellNet: A Computational Tool to Classify Single Cell RNA-Seq Data Across Platforms and Across Species. *Cell Systems*, *9*, 207–213. doi: 10.1016/j.cels.2019.06.004

Tasnim, H., Fricke, G. M., Byrum, J. R., Sotiris, J. O., Cannon, J. L., & Moses, M. E. (2018). Quantitative Measurement of Naive T Cell Association With Dendritic Cells, FRCs, and Blood Vessels in Lymph Nodes. *Front. Immunol.*, *9*. doi: 10.3389/fimmu.2018.01571

Tay, M. Z., Poh, C. M., Renia, L., MacAry, P. A., & Ng, L. F. P. (2020). The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol*, *20*, 363–374. doi: 10.1038/s41577-020-0311-8

The HDF Group. (1997-2018). *Hierarchical Data Format, version 5*. http://www.hdfgroup.org/HDF5/.

The Tabula Muris Consortium. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*, *562*, 367–372. doi: 10.1038/s41586-018-0590-4

Tian, Z., Liu, M., Zhang, Y., & Wang, X. (2021). Bispecific T cell engagers: an emerging therapy for management of hematologic malignancies. *J Hematol Oncol*, *14*. doi: 10.1186/s13045-021-01084-4

Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., ... Garraway, L. A. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *American Association for the Advancement of Science*, *352*, 189–196. doi: 10.1126/science.aad0501

Tosolini, M., Pont, F., Poupot, M., Vergez, F., Nicolau-Travers, M.-L., Vermijlen, D., ... Fournie, J.-J. (2017). Assessment of tumor-infiltrating TCRVγ9Vδ2 $\gamma\delta$ lymphocyte abundance by deconvolution of human cancers microarrays. *Oncoimmunology*, *6*. doi: 10.1080/2162402X.2017.1284723

Tzedakis, G., Tzamali, E., Marias, K., & Sakkalis, V. (2015). The Importance of Neighborhood Scheme Selection in Agent-based Tumor Growth Modeling. *Cancer Inform*, *14*(4), 67–81. doi: 10.4137/CIN.S19343

Valkiers, S., de Vrij, N., Gielis, S., Verbandt, S., Ogunjimi, B., Laukens, K., & Meysman, P. (2022). Recent advances in T-cell receptor repertoire analysis: Bridging the gap with multimodal single-cell RNA sequencing. *ImmunoInformatics*, *5*. doi: 10.1016/j.immuno.2022.100009

Valpione, S., Mundra, P. A., Galvani, E., Campana, L. G., Lorigan, P., Rosa, F. D., ... Marais, R. (2021). The T cell receptor repertoire of tumor infiltrating T cells is predictive and prognostic for cancer survival. *Nature Communications*, *12*. doi: 10.1038/s41467-021-24343-x

Victori, P., & Buffa, F. M. (2019). The many faces of mathematical modelling in oncology. *Br J Radiol*, *92*. doi: 10.1259/bjr.20180856

Victori, P., & Buffa, F. M. (2022). Ingres: from single-cell RNA-seq data to single-cell probabilistic Boolean networks. *bioRxiv*. doi: 10.1101/2022.09.04.506528

Vignesh, K. S., & Deepe, G. S. (2017). Metallothioneins: Emerging Modulators in Immunity and Infection. *Int J Mol Sci*, *18*. doi: 10.3390/ijms18102197

Vistain, L. F., & Tay, S. (2021). Single-Cell Proteomics. *Trends in Biochemical Sciences*, *46*. doi: 10.1016/j.tibs.2021.01.013

Vorkas, C. K., Krishna, C., Li, K., Aube, J., Fitzgerald, D. W., Mazutis, L., ... Glickman, M. S. (2022). Single-Cell Transcriptional Profiling Reveals Signatures of Helper, Effector, and Regulatory MAIT Cells during Homeostasis and Activation. *J Immunol*, *208*, 1042–1056. doi: 10.4049/jimmunol.2100522

Voukantsis, D., Kahn, K., Hadley, M., Wilson, R., & Buffa, F. M. (2019). Modeling genotypes in their microenvironment to predict single- and multi-cellular behavior. *GigaScience*, *8*. doi: 10.1093/gigascience/giz010

Wang, H., Hu, S., Chen, X., Shi, H., Chen, C., Sun, L., & Chena, Z. J. (2017). cGAS is essential for the antitumor effect of immune checkpoint blockade. *Proc Natl Acad Sci U S A*, *114*, 1637–1642. doi: 10.1073/pnas.1621363114

Wang, X., Shen, X., Chen, S., Liu, H., Hong, N., Zhong, H., ... Jin, W. (2022). Reinvestigation of Classic T Cell Subsets and Identification of Novel Cell Subpopulations by Single-Cell RNA Sequencing. *J Immunol*, *208*, 396–406. doi: doi.org/10.4049/jimmunol.2100581

Wang, Y., Jiang, H., Luo, H., Sun, Y., Shi, B., Sun, R., & Li, Z. (2019). An IL-4/21 Inverted Cytokine Receptor Improving CAR-T Cell Potency in Immunosuppressive Solid-Tumor Microenvironment. *Front Immunol*, *10*. doi: 10.3389/fimmu.2019.01691

Wang, Y., Luo, M., Chen, Y., Wang, Y., Zhang, B., Ren, Z., ... Wang, Y. (2021). ZMYND8 Expression in Breast Cancer Cells Blocks T-Lymphocyte Surveillance to Promote Tumor Growth. *Tumor Biology and Immunology*, *81*. doi: 10.1158/0008-5472.CAN-20-1710

Wang, Y., Wang, H., Yao, H., Li, C., Fang, J.-Y., & Xu, J. (2018). Regulation of PD-L1: Emerging Routes for Targeting Tumor Immune Evasion. *Front. Pharmacol.*, *1*. doi: 10.3389/fphar.2018.00536

Wei, H., Wang, Z., Kuang, Y., Wu, Z., Zhao, S., Zhang, Z., ... Tong, A. (2020). Intercellular Adhesion Molecule-1 as Target for CAR-T-Cell Therapy of Triple-Negative Breast Cancer. *Front Immunol*, *11*. doi: 10.3389/fimmu.2020.573823

Wei, Y., Hu, Z., Gu, W., Liu, G., Shi, B., Liu, E., & Liu, T. (2017). CD117+CD44+ Stem T Cells Develop in the Thymus and Potently Suppress T-cell Proliferation by Modulating the CTLA-4 Pathway. *Stem Cell Res Ther*, *8*. doi: 10.1186/s13287-017-0495-4

Weisstein, E. W. (2022). *Moore Neighborhood From MathWorld – A Wolfram Web Resource.* https://mathworld.wolfram.com/MooreNeighborhood.html. (Accessed: 09-28-2022)

Wilensky, U. (1997). *NetLogo Wolf Sheep Predation model.* http://ccl.northwestern.edu/netlogo/models/WolfSheepPredation. (Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL)

Williams, C. G., Lee, H. J., Asatsuma, T., Vento-Tormo, R., & Haque, A. (2022). An introduction to spatial transcriptomics for biomedical researc. *Genome Med*, *14*. doi: 10.1186/s13073-022-01075-1

Wu, S. Y., Xiao, Y., Wei, J.-L., Xu, X.-E., Jin, X., Hu, X., ... Shao, Z.-M. (2021). MYC suppresses STING-dependent innate immunity by transcriptionally up-

regulating DNMT1 in triple-negative breast cancer. *Journal for ImmunoTherapy of Cancer*, *9*. doi: 10.1136/jitc-2021-002528

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., ... Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*, *2*(3), 100141. doi: https://doi.org/10.1016/j.xinn.2021.100141

Xia, S., Tang, Q., Wang, X., Zhang, L., Jia, L., Wu, D., ... Lu, L. (2019). Overexpression of PSMA7 predicts poor prognosis in patients with gastric cancer. *Oncol Lett.*, 5341–5349. doi: 10.3892/ol.2019.10879

Xiao, Y. (2009). A tutorial on analysis and simulation of boolean gene regulatory network models. *Current Genomics*, *10*, 511–525.

Xiao, Y., & Dougherty, E. R. (2007). The impact of function perturbations in Boolean networks. *Bioinformatics*, *23*, 1265–1273.

Xie, B., Jian, Q., Mora, A., & Li, X. (2021). Automatic cell type identification methods for single-cell RNA sequencing. *Computational and Structural Biotechnology Journal*, *19*, 5874–5887. doi: 10.1016/j.csbj.2021.10.027

Xin, Z., Jinyu, L., Shasha, G., Chen, C., Shuang, L., Xiaoyu, W., ... Dayong, C. (2019). Saikosaponin A Inhibits Breast Cancer by Regulating Th1/Th2 Balance. *Frontiers in Pharmacology*, *10*. doi: 10.3389/fphar.2019.00624

Xu, N., Palmer, D. C., Robeson, A. C., Shou, P., Bommiasamy, H., Laurie, S. J., ... Serody, J. S. (2021). STING agonist promotes CAR T cell trafficking and persistence in breast cancer. *J Exp Med*, *218*. doi: 10.1084/jem.20200844

Xu, S., Xu, H., Wang, W., Li, S., Li, H., Li, T., ... Liu, L. (2019). The role of collagen in cancer: from bench to bedside. *Journal of Translational Medicine*, *17*. doi: 10.1186/s12967-019-2058-1

Yanai, H., Chiba, S., Hangai, S., Kometani, K., Inoue, A., Kimura, Y., ... Taniguchi, T. (2018). Revisiting the role of IRF3 in inflammation and immunity by conditional and specifically targeted gene ablation in mice. *Proceedings of the National Academy of Sciences*, *115*(20), 5253-5258. doi: 10.1073/pnas.1803936115

Yang, E., & Li, M. M. H. (2020). All About the RNA: Interferon-Stimulated Genes

That Interfere With Viral RNA Processes. *Front. Immunol.*, *11*. doi: 10.3389/fimmu.2020.605024

Yi, M., Nissley, D. V., McCormick, F., & Stephens, R. M. (2020). ssGSEA score-based Ras dependency indexes derived from gene expression data reveal potential Ras addiction mechanisms with possible clinical implications. *Scientific Reports*, *10*. doi: 10.1038/s41598-020-66986-8

Yin, R., Kwoh, C. K., & Zheng, J. (2019). Whole Genome Sequencing Analysis. *Encyclopedia of Bioinformatics and Computational Biology*, 176-183. doi: 10.1016/B978-0-12-809633-8.20095-2

Yuanyuan, L., & Tollefsbol, T. O. (2011). DNA methylation detection: bisulfite genomic sequencing analysis. *Methods in molecular biology*, *791*, 11–21. doi: 10.1007/978-1-61779-316-5_2

Yue, X., Wu, W.-Y., Dong, M., & Guo, M. (2021). LncRNA MALAT1 promotes breast cancer progression and doxorubicin resistance via regulating miR-570-3p. *Biomedical Journal*, *44*, S296–S304. doi: 10.1016/j.bj.2020.11.002

Yunshuo, C., Ranran, L., Siqi, S., Xuejiao, Y., Lei, L., Wenbo, W., & Yueying, W. (2021). Therapeutic Potential of TNFa and IL1B Blockade for CRS/ICANS in CAR-T Therapy via Ameliorating Endothelial Activation. *Frontiers in Immunology*, *12*. doi: 10.3389/fimmu.2021.623610

Zadeh, J., Steenberg, C., Bois, J., Wolfe, B., Pierce, M., & Khan, A. (2011). NUPACK: analysis and design of nucleic acid systems. *J. Comput. Chem.*, *32*, 170–173.

Zanker, D. J., Owen, K. L., Baschuk, N., Spurling, A. J., & Parker, B. S. (2021). Loss of type I IFN responsiveness impairs natural killer cell antitumor activity in breast cancer. *Cancer Immunology Immunotherapy*, *70*, 2125 – 2138. doi: https://doi.org/10.1007/s00262-021-02857-z

Zhang, N., Hartig, H., Dzhagalov, I., Draper, D., & He, Y. W. (2005). The role of apoptosis in the development and function of T lymphocytes. *Cell Res*, *15*, 749–769. doi: 10.1038/sj.cr.7290345

Zhang, Q., Bastard, P., Bolze, A., Jouanguy, E., Zhang, S.-Y., Effort, C. H. G., ...

Casanova, J.-L. (2020). Life-Threatening COVID-19: Defective Interferons Unleash Excessive Inflammation. *Med (N Y)*, *1*, 14–20. doi: 10.1016/j.medj.2020.12.001

Zhang, Z., Liu, S., Zhang, B., Qiao, L., Zhang, Y., & Zhang, Y. (2020). T Cell Dysfunction and Exhaustion in Cancer. *Front. Cell Dev. Biol.*, *8*. doi: 10.3389/fcell.2020.00017

Zhao, F., Liu, A., Gong, X., Chen, H., Wei, J., Chen, B., ... Mao, R. (2022). Hypoxia-induced RNASEH2A limits activation of cGAS-STING signaling in HCC and predicts poor prognosis. *Tumori Journal*, *108*, 63–76. doi: 10.1177/03008916211026019

Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., ... Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, *8*. doi: 10.1038/ncomms14049

Zhou, L., Adrianto, I., Wang, J., Wu, X., Datta, I., & Mi, Q.-S. (2020). Single-Cell RNA-Seq Analysis Uncovers Distinct Functional Human NKT Cell Sub-Populations in Peripheral Blood. *Front Cell Dev Biol*, *8*. doi: 10.3389/fcell.2020.00384

Zhou, S., Liu, M., Ren, F., Meng, X., & Yu, J. (2021). The landscape of bispecific T cell engager in cancer treatment. *Biomark Res*, *9*. doi: 10.1186/s40364-021-00294-9

Zhu, Y., An, X., Zhang, X., Qiao, Y., Zheng, T., & Li, X. (2019). STING: a master regulator in the cancer-immunity cycle. *Mol Cancer*, *18*. doi: 10.1186/s12943-019-1087-y

APPENDIX A

Predicting Base-Calling Errors from the Oxford Nanopore Sequencer

A.1 Overview

Oxford Nanopore Technologies (ONT) is revolutionising genome sequencing with the development of portable high-throughput DNA/RNA sequencers. Nucleotide bases are identified by a base-caller as they pass through a biological pore by measuring the changes in electrical conductivity they produce. However, the presence of homopolymers, consecutive signal overlaps and secondary structures, which are investigated in this chapter, affect the accuracy of the readings. Training a Hidden Markov Model (HMM) capable of recognising base analogues such as bromodeoxyuridine (BrdU) requires high read accuracy data. The high error rates associated with Nanopore reads inhibit accurate downstream analysis. As such, data collected from the MinION sequencing platform was used to predict positions in the sequence that will be miscalled by the base-caller. This work was conducted at the Sir William Dunn School of Pathology under the supervision of Dr. Michael A. Boemo (Nieduszynski Lab) in 2018.

A.2 Introduction

Advances in the field of genomics have been facilitated by the ability to sequence DNA at a lower cost and higher throughput. State-of-the-art technologies continue

to be developed and are transforming biological and biomedical research. The ONT MinION is the first pocket-sized real-time sequencing device developed (H. Lu et al., 2016). It directly sequences DNA or RNA based on the characteristic current produced by each nucleotide as a single-strand of nucleotides threads through a lipid-embedded protein channel (Fig. A.1). An electrical connection is formed between two salt solutions, enabling the generation of an ion current when voltage is applied across the membrane (Jain et al., 2016). Nanopore technology is suitable for long-reads and fully scalable, from miniature devices to ultra-high throughput implementations.

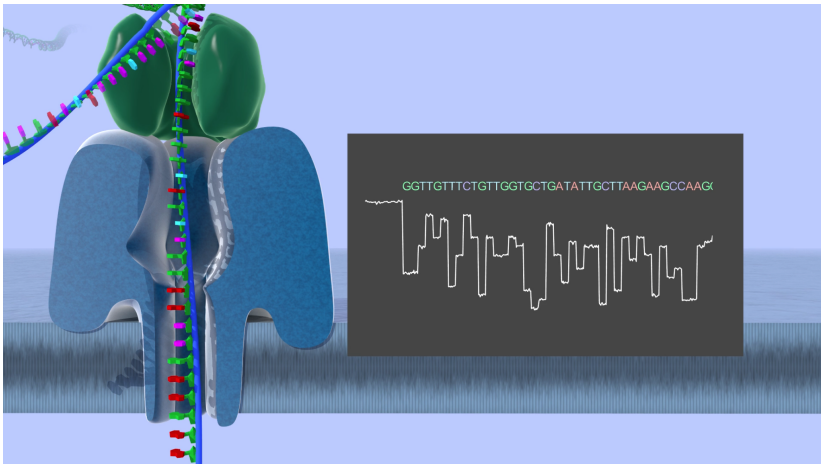


Figure A.1: Oxford Nanopore Technologies DNA Sequencing
Graphical representation of the ONT DNA sequencing method which measures changes in electrical conductivity as nucleotide bases pass through a biological pore. Retrieved from https://nanoporetech.com/.

Base analogues can substitute for normal bases during DNA replication, resulting in altered base pairing or structural changes. Others, like BrdU, will mimic base pairing of their natural base analogue (Fig. A.2A). BrdU is characterised by a very similar chemical structure to thymidine, and is commonly used in cell proliferation assays (Salic & Mitchison, 2008). It is incorporated into the newly synthesised DNA during S phase of the cell cycle, with subsequent detection using BrdU-specific antibodies. With base analogues producing their own characteristic current that distinguishes them from normal bases (Figs. A.2B–A.2C), Nanopore sequencing could offer an instantaneous, quantitative method, to pinpoint the positions of BrdU incorporation. Since the current Nanopore model is not adapted for their detection, a HMM was developed

by the Nieduszynski Lab (University of Oxford) to identify regions of incorporation (Muller et al., 2019). Accuracy of the HMM, however, relies on the accuracy of the training data it receives (Eddy, 2004). At the time of writing, Nanopore technology is well known for its high error rate and the difficulty to achieve good signal-to-noise ratios for longer reads (Rang et al., 2018). This is not a problem for genome assembly analysis where an ensemble method is acceptable but is an issue when position-specific resolution is required (e.g. single nucleotide variation or indel detection).

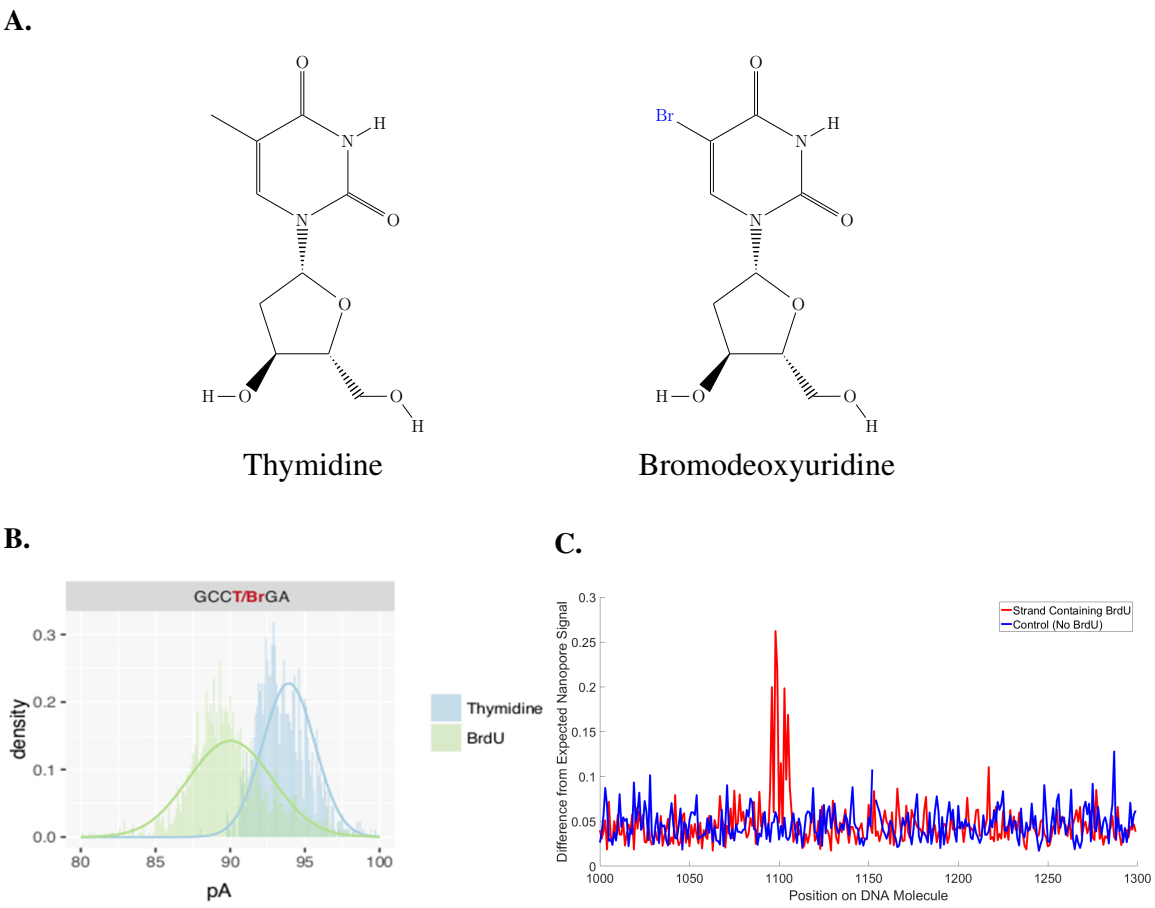


Figure A.2: Signal Events for Thymidine and Bromodeoxyuridine.
A. Structure of thymidine (PubChem CID: 5789) and synthetic nucleoside analogue BrdU (PubChem CID: 6035). **B.** BrdU-containing 6-mer signal deviates significantly from what would be expected if it were thymidine (Nieduszynski Lab). **C.** Difference from expected Nanopore signal for DNA molecule containing BrdU in red against control in blue (Nieduszynski Lab).

This chapter explores error-prone regions in sequencing data collected using the Min-

ION platform. Sequence-specific contexts which may result in signal-space overlaps or secondary structures were found to be harder on the base-caller. These findings were used to develop a predictive model to forecast positions that may result in unreliable reads from a given DNA input sequence. The aim of this work was to help design DNA sequences to be sequenced using Nanopore for use as training data for the HMM.

A.3 Materials and Methods

A.3.1 MinION-Sequenced Data

The dataset received was of de Bruijn DNA sequences, 500 bases long, PCR amplified with different quantities of the base analogue BrdU, ranging from 0% to 20%. De Bruijn sequences, denoted by $B(k, n)$, are contiguous sequences that using an alphabet of size k ensure that the entire sequence consists of different combinations of subsequences of size n, for which there are

$$\frac{(k!)^{k^{n-1}}}{k^n}$$

distinct solutions. For our alphabet of size $k = 4$ $\{A, C, T, G\}$, this ensured that different 5-mers ($n = 5$) be present. The DNA sequences had been sequenced using the MinION platform, and the resulting BAM files contained the aligned base-called sequences.

A.3.2 Analysing Miscalled Positions & Regions

The sequences in the base-called FASTQ files were mapped to a reference FASTA sequence, resulting in an alignment in BAM format. Using the command line version of IGVtools the base count at each position was retrieved in an easily parseable file format, wiggle (.wig) (Robinson et al., 2011). The data was imported as an array containing the position index, the number of times A, C, G, T, N were called, and the

number of deletions and insertions. Reads were considered of poor quality at positions where the called base differed from the reference in greater than 20% of the reads. It was initially expected that higher BrdU concentrations would correlate with a higher error-rate due to signal disruption. Since this was not the case, datasets with 0% BrdU were used for downstream analysis, and the other ones were kept as a control for polymerase-associated errors during PCR. Errors across different PCR samples of the same sequence were relatively consistent, they had a similar spread and clustered in shared sequence contexts.

Having isolated problematic regions and positions, corresponding overlapping 5-mers were reconstructed at these positions by extracting bases on the 5' and 3' ends of the positions. They were subsequently mapped to their corresponding emission values based on characteristic mean and standard deviations from the pore model. Error positions could then be classified as having consecutive 5-mers close in signal-space that may be interfering with each other's signal, specifically those where the mean +/- standard deviation was within the range of flanking positions. The analysis was rerun using the second most called base at those problematic sites, one which do not match the base on the reference. These simulated sequence of events were very similar to ground truth, which may explain certain difficulty encountered by the base-caller when distinguishing between bases with similar 5-mer emission value contexts.

Error positions which could not be explained by overlapping flanking signal-spaces or overlapping emission value contexts were further analysed. The Nucleic Acid Package (NUPACK) algorithm (Dirks et al., 2007) was used to search for potential base-pairing occurring at the 5' end and within 3 to 10 bases of the problematic position. A threshold of 3 to 10 bases was chosen arbitrarily but fit the model adequately. The NUPACK algorithm is formulated in terms of nucleic acid secondary structure. It is a nearest-neighbour DNA thermodynamics model which predicts the most stable secondary structure, and gives a list of base pairing probabilities based on the minimum free energy (MFE) secondary structure of the sequence.

A.3.3 Developing Predictive Model

Preprocessing Input Data

The model was developed in Python and accepts any sequence of bases, $\{A, C, T, G\}$, of any length. It is important to note, however, that the computational complexity is $\mathcal{O}(n)$, thus run time increases linearly with sequence length. The model initially calculates basic statistics on the input sequence including base count, base percentage, sequence length, and searches for homopolymer runs. It collects associated rolling 5-mers and fetches their corresponding emission values from the pore model before running the prediction.

Scoring Bases Using Kullback-Leibler Divergence

It starts by analysing the sequence for overlapping signal-space regions. Since 5-mer emission values fit a normal distribution, $\mathcal{N}(\mu,\ \sigma^2)$, Kullback-Leibler (KL) divergence for continuous probabilities is used to measure how one probability distribution diverges from another.

Let $p(x) = \mathcal{N}(\mu_1,\ {\sigma^2}_1)$ and $q(x) = \mathcal{N}(\mu_2,\ {\sigma^2}_2)$:

$$\mathcal{N} = \frac{1}{\sigma\,\sqrt{2\,\pi}}\,\mathrm{e}^{\,-\frac{(x-\mu)^2}{2\,\sigma^2}}$$

Kullback-Leibler divergence for continuous probabilities (Joyce & James, 2011):

$$\begin{aligned}
D(p,q) &= \int p(x)\log\frac{p(x)}{q(x)}\mathrm{d}x \\
&= \int p(x)\ \ln p(x)\mathrm{d}x - \int p(x)\ \ln q(x)\mathrm{d}x \\
&= \frac{1}{2}\ln\left(2\,\pi\,\sigma_2^2\right) + \frac{\sigma_1^2 + (\mu_1-\mu_2)^2}{2\,\sigma_2^2} - \frac{1}{2}\left(1 + \ln 2\,\pi\,\sigma_1^2\right) \\
&= \ln\frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1-\mu_2)^2}{2\,\sigma_2^2} - \frac{1}{2}
\end{aligned}$$

Note how $D(p, q)$ becomes 0 when $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$. Values are then normalised to be between 0 and 1 through a nonlinear transformation:

$$1 - \mathrm{e}^{-D}$$

A value of 1 means that flanking 5-mers completely differ in signal-space whilst a value of 0 ensures complete overlap. For each base the model then assigns a score based on the sum of their KL divergence. The lower the score, the greater the probability that this position is affected by signal-space overlap.

Calculating Minimum Free Energy Structures with NUPACK

The model runs the sequence through the NUPACK algorithm (Dirks et al., 2007) which results in a list of base-pairing probabilities that is used to assign a score between 0 and 1 to each base, based on the average probability 3 to 10 bases upstream. A polymer graph representation is also constructed, ordering the strand around the circumference of a circle and showing equilibrium probabilities (Zadeh et al., 2011).

Scoring Metric

The final score for each individual base in the sequence is the sum of the values it received in the KL divergence section and NUPACK section. This method provides a scoring metric to identify error-prone sites. The sum of all these scores gives insight into the overall quality of the input sequence of length i, with higher scores representing more favourable sequences:

$$S = \sum_{b=1}^{i} (D_{\mathrm{b}} + N_{\mathrm{b}})$$

A.4 Results and Discussion

A.4.1 Overlapping Signal-Spaces Confuse the Base-Caller

The current Nanopore model maps the measured current as the DNA transits through the pore to five-base-long subsequences called 5-mers. Initially, the mean current in picoamps (pA) measured by the instrument over time is recorded and written to an HDF5 file (The HDF Group, 1997-2018). A base-caller (Albacore) subsequently translates these events into a DNA sequence, directly to FASTQ file. The call is based on the mean and standard deviation emission values characteristic to each possible 1024 5-mer (Table A.1) which fit a normal distribution (Cock et al., 2010; David et al., 2017). The resulting FASTQ files are then aligned to a reference in the creation of a BAM file (H. Li et al., 2009).

Table A.1: Pore Model Provides Expected Signal Distributions.
Preview of the pore model giving the μ & σ for all 1024 possible 5-mers.

5-Mer	μ	σ
CCCCC	97.756	1.613
CCCCA	99.211	1.561
...	...	...
CAAAA	92.086	2.787
AAAAA	86.643	1.982

On average, current signals typically lie between 50 to 150 pA, as a result of this narrow range, many of the 5-mers share similar distributions. Interestingly, more than 60% of the error-prone positions in all samples were of miscalled Adenine bases (Fig. A.3B). When consecutive 5-mers find themselves overlapping in signal-space, it becomes difficult for the base-caller to distinguish between events, explaining why positions in these regions are miscalled (Fig. A.3C). Similarly, homopolymer runs are difficult to analyse, with changes in current becoming harder to detect, making it difficult for the base-caller to identify the number of repeated bases (Fig. A.3D). In some cases, as seen in Figure A.3E, changing just one of the base does not change the resulting signal enough. In addition, noise may affect event recognition and cause the

detector to split single events into different events.

A.4.2 Local Secondary Structures Affect 3' End Base-Calling

The presence of a helicase at the opening of the pore ensures that only single-stranded DNA (ssDNA) threads through it. As the ssDNA is fed through the pore from 5' to 3', secondary structures may start forming at the 5' end. These secondary structures are believed to interfere with the reading of the nucleotides passing through the pore by adding additional tension force on the DNA, affecting current measurements. By analysing the equilibrium base-pairing properties of nucleotides in nanopore-sequenced data, error positions in the sequences could be matched to pair-wise interactions potentially occurring on the other side of the membrane (Figs. A.4A–A.4B).

A.4.3 Predicting Error-Prone Positions

The high inaccuracy in nanopore-sequenced data is problematic for downstream analysis that requires position-specific resolution. As such, a preliminary predictive model was developed to identify regions and positions in any input sequence of DNA that are likely to be miscalled during nanopore sequencing. The model searches for homopolymers, signal-space overlaps, and analyses the 5' end of 5-mers for secondary structures that could arise. The aim of this model is to offer valuable information for the design of DNA sequences that will ensure accurate reads and lead to high quality datasets. This is of great importance to the HMM which requires error-free training data to learn how to recognise base analogues. The HMM can cope with some degree of error like insertions or deletions, but context-based systematic shifts in the signal at certain positions can be very problematic. The next step would be to evaluate the accuracy of the model, and test it by synthesising and sequencing DNA sequences that were designed based on model predictions.

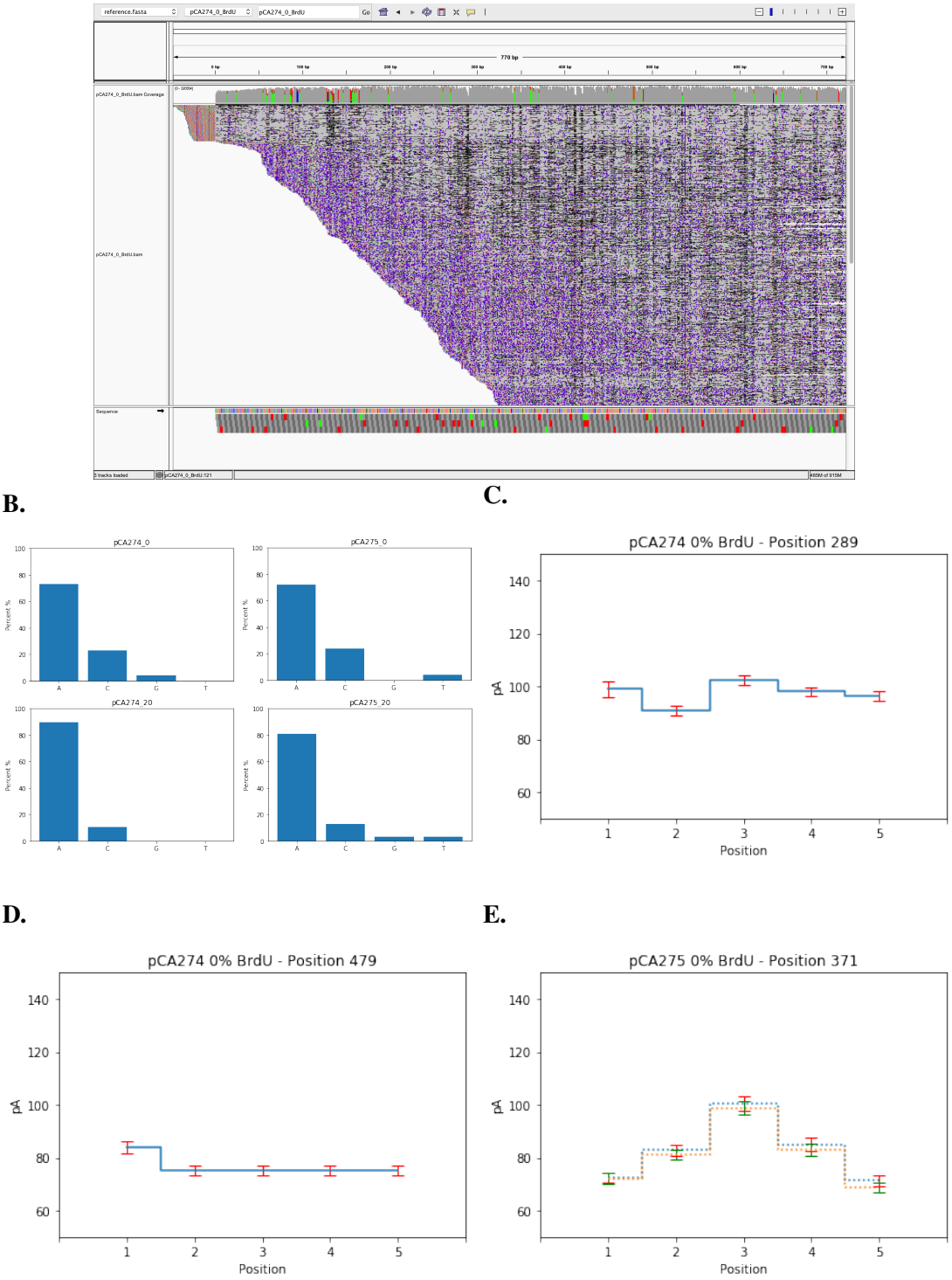
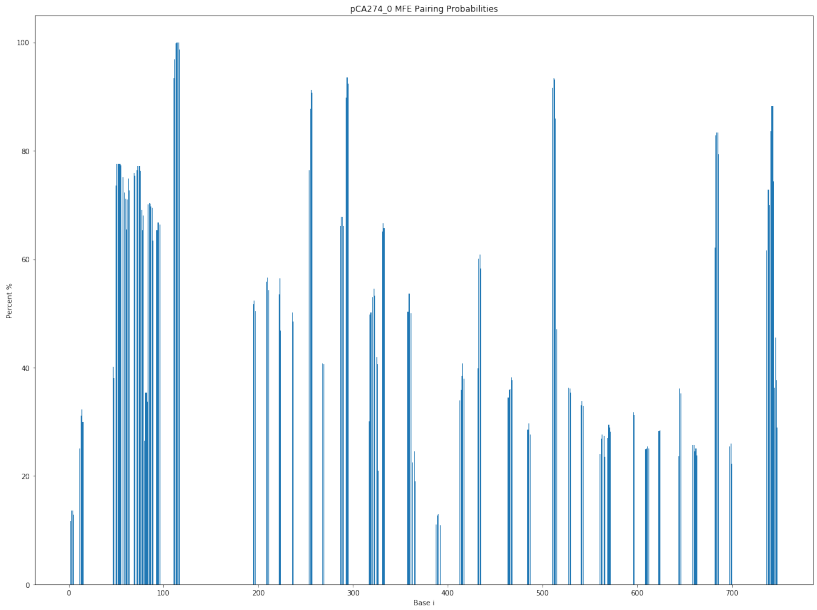


Figure A.3: Overlapping Signal-Spaces Confuse Base-Caller.
A. Overview of the IGV GUI which displays a coverage track of the depth of the reads, displayed at each locus as a grey bar chart. Coloured bars depict high error-rate positions in proportion to the read count of each base. **B.** Adenine is found to be the most miscalled base. **C.** Signal-space overlap of consecutive 5-mers at miscalled positions. **D.** Homopolymer run results in indistinguishable events as seen at miscalled positions. **E.** Similar signals arise when one of the bases at an error position is replaced by the next most called base.

A.



B.

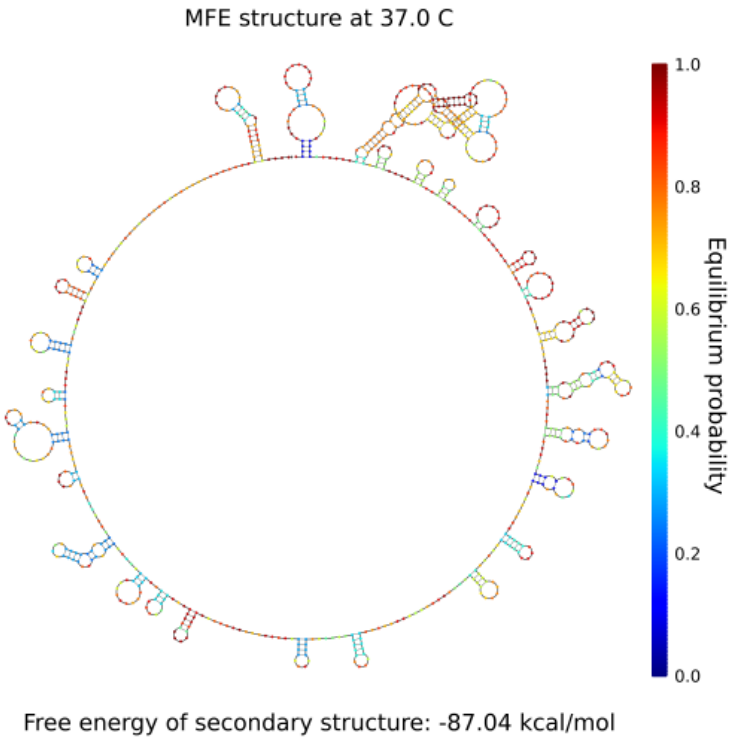


Figure A.4: Sequence Minimum Free Energy Secondary Structure Probabilities. **A.** Base-pairing probabilities for base i in the proximity of error-prone positions. **B.** Schematic of the minimum free energy DNA secondary structure generated by the NUPACK software suite.

APPENDIX B

Comparing T Cell-Intrinsic Interferon Profiles in Cancer and Infectious Diseases

B.1 Overview

Cancer patients are more susceptible to infections and mount less effective responses to vaccines. They are at higher risk of developing serious complications from respiratory infections caused by influenza A viruses (H3N2 and H1N1 strains) or severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Several factors may contribute to the immunosuppressive burden in these patients ranging from therapy regimen to the tumour itself but their relative contributions is a continuous debate. Tumour development alters both local and systemic immunity causing changes in inflammatory parameters and immunomodulatory activities. Impaired type I interferon (IFN) signalling and T cell activity have previously been detected in the peripheral blood mononuclear cells (PBMCs) of breast cancer patients. However, comprehensive analysis of the relationship between systemic cancer-associated defects in T cell immunity and IFN pathways, and susceptibility to infectious diseases are lacking, particularly in the context of solid tumours and treatment naive patients. An ongoing project aims to identify potential parallels in T cell profiles between breast cancer, COVID-19, influenza, sepsis and healthy controls.

B.2 Introduction

T cells are central players in the immune response to SARS-CoV-2 and influenza infection by coordinating anti-viral immune responses, driving immune cell recruitment, supporting humoral immunity and helping limit viral replication. Interferon signalling is vital to intercellular communication, alerting the host innate and adaptive immune system to potential danger and supporting T cell priming and activity. Tumour development results in systemic alterations that impair new immune responses by causing changes in inflammatory parameters and immunomodulatory activities (Chechlinska et al., 2010). Defects in IFN activity extend beyond the local tumour microenvironment and have been detected in the peripheral blood lymphocytes of breast cancer patients. Critchley-Thorne et al. (2009) investigated the function of IFN signalling pathways in peripheral lymphocytes of breast cancer and melanoma patients at different cancer stages. Real-time quantitative PCR measurement of IFN stimulated genes (ISGs) including *STAT1*, *IFI44*, *IFIT1*, *IFIT2* and *MX1* indicated a reduced expression in cancer patients across all stages compared to age-matched healthy controls. This was apparent in both treatment-naive patients and those treated with systemic therapy suggesting that defects in ISG expression is independent of cancer treatment. To further investigate downstream functional responses in T cells, the authors stimulated naive and tumour associated antigen experienced memory and effector T cells with IFNα. All populations showed reduced pSTAT1 expression as well as reduced expression of activation markers HLA-DR, CD54, CD95 and CD25 when co-stimulated with anti-CD3 and anti-CD28. More recently, Allen et al. (2020) reported that tumour-experienced mouse breast cancer models exhibit reduced T cell activation, proliferation and differentiation in response to infection when compared to tumour-free controls even to highly immunogenic pathogens. Their study revealed impaired type 1 immune responses in the context of tumour burden, negatively affecting initial coordination between antigen-presenting cells and T cells.

During the COVID-19 pandemic, defects in early T cell responses were found to contribute to viral persistence and exacerbate tissue damage by allowing uncontrolled

viremia to drive a stronger subsequent immune response (Y. Peng et al., 2020). Lymphopenia, reduced peripheral T cell levels and function, was also reported to be a marker for poor prognosis (Tay et al., 2020). Type I IFNs are pleiotropic cytokines that regulate T cell expansion, function and differentiation both through direct signalling and indirectly through accessory cells. Similarly, type II IFN-γ which is produced by activated cytotoxic $CD8^+$ T cells and Th1 $CD4^+$ T cells, promotes the expression of MHC class I and II on APCs and is vital to sustain proper cytotoxic and effector responses. As a result, tumours and pathogenic viruses tend to evolve strategies to inhibit or impair host IFN response which results in impaired T cell immunity. This is of utmost relevance to COVID-19 severity as host anti-viral defence mechanisms play an important role early in the infection, and mediators of inflammatory parameters are important determinants of host-driven inflammatory lung injury in late and life-threatening COVID-19. Indeed, patients with severe COVID-19 were found to have an IFN deficiency in the blood and a dysregulated inflammatory host response (Hadjadj et al., 2020). Genes that have been associated to the development of life-threatening disease include *IFNAR2*, *DPP9*, *OAS1*, *OAS3*, *IFNAR2* and *TYK2* (Pairo-Castineira et al., 2021; Baranova et al., 2021). In addition, mucosal-associated invariant T which contribute to a protective anti-viral innate response had altered frequencies and activation profiles due to IFNα-IL-18 imbalance (Flament et al., 2021).

Pre-existing medical conditions and comorbidities that impair immunity and inflammatory responses are risk factors for COVID-19. Likewise, inborn errors of type I IFN immunity have been found to underlie critical COVID-19 pneumonia (Q. Zhang et al., 2020). In the context of cancer, hematological, lung and breast cancer have also been associated with higher vulnerability to COVID-19 (Derosa et al., 2020). Access to the University of Oxford COvid-19 Multi-omics Blood ATlas Consortium containing data on PBMCs from patients with COVID-19, influenza and sepsis was granted prior to its publication in Cell (COMBAT, 2022) in order to compare T cell-intrinsic IFN signatures from Chapter 4 across disease conditions.

B.3 Materials and Methods

COMBAT Data Access and Sharing Agreement

Draft Version 1.0, 2/7/20

This agreement governs the terms on which access will be granted to the data generated by the COMBAT consortium at the University of Oxford. In signing this agreement, you are agreeing to be bound by the terms and conditions of access set out in this agreement.

Definitions:

Agreement means this Data Access and Sharing Agreement, as may be amended from time to time.

Arising Intellectual Property means any Intellectual Property Rights, invented or developed in the course of performing work with the Data in the context of this agreement.

Consortium means the COMBAT consortium, a group of researchers within the University of Oxford who are working on the deep phenotyping of COVID-19 patients.

Data means all and any data generated by the Consortium including clinical, phenotypic and molecular.

Data Access Committee means the committee responsible for reviewing requests for access to the Data.

Data Management Plan means the formal document which provides details of how Data are to be handled during and after the Project.

Management Team means members of the Consortium who are responsible for the overall management of the Project and its outputs.

Project means the work associated with the aims of the Consortium.

Research Participant means an individual whose data form part of the Data.

Researcher means the individual requesting access to the data who has signed this Agreement.

Terms of the Agreement

1. The Researcher agrees to only use the Data for the purpose of the Project.
2. The Researcher may make a copy of relevant non-clinical data for use locally in the context of the Project but must not share it further in any form with any other parties without explicit consent from the Data Access Committee. The Researcher must inform the Management Team of the creation of any copy in terms of what files have been copied and where they have been copied to.
3. The Researcher must not copy any clinical data outside of the Project's data warehouse without explicit prior written agreement by the Management Team. Further redistribution is similarly not permitted without explicit prior written agreement by the Management Team.
4. The Researcher agrees to preserve, at all times, the confidentiality of the Data unless it has previously been made publicly available by the Consortium. This includes taking appropriate technical and organisational measures against the unauthorised use of the Data or the processing of it beyond the aims of the Project.
5. The Researcher agrees to protect the confidentiality of Research Participants in any research papers or publication that they prepare by taking all reasonable care to limit the possibility of identification.
6. The Researcher agrees to comply with the University of Oxford IT Regulations and the Data Protection Policy (https://compliance.admin.ox.ac.uk/data-protection-policy) when accessing the Data.
7. The Researcher agrees to share Project-specific generated raw datasets and derived datasets along with metadata on their generation with the Consortium, in accordance with the Data Management Plan.
8. The Researcher agrees to submit to the Project research plans for use of Data to address specific research questions where these do not form part of agreed prioritised research objectives and associated research

questions of the Consortium. This should be done in a timely manner with the goal of avoiding duplication of work, ensuring alignment with objectives of primary data generators and analysts, and the Consortium.

9. When the Researcher wishes to submit a publication which uses the Data before it has been made available publicly, the Researcher will submit the proposed publication to the Management Team for review and comment not less than fourteen (14) days in advance of the submission for publication.
10. The Researcher agrees to destroy/discard Data copied locally once it is no longer used for the Project or when notified to do so by the Project or the University of Oxford.
11. If the Researcher is not affiliated with the University of Oxford, the Researcher agrees that all Arising Intellectual Property from the Data can be freely used by the Project for the purposes of academic research.
12. The Researcher accepts that it may be necessary for the Consortium to alter the terms of this Agreement from time to time. In the event that changes are required, the Consortium will contact the Researcher to inform the changes whereupon the Researcher may accept them or terminate the agreement.

Agreement

The Researcher acknowledges the contents of this Agreement and agrees to comply with the obligations within.

Signature	
Name	Clara Eleonore Pavillet
Email address	Clara.pavillet@msdtc.ox.ac.uk
Institution/Dept.	Medical Sciences DTC
Date	10/05/21

APPENDIX C

Supplementary Material for Chapter 3

C.1 T Cell Signatures

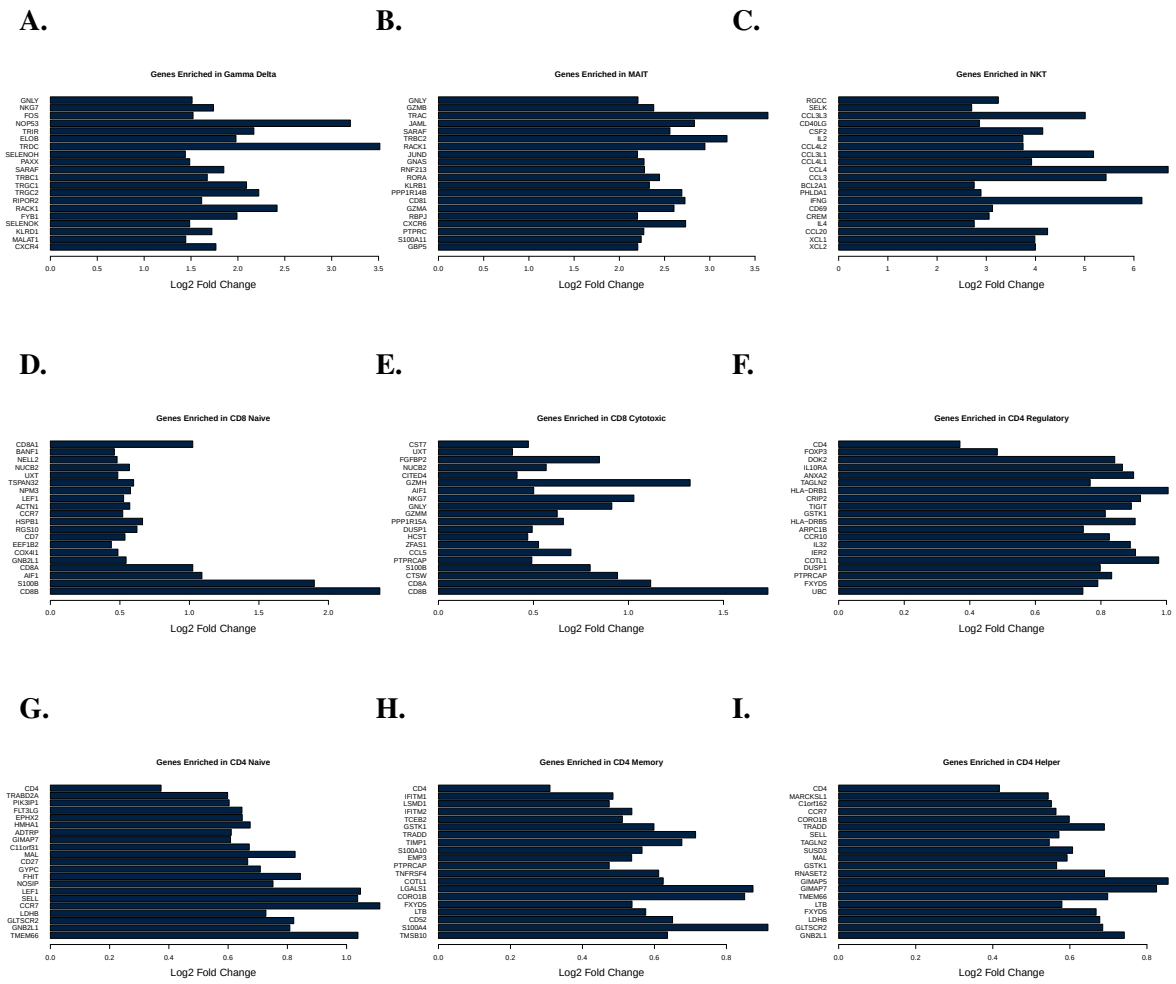


Figure C.1: Fold Changes of Genes Enriched in T Cell Subtypes.
A–I. Log base 2 fold changes of selected genes for each T cell subtype.

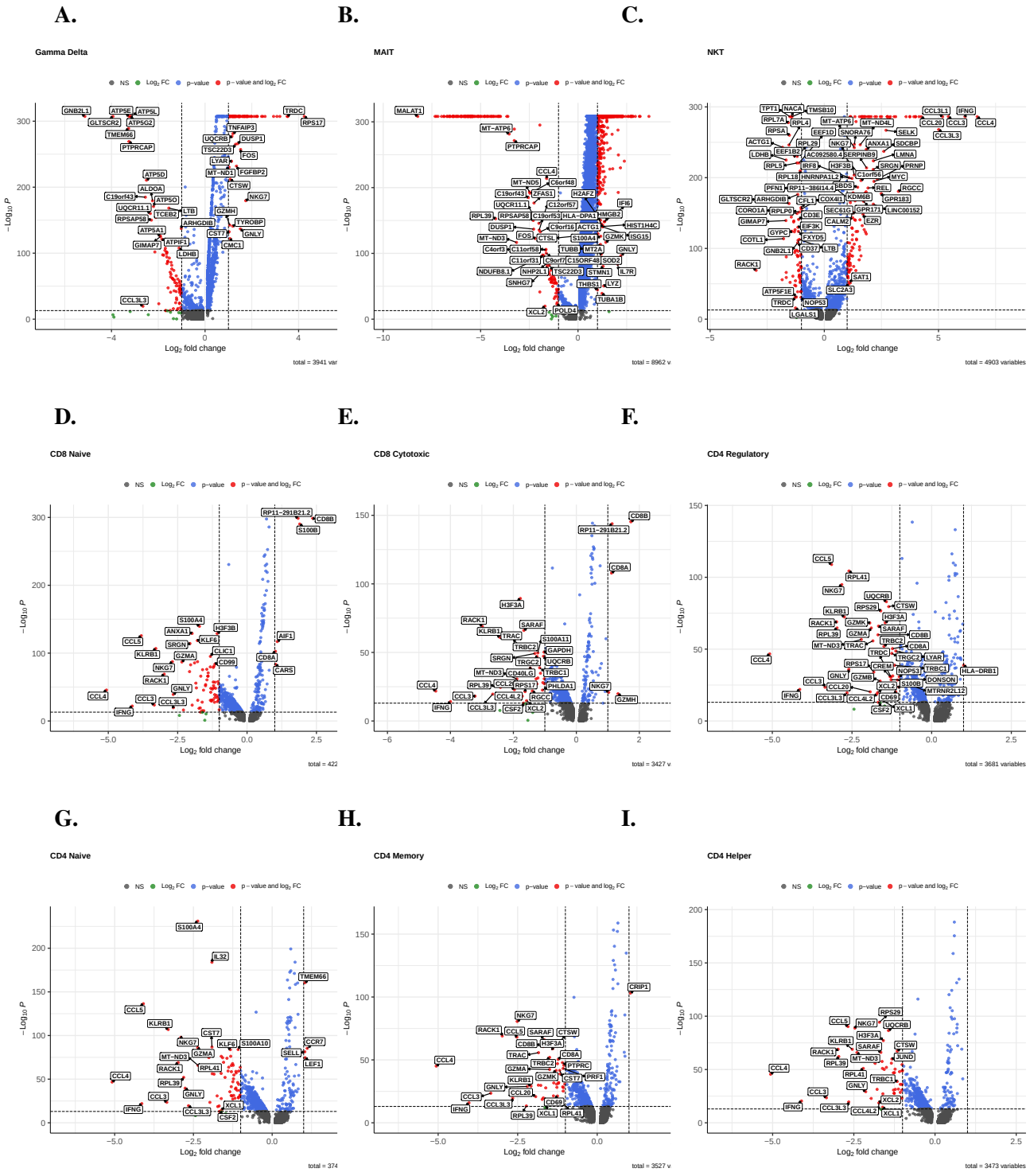


Figure C.2: Volcano Plots of Differential Gene Expression.
A–I. Volcano plots showing statistical significance (p-value) versus magnitude of change (log base 2 fold change) of genes differentially expressed for each T cell subtype. Outlined points represent top differentially expressed genes that pass a log 2 fold change cut-off of 1 and a p-value threshold of 10E-14.

APPENDIX D

Supplementary Material for Chapter 4

D.1 Tissue Processing

The protocol below comes from the Tuveson Lab (Cold Spring Harbor Laboratory) and was adapted by Dr. Karen Sayal to process the BRO breast tissue samples in preparation for scRNA-seq.

Protocol : Digestion of Pancreatic Tumour

Materials

Digestion Buffer

Dissolve in 10% FBS/DMEM:

- $2.5\,\mathrm{mg/ml}$ Collagenase D (Sigma-Aldrich $\#11088866001$)
- $0.5\,\mathrm{mg/ml}$ Liberase DL (Sigma-Aldrich $\#5466202001$)
- $0.2\,\mathrm{mg/ml}$ DNase I (Sigma-Aldrich $\#D5025$)

ACK Lysis Buffer (Thermo Fisher $\#A1049201$)

Flow Buffer

2% FBS in PBS

Procedure

1. Mince tumour with razor blade and put pieces into 4-$5\,\mathrm{ml}$ of digestion buffer on ice.

2. Digest in **digestion buffer** $37\,°\mathrm{C}$ for $45\,\mathrm{min}$, shaking or rotating.
3. Strain through $100\,\mu\mathrm{M}$ cell strainer over $50\,\mathrm{ml}$ tube.
4. Transfer liquid to $15\,\mathrm{ml}$ tube, and add $\approx 5\,\mathrm{ml}$ of **flow buffer**, mix by inverting the tube.
5. Spin 1000 rpm, $4\,°\mathrm{C}$ for 5 mins.
6. Aspirate supernatant and resuspend in $3\,\mathrm{ml}$ of **ACK lysis buffer**, let stand for $3\,\mathrm{min}$ on ice and quench immediately by adding $10\,\mathrm{ml}$ of **flow buffer**, mix by inverting the tube.
7. Spin $1000\,\mathrm{rpm}$, $4\,°\mathrm{C}$ for $5\,\mathrm{min}$.
8. Aspirate supernatant and resuspend in **flow buffer** $10\,\mathrm{ml}$ for big pellet, less for smaller pellets.
9. Count cells and document viability.

D.2 Probes for Spatial Transcriptomics

Table D.1: Panel of Gene-Specific Probes for CARTANA.

CARTANA	
Gene	**Type**
ACTB	CELL FUNCTION
ADAM17	HYPOXIA-EPITHELIAL-ENDOTHELIAL
APOE	MACROPHAGES-MONOCYTES
AREG	HYPOXIA-EPITHELIAL-ENDOTHELIAL
BNIP3	HYPOXIA-EPITHELIAL-ENDOTHELIAL
C1QA	MACROPHAGES-MONOCYTES
C1QB	MACROPHAGES-MONOCYTES
CA9	HYPOXIA-EPITHELIAL-ENDOTHELIAL
CCR4	T CELLS

CD14	MACROPHAGES-MONOCYTES
CD163	MACROPHAGES-MONOCYTES
CD19	B CELLS
CD1C	DENDRITIC CELLS
CD209	DENDRITIC CELLS
CD28	T CELLS
CD34	IMMUNE MARKER
CD3D	T CELLS
CD3E	T CELLS
CD4	T CELLS
CD40	IMMUNE MARKER
CD68	MACROPHAGES-MONOCYTES
CD7	NK CELLS
CD79A	B CELLS
CD79B	B CELLS
CD80	IMMUNE MARKER
CD8A	T CELLS
CD8B	T CELLS
CDH1	HYPOXIA-EPITHELIAL-ENDOTHELIAL
CEACAM8	IMMUNE MARKER
CHCHD2	HYPOXIA-EPITHELIAL-ENDOTHELIAL
CLEC10A	DENDRITIC CELLS
CLEC9A	DENDRITIC CELLS
CORO1C	CELL FUNCTION
CXCR3	T CELLS
DDIT4	HYPOXIA-EPITHELIAL-ENDOTHELIAL
EGFR	HYPOXIA-EPITHELIAL-ENDOTHELIAL
ENO1	HYPOXIA-EPITHELIAL-ENDOTHELIAL
ERBB2	HYPOXIA-EPITHELIAL-ENDOTHELIAL
ERBB3	HYPOXIA-EPITHELIAL-ENDOTHELIAL
EREG	HYPOXIA-EPITHELIAL-ENDOTHELIAL
ESR1	HYPOXIA-EPITHELIAL-ENDOTHELIAL
ESRP1	HYPOXIA-EPITHELIAL-ENDOTHELIAL
FCGR1A	MACROPHAGES-MONOCYTES
FCGR3A	MACROPHAGES-MONOCYTES

FGF13	HYPOXIA-EPITHELIAL-ENDOTHELIAL
FGF7	HYPOXIA-EPITHELIAL-ENDOTHELIAL
FGFR1	HYPOXIA-EPITHELIAL-ENDOTHELIAL
FOLR2	MACROPHAGES-MONOCYTES
FOXP3	T CELLS
FUT4	CELL FUNCTION
GAPDH	CELL FUNCTION
GPI	CELL FUNCTION
GPR183	SIGNALLING
HBEGF	HYPOXIA-EPITHELIAL-ENDOTHELIAL
HK2	HYPOXIA-EPITHELIAL-ENDOTHELIAL
HPRT1	CELL FUNCTION
IDO1	DENDRITIC CELLS
IGKC	B CELLS
IL2RA	T CELLS
IL3RA	DENDRITIC CELLS
ITGAX	DENDRITIC CELLS
JCHAIN	B CELLS
KLRB1	T CELLS
KLRC2	NK CELLS
KLRD1	NK CELLS
KRT19	HYPOXIA-EPITHELIAL-ENDOTHELIAL
KRT7	HYPOXIA-EPITHELIAL-ENDOTHELIAL
LDHA	HYPOXIA-EPITHELIAL-ENDOTHELIAL
LST1	IMMUNE MARKER
MIF	HYPOXIA-EPITHELIAL-ENDOTHELIAL
MRPL15	CELL FUNCTION
MS4A1	B CELLS
MS4A2	B CELLS
MUC1	HYPOXIA-EPITHELIAL-ENDOTHELIAL
NDRG1	HYPOXIA-EPITHELIAL-ENDOTHELIAL
NKG7	NK CELLS
P4HA1	HYPOXIA-EPITHELIAL-ENDOTHELIAL
PECAM1	HYPOXIA-EPITHELIAL-ENDOTHELIAL
PFKP	HYPOXIA-EPITHELIAL-ENDOTHELIAL

PGAM1	CELL FUNCTION
PGK1	CELL FUNCTION
PNP	CELL FUNCTION
PSMA7	CELL FUNCTION
PTPRC	IMMUNE MARKER
RPL11	CELL FUNCTION
S100A8	SIGNALLING
SEC61G	HYPOXIA-EPITHELIAL-ENDOTHELIAL
SLC16A1	HYPOXIA-EPITHELIAL-ENDOTHELIAL
SLC25A32	CELL FUNCTION
SLC2A1	HYPOXIA-EPITHELIAL-ENDOTHELIAL
TGFA	SIGNALLING
TGFBR1	SIGNALLING
TPI1	CELL FUNCTION
TPSAB1	CELL FUNCTION
TUBA1A	CELL FUNCTION
TYROBP	NK CELLS
VEGFA	HYPOXIA-EPITHELIAL-ENDOTHELIAL
YKT6	CELL FUNCTION

Table D.2: Panel of Gene-Specific Probes for Molecular Cartography.

Molecular Cartography			
Gene	**ENSG**	**Alias**	**Type**
CD34	ENSG00000174059		ENDOTHELIAL
PECAM1	ENSG00000261371		ENDOTHELIAL
ADGRL4	ENSG00000162618	ELTD1	ENDOTHELIAL
CDH5	ENSG00000179776	VE-CADHERIN	ENDOTHELIAL
ECM1	ENSG00000143369		ENDOTHELIAL
ESR1	ENSG00000091831		EPITHELIAL
KRT19	ENSG00000171345		EPITHELIAL
KRT7	ENSG00000135480		EPITHELIAL
ADAM17	ENSG00000151694		EPITHELIAL
AREG	ENSG00000109321		EPITHELIAL
EGFR	ENSG00000146648		EPITHELIAL; CAR TARGET
ERBB3	ENSG00000065361		EPITHELIAL
EREG	ENSG00000124882		EPITHELIAL
FGF7	ENSG00000140285		EPITHELIAL
FGFR1	ENSG00000077782		EPITHELIAL
HBEGF	ENSG00000113070		EPITHELIAL
TGFBR1	ENSG00000106799		EPITHELIAL
CA9	ENSG00000107159		HYPOXIA
DDIT4	ENSG00000168209		HYPOXIA
HK2	ENSG00000159399		HYPOXIA
LDHA	ENSG00000134333		HYPOXIA
NDRG1	ENSG00000104419		HYPOXIA
P4HA1	ENSG00000122884		HYPOXIA
PGK1	ENSG00000102144		HYPOXIA
SLC16A1	ENSG00000155380		HYPOXIA
SLC2A1	ENSG00000117394		HYPOXIA
VEGFA	ENSG00000112715		HYPOXIA
ADM	ENSG00000148926		HYPOXIA
ANGPTL4	ENSG00000167772		HYPOXIA
SLC16A3	ENSG00000141526		HYPOXIA
PDHA1	ENSG00000131828		HYPOXIA
FGFR2	ENSG00000066468		HYPOXIA
MET	ENSG00000105976		HYPOXIA
HGF	ENSG00000019991		HYPOXIA
LUM	ENSG00000139329		FIBROBLAST
FBLN1	ENSG00000077942		FIBROBLAST

ACTA2	ENSG00000107796		FIBROBLAST
DCN	ENSG00000011465		FIBROBLAST
S100A4	ENSG00000196154	FSP1	FIBROBLAST
ICAM1	ENSG00000090339	CD54	CAR TARGET
ANTXR1	ENSG00000169604	TEM8	CAR TARGET
MYC	ENSG00000136997		IMMUNE SUPPRESSION
DNMT1	ENSG00000130816		IMMUNE SUPPRESSION
RNASEH2A	ENSG00000104889		IMMUNE SUPPRESSION
PDCD1	ENSG00000188389	PD-1	IMMUNE CHECKPOINT
CD274	ENSG00000120217	PD-L1	IMMUNE CHECKPOINT
CTLA4	ENSG00000163599		IMMUNE CHECKPOINT
LGALS9	ENSG00000168961	GAL9	IMMUNE CHECKPOINT
CGAS	ENSG00000164430	C6orf150	IFN SIGNALLING
STING1	ENSG00000184584	TMEM173	IFN SIGNALLING
TBK1	ENSG00000183735		IFN SIGNALLING
IRF3	ENSG00000126456		IFN SIGNALLING
DDX58	ENSG00000107201	RIG-1	IFN SIGNALLING
NFKB1	ENSG00000109320	NF-κB	IFN SIGNALLING
IRF5	ENSG00000128604		IFN SIGNALLING
IFNAR2	ENSG00000159110		IFN SIGNALLING
OAS1	ENSG00000089127		IFN SIGNALLING
TNF	ENSG00000232810	TNF-α	IFN SIGNALLING
STAT3	ENSG00000168610		IFN SIGNALLING
JAK2	ENSG00000096968		IFN SIGNALLING
CCL5	ENSG00000271503		IFN SIGNALLING
MX1	ENSG00000157601		IFN SIGNALLING
IL6R	ENSG00000160712		IFN SIGNALLING
TYK2	ENSG00000105397		IFN SIGNALLING
IFIT1	ENSG00000185745		IFN SIGNALLING
IFI16	ENSG00000163565		IFN SIGNALLING
STAT6	ENSG00000166888		IFN SIGNALLING
JAK1	ENSG00000162434		IFN SIGNALLING
PRKDC	ENSG00000253729		IFN SIGNALLING
IFNG	ENSG00000111537		IFN SIGNALLING
STAT1	ENSG00000115415		IFN SIGNALLING
IL6	ENSG00000136244		IFN SIGNALLING
ZMYND8	ENSG00000101040		IFN SIGNALLING
HAVCR2	ENSG00000135077	TIM3	T CELL IMMUNITY
HLA-DRB1	ENSG00000196126	MHC-II	T CELL IMMUNITY

TGFB1	ENSG00000105329		T CELL IMMUNITY
CD38	ENSG00000004468		T CELL IMMUNITY
CD69	ENSG00000110848		T CELL IMMUNITY
CD28	ENSG00000178562		T CELL IMMUNITY
IL7R	ENSG00000168685		T CELL IMMUNITY
TRGC1	ENSG00000211689		T CELL IMMUNITY
TRAC	ENSG00000277734		T CELL IMMUNITY
KLRB1	ENSG00000111796		T CELL IMMUNITY
KLRG1	ENSG00000139187		T CELL IMMUNITY
CD44	ENSG00000026508		T CELL IMMUNITY
GZMK	ENSG00000113088		T CELL IMMUNITY
CCR5	ENSG00000160791		T CELL IMMUNITY
PTPRC	ENSG00000081237	CD45RA	T CELL IMMUNITY
FAS	ENSG00000026103	CD95	T CELL IMMUNITY
CD27	ENSG00000139193		T CELL IMMUNITY
CD3D	ENSG00000167286		T CELL IMMUNITY
CD4	ENSG00000010610		T CELL IMMUNITY
CD8A	ENSG00000153563		T CELL IMMUNITY
IDO1	ENSG00000131203		T CELL IMMUNITY
FOXP3	ENSG00000049768		T CELL IMMUNITY
DUSP2	ENSG00000158050		T CELL IMMUNITY
TRAF3	ENSG00000131323		T CELL IMMUNITY
MS4A2	ENSG00000149534		MAST
MARCO	ENSG00000019169		TAM
CD68	ENSG00000129226		TAM

D.3 Scope of Work

Exhibit – Scope of Work

1. Background and description of Work.

   Analysis of the expression and localization of immune cell infiltrates in triple negative breast cancer

2. Timelines

Between 6 to 8 weeks from synthesizing probesets and receiving samples.

3. Acceptance criteria

Electronic data provided with spot coordinates of 100 genes (as listed in 5a) together with DAPI images

4. Project managers

For University of Oxford: Helen Sheldon
For Resolve BioSciences: Jeroen Aerts

5. List of material, equipment and/or tools to be provided by University of Oxford
   a. Complete list of genes to be used within the scope of this project: Supplied by Clara Pavillet

Gene ID	Gene Name
CD34	ENSG00000174059
PECAM1	ENSG00000261371
ADGRL4	ENSG00000162618
CDH5	ENSG00000179776
ECM1	ENSG00000143369
ESR1	ENSG00000091831
KRT19	ENSG00000171345
KRT7	ENSG00000135480
ADAM17	ENSG00000151694
AREG	ENSG00000109321
EGFR	ENSG00000146648

ERBB3	ENSG00000065361
EREG	ENSG00000124882
FGF7	ENSG00000140285
FGFR1	ENSG00000077782
HBEGF	ENSG00000113070
TGFBR1	ENSG00000106799
CA9	ENSG00000107159
DDIT4	ENSG00000168209
HK2	ENSG00000159399
LDHA	ENSG00000134333
NDRG1	ENSG00000104419
P4HA1	ENSG00000122884
PGK1	ENSG00000102144
SLC16A1	ENSG00000155380
SLC2A1	ENSG00000117394
VEGFA	ENSG00000112715
ADM	ENSG00000148926
ANGPTL4	ENSG00000167772
SLC16A3	ENSG00000141526
PDHA1	ENSG00000131828
FGFR2	ENSG00000066468
MET	ENSG00000105976
HGF	ENSG00000019991
LUM	ENSG00000139329
FBLN1	ENSG00000077942
ACTA2	ENSG00000107796
DCN	ENSG00000011465
S100A4	ENSG00000196154
ICAM1	ENSG00000090339
ANTXR1	ENSG00000169604
MYC	ENSG00000136997
DNMT1	ENSG00000130816
RNASEH2A	ENSG00000104889
PDCD1	ENSG00000188389
CD274	ENSG00000120217
CTLA4	ENSG00000163599
LGALS9	ENSG00000168961
CGAS	ENSG00000164430
STING1	ENSG00000184584

CONFIDENTIAL – FOR DRAFT PURPOSE ONLY

TBK1	ENSG00000183735
IRF3	ENSG00000126456
DDX58	ENSG00000107201
NFKB1	ENSG00000109320
IRF5	ENSG00000128604
IFNAR2	ENSG00000159110
OAS1	ENSG00000089127
TNF	ENSG00000232810
STAT3	ENSG00000168610
JAK2	ENSG00000096968
CCL5	ENSG00000271503
MX1	ENSG00000157601
IL6R	ENSG00000160712
TYK2	ENSG00000105397
IFIT1	ENSG00000185745
IFI16	ENSG00000163565
STAT6	ENSG00000166888
JAK1	ENSG00000162434
PRKDC	ENSG00000253729
IFNG	ENSG00000111537
STAT1	ENSG00000115415
IL6	ENSG00000136244
ZMYND8	ENSG00000101040
HAVCR2	ENSG00000135077
HLA-DRB1	ENSG00000196126
TGFB1	ENSG00000105329
CD38	ENSG00000004468
CD69	ENSG00000110848
CD28	ENSG00000178562
IL7R	ENSG00000168685
TRGC1	ENSG00000211689
TRAC	ENSG00000277734
KLRB1	ENSG00000111796
KLRG1	ENSG00000139187
CD44	ENSG00000026508
GZMK	ENSG00000113088
CCR5	ENSG00000160791
PTPRC	ENSG00000081237
FAS	ENSG00000026103

CONFIDENTIAL – FOR DRAFT PURPOSE ONLY

CD27	ENSG00000139193
CD3D	ENSG00000167286
CD4	ENSG00000010610
CD8A	ENSG00000153563
IDO1	ENSG00000131203
FOXP3	ENSG00000049768
DUSP2	ENSG00000158050
TRAF3	ENSG00000131323
MS4A2	ENSG00000149534
MARCO	ENSG00000019169
CD68	ENSG00000129226

b. Tissue/Cell Culture Type: breast tumour sections
c. Tissue/Cell culture Source: human
d. Number of wells and slides (NOTE: each slide contains eight wells): 8
e. Dimensions of the combined area of all samples on a slide to be imaged: 26 mm² distributed across all slices.
f. Any additional dye labeling, if any (DAPI is done by default): none

University of Oxford – Customer

H. Sheldon

Signature_____ _____ Date___21/1/22_____

Name: Helen Sheldon
Title: Dr

CONFIDENTIAL – FOR DRAFT PURPOSE ONLY

Resolve BioSciences GmbH – Provider

Signature_____ Date____21/1/22_____

Name: Chris Barbazette
Title: Chief Commercial Officer

CONFIDENTIAL – FOR DRAFT PURPOSE ONLY

D.4 Biosafety Declaration

Resolve Biosciences
Creative Campus
Building A03
Alfred-Nobel-Str. 10,
40789 Monheim,
Germany

Biosafety Level Declaration

Date 21/1/22

Project Number 32676

Please fill in all fields marked in grey, sign and scan the form, and send back the scanned file to Resolve Biosciences by email.

1 Address

Name	Dr Helen Sheldon
Institute/ Company	Dept Oncology, University of Oxford
Street	Old Road Campus Research Building, Roosevelt Drive
Postcode/ City	Oxford, OX3 7DQ
Country	UK
Email	helen.sheldon@oncology.ox.ac.uk
Tel. (+ country code)	+44 01865 617331

2 Samples were already shipped with a former project

☐ Yes	☒ No

If yes, please specify the former project number and sign (see section 7).

If no, answer all questions in the declaration form and sign.

3 Organism/Species of the sample

Species of organism: Human

Organism is grown under controlled conditions (e.g. green house, animal facility) or in natural environment

☐ Controlled conditions	☒ Natural environment

If organism is grown under controlled conditions, please specific (e.g. animal facility)

Address of e.g. animal facility, green house etc.

4 Sample type

☐ Whole organism	☐ Organ	☐ Cells	☒ Other

Please specify (e.g. type of organ, cells)

breast tumour

Is the sample fixed/stabilized by reagents (e.g. formaldehyde, PAXgene, alcohols)?

Yes	☐	☐ Formaldehyde
No	☒	☐ PAXgene
		☐ Alcohol

How large was the dimension of the sample during fixation (e.g. 10 x 10 x 10 mm)?

_____ x _____ x _____ mm

How is the sample fixed/stabilized (e.g. type of fixative, concentration of fixative, time and temperature during fixation)?

How and how long was the sample stored until shipment (Temperature, Time etc.)?

Snap frozen liquid nitrogen and stored in liquid nitrogen between 1-10 years

5 Infectious status

Was the organism or the part of the organism subject to an infection?

Yes	
No	X

If yes, which pathogen infected the sample? Please, specify (complete name)?

Was the organism or the part of the organism subject to a pathogen inactivation process?

Yes	
No	X

If yes, which pathogen inactivation process was used? Please, describe the process and attach validation data?

Was the organism or the part of the organism tested for viral infections?

Yes	
No	X

If yes, what was the diagnostic result for the following viruses?

	Positive	Negative	Not tested
HIV			
HCV			
HBV			

6 Genetic engineering

Was the organism or the part of the organism subject to genetic engineering[1)]?

Yes	
No	X

If yes, what type of genetic engineering was performed? Please, describe the process and attach validation data?

[1)] A genetically modified organism is an organism whose genetic material has changed in a way that cannot occur under natural conditions by mutuations, crossing or natural recombination. The descendants of such an organism are also defined as genetically modified organisms.

7 Viability

Is the organism, the part of the organism, or pathogens within the sample still viable?

Yes	
No	X

Are parts of the organism or pathogens within the sample capable for cell division or replication?

Yes	
No	X

8 Signing by customer

I certify that I have read and understood all the questions set forth in this application and the answers I have furnished on this form are true and correct to the best of my knowledge and belief.

Date/signing

D.5 Tender Exemption

UNIVERSITY OF OXFORD

PURCHASING DEPARTMENT
Finance Division

Tender Exemption Request Form

To be read in conjunction with the Tender Exemption Guidance Notes. Please **attach** all additional information in support of your request e.g. quote, specification, grant award.

Title of purchase/ project & ref no.	Spatial Transcriptomics – DPhil Project		**Contract duration**	Once
			Grant or X5 ref no.	H800.02 BST00100
Value of purchase / £ (ex VAT)	*24 000 EUR*		**Source of funds**	Grant ☒ Chest ☐ Both ☐
Proposed supplier	Resolve Bioscience		**University preferred supplier?**	Yes ☐ No ☒
Reason(s) for exemption: (Please tick all relevant boxes)	Technical compatibility	☒	Protection of exclusive rights	☐
	Single provider	☐	One-off, unique opportunity of significant value	☒
	Extension of an existing contract	☐	Extreme urgency, unforeseeable and not attributable to requester	☐
	Specified in grant award	☐	Overriding University interest (not meeting the requirements for any other dispensation)	☐
Construction line registered?	For Estates Services only ☐ Yes ☐ No			
Justification: (Please refer to points for consideration in the Guidance Notes)	**Supplier selection (including due diligence):** The supplier was decided in a conversation between my professor, a postdoc and I. Resolve Bioscience uses a method we wanted to try for my DPhil project and we received a discount from them which made it cost effective for the number of samples we wanted to send for Spatial Transcriptomics. **Evaluation criteria:** Using my Wellcome Trust funding for experiments required for my DPhil project and a paper I am working on. **Market research:** Another student in my lab did Spatial Transcriptomics on the same samples with another company – CARTANA and we were not happy with their service. As such, we started looking at other companies that do Spatial Transcriptomics. **Demonstration of value for money (including whole-life cost & sustainability):** We received a 20% reduction because of our lab's connection with Qiagen. **Specification alteration:** **Acceptance of University terms and conditions / previously agreed terms:** *(Confirm which terms have been provisionally agreed)*			

	Name	**Department**	**Date**
Requested by	**Clara Eleonore Pavillet**	**Medical Sciences DTC/MRC WIMM/Biochemistry**	**25/01/22**
The request for exemption is: APPROVED ☒ / NOT APPROVED ☐			
Head of Department / Administrator	S Keeble	WIMM	3/2/22
Purchasing Department (if required)			

Purchasing Department approval must be given by the Director of Purchasing or the Director of Finance for any purchase exceeding £25,000 (ex-VAT) under the 'overriding University interest' exemption.

Visit our website for further details: https://finance.admin.ox.ac.uk/purchasing

Purchasing Department approval is required for any purchase exceeding £100,000 (ex-VAT) or any purchase between £25,000 and £100,000 (ex-VAT) that is not on the University's terms and conditions. See Guidance Notes for full details.

For Estates Services ONLY: Between £25,000 and £100,000 (ex-VAT), approvers are Purchasing Department and Director of Estates. Above £100,000 (ex-VAT), approvers are Purchasing Department, Director of Estates and BESC.

STANDARD PURCHASE ORDER: H82072087

CONFIRMATION ORDER - COPY

Revision: 0
Document Date: 08-FEB-2022

Resolve BioSciences GmbH
Alfred-Nobel-Str. 10
Monheim 40789
ATTN OF:

Delivery Address
H802
Industrial Block for WIMM
University of Oxford
John Radcliffe Hospital
Headington, Oxford, Oxfordshire OX3 9DU

ATTN OF: Weatherall Institute Of Molecular Medicine
accountspayable@imm.ox.ac.uk
01865 222366

Billing Address
Weatherall Institute of Molecular Medicine
University of Oxford
John Radcliffe Hospital
Headington, Oxford, OX3 9DS

University of Oxford
Weatherall Institute of
Molecular Medicine
VAT No. GB 125 5067 30

Special Instructions: Quote - OPP-32676_Sheldon_Oxford.pdf

EORI No: GB 125 5067 30 065

Supplier Number	Supplier VAT Number	Customer Account No.	Payment Terms	Supplier Contact Phone	Supplier Contact Fax
10236911	DE327888119		30 Days	+49 21732975311	

Line No.	Part Number/Description	Delivery Date	Qty	UOM	Unit Price EUR	Total EUR	VAT Reference
1	Sample service as per Quote RBQ-32688	Needed: 09-FEB-2022	1	Each	24,000.00	24,000.00	
					Total Excl. VAT:	24,000.00	

Registered Office: Council Sec. University of Oxford Wellington Sq Oxford OX1 2JD United Kingdom

Terms and Conditions

Unless the University has entered into a separate written agreement with you for the supply of goods and services, acceptance of this order implies your acceptance of the University's Standard Conditions of Purchase (See Terms and Conditions). The University's Standard Conditions of Purchase or, as the case may be, the separate written agreement entered into between the University and you shall govern the agreement for your supply of goods and/or services to the entire exclusion of all other terms and conditions, including, but without limitation, any terms and conditions which you purport to apply under any acknowledgement or confirmation of order, quotation, specification, delivery note, or similar document.

D.7 Material Transfer Agreement

MATERIAL TRANSFER AGREEMENT

between

The Chancellor, Masters and Scholars of the University of Oxford, whose administrative offices are located at University Offices, Wellington Square, Oxford, OX1 2JD, United Kingdom (the "Provider Institution" or the "University") and

Resolve Biosciences GmbH of Alfred-Nobel Str 10, 40789 Monheim am Rhein, Germany ("Recipient")

hereinafter referred to as "the Parties" and each of them being "a Party"

BACKGROUND

(A) The Recipient is providing a service entitled "Spatial transcriptomics of human breast tumour tissue samples" as described in more detail in the Standard Purchase Order H82072087 (the "Services"). The Provider Institution has obtained NRES approval with reference number HCRW 19/SC/0025 and OCHRE 21A066 which covers the Services.

(B) The Recipient wishes to access and utilise the material identified in Appendix 2 (the "Material") for the purpose of the Services. The term "Material" includes but is not limited to material, other than human gametes or embryos, which consists of, or includes human cells and which is considered "Relevant Material" for the purposes of the Human Tissue Act 2004[1] together with related data.

(C) The Provider Institution is willing to supply the Material to the Recipient and the Recipient is willing to receive the Material in accordance with the terms and conditions contained within this agreement (the "Agreement");

TERMS AND CONDITIONS

1. The Recipient will hold the Material on the terms of this Agreement and solely for the purpose of the Services.

2. The Recipient hereby agrees to comply and procure that all personnel who work with the Material comply with the terms and conditions of this Agreement. The Recipient will not use the Material for (a) administration to human subjects or (b) "human application" as that term is defined in the Human Tissue (Quality and Safety for Human Application) Regulations 2007 (or equivalent as may be replaced or amended from time to time), or for clinical or diagnostic purposes.[2]

[1] The Human Tissue Act 2004 applies to the "authorised activities" principally the removal, storage and use of "Relevant Materials" (as defined under the Act, including human cells, tissue and organs, but not cell lines) which come from a living or deceased person for "Scheduled Purposes" (these are set out in Schedule 1 of the Act, including, but not limited to, "research in connection with disorders, or the function of the human body", "education or training relating to human health", and "transplantation").

[2] The Human Tissue (Quality and Safety for Human Application) Regulations 2007 apply to the procurement, testing, processing, storage, distribution, and import or export of tissues and cells (including cell lines). "Cells" mean human cells (whether individually or in an unbound collection) including cell lines, but not including gametes, embryos outside the body, blood or blood components. "Tissue" for the Regulations, means all constituent parts of the human body formed by cells, but not including gametes and embryos outside the body (which are regulated by the Human Fertilisation and Embryology Authority pursuant to the Human Fertilisation and Embryology Act 1990), or organs.

Provider Ref: R80564/CN001
Recipient Ref:

3. The Recipient may use the Material for the purposes of the Services as described in Appendix 1, from the date of receipt of the Material until the conclusion of the Services. With respect to its use of the Material (including, not limited to disposal or return), the Recipient will comply fully with all applicable environmental, health and safety laws, and other applicable laws which means all laws, rules, regulations, codes of practice, research governance or ethical guidelines, or other requirements of any regulatory authority, that may apply to the use of the Material by the Recipient from time to time, including (but not limited to) the Human Tissue Act 2004 or the Human Tissue (Scotland) Act 2006 (as applicable), the Human Tissue (Quality and Safety for Human Application) Regulations 2007, the Human Fertilisation and Embryology Act 1990 (as amended), the EU Tissues and Cells Directive (2004/23/EC) and Commission Directives 2006/17/EC and 2006/86/EC, The Human Tissue Authority Directions and Codes of Practice, and the Medicines for Human Use (Clinical Trials) Regulations 2004, as updated and amended from time to time and, where relevant, the national implementations of the same ("Applicable Laws").

4. The Provider Institution will arrange for the transport of the Material. Risk in and responsibility for the Material shall pass to the Recipient once it is delivered by the courier to the Recipient. If so requested by the Provider Institution the Recipient shall provide it with written confirmation of the safe receipt of the Materials promptly after their delivery to the Recipient's laboratory.

5. The Recipient understands that the Material may have hazardous properties, contain infectious agents or pose other health and safety risks. Subject to clause 10, the Provider Institution makes no representations and gives no warranties either express or implied in relation to it: for example (without limitation), no warranties are given about quality or fitness for a particular purpose, or freedom from infection. The Provider Institution will not be liable for any use made of the Material by the Recipient. The Recipient will use the Material in accordance with good laboratory practice standards, all due skill and care and with dignity, sensitivity and respect. The Recipient will comply with all Applicable Laws, approvals, rules, codes of practice and regulations governing the transportation, storage, use and disposal of the Material. The Recipient warrants that it will only use, or permit the use of the Material in work that has ethical approval.

6. Except to the extent prohibited by law and subject to clause 10, the Recipient assumes all liability for damages which may arise from its receipt, use, storage or disposal of the Material. The Provider Institution will not be liable to the Recipient for any loss, claim or demand made by the Recipient, or made against the Recipient by any other party, due to or arising from its use, storage or disposal of the Material by the Recipient, except to the extent the law otherwise requires. The Recipient agrees to indemnify the Provider Institution against all and any losses, costs, fees, claims, demands and liabilities which may arise out of or in connection with this Agreement or the Recipient's use of the Material unless such claim relates to death or personal injury caused by the Provider Institution's negligence.

7. The liability of either Party for any breach of this Agreement, or arising in any other way out of the subject matter of this Agreement, will not extend to loss of business or profit, or to any indirect or consequential damages or losses. The maximum liability of the Provider Institution for any negligence or arising in any other way out of the subject matter of this Agreement shall not exceed £10,000.

8. The Recipient agrees to obtain the written consent of the Provider Institution if there is any material change to the proposed use of the Material in the Services as described in Appendix 1.

9. The Recipient acknowledges that the results of the Services shall belong to the Provider Institution. Nothing in this Agreement grants the Recipient any rights over the Material (other than as specifically granted by this Agreement) or under any patents, nor any right to use, or permit the use of, any products or processes containing, using, or directly derived from the Material for profit-making or commercial purposes ("Commercial Use") which are not covered under this Agreement. Always subject to prior ethical approval, should the Recipient wish to make Commercial Use of the Material and should the Provider Institution be willing and able to grant a licence for such purposes, the Parties shall negotiate in good faith to agree an appropriate licence or revenue sharing agreement on fair and reasonable terms.

10. The Provider Institution warrants that where required by Applicable Laws the Material has been obtained from humans with the appropriate consent as required by the Human Tissue Act 2004 or the Human Tissue (Scotland) Act 2006 (as applicable) and with ethical approval and the Provider Institution shall be liable for any claims arising due to the breach of this warranty. The Provider Institution hereby grants to the Recipient a non-exclusive licence to use the Material for the Services only. The Provider Institution will not provide any information (and does not intend to provide any information) which has led or may lead to the Recipient being able to identify the person from whom the relevant material came. In the event that any data is or becomes personal data when held or processed by the Recipient the Recipient will comply with all Applicable Laws including the Data Protection Act 2018.

11. The Recipient undertakes to store the Material in accordance with all Applicable Laws and not to attempt to identify or contact the donor of the Material or to compromise or otherwise infringe the confidentiality of information on the donors and their right to privacy.

12. If, as per the details included in Appendix 1, the Material is to be transferred to another organisation for the purposes of the Services, the responsibility for compliance with the terms of this Agreement rests with the Recipient.

13. The Recipient shall not publish any confidential or proprietary information belonging to the Provider Institution without its prior written consent, including such information contained within the Material provided. Confidential and proprietary information shall be deemed to include information which was described as such at the point of disclosure and/or was marked as either "confidential" or "proprietary". The confidentiality obligations in this clause shall not apply where the confidential or proprietary information:

    (a) has become public knowledge, other than through an unauthorised disclosure by the Recipient;

    (b) was already known to the Recipient, prior to disclosure by the Provider Institution;

    (c) was disclosed to the Recipient by a third party, whom to the Recipient's knowledge, was not under any obligation of confidence to the Provider Institution;

    (d) was released from confidential status by written authorisation of the Provider Institution; or

    (e) is required to be disclosed by law or by requirement of a regulatory body or court order.

15. The Provider Institution has the right to terminate this agreement forthwith at any time by means of written notice to Recipient if the ethical approval is withdrawn or if the

Recipient is in breach of this Agreement. In the case of any termination, the Recipient shall immediately discontinue all use of the Material and, at the Provider Institution's discretion, promptly return or destroy (at the Recipient's own cost) all unused Material and provide written confirmation that this has been completed. If requested, the Recipient must certify that it has complied in full with any such requirement of the Provider Institution. Should an individual donor or their next of kin rescind their consent, the Provider Institution will require and the Recipient agrees to discontinue using the appropriately identified sample and return or destroy it in accordance with the Provider Institution's instructions.

16. No person except a Party to this Agreement has any right to prevent the amendment or termination of this Agreement, and no person except a Party to this Agreement may enforce any benefit conferred by this Agreement, unless this Agreement expressly provides otherwise.

17. Nothing in this Agreement shall prevent or impede the Provider Institution from being able to use the Material for any purpose, including but not limited to distribution and licensing of the Material to third parties, whether public, private or third sector, for any purpose.

18. Nothing in this Agreement shall create, evidence or imply any agency, partnership or joint venture between the Parties. No Party shall act or describe itself as the agent of the other Party nor shall it represent that it has any authority to make commitments on behalf of the other Party

19. This Agreement shall be governed and construed in accordance with the laws of England and Wales and the Parties agree to the exclusive jurisdiction of the English Courts.

Accepted and Agreed by an authorised signatory on behalf of	Accepted and Agreed on behalf of
Resolve Biosciences GmbH	The Chancellor Masters and Scholars of the University of Oxford
Name: Mario Scheel-Werner	Name: Gavin Brown
Position: Head Commercial Operations EMEA	Position: Research Contracts Lead
Signature	Signature: DocuSigned by: Gavin Brown 8B6685D14BD9487...
Date: 10 March 2022	Date: 10 March 2022

Provider Ref: R80564/CN001
Recipient Ref:

APPENDIX 1:

Services description and details of Materials

1. DETAILS OF MATERIALS: 8 breast tissue samples

2. LOCATION OF LABORATORY WHERE MATERIALS ARE TO BE HELD/USED:

Resolve BioSciences GmbH
Creative Campus Monheim, Geb. A03
Alfred-Nobel-Str. 10
40789 Monheim

HTA LICENCE / ETHICS APPROVAL:

Complete the following, as applicable:

✓☐ The Services have been given a favourable opinion by an ethics committee which, within the UK, is recognised under the Human Tissue Act 2004 or the Human Tissue (Scotland) Act 2006. Please provide the reference of the opinion and name of the committee: HCRW 19/sc/0025 and OCHRE 21A066

APPENDIX 2:

Delivery and Storage of Materials

TO BE COMPLETED BY THE PROVIDER INSTITUTION:

CONDITIONS OF STORAGE: -80º celsius

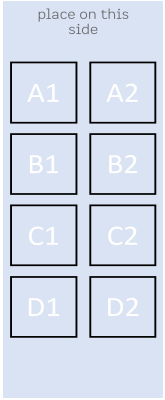
4. RETURN/DESTRUCTION OF SURPLUS MATERIALS ON COMPLETION OF THE SERVICES

If there are any Materials left over at the completion of the Services, the Recipient needs to provide confirmation to the Provider Institution that any remaining Material has been destroyed.

Sample details

The order of each well is defined by its position relative to the "place on this side" marking.
Please make sure the "place on this side" markings are readable to ensure placement of samples on the correct side (see tab "appendix")
Slides are not numbered. Use a permanent marker to label each slide in one of the lower corners. This can be done on either side.
Please fill in details for each individual glass slide and email this file to your Customer Technology Advisor.

Slide 1



Well ID	Species	Tissue type	Fixation	Sample ID / description
A1	Human	breast tumour	Fresh frozen	BC_18_3
A2	Human	breast tumour	Fresh frozen	BC_20_2
B1	Human	breast tumour	Fresh frozen	BC_23_4
B2	Human	breast tumour	Fresh frozen	BC_20_4
C1	Human	breast tumour	Fresh frozen	TRN-01
C2	Human	breast tumour	Fresh frozen	TRN-02
D1	Human	breast tumour	Fresh frozen	TRN-03
D2	Human	breast tumour	Fresh frozen	TRN-04

Purpose	

Priority	

(in case of multiple slides)

D.9 Cell Ranger Metrics

10x GENOMICS Cell Ranger • aggr

BRO18

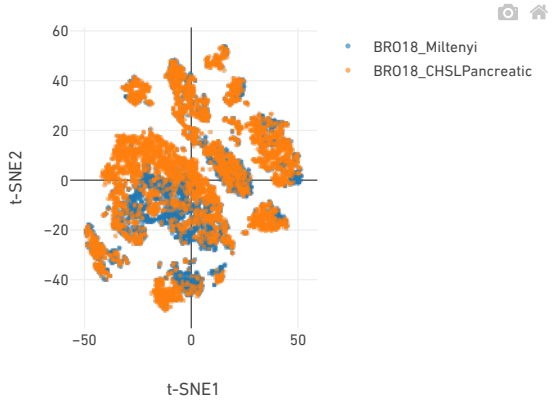
Count Summary | Count Analysis

11,951
Estimated Number of Cells

Aggregation

Pre-Normalization Total Number of Reads	420,900,348
Post-Normalization Total Number of Reads	420,900,348
Pre-Normalization Mean Reads per Cell	35,219
Post-Normalization Mean Reads per Cell	35,219
Fraction of Reads Kept (BRO18_Miltenyi)	100.0%
Fraction of Reads Kept (BRO18_CHSLPancreatic)	100.0%
Pre-Normalization Total Reads per Cell (BRO18_Miltenyi)	48,882
Pre-Normalization Total Reads per Cell (BRO18_CHSLPancreatic)	26,456
Pre-Normalization Confidently Mapped Barcoded Reads per Cell (BRO18_Miltenyi)	17,036
Pre-Normalization Confidently Mapped Barcoded Reads per Cell (BRO18_CHSLPancreatic)	11,271

t-SNE Projection Colored by Library ID



Cells

Estimated Number of Cells	11,951
Fraction Reads in Cells	84.8%
Median Genes per Cell	1,217
Median UMI Counts per Cell	3,474

Sample

Sample ID	BRO18
Sample Description	
Chemistry	Single Cell 3' v3
Transcriptome	GRCh38
Pipeline Version	cellranger-5.0.0

10x GENOMICS Cell Ranger • count

gene_expression_scs_004_3p

Summary | Analysis

9,095
Estimated Number of Cells

20,225
Mean Reads per Cell

1,310
Median Genes per Cell

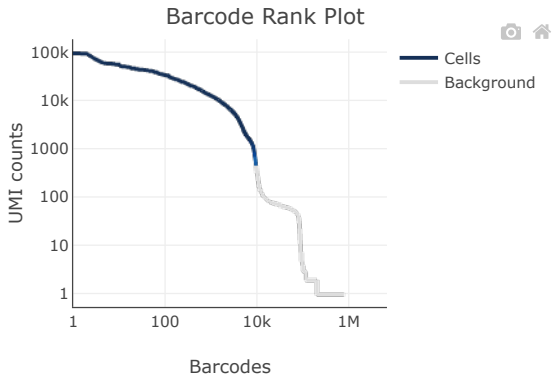
Sequencing

Number of Reads	183,948,312
Number of Short Reads Skipped	0
Valid Barcodes	97.6%
Valid UMIs	99.9%
Sequencing Saturation	30.1%
Q30 Bases in Barcode	95.9%
Q30 Bases in RNA Read	85.5%
Q30 Bases in UMI	94.9%

Mapping

Reads Mapped to Genome	94.4%
Reads Mapped Confidently to Genome	89.0%
Reads Mapped Confidently to Intergenic Regions	5.5%
Reads Mapped Confidently to Intronic Regions	31.9%
Reads Mapped Confidently to Exonic Regions	51.6%
Reads Mapped Confidently to Transcriptome	47.7%
Reads Mapped Antisense to Gene	1.5%

Cells



Estimated Number of Cells	9,095
Fraction Reads in Cells	89.7%
Mean Reads per Cell	20,225
Median Genes per Cell	1,310
Total Genes Detected	23,762
Median UMI Counts per Cell	3,243

Sample

Sample ID	gene_expression_scs_004_3p
Sample Description	
Chemistry	Single Cell 3' v3
Include introns	False
Reference Path	…nger/refdata-cellranger-GRCh38-3.0.0
Transcriptome	GRCh38-3.0.0
Pipeline Version	cellranger-5.0.0

10x GENOMICS Cell Ranger • count

gene_expression_scs_004_5p

Summary | Analysis

5,287
Estimated Number of Cells

21,114
Mean Reads per Cell

962
Median Genes per Cell

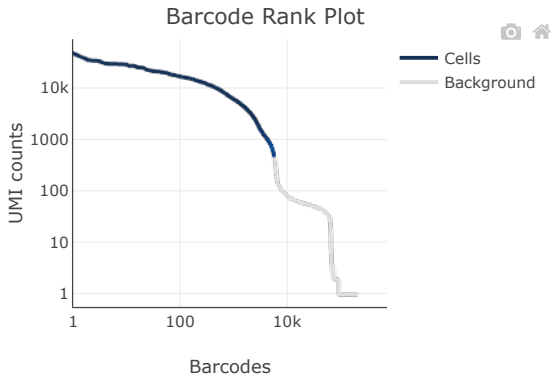
Sequencing

Number of Reads	111,632,166
Number of Short Reads Skipped	0
Valid Barcodes	77.3%
Valid UMIs	99.9%
Sequencing Saturation	36.4%
Q30 Bases in Barcode	95.2%
Q30 Bases in RNA Read	81.7%
Q30 Bases in UMI	94.0%

Mapping

Reads Mapped to Genome	82.7%
Reads Mapped Confidently to Genome	57.3%
Reads Mapped Confidently to Intergenic Regions	5.4%
Reads Mapped Confidently to Intronic Regions	4.8%
Reads Mapped Confidently to Exonic Regions	47.1%
Reads Mapped Confidently to Transcriptome	36.5%
Reads Mapped Antisense to Gene	8.4%

Cells



Estimated Number of Cells	5,287
Fraction Reads in Cells	85.2%
Mean Reads per Cell	21,114
Median Genes per Cell	962
Total Genes Detected	20,975
Median UMI Counts per Cell	2,309

Sample

Sample ID	gene_expression_scs_004_5p
Sample Description	
Chemistry	Single Cell 5' R2-only
Include introns	False
Reference Path	…nger/refdata-cellranger-GRCh38-3.0.0
Transcriptome	GRCh38-3.0.0
Pipeline Version	cellranger-5.0.0

10x GENOMICS Cell Ranger • count

gene_expression_scs_005_5p

Summary | Analysis

7,392
Estimated Number of Cells

21,580
Mean Reads per Cell

830
Median Genes per Cell

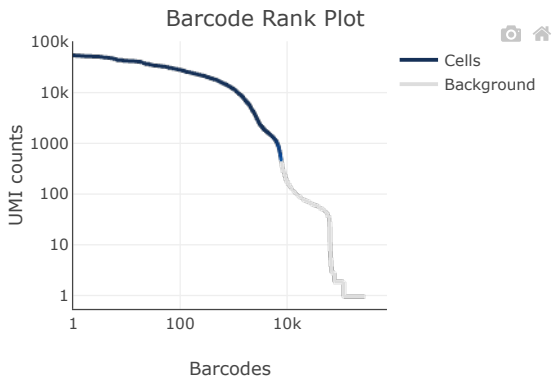
Sequencing

Number of Reads	159,516,886
Number of Short Reads Skipped	0
Valid Barcodes	80.7%
Valid UMIs	99.9%
Sequencing Saturation	39.0%
Q30 Bases in Barcode	95.2%
Q30 Bases in RNA Read	83.2%
Q30 Bases in UMI	94.1%

Mapping

Reads Mapped to Genome	87.8%
Reads Mapped Confidently to Genome	69.6%
Reads Mapped Confidently to Intergenic Regions	7.8%
Reads Mapped Confidently to Intronic Regions	5.1%
Reads Mapped Confidently to Exonic Regions	56.7%
Reads Mapped Confidently to Transcriptome	46.0%
Reads Mapped Antisense to Gene	7.7%

Cells



Estimated Number of Cells	7,392
Fraction Reads in Cells	89.6%
Mean Reads per Cell	21,580
Median Genes per Cell	830
Total Genes Detected	22,193
Median UMI Counts per Cell	1,964

Sample

Sample ID	gene_expression_scs_005_5p
Sample Description	
Chemistry	Single Cell 5' R2-only
Include introns	False
Reference Path	…nger/refdata-cellranger-GRCh38-3.0.0
Transcriptome	GRCh38-3.0.0
Pipeline Version	cellranger-5.0.0

10x GENOMICS Cell Ranger • vdj

bcr_count_scs_004

Alerts

The analysis detected ⚠ 1 warning.

	Alert	Value	Detail
⚠	Low RNA Read Q30 Fraction (Illumina R1 for Single Cell V(D)J)	84.6%	Ideal > 85%. Application performance may be affected.

Summary | Analysis

515
Estimated Number of Cells

29,225
Mean Reads per Cell

482
Number of Cells With Productive V-J Spanning Pair

Enrichment

Reads Mapped to Any V(D)J Gene	94.3%
Reads Mapped to IGH	18.7%
Reads Mapped to IGK	65.6%
Reads Mapped to IGL	10.0%

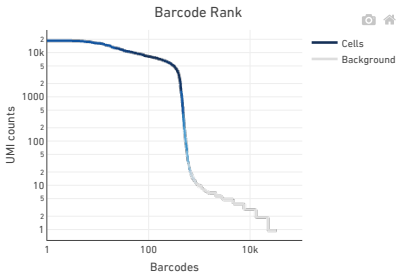
V(D)J Expression

Median IGH UMIs per Cell	2,370.00
Median IGK UMIs per Cell	11,789.00
Median IGL UMIs per Cell	8,956.00

V(D)J Annotation

Cells With Productive V-J Spanning Pair	93.6%
Cells With Productive V-J Spanning (IGK, IGH) Pair	76.7%
Cells With Productive V-J Spanning (IGL, IGH) Pair	17.1%
Paired Clonotype Diversity	37.83
Cells With IGH Contig	96.7%
Cells With IGK Contig	83.1%
Cells With IGL Contig	29.7%
Cells With CDR3-annotated IGH Contig	96.1%
Cells With CDR3-annotated IGK Contig	82.9%
Cells With CDR3-annotated IGL Contig	23.1%
Cells With V-J Spanning IGH Contig	96.1%
Cells With V-J Spanning IGK Contig	81.7%
Cells With V-J Spanning IGL Contig	23.5%
Cells With Productive IGH Contig	95.0%
Cells With Productive IGK Contig	80.4%
Cells With Productive IGL Contig	18.4%

Cells



Estimated Number of Cells	515
Mean Reads per Cell	29,225
Mean Used Reads per Cell	14,829
Fraction Reads in Cells	84.8%

Sequencing

Number of Reads	15,050,934
Number of Short Reads Skipped	0
Valid Barcodes	96.9%
Q30 Bases in Barcode	93.7%
Q30 Bases in RNA Read 1	84.6%
Q30 Bases in UMI	92.7%

Sample

Sample ID	bcr_count_scs_004
Sample Description	
Chemistry	Single Cell V(D)J R2-only
V(D)J Reference	vdj_GRCh38_alts_ensembl-5.0.0
Pipeline Version	cellranger-5.0.0

10x GENOMICS Cell Ranger • vdj

tcr_count_scs_005

Alerts

The analysis detected ⚠ 1 warning.

	Alert	Value	Detail
⚠	Low Fraction Valid Barcodes	78.2%	Ideal > 85%. This usually indicates a quality issue with the Ilumina R1 read. Application performance may be affected.

Summary | Analysis

3,550
Estimated Number of Cells

11,865
Mean Reads per Cell

3,017
Number of Cells With Productive V-J Spanning Pair

Enrichment

Reads Mapped to Any V(D)J Gene	59.2%
Reads Mapped to TRA	16.6%
Reads Mapped to TRB	40.1%

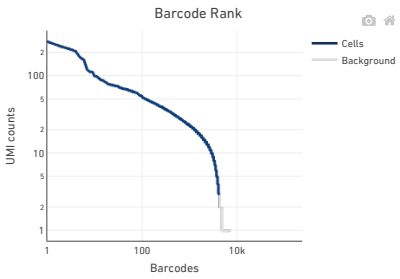
V(D)J Expression

Median TRA UMIs per Cell	4.00
Median TRB UMIs per Cell	12.00

V(D)J Annotation

Cells With Productive V-J Spanning Pair	85.0%
Cells With Productive V-J Spanning (TRA, TRB) Pair	85.0%
Paired Clonotype Diversity	305.16
Cells With TRA Contig	90.2%
Cells With TRB Contig	98.6%
Cells With CDR3-annotated TRA Contig	88.8%
Cells With CDR3-annotated TRB Contig	98.2%
Cells With V-J Spanning TRA Contig	89.5%
Cells With V-J Spanning TRB Contig	98.3%
Cells With Productive TRA Contig	88.0%
Cells With Productive TRB Contig	97.0%

Cells



Estimated Number of Cells	3,550
Mean Reads per Cell	11,865
Mean Used Reads per Cell	4,935
Fraction Reads in Cells	61.7%

Sequencing

Number of Reads	42,120,829
Number of Short Reads Skipped	0
Valid Barcodes	78.2%
Q30 Bases in Barcode	95.0%
Q30 Bases in RNA Read 1	87.4%
Q30 Bases in UMI	94.2%

Sample

Sample ID	tcr_count_scs_005
Sample Description	
Chemistry	Single Cell V(D)J R2-only
V(D)J Reference	vdj_GRCh38_alts_ensembl-5.0.0
Pipeline Version	cellranger-5.0.0

10x GENOMICS Cell Ranger • vdj

bcr_count_scs_005

Summary | Analysis

808
Estimated Number of Cells

15,852
Mean Reads per Cell

676
Number of Cells With Productive V-J Spanning Pair

Enrichment

Reads Mapped to Any V(D)J Gene	93.8%
Reads Mapped to IGH	21.3%
Reads Mapped to IGK	48.0%
Reads Mapped to IGL	24.5%

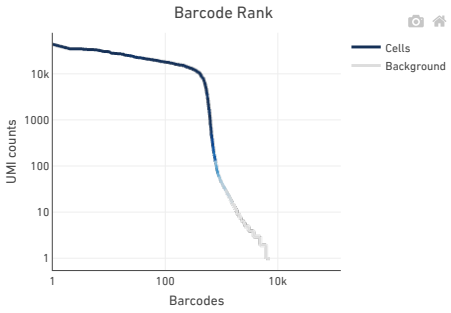
V(D)J Expression

Median IGH UMIs per Cell	1,701.00
Median IGK UMIs per Cell	8,461.00
Median IGL UMIs per Cell	7,129.00

V(D)J Annotation

Cells With Productive V-J Spanning Pair	83.7%
Cells With Productive V-J Spanning (IGK, IGH) Pair	52.2%
Cells With Productive V-J Spanning (IGL, IGH) Pair	31.8%
Paired Clonotype Diversity	119.14
Cells With IGH Contig	90.8%
Cells With IGK Contig	63.1%
Cells With IGL Contig	43.4%
Cells With CDR3-annotated IGH Contig	89.7%
Cells With CDR3-annotated IGK Contig	60.1%
Cells With CDR3-annotated IGL Contig	42.0%
Cells With V-J Spanning IGH Contig	90.3%
Cells With V-J Spanning IGK Contig	61.8%
Cells With V-J Spanning IGL Contig	42.2%
Cells With Productive IGH Contig	85.1%
Cells With Productive IGK Contig	59.3%
Cells With Productive IGL Contig	39.6%

Cells



Estimated Number of Cells	808
Mean Reads per Cell	15,852
Mean Used Reads per Cell	12,961
Fraction Reads in Cells	92.8%

Sequencing

Number of Reads	12,808,657
Number of Short Reads Skipped	0
Valid Barcodes	97.0%
Q30 Bases in Barcode	93.9%
Q30 Bases in RNA Read 1	85.3%
Q30 Bases in UMI	92.7%

Sample

Sample ID	bcr_count_scs_005
Sample Description	
Chemistry	Single Cell V(D)J R2-only
V(D)J Reference	vdj_GRCh38_alts_ensembl-5.0.0
Pipeline Version	cellranger-5.0.0

10x GENOMICS Cell Ranger • vdj

tcr_count_scs_005

Alerts

The analysis detected ⚠ 1 warning.

	Alert	Value	Detail
⚠	Low Fraction Valid Barcodes	78.2%	Ideal > 85%. This usually indicates a quality issue with the Illumina R1 read. Application performance may be affected.

Summary | Analysis

3,550
Estimated Number of Cells

11,865
Mean Reads per Cell

3,017
Number of Cells With Productive V-J Spanning Pair

Enrichment

Reads Mapped to Any V(D)J Gene	59.2%
Reads Mapped to TRA	16.6%
Reads Mapped to TRB	40.1%

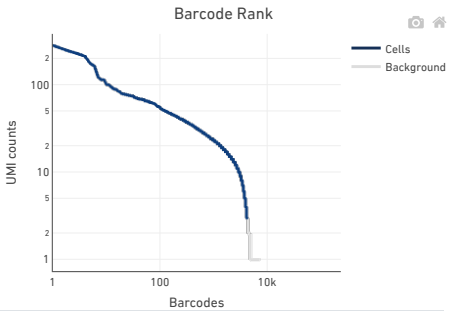
V(D)J Expression

Median TRA UMIs per Cell	4.00
Median TRB UMIs per Cell	12.00

V(D)J Annotation

Cells With Productive V-J Spanning Pair	85.0%
Cells With Productive V-J Spanning (TRA, TRB) Pair	85.0%
Paired Clonotype Diversity	305.16
Cells With TRA Contig	90.2%
Cells With TRB Contig	98.6%
Cells With CDR3-annotated TRA Contig	88.8%
Cells With CDR3-annotated TRB Contig	98.2%
Cells With V-J Spanning TRA Contig	89.5%
Cells With V-J Spanning TRB Contig	98.3%
Cells With Productive TRA Contig	88.0%
Cells With Productive TRB Contig	97.0%

Cells



Estimated Number of Cells	3,550
Mean Reads per Cell	11,865
Mean Used Reads per Cell	4,935
Fraction Reads in Cells	61.7%

Sequencing

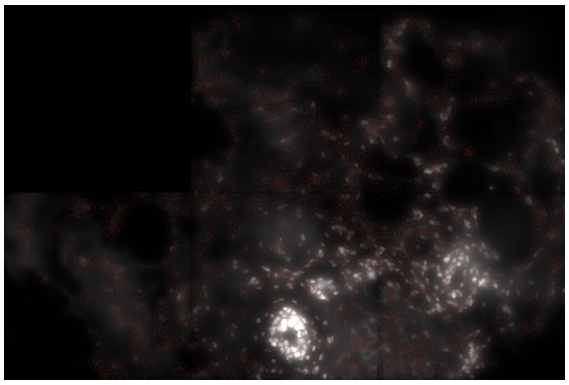
Number of Reads	42,120,829
Number of Short Reads Skipped	0
Valid Barcodes	78.2%
Q30 Bases in Barcode	95.0%
Q30 Bases in RNA Read 1	87.4%
Q30 Bases in UMI	94.2%

Sample

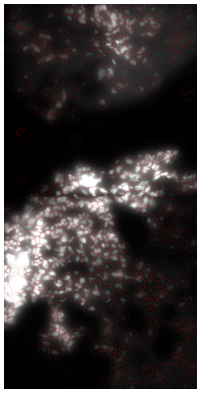
Sample ID	tcr_count_scs_005
Sample Description	
Chemistry	Single Cell V(D)J R2-only
V(D)J Reference	vdj_GRCh38_alts_ensembl-5.0.0
Pipeline Version	cellranger-5.0.0

D.10 Cell Segmentations

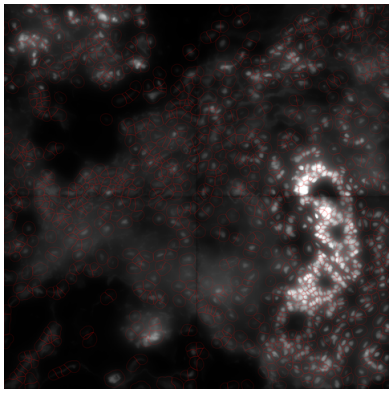
Table D.3: Mosaic of Cell Segmentations.



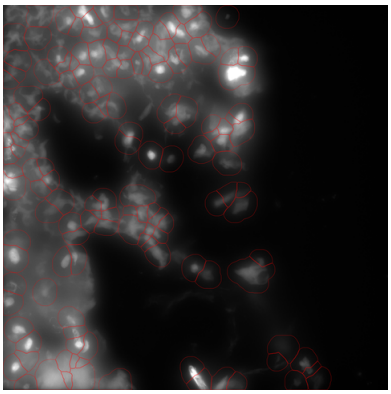
$A1-1$



$A1-2$



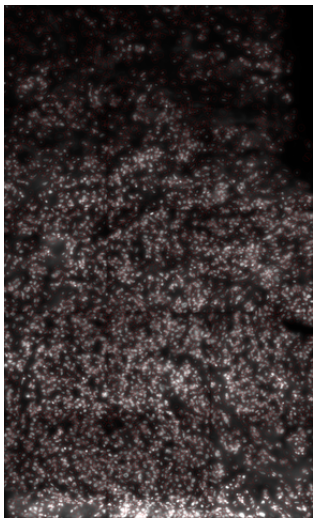
$A1-3$



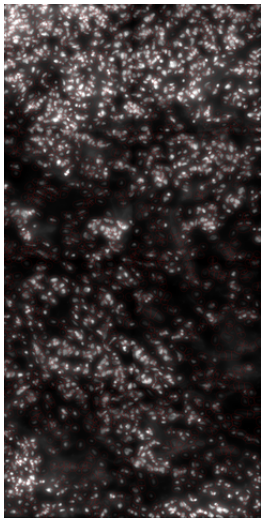
$A1-4$



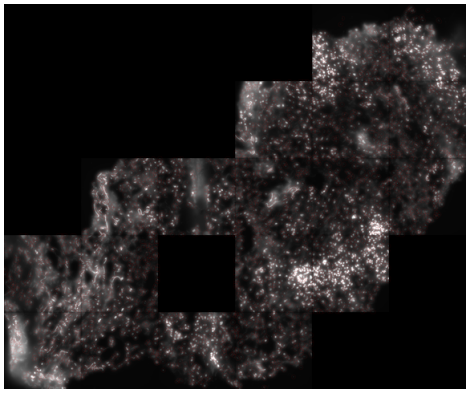
$A2-1$



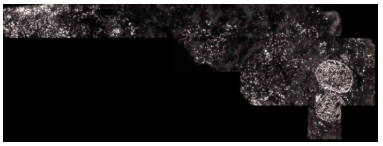
$A2-2$



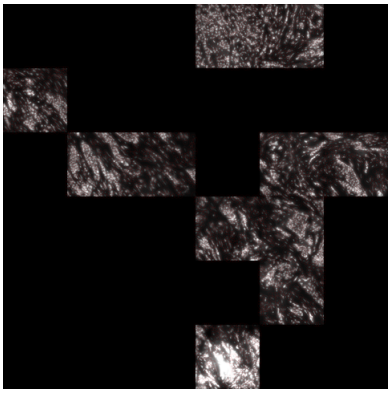
$A2-3$



$B1-1$



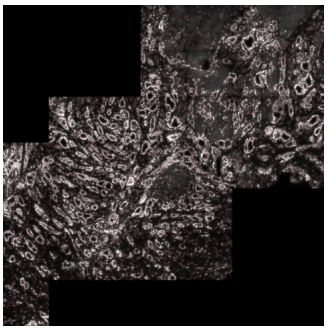
$B2-1$



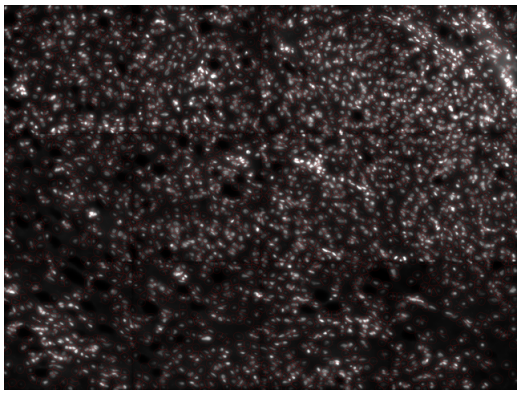
$C1-1$



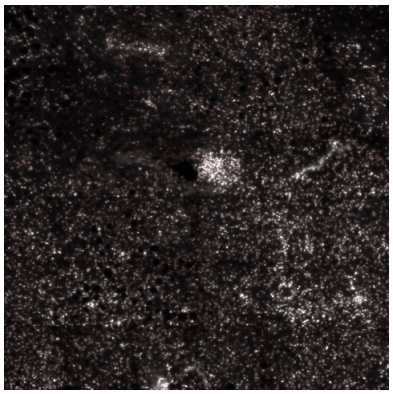
$C1-2$



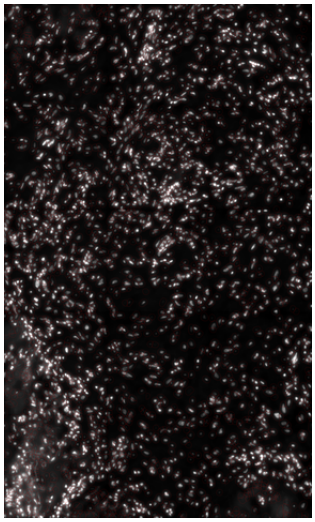
$C2-1$



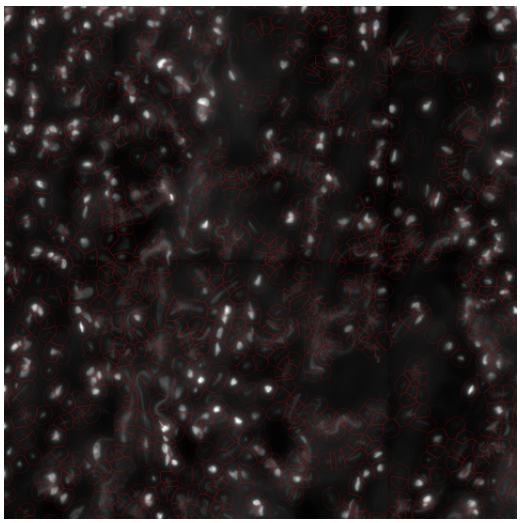
$D1-1$



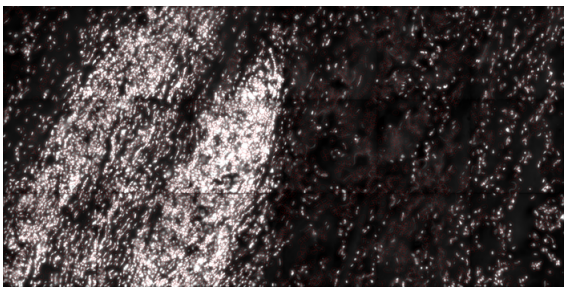
$D1-2$



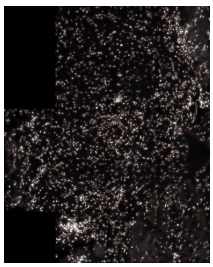
$D2-1$



$D2-2$



$D2-3$



$D2-4$

D.11 Cluster Markers

#		pVal	avgLog2FC	pct1	pct2	pValAdj	cluster	gene
1		0	2.90202142840733	0.898	0.139	0	0	GZMK
2		0	2.32908721269346	0.846	0.23	0	0	CST7
3		0	2.09382539628455	0.932	0.262	0	0	CCL5
4		0	2.01312949630155	0.764	0.172	0	0	GZMA
5		0	2.00732205684697	0.736	0.166	0	0	NKG7
6		0	2.00256624142963	0.663	0.239	0	0	TUBA4A
7		0	1.9807269232457	0.501	0.087	0	0	CRTAM
8		0	1.92423527925786	0.501	0.089	0	0	CD8A
9		0	1.78956005620689	0.39	0.064	0	0	CD8B
10		0	1.52966098231092	0.968	0.564	0	0	CXCR4
11		0	1.47613449503001	0.517	0.225	0	0	DUSP2
12		0	1.43424152638416	0.698	0.366	0	0	LEPROTL1
13		0	1.4277616385203	0.422	0.109	0	0	GZMM
14		0	1.36056234180061	0.38	0.099	0	0	SH2D1A
15		0	1.32968448121638	0.464	0.189	0	0	RUNX3
16		0	1.26475861241353	0.846	0.578	0	0	CREM

17	0	1.25702156042335	0.524	0.219	0	0	SYTL3
18	0	1.24926428118504	0.429	0.165	0	0	RNF125
19	0	1.2451123168067	0.92	0.535	0	0	PTPRC
20	0	1.23161111324143	0.605	0.33	0	0	FYN
21	0	1.22287882053321	0.293	0.073	0	0	GZMH
22	0	1.22248928079385	0.59	0.241	0	0	CD3E
23	0	1.22184210608146	0.311	0.077	0	0	TRAT1
24	0	1.19468747972472	0.692	0.409	0	0	ISG20
25	0	1.1813354413548	0.603	0.302	0	0	DUSP4
26	0	1.17501812844049	0.342	0.1	0	0	CTSW
27	0	1.1648225943216	0.645	0.321	0	0	HCST
28	0	1.15284336589288	0.64	0.369	0	0	LCP1
29	0	1.12008198256795	0.673	0.455	0	0	RNF19A
30	0	1.11929502051585	0.86	0.615	0	0	EZR
31	0	1.09860411312911	0.824	0.605	0	0	ARHGDIB
32	0	1.0934658200038	0.889	0.738	0	0	SARAF
33	0	1.04304504757273	0.757	0.619	0	0	SRSF7

34	0	1.03745442775318	0.591	0.257	0	0	CD3D
35	2.26632751009579E-304	1.02142817767328	0.463	0.194	7.60080920335925E-300	0	CD3G
36	3.05302358496763E-304	1.153557810174	0.683	0.508	1.02392304992644E-299	0	YPEL5
37	1.66930432766347E-300	1.02149145037867	0.538	0.267	5.59851285411774E-296	0	SLA
38	3.35954719488106E-300	1.24576754696436	0.435	0.194	1.12672493821921E-295	0	AKNA
39	4.41365051868166E-286	1.36732617476419	0.561	0.35	1.48025011095546E-281	0	PIK3R1
40	8.2570897341801E-271	1.21825470784184	0.567	0.345	2.76926275504932E-266	0	CNOT6L
41	9.73297980673464E-249	1.20405639016793	0.601	0.439	3.26424676758266E-244	0	FAM177A1
42	1.37207589056032E-230	1.17039115383341	0.506	0.301	4.60166812176119E-226	0	CEMIP2
43	1.07427431134263E-220	1.12136074016666	0.523	0.336	3.6029011853809E-216	0	PPP2R5C
44	3.96769009729088E-206	1.07253583865314	0.499	0.292	1.33068390482941E-201	0	ITM2A
45	1.14088168742657E-202	1.02053187723822	0.289	0.112	3.82628900329124E-198	0	SH2D2A
46	1.84471192636243E-194	0.990976520103249	0.36	0.172	6.18679485863432E-190	0	PARP8
47	7.68663584313367E-167	1.02267730972701	0.58	0.457	2.57794392907017E-162	0	LITAF
48	5.81963519427292E-164	1.03054664122911	0.541	0.403	1.95178925145525E-159	0	PDCD4
49	1.15028665685075E-129	0.965816446138302	0.35	0.199	3.85783138974604E-125	0	TENT5C
50	9.81604944132348E-72	0.937127672363046	0.194	0.105	3.29210666163107E-67	0	ENC1

51	0	2.33988214145301	0.931	0.261	0	1	IL7R
52	0	1.66349540256444	0.476	0.118	0	1	CCR7
53	0	1.36129568289033	0.58	0.238	0	1	CRYBG1
54	0	1.23379291072358	0.934	0.74	0	1	SARAF
55	0	1.08353103575297	0.535	0.197	0	1	SPOCK2
56	0	1.07601270924393	0.822	0.408	0	1	CD52
57	0	1.06298876479937	0.706	0.38	0	1	LEPROTL1
58	0	0.957423591275444	0.988	0.967	0	1	RPS12
59	0	0.950631927611137	0.522	0.177	0	1	LTB
60	0	0.923183801784647	0.986	0.954	0	1	RPL30
61	0	0.917640953536417	0.982	0.948	0	1	RPL34
62	0	0.899238644224345	0.991	0.967	0	1	RPS27
63	0	0.880616172884022	0.929	0.855	0	1	RPSA
64	0	0.87917228929996	0.986	0.955	0	1	RPL11
65	0	0.873772162376857	0.987	0.954	0	1	RPL32
66	0	0.870465741056462	0.98	0.944	0	1	RPS3
67	0	0.870405622013593	0.986	0.955	0	1	RPS27A

68	0	0.863605694929294	0.942	0.882	0	1	RPS21
69	0	0.845465776881082	0.967	0.93	0	1	RPS25
70	0	0.843127049671598	0.966	0.928	0	1	RPL21
71	0	0.826980984099952	0.979	0.952	0	1	RPS14
72	0	0.821698336180737	0.983	0.945	0	1	RPS3A
73	0	0.821432150567392	0.943	0.584	0	1	CXCR4
74	2.75257332540543E-303	0.957688829961609	0.894	0.804	9.23158041874474E-299	1	PABPC1
75	7.47874518233654E-288	1.58902561639449	0.805	0.597	2.50822155925203E-283	1	ANXA1
76	7.0632889400028E-264	0.954439739724141	0.345	0.102	2.36888584469814E-259	1	ITK
77	7.37365720231792E-258	0.914610769803682	0.851	0.612	2.47297715251338E-253	1	ARHGDIB
78	6.46578599273931E-256	0.907724548593659	0.841	0.59	2.16849530624491E-251	1	CREM
79	1.25623865313262E-254	0.881493458945788	0.595	0.256	4.21317319487618E-250	1	CD3E
80	2.40215068759746E-239	1.028043416837	0.241	0.059	8.05633297606436E-235	1	GPR171
81	7.40021195129643E-239	1.0244726220413	0.861	0.626	2.4818830842258E-234	1	EZR
82	4.79014098072097E-235	0.947970968772222	0.558	0.268	1.6065174821142E-230	1	CD48
83	2.14492888479307E-220	1.0776747879964	0.218	0.052	7.193662493819E-216	1	KLRB1
84	6.32155035941041E-209	0.880141735744962	0.861	0.701	2.12012155953906E-204	1	TSC22D3

85	8.80818710059173E-185	1.08416789619142	0.747	0.59	2.95408978979645E-180	1	ZFP36L2
86	6.39473313545735E-183	0.824430971559515	0.616	0.341	2.14466559896969E-178	1	FYB1
87	4.14451139453129E-182	0.818476895206857	0.828	0.788	1.3899862314979E-177	1	RPS20
88	1.32003262540061E-177	0.829327644954646	0.413	0.178	4.42712541906858E-173	1	RNF125
89	4.19174327690986E-177	0.938500069806838	0.394	0.17	1.40582686021003E-172	1	GIMAP7
90	2.06697723170821E-171	1.02858320101757	0.896	0.862	6.93222823970298E-167	1	RPS29
91	3.92315535550616E-169	0.836083810064148	0.84	0.818	1.31574784312966E-164	1	RPL38
92	1.18283725662185E-163	1.03099003338336	0.51	0.297	3.96699959125837E-159	1	FKBP5
93	8.83175620231527E-162	0.993335798295275	0.671	0.482	2.9619943951325E-157	1	FXYD5
94	4.71436628949207E-154	0.895598569336197	0.349	0.154	1.58110416616985E-149	1	SC5D
95	1.75819813123095E-147	0.932605610979387	0.32	0.135	5.89664489252236E-143	1	SELL
96	1.84610533380489E-133	0.873047617000312	0.611	0.445	6.19146806851485E-129	1	FAM177A1
97	6.58003681426632E-127	0.828014971456976	0.547	0.345	2.20681274676864E-122	1	ARL4C
98	1.03075994758413E-111	0.816543638974104	0.316	0.153	3.45696271220766E-107	1	PTGER4
99	2.02863998703794E-87	1.07628213000276	0.451	0.312	6.80365278852785E-83	1	KLF2
100	8.95416595656163E-58	0.834204425279165	0.398	0.292	3.00304817851164E-53	1	GLIPR1
101	0	2.62856927976423	0.704	0.161	0	2	LTB

102	0	2.62192352705931	0.632	0.084	0	2	BATF
103	0	2.24032955834178	0.411	0.028	0	2	IL2RA
104	0	2.22260238776622	0.954	0.526	0	2	IL32
105	0	2.13353309572999	0.622	0.187	0	2	CARD16
106	0	2.13149460763896	0.563	0.075	0	2	CTLA4
107	0	2.12785858098359	0.39	0.064	0	2	TNFRSF4
108	0	2.06804713966104	0.38	0.01	0	2	FOXP3
109	0	2.04366478428133	0.545	0.087	0	2	TNFRSF9
110	0	1.94927741775349	0.572	0.1	0	2	TIGIT
111	0	1.84081076622405	0.827	0.288	0	2	CD2
112	0	1.73303013412451	0.376	0.055	0	2	TNFRSF18
113	0	1.70759960599711	0.406	0.09	0	2	TBC1D4
114	0	1.70522538081956	0.327	0.022	0	2	RTKN2
115	0	1.61637533770577	0.572	0.176	0	2	TRAC
116	0	1.56132811861104	0.463	0.164	0	2	PMAIP1
117	0	1.55671778504299	0.428	0.121	0	2	GBP5
118	0	1.51986755685176	0.514	0.197	0	2	PBXIP1

119	0	1.51561978794252	0.314	0.064	0	2	LAYN
120	0	1.51362892388663	0.455	0.12	0	2	CD27
121	0	1.50746542230451	0.343	0.037	0	2	IKZF2
122	0	1.50650483267022	0.412	0.078	0	2	ICOS
123	0	1.50045468753842	0.367	0.094	0	2	GK
124	0	1.47802640393126	0.211	0.004	0	2	CCR8
125	0	1.45244215929546	0.607	0.191	0	2	SPOCK2
126	0	1.43786399155127	0.913	0.614	0	2	S100A4
127	0	1.38551795233563	0.569	0.217	0	2	TRBC2
128	0	1.33203793992571	0.346	0.1	0	2	STAM
129	0	1.3308001720069	0.722	0.261	0	2	CD3D
130	0	1.3258225239051	0.661	0.325	0	2	SAMSN1
131	0	1.304620938593	0.243	0.034	0	2	CCR6
132	0	1.29178274949511	0.347	0.074	0	2	DUSP16
133	0	1.24930882533672	0.192	0.018	0	2	IL1R2
134	0	1.23907176642762	0.574	0.21	0	2	CD7
135	0	1.13868277299564	0.229	0.032	0	2	LINC01943

136	0	1.08530939514191	0.262	0.047	0	2	ZC3H12D
137	1.16273872893046E-295	1.50333814313531	0.436	0.148	3.89959314908698E-291	2	TRBC1
138	2.95201492899267E-282	1.16225652869987	0.676	0.357	9.90046766885562E-278	2	CYTIP
139	1.13262013654596E-270	1.09424896555637	0.347	0.104	3.79858141394784E-266	2	P2RY10
140	3.47136708417776E-270	1.0817058120597	0.402	0.132	1.16422709269154E-265	2	SKAP1
141	5.81112468758904E-267	1.25540256390509	0.664	0.392	1.94893499772361E-262	2	STK17B
142	8.61811092916006E-250	1.16546376758363	0.379	0.13	2.8903420434217E-245	2	SELL
143	6.63343334220392E-228	1.13500448993134	0.566	0.297	2.22472087430835E-223	2	CDKN1B
144	1.3387920066241E-214	1.45670114414108	0.513	0.295	4.49004063181591E-210	2	UGP2
145	1.1765702693645E-182	1.12518296496486	0.286	0.105	3.94598136939464E-178	2	HTATIP2
146	4.80159102457585E-181	1.17265317080858	0.405	0.186	1.61035759782225E-176	2	RAB11FIP1
147	1.96060418841393E-99	1.09109369156909	0.357	0.209	6.57547432710264E-95	2	CORO1B
148	1.42105743318553E-90	1.21829318867538	0.224	0.103	4.76594241941763E-86	2	MAGEH1
149	6.68274071118783E-66	1.39497223319592	0.047	0.008	2.24125757971817E-61	2	TRBV20-1
150	5.36475216752495E-41	1.47944402009437	0.032	0.006	1.79923058194452E-36	2	TRBV7-2
151	0	4.04350662357925	0.935	0.181	0	3	ACTA2
152	0	3.97090881200576	0.976	0.232	0	3	TAGLN

153	0	3.78225302432055	0.583	0.029	0	3	MYH11
154	0	3.72155963076353	0.98	0.197	0	3	MYL9
155	0	3.44403126614784	0.172	0.019	0	3	CCL19
156	0	3.41146425957241	0.917	0.123	0	3	ADIRF
157	0	3.38498790239625	0.966	0.199	0	3	TPM2
158	0	3.37717517850593	0.604	0.134	0	3	CCL2
159	0	3.23410474309721	0.714	0.152	0	3	C11orf96
160	0	3.08096633036996	0.998	0.287	0	3	CALD1
161	0	2.86418007993651	0.714	0.081	0	3	RGS5
162	0	2.8077218228513	0.94	0.243	0	3	TPM1
163	0	2.79932741884793	0.926	0.308	0	3	IGFBP7
164	0	2.79091340318686	0.578	0.044	0	3	MT1A
165	0	2.75885294365862	0.674	0.07	0	3	MT1M
166	0	2.706523480959	0.541	0.044	0	3	STEAP4
167	0	2.64097551230044	0.942	0.244	0	3	SPARCL1
168	0	2.6335036146426	0.828	0.11	0	3	MYLK
169	0	2.62718030433632	0.812	0.125	0	3	ADAMTS1

170	0	2.56010198361427	0.892	0.175	0	3	MAP1B
171	0	2.52398435787697	0.73	0.052	0	3	NDUFA4L2
172	0	2.49628487544133	0.934	0.427	0	3	DSTN
173	0	2.48722551078478	0.722	0.071	0	3	SYNPO2
174	0	2.47031788260114	0.953	0.581	0	3	MT2A
175	0	2.44618357800955	0.616	0.124	0	3	IGFBP5
176	0	2.358821737651	0.671	0.089	0	3	ADAMTS4
177	0	2.35678063967577	0.794	0.106	0	3	MCAM
178	0	2.30362205706031	0.917	0.192	0	3	CAV1
179	0	2.28177145192773	0.573	0.075	0	3	ID4
180	0	2.27111140950792	0.649	0.053	0	3	LMOD1
181	0	2.26161719819453	0.672	0.05	0	3	PPP1R14A
182	0	2.26093192200665	0.788	0.143	0	3	CRISPLD2
183	0	2.17382455472496	0.854	0.273	0	3	LPP
184	0	2.12858298559609	0.67	0.229	0	3	CEBPD
185	0	2.09620581561261	0.84	0.168	0	3	LHFPL6
186	0	2.094705526991	0.822	0.154	0	3	EPS8

187	0	2.03377150387358	0.626	0.044	0	3	GJA4
188	0	2.02751626689386	0.646	0.091	0	3	CSRP2
189	0	2.02492292661158	0.732	0.06	0	3	NOTCH3
190	0	2.02211917257448	0.195	0.013	0	3	CCL21
191	0	2.00977983967787	0.287	0.021	0	3	CCL8
192	0	2.00310184768784	0.718	0.114	0	3	CPE
193	0	1.98820493431768	0.646	0.169	0	3	CSRP1
194	0	1.98400125778574	0.336	0.032	0	3	SORBS2
195	0	1.95983765905957	0.793	0.215	0	3	FILIP1L
196	0	1.93816853555266	0.739	0.155	0	3	COL18A1
197	0	1.9048499339798	0.785	0.132	0	3	PDGFRB
198	0	1.85979764880154	0.984	0.432	0	3	IFITM3
199	0	1.85590082324568	0.732	0.149	0	3	MFGE8
200	2.90979389186717E-260	1.88558574290444	0.731	0.386	9.75886675454413E-256	3	GADD45B
201	0	5.14050021323869	0.966	0.043	0	4	C1QB
202	0	5.04847916022665	0.973	0.041	0	4	C1QA
203	0	4.69414483712644	0.835	0.068	0	4	APOC1

204	0	4.68962645330662	0.955	0.031	0	4	C1QC
205	0	4.55701117112165	0.861	0.12	0	4	APOE
206	0	3.91167114361127	1	0.898	0	4	FTL
207	0	3.76674033494965	0.912	0.07	0	4	LYZ
208	0	3.65395051107252	0.995	0.37	0	4	HLA-DRA
209	0	3.3899139775895	0.975	0.32	0	4	CTSB
210	0	3.3574134889517	0.982	0.102	0	4	TYROBP
211	0	3.31474149539234	0.319	0.01	0	4	CCL18
212	0	3.30190738178524	0.942	0.077	0	4	CD68
213	0	3.27336130723174	0.541	0.083	0	4	RNASE1
214	0	3.26982071957693	0.983	0.47	0	4	PSAP
215	0	3.26398030830677	0.97	0.081	0	4	FCER1G
216	0	3.19206591279674	0.986	0.419	0	4	HLA-DPA1
217	0	3.18069929169964	0.911	0.034	0	4	MS4A7
218	0	3.18012061353664	0.793	0.07	0	4	HMOX1
219	0	3.16403928607875	0.89	0.04	0	4	CD14
220	0	3.11561474401526	0.951	0.058	0	4	AIF1

221	0	3.08997613849033	0.995	0.644	0	4	CD74
222	0	3.08365850962147	0.924	0.21	0	4	CTSL
223	0	3.03475375580182	0.97	0.318	0	4	HLA-DRB1
224	0	3.00191371795291	0.492	0.039	0	4	CCL3
225	0	2.996648027069	0.973	0.383	0	4	HLA-DPB1
226	0	2.99411030286513	0.916	0.043	0	4	MS4A6A
227	0	2.98562180316832	0.647	0.184	0	4	SELENOP
228	0	2.96691949400932	0.243	0.029	0	4	SPP1
229	0	2.95554031275134	0.882	0.224	0	4	CTSD
230	0	2.89880050603453	0.905	0.185	0	4	HLA-DQB1
231	0	2.8969704717053	0.916	0.24	0	4	HLA-DRB5
232	0	2.89342449365327	0.887	0.241	0	4	CTSZ
233	0	2.86461712390215	0.825	0.031	0	4	FCGR3A
234	0	2.80305362673668	0.848	0.029	0	4	C5AR1
235	0	2.78601885668188	0.845	0.154	0	4	GPNMB
236	0	2.77723560427902	0.809	0.168	0	4	LGMN
237	0	2.76232393613904	0.521	0.026	0	4	F13A1

238	0	2.75698243902524	0.919	0.174	0	4	HLA-DQA1
239	0	2.75097556594314	0.809	0.02	0	4	CD163
240	0	2.74447114328384	0.88	0.161	0	4	PLAUR
241	0	2.74252786724657	0.768	0.103	0	4	PLTP
242	0	2.73240164275671	0.864	0.035	0	4	FCGR2A
243	0	2.66555351177899	0.938	0.305	0	4	CTSS
244	0	2.63736193064945	0.885	0.216	0	4	SGK1
245	0	2.6369501623772	0.932	0.272	0	4	GRN
246	0	2.56070189793722	0.872	0.182	0	4	HLA-DQA2
247	0	2.53867218105021	0.853	0.056	0	4	CYBB
248	0	2.51683163357981	0.93	0.257	0	4	GLUL
249	0	2.45273769204815	0.964	0.354	0	4	NPC2
250	0	2.45102176912458	0.903	0.165	0	4	HLA-DMA
251	0	7.88498265595441	0.958	0.67	0	5	IGKC
252	0	6.68565291428108	0.828	0.244	0	5	IGHG1
253	0	5.76769238275568	0.327	0.06	0	5	IGHV3-30
254	0	5.67346448549741	0.286	0.016	0	5	IGKV1-39

255	0	5.66903024386227	0.765	0.086	0	5	JCHAIN
256	0	5.55960351195181	0.214	0.027	0	5	IGHV3-23
257	0	5.45654854102214	0.66	0.089	0	5	IGHG3
258	0	5.33960599632733	0.56	0.054	0	5	IGHG4
259	0	4.99142438010677	0.417	0.065	0	5	IGLV3-1
260	0	4.26972510567039	0.971	0.037	0	5	MZB1
261	0	4.23217141446503	0.178	0.005	0	5	IGKV1D-39
262	0	4.10461098305969	0.514	0.018	0	5	IGHGP
263	0	3.78638141016197	0.9	0.03	0	5	DERL3
264	0	3.35595271303773	0.961	0.589	0	5	SSR4
265	0	3.10336361688095	0.61	0.015	0	5	IGHG2
266	4.97998850825338E-264	5.41124590253948	0.128	0.011	1.67018854589802E-259	5	IGHV1-69D
267	6.87259128190504E-207	6.61357497677444	0.502	0.198	2.30492966412531E-202	5	IGHA1
268	1.94413871013438E-194	5.07119991186309	0.164	0.026	6.52025240604867E-190	5	IGHV4-59
269	7.18891930735407E-169	6.51645991843918	0.215	0.05	2.41101975730041E-164	5	IGHV4-61
270	1.79649932394519E-168	3.7306178639005	0.054	0.002	6.02509943264737E-164	5	IGHV1-69
271	2.03040117702829E-160	4.52934992149842	0.29	0.086	6.80955946751749E-156	5	IGHM

272	3.89233839003856E-143	5.71137173274499	0.224	0.061	1.30541244925113E-138	5	IGLC3
273	2.69041095143233E-131	5.67278686948481	0.122	0.021	9.02310024891375E-127	5	IGHV4-39
274	4.52278312135462E-127	3.23056855612735	0.031	0.001	1.51685100323991E-122	5	IGKV1D-13
275	7.13430556441945E-127	4.34738522186716	0.088	0.011	2.392703400195E-122	5	IGHV3-49
276	2.14410379624268E-125	5.66324518244079	0.249	0.077	7.19089531183869E-121	5	IGKV1-5
277	7.67198044620747E-117	3.92836611914662	0.058	0.005	2.57302880204906E-112	5	IGHV3-15
278	5.28198670325307E-108	6.42945768884889	0.512	0.272	1.77147270053702E-103	5	IGKV4-1
279	7.86672628500156E-97	7.24847312776248	0.412	0.201	2.63834266146382E-92	5	IGLC2
280	2.56399127861382E-70	4.82731677713256	0.136	0.04	8.59911395021503E-66	5	IGKV3-11
281	4.96508785280646E-61	3.21103178867259	0.065	0.012	1.66519116407423E-56	5	IGKV1-9
282	8.91290895214873E-47	3.59623650859069	0.03	0.004	2.98921140437164E-42	5	IGHV3-74
283	2.23596592550586E-40	4.8918182750508	0.236	0.121	7.49898252096157E-36	5	IGKV3-20
284	5.01370882997089E-40	5.88168712616855	0.166	0.073	1.68149766739564E-35	5	IGKV3-15
285	6.91747797671207E-37	5.60142533369372	0.101	0.036	2.31998376382969E-32	5	IGLV3-21
286	9.60426851185943E-34	3.76979129262033	0.056	0.015	3.22107957350742E-29	5	IGKV1-27
287	2.1879795447931E-32	5.80033969903387	0.078	0.027	7.33804579732711E-28	5	IGHV2-5
288	7.54855129933133E-28	5.28782025770274	0.065	0.022	2.53163313476974E-23	5	IGLV2-8

289	3.13158912884996E-24	3.80751516270282	0.047	0.014	1.0502723620337E-19	5	IGHV1-18
290	2.89598612070092E-21	3.51931380235487	0.018	0.003	9.71255825160676E-17	5	IGHV1-46
291	4.21523538999294E-17	4.65523694047554	0.059	0.024	1.41370564509583E-12	5	IGHV3-11
292	6.49522308405411E-16	3.96256958159007	0.052	0.021	2.17836791793007E-11	5	IGLV2-11
293	4.12021822147922E-13	4.73249156343457	0.067	0.033	1.3818387871197E-08	5	IGLV1-51
294	4.48479377373253E-13	4.69007520532428	0.096	0.053	1.50411013583442E-08	5	IGLV2-14
295	1.79542595363059E-09	4.34418054591418	0.04	0.018	6.02149956328627E-05	5	IGHV5-51
296	5.35519125503207E-08	3.77854994592448	0.053	0.029	0.001796024043113	5	IGLV1-44
297	3.41182303494302E-07	4.22928187083006	0.049	0.027	0.011442572094592	5	IGLV1-47
298	8.33533571554932E-06	3.02851535349455	0.009	0.003	0.279550489228093	5	IGLV3-10
299	0.000143052343324	3.32054569566579	0.044	0.028	1	5	IGLV3-19
300	0.001887220375473	3.08342475990923	0.011	0.005	1	5	IGLV8-61
301	0	3.82443971188665	0.904	0.028	0	6	VWF
302	0	3.53104563107433	0.838	0.025	0	6	PLVAP
303	0	3.30780497287445	0.909	0.089	0	6	TM4SF1
304	0	3.27072835920102	0.948	0.13	0	6	PECAM1
305	0	3.20426259302355	0.901	0.102	0	6	HSPG2

306	0	3.04698017809026	0.45	0.013	0	6	ACKR1
307	0	3.03027293172268	0.902	0.117	0	6	ENG
308	0	2.92285518636699	0.873	0.035	0	6	RAMP2
309	0	2.90808218061498	0.871	0.019	0	6	ADGRL4
310	0	2.90052640053785	0.76	0.029	0	6	AQP1
311	0	2.88023060147621	0.765	0.039	0	6	FLT1
312	0	2.76380814624459	0.684	0.06	0	6	INSR
313	0	2.71915939473919	0.872	0.197	0	6	COL4A1
314	0	2.71688170036305	0.847	0.064	0	6	CALCRL
315	0	2.68545881229731	0.844	0.049	0	6	CD93
316	0	2.66756057781215	0.949	0.25	0	6	A2M
317	0	2.54743897590932	0.887	0.137	0	6	GNG11
318	0	2.49487022352558	0.85	0.181	0	6	COL4A2
319	0	2.49437190215174	0.685	0.02	0	6	PODXL
320	0	2.48532783030844	0.915	0.248	0	6	APP
321	0	2.38829985134943	0.807	0.035	0	6	PALMD
322	0	2.38413194665151	0.484	0.013	0	6	STC1

323	0	2.38191726039961	0.7	0.015	0	6	RAMP3
324	0	2.34288433732839	0.782	0.016	0	6	EMCN
325	0	2.34195757823808	0.902	0.248	0	6	EMP1
326	0	2.33264092835022	0.775	0.016	0	6	PTPRB
327	0	2.33064262116361	0.664	0.052	0	6	COL15A1
328	0	2.32761128791837	0.867	0.132	0	6	CRIP2
329	0	2.31841907632327	0.951	0.342	0	6	IFI27
330	0	2.28638551506559	0.888	0.229	0	6	SPTBN1
331	0	2.26983460672525	0.348	0.007	0	6	SELE
332	0	2.26757267375073	0.776	0.016	0	6	CDH5
333	0	2.26718649676778	0.739	0.057	0	6	ITGA6
334	0	2.22437454119687	0.656	0.062	0	6	SLC9A3R2
335	0	2.21802253428404	0.629	0.012	0	6	PCDH17
336	0	2.21507504552006	0.786	0.041	0	6	CD34
337	0	2.1923303232516	0.766	0.119	0	6	WWTR1
338	0	2.18806070073818	0.52	0.018	0	6	RBP7
339	0	2.18116216760585	0.73	0.011	0	6	CLEC14A

340	0	2.17451369718599	0.645	0.12	0	6	ID1
341	0	2.13895416975326	0.917	0.284	0	6	TCF4
342	0	2.13144971757901	0.754	0.069	0	6	NPDC1
343	0	2.12158630973872	0.826	0.155	0	6	EPAS1
344	0	2.09687374757665	0.653	0.047	0	6	MCTP1
345	0	2.06710894711374	0.742	0.157	0	6	NRP1
346	0	2.06607282971439	0.896	0.348	0	6	FKBP1A
347	0	2.04439592402257	0.708	0.187	0	6	PRCP
348	0	2.02111457034164	0.706	0.121	0	6	SPRY1
349	0	2.00638446730736	0.694	0.05	0	6	HYAL2
350	1.59769022394052E-284	2.91819258499721	0.335	0.075	5.35833347305171E-280	6	FABP4
351	0	4.6370543689673	0.99	0.198	0	7	CCL4
352	0	4.35611384344741	0.752	0.081	0	7	CCL4L2
353	0	2.57723702927793	0.313	0.027	0	7	IFNG
354	0	2.35537822814497	0.555	0.115	0	7	CRTAM
355	0	2.23663015148808	0.95	0.311	0	7	CCL5
356	0	1.93379051501159	0.785	0.202	0	7	GZMK

357	0	1.78681063699347	0.783	0.206	0	7	NKG7
358	0	1.72076391057662	0.726	0.219	0	7	GZMA
359	0	1.49010061691006	0.719	0.284	0	7	CST7
360	2.9325223621236E-300	1.4026568926639	0.453	0.122	9.83509349809013E-296	7	CD8A
361	2.55941759797986E-269	2.23211614657672	0.243	0.042	8.58377474010486E-265	7	XCL2
362	3.08966536302302E-268	1.90032538195515	0.533	0.2	1.03621196945066E-263	7	CD69
363	9.8187790368441E-254	1.18080879393914	0.893	0.599	3.29302211337678E-249	7	CXCR4
364	2.54746747133622E-252	1.44749286119933	0.347	0.086	8.54369640536742E-248	7	GZMH
365	6.63511968266242E-229	1.61514277807935	0.594	0.284	2.22528643917132E-224	7	SLA
366	1.59656300525385E-214	1.07139189851427	0.199	0.035	5.35455300702037E-210	7	DTHD1
367	1.03593736733511E-187	1.24256523530409	0.323	0.093	3.47432674256848E-183	7	CD8B
368	1.1581385460471E-183	0.965980039808839	0.933	0.918	3.88416505573275E-179	7	H3F3B
369	3.59293994184541E-182	1.25391262582668	0.217	0.049	1.20500019769611E-177	7	TNFSF9
370	1.04573098641594E-170	1.28410067852858	0.745	0.527	3.50717258224177E-166	7	TNFAIP3
371	7.84888404242515E-170	1.22119250671058	0.356	0.121	2.63235873014855E-165	7	SH2D1A
372	2.97617492034689E-168	1.13259714472228	0.227	0.055	9.98149544785939E-164	7	CCL3L1
373	2.5803123826495E-154	1.47029202266988	0.19	0.043	8.65385166892989E-150	7	TNF

374	2.63708120670742E-150	1.11581064229455	0.607	0.347	8.84424295105534E-146	7	HCST
375	1.72191657559447E-149	1.19360153948285	0.743	0.63	5.77496381122874E-145	7	SRSF7
376	3.6052903894094E-149	1.40038584707254	0.495	0.248	1.20914229080012E-144	7	DUSP2
377	1.55243604542706E-132	1.10675867046961	0.798	0.71	5.20656000915326E-128	7	TSC22D3
378	3.79695077480291E-132	1.52913531790765	0.411	0.194	1.2734213508534E-127	7	TAGAP
379	6.61670816547373E-129	1.0810634172881	0.326	0.119	2.21911158453658E-124	7	CTSW
380	2.57939361269393E-122	1.39166847879027	0.62	0.447	8.65077029825291E-118	7	NR4A2
381	8.73779786146848E-118	1.0800113800055	0.211	0.061	2.9304826467793E-113	7	KLRD1
382	9.94476631195397E-115	0.982448741945767	0.338	0.137	3.33527572570312E-110	7	GZMM
383	2.32201594232183E-112	1.16641410496404	0.627	0.474	7.78757706735896E-108	7	RNF19A
384	5.89597829832774E-112	1.12332553184283	0.503	0.28	1.97739320169316E-107	7	TUBA4A
385	1.30827071461194E-102	1.25516193090072	0.772	0.682	4.38767832266552E-98	7	KLF6
386	7.28286719841636E-99	1.01644337276761	0.228	0.079	2.44252800100488E-94	7	PYHIN1
387	2.67019392890134E-92	1.02796291348099	0.404	0.213	8.95529639874931E-88	7	RUNX3
388	1.18574984445955E-77	1.05922038919647	0.306	0.151	3.97676782834842E-73	7	APOBEC3G
389	3.80002636998094E-75	1.1836652236688	0.175	0.06	1.27445284396421E-70	7	CCL3
390	9.01547851205192E-66	1.01539966473438	0.469	0.324	3.02361118337197E-61	7	ADGRE5

391	6.48674757311765E-57	1.24834406761738	0.619	0.547	2.1755254010722E-52	7	DNAJB1
392	1.33527969065289E-53	1.19774904027945	0.471	0.361	4.47826102651166E-49	7	RGCC
393	2.57831325116058E-53	1.04317328679707	0.329	0.204	8.64714698174235E-49	7	LYST
394	1.27327240219318E-49	1.03733802063617	0.255	0.142	4.27030098247548E-45	7	CMC1
395	6.1252861682678E-42	0.96415669272006	0.352	0.246	2.05429847511366E-37	7	CBLB
396	2.34891330272229E-37	1.46955212544134	0.099	0.036	7.87778543467E-33	7	XCL1
397	4.57537910902616E-36	1.07277949863481	0.44	0.363	1.53449064558519E-31	7	ID2
398	2.98742630631641E-28	1.05286526616019	0.325	0.254	1.0019230346124E-23	7	UBE2S
399	4.46224725261985E-26	0.99475297280699	0.399	0.346	1.49654848358365E-21	7	CLEC2B
400	8.22429465188643E-12	0.968965179206625	0.288	0.26	2.75826394034967E-07	7	NEU1
401	0	4.90656416804316	0.883	0.093	0	8	APOD
402	0	4.57793773868447	0.999	0.166	0	8	DCN
403	0	4.28908092161808	0.702	0.047	0	8	CXCL14
404	0	4.22899639244965	0.987	0.153	0	8	CCDC80
405	0	4.05208405823292	0.996	0.36	0	8	MGP
406	0	3.90210230445954	0.745	0.063	0	8	SFRP4
407	0	3.88478150820041	0.927	0.088	0	8	C3

408	0	3.79632654865651	0.922	0.093	0	8	SFRP2
409	0	3.74888822275475	0.667	0.076	0	8	CFD
410	0	3.57028725308099	0.984	0.156	0	8	LUM
411	0	3.34025268688152	0.989	0.145	0	8	C1S
412	0	3.2603536280393	0.934	0.066	0	8	FBLN1
413	0	3.11095214542804	0.96	0.145	0	8	C1R
414	0	3.10861694410395	0.344	0.008	0	8	PLA2G2A
415	0	3.10730783955085	0.833	0.064	0	8	IGF1
416	0	3.04043479962497	0.975	0.207	0	8	SERPINF1
417	0	2.98615282645806	0.844	0.043	0	8	EFEMP1
418	0	2.9098062460812	0.958	0.124	0	8	CTSK
419	0	2.82248895964356	0.946	0.146	0	8	COL6A3
420	0	2.80584022172949	0.372	0.005	0	8	PI16
421	0	2.7978769772195	0.733	0.036	0	8	DPT
422	0	2.74656273949729	0.838	0.067	0	8	MEG3
423	0	2.71785824578835	0.918	0.12	0	8	MMP2
424	0	2.70433718490072	0.77	0.12	0	8	CLU

425	0	2.69637294928871	0.941	0.203	0	8	NNMT
426	0	2.68162205541258	0.485	0.038	0	8	PTGDS
427	0	2.67180495473681	0.606	0.012	0	8	ADH1B
428	0	2.66226678087942	0.908	0.21	0	8	IGFBP4
429	0	2.57647624982693	0.949	0.248	0	8	COL3A1
430	0	2.56864531820216	0.991	0.261	0	8	COL1A2
431	0	2.54612919822871	0.945	0.295	0	8	GSN
432	0	2.53027303996808	0.97	0.247	0	8	COL1A1
433	0	2.50523865703929	0.492	0.024	0	8	WISP2
434	0	2.46525178461795	0.897	0.125	0	8	COL14A1
435	0	2.43646907929322	0.675	0.053	0	8	IGFBP6
436	0	2.43301066340133	0.559	0.05	0	8	MFAP5
437	0	2.4047510908696	0.921	0.179	0	8	FSTL1
438	0	2.39855922799084	0.789	0.043	0	8	FGF7
439	0	2.37991135141615	0.943	0.184	0	8	COL6A1
440	0	2.37810205476631	0.333	0.017	0	8	CXCL1
441	0	2.369382943157	0.725	0.021	0	8	OGN

442	0	2.36164973490626	0.926	0.134	0	8	LRP1
443	0	2.34039355021046	0.777	0.037	0	8	MFAP4
444	0	2.30752835377688	0.846	0.105	0	8	CXCL12
445	0	2.30279911455637	0.833	0.138	0	8	FBN1
446	0	2.28695266939786	0.907	0.123	0	8	RARRES2
447	0	2.27868006759664	0.979	0.4	0	8	TIMP1
448	0	2.26889974475092	0.964	0.266	0	8	SERPING1
449	0	2.2468191866518	0.963	0.194	0	8	COL6A2
450	0	2.23851715649423	0.656	0.109	0	8	CXCL2
451	0	5.36259466882885	0.746	0.023	0	9	GNLY
452	0	3.3781442538947	0.78	0.085	0	9	GZMB
453	0	2.51352236252845	0.878	0.201	0	9	NKG7
454	0	2.46314535366958	0.569	0.041	0	9	KLRD1
455	0	2.27599293968047	0.515	0.068	0	9	PRF1
456	0	2.24959968261526	0.593	0.104	0	9	CTSW
457	0	1.9979406170706	0.34	0.037	0	9	XCL2
458	0	1.98744109456785	0.304	0.024	0	9	XCL1

459	0	1.82617295426443	0.423	0.082	0	9	GZMH
460	0	1.79861314630376	0.267	0.009	0	9	KLRC1
461	0	1.77290623556136	0.625	0.218	0	9	CD7
462	0	1.71395532859086	0.863	0.317	0	9	CCL5
463	0	1.54329644704126	0.624	0.237	0	9	SYTL3
464	0	1.50146179131451	0.696	0.221	0	9	GZMA
465	0	1.45102629695272	0.181	0.016	0	9	KRT86
466	0	1.42675166216864	0.224	0.02	0	9	ZNF683
467	0	1.21421655303362	0.116	0.004	0	9	FGFBP2
468	0	1.01019695078213	0.145	0.008	0	9	TRDC
469	7.46415613878218E-306	1.60558638501201	0.29	0.053	2.50332868582477E-301	9	KLRB1
470	1.82649533846416E-290	1.19024108241674	0.717	0.284	6.12570006614109E-286	9	CST7
471	7.39887172035193E-270	1.59821963165946	0.439	0.138	2.48143359757163E-265	9	HOPX
472	1.67563750291259E-243	1.10634380565934	0.236	0.042	5.61975305726825E-239	9	LINC01871
473	8.1702115499267E-241	1.24140166618797	0.688	0.343	2.74012554961441E-236	9	HCST
474	6.81842765212668E-239	1.30665639489635	0.315	0.076	2.28676426597025E-234	9	IL2RB
475	8.5628793988474E-234	2.15003341310186	0.455	0.165	2.87181849278544E-229	9	AREG

476	7.22957073520617E-199	0.97010419559901	0.56	0.223	2.42465343317344E-194	9	CCL4
477	1.68184186843896E-173	1.01359300031232	0.412	0.14	5.64056125837059E-169	9	TYROBP
478	2.31861650782282E-167	1.01674640015308	0.225	0.053	7.77617604393618E-163	9	SLA2
479	1.32396440901631E-157	0.977485326553976	0.726	0.427	4.44031183495891E-153	9	CD52
480	1.79063840125012E-151	0.999110460003379	0.66	0.388	6.00544307011265E-147	9	LCP1
481	2.53158855365888E-150	1.15708330332458	0.328	0.116	8.49044169126117E-146	9	GPR65
482	1.71379460727218E-145	1.23710932691975	0.751	0.582	5.74772435386945E-141	9	REL
483	8.42108101834439E-144	1.03027298537911	0.407	0.164	2.82426215193234E-139	9	CD96
484	5.62018898867208E-142	1.08922236384935	0.427	0.185	1.88489898302084E-137	9	RNF125
485	4.24030234137758E-134	1.1544721155589	0.577	0.345	1.42211259925121E-129	9	DDIT4
486	2.7785419094813E-130	1.20413655277231	0.547	0.319	9.31867385601839E-126	9	ADGRE5
487	8.75727857630769E-128	1.06570817494344	0.443	0.212	2.93701608892207E-123	9	AKNA
488	3.04461111611061E-127	0.967249008227174	0.332	0.123	1.02110167612118E-122	9	SH2D2A
489	2.60741261439743E-125	1.09380578911378	0.316	0.118	8.7447404261661E-121	9	CD247
490	3.57929172437953E-119	0.990122045131842	0.439	0.211	1.20042285852241E-114	9	RUNX3
491	3.80385443620621E-116	0.973738451507321	0.353	0.147	1.27573670081484E-111	9	STAT4
492	1.53944559770077E-113	1.02609802944266	0.728	0.596	5.16299264556884E-109	9	ZFP36L2

493	8.84226184411908E-112	1.03739825377421	0.392	0.184	2.96551777728066E-107	9	PARP8
494	1.16811700034278E-101	1.09363587777638	0.541	0.358	3.91763079574961E-97	9	ID2
495	9.31238048906146E-99	1.11146133121612	0.548	0.366	3.12318616842143E-94	9	PIK3R1
496	4.64187322016694E-97	1.07859328079254	0.599	0.451	1.55679144057959E-92	9	FAM177A1
497	1.06166280985954E-89	1.37780655674794	0.298	0.137	3.56060473170691E-85	9	LDLRAD4
498	8.30707983838612E-81	0.992975076090012	0.513	0.307	2.78602843619794E-76	9	IL7R
499	7.39580636170677E-59	0.969825927114988	0.393	0.252	2.48040553758922E-54	9	VPS37B
500	8.45137287735295E-50	0.978368996726405	0.346	0.223	2.83442143560663E-45	9	METRNL
501	0	4.84206183388055	0.968	0.13	0	10	POSTN
502	0	4.33796965812667	0.999	0.248	0	10	COL1A1
503	0	3.99173070946048	0.999	0.264	0	10	COL1A2
504	0	3.89462631337168	0.991	0.306	0	10	FN1
505	0	3.8008405254411	0.777	0.042	0	10	MMP11
506	0	3.77732272281985	0.999	0.249	0	10	COL3A1
507	0	3.67055813278202	0.939	0.106	0	10	CTHRC1
508	0	3.59446813038965	0.992	0.16	0	10	LUM
509	0	3.40965868979071	0.934	0.125	0	10	RARRES2

510	0	3.25853261947658	0.976	0.157	0	10	AEBP1
511	0	3.24288351749374	0.98	0.148	0	10	COL6A3
512	0	3.18537719770556	0.983	0.176	0	10	VCAN
513	0	3.15004610218262	0.917	0.106	0	10	COL12A1
514	0	3.08163411557501	0.916	0.097	0	10	SFRP2
515	0	2.98381615238669	0.952	0.128	0	10	CTSK
516	0	2.94705675006769	0.976	0.155	0	10	COL5A2
517	0	2.94034063695724	0.694	0.01	0	10	COL11A1
518	0	2.84947059406766	0.846	0.089	0	10	SULF1
519	0	2.79733236286812	0.868	0.067	0	10	INHBA
520	0	2.65634676365762	0.814	0.037	0	10	COL8A1
521	0	2.64977049971299	0.912	0.076	0	10	THBS2
522	0	2.61162472794528	0.984	0.171	0	10	DCN
523	0	2.60281875706551	0.965	0.187	0	10	COL6A1
524	0	2.5765434572195	0.759	0.174	0	10	TIMP3
525	0	2.54059306133188	0.882	0.072	0	10	MXRA5
526	0	2.50598373462712	0.994	0.328	0	10	SPARC

527	0	2.4664941422576	0.816	0.072	0	10	MEG3
528	0	2.43528322953799	0.734	0.014	0	10	COL10A1
529	0	2.40872155716063	0.854	0.144	0	10	HTRA1
530	0	2.40613543773854	0.882	0.079	0	10	CDH11
531	0	2.36027231775877	0.836	0.07	0	10	ISLR
532	0	2.35175218391766	0.972	0.197	0	10	COL6A2
533	0	2.33400401533598	0.731	0.125	0	10	IGFBP5
534	0	2.33293491814663	0.961	0.158	0	10	CCDC80
535	0	2.32742159930426	0.763	0.131	0	10	CTGF
536	0	2.29042319338728	0.726	0.135	0	10	RGS16
537	0	2.28201325975335	0.976	0.57	0	10	LGALS1
538	0	2.27449010347336	0.858	0.144	0	10	PALLD
539	0	2.25728083423746	0.867	0.127	0	10	MMP2
540	0	2.22350868474853	0.883	0.151	0	10	MDK
541	0	2.19217648737422	0.686	0.069	0	10	SFRP4
542	0	2.18654071834089	0.961	0.219	0	10	BGN
543	0	2.17554903877261	0.871	0.073	0	10	FAP

544	0	2.1676114332451	0.564	0.07	0	10	ASPN
545	0	2.16414543271375	0.789	0.1	0	10	NBL1
546	0	2.14463426433699	0.871	0.084	0	10	COL5A1
547	0	2.13537977550214	0.927	0.245	0	10	TAGLN
548	0	2.09478057252596	0.935	0.254	0	10	TPM1
549	0	2.06606938332026	0.992	0.298	0	10	CALD1
550	0	2.01715018658654	0.645	0.048	0	10	MFAP5
551	0	3.53038718469993	0.861	0.014	0	11	MS4A1
552	0	3.09454965940327	0.819	0.058	0	11	CD79A
553	0	2.65221172160849	0.713	0.013	0	11	BANK1
554	0	2.21523594894286	0.734	0.167	0	11	CD83
555	0	2.02807833828931	0.989	0.381	0	11	HLA-DRA
556	0	1.94561658531275	0.89	0.324	0	11	CD37
557	0	1.90465129537164	0.424	0.019	0	11	BCL11A
558	0	1.8657610253839	0.452	0.013	0	11	TNFRSF13C
559	0	1.83212092772663	0.859	0.199	0	11	HLA-DQB1
560	0	1.81129279624929	0.667	0.231	0	11	MEF2C

561	0	1.7419236760107	0.998	0.649	0	11	CD74
562	0	1.7251613766205	0.491	0.077	0	11	RALGPS2
563	0	1.71976360976341	0.814	0.191	0	11	HLA-DQA1
564	0	1.69077840145065	0.931	0.331	0	11	HLA-DRB1
565	0	1.56546843866642	0.368	0.041	0	11	MARCH1
566	0	1.50917229796017	0.371	0.05	0	11	LY9
567	0	1.50286410810514	0.411	0.05	0	11	ADAM28
568	0	1.47894281642405	0.44	0.103	0	11	LYN
569	0	1.46527926670463	0.389	0.07	0	11	ARHGAP24
570	0	1.40534665729115	0.96	0.429	0	11	HLA-DPA1
571	0	1.40257529921945	0.394	0.078	0	11	POU2F2
572	0	1.35380084419974	0.942	0.395	0	11	HLA-DPB1
573	0	1.31722248344789	0.337	0.052	0	11	CD79B
574	0	1.23141338262264	0.947	0.818	0	11	RPS11
575	0	1.20005639929943	0.204	0.002	0	11	IGHD
576	0	1.18309063883787	0.98	0.929	0	11	RPL21
577	0	1.18066125341278	0.255	0.003	0	11	VPREB3

578	0	1.1412824373717	0.252	0.005	0	11	CD22
579	0	1.13982928477411	0.989	0.968	0	11	RPS27
580	0	1.10747356056393	0.239	0.002	0	11	FCER2
581	0	1.10414095849106	0.293	0.005	0	11	BLK
582	0	1.07084944059094	0.217	0.002	0	11	LINC02397
583	9.83370605355813E-301	1.36902159848038	0.85	0.385	3.29802833624233E-296	11	LAPTM5
584	6.4376807999291E-297	1.38302282144172	0.514	0.132	2.15906938668022E-292	11	CCR7
585	1.08645472578272E-266	1.05500513667674	0.256	0.041	3.6437518593301E-262	11	HVCN1
586	2.94885181608242E-263	1.22206664702477	0.918	0.785	9.88985922077721E-259	11	RPS20
587	7.25757488717642E-251	1.34726975218167	0.955	0.861	2.43404546566123E-246	11	RPS29
588	5.25778688140271E-244	1.06000290130342	0.877	0.423	1.76335656428484E-239	11	CD52
589	1.02369381832304E-235	1.23153970008262	0.337	0.078	3.43326432789183E-231	11	IRF8
590	2.38992677876166E-235	1.11194180183795	0.432	0.117	8.01533643061084E-231	11	HLA-DMB
591	4.84949493009685E-218	1.27436772593879	0.74	0.389	1.62642360965588E-213	11	CD55
592	3.04341627022099E-212	1.13856896221631	0.342	0.087	1.02070094870672E-207	11	NFKBID
593	2.39116446004872E-204	1.32377254774447	0.76	0.506	8.0194873661114E-200	11	RPS10
594	3.41183934332662E-186	1.09215654204084	0.864	0.578	1.14426267896488E-181	11	REL

595	1.48761034221481E-185	1.08810459510791	0.748	0.407	4.98914756572004E-181	11	SMAP2
596	7.97027325581253E-163	1.10567340524036	0.581	0.286	2.67307024453441E-158	11	PRDM2
597	1.46192701916445E-142	1.150126656699	0.475	0.22	4.90301083687373E-138	11	SNX2
598	2.43311994596674E-136	1.12689590845538	0.389	0.159	8.16019767478325E-132	11	SMIM14
599	6.38588539299877E-99	1.19019402966496	0.46	0.219	2.14169824310393E-94	11	MTRNR2L12
600	2.02238082671812E-60	1.06561883663759	0.272	0.121	6.78266081664722E-56	11	MTRNR2L8
601	0	1.29733664277472	0.993	0.99	0	12	MALAT1
602	9.79974507601953E-186	0.954347580113622	0.97	0.972	3.28663850359543E-181	12	MT-CYB
603	1.49336670073924E-173	0.969963409323809	0.956	0.94	5.00845324093925E-169	12	MT-ND1
604	1.52058308336386E-169	0.915428506180311	0.975	0.98	5.09973154498573E-165	12	MT-CO3
605	4.68296209242402E-161	1.02085929476994	0.911	0.911	1.57057182655717E-156	12	MT-ND2
606	8.09651474155657E-157	1.04610149209116	0.961	0.95	2.71540911402324E-152	12	MT-ATP6
607	2.44318015910001E-122	0.952365403045435	0.874	0.885	8.19393761758961E-118	12	MT-ND5
608	1.70202647414789E-115	0.875486143788788	0.92	0.913	5.7082563889972E-111	12	MT-ND4
609	3.32970451851618E-107	0.904388231754318	0.938	0.941	1.11671630141995E-102	12	MT-ND3
610	2.89387710530262E-77	1.6554405646105	0.599	0.507	9.70548503576392E-73	12	XIST
611	6.76947502575596E-66	1.40887731119883	0.819	0.753	2.27034653413803E-61	12	NEAT1

612	9.9347945955285E-36	1.26164339620661	0.503	0.434	3.33193141144835E-31	12	SYNE2
613	1.96717304161475E-34	0.838407246715814	0.08	0.024	6.59750494696753E-30	12	LINC00861
614	5.71265671734918E-33	0.939392411530481	0.156	0.071	1.91591080986457E-28	12	GOLGA8B
615	3.52692099406938E-31	0.890366653625838	0.131	0.055	1.18285876299099E-26	12	SLFN12L
616	3.75456891793072E-31	1.04846121587949	0.529	0.527	1.25920732369561E-26	12	RSRP1
617	5.58175777838969E-24	1.1521886516909	0.522	0.512	1.87200992371633E-19	12	TXNIP
618	1.22767983220781E-23	0.932889852277676	0.141	0.072	4.11739262125856E-19	12	GOLGA8A
619	2.72892454879913E-23	1.41667233009908	0.412	0.339	9.15226715176253E-19	12	HSPA1B
620	5.43356508378416E-23	1.23903911760701	0.205	0.125	1.82230905779953E-18	12	MTRNR2L8
621	8.95551463215135E-23	1.58344049315237	0.304	0.227	3.00350049733092E-18	12	MTRNR2L12
622	2.08754267730859E-20	1.07357941041889	0.505	0.521	7.00120063115753E-16	12	RNF213
623	3.77120966678542E-20	1.79730275400459	0.183	0.113	1.2647882980465E-15	12	PLCG2
624	6.19845589475003E-20	1.1866518308347	0.297	0.215	2.07883813798126E-15	12	CD69
625	2.09634650361302E-19	0.984679152428024	0.104	0.049	7.03072690381735E-15	12	TNF
626	1.06096298682416E-15	1.17273898346671	0.374	0.346	3.55825766521087E-11	12	NKTR
627	1.33689567350657E-15	0.948628246993051	0.489	0.525	4.48368070980635E-11	12	CCNL1
628	1.82635017908837E-14	0.901932295721393	0.473	0.499	6.12521323062656E-10	12	ARGLU1

629	4.58634702966231E-14	0.968665414817666	0.456	0.486	1.53816906680815E-09	12	PNISR
630	1.66843220156831E-13	0.979006945744229	0.37	0.339	5.59558791761981E-09	12	KIAA1551
631	6.00030470227648E-13	0.962480575647312	0.62	0.69	2.01238219104949E-08	12	KLF6
632	3.3253016604939E-12	1.00609308672898	0.186	0.135	1.11523967089644E-07	12	ANKRD36C
633	1.18337305442678E-11	1.06271133949894	0.357	0.35	3.96879654993653E-07	12	LUC7L3
634	4.94175882605625E-11	1.20305320837807	0.367	0.367	1.65736707508274E-06	12	VMP1
635	1.0769494133165E-10	1.22683512082852	0.124	0.079	3.61187294238088E-06	12	TTN
636	2.94287676266073E-10	1.14580682670024	0.389	0.405	9.86982008661154E-06	12	MACF1
637	8.5050958616162E-10	1.11659702223899	0.462	0.454	2.85243905006884E-05	12	HSPA1A
638	1.41756300002075E-09	0.887341947994287	0.433	0.47	4.75422278946958E-05	12	N4BP2L2
639	9.82221204819517E-08	0.993436170454842	0.447	0.518	0.003294173476724	12	DDX17
640	4.24454215445819E-07	0.863233626667678	0.481	0.508	0.014235345477622	12	JUN
641	1.43681248144948E-06	0.846568619159203	0.182	0.15	0.048187817002853	12	AC245297.3
642	3.88474948555771E-06	1.01144475257484	0.251	0.236	0.130286728246634	12	POLR2J3.1
643	9.62960354391176E-06	0.842605278560289	0.356	0.373	0.322957643655713	12	UTRN
644	1.68751640590037E-05	0.906829455674248	0.314	0.325	0.565959252210865	12	ASH1L
645	3.11962237948806E-05	0.826933213645015	0.146	0.12	1	12	NPIPB5

646	5.99279424460636E-05	0.87591232093267	0.304	0.314	1	12	KMT2A
647	0.000905953619445	0.856354994765072	0.215	0.206	1	12	PCSK7
648	0.002767892421435	1.16300658856506	0.291	0.313	1	12	XAF1
649	0.004479933320692	0.851065998939063	0.232	0.235	1	12	ATM
650	0.005415325649272	0.894157946748463	0.186	0.179	1	12	NABP1
651	0	4.67835598364023	0.632	0.018	0	13	CXCL13
652	0	1.74911373269375	0.804	0.322	0	13	DUSP4
653	0	1.67811218312095	0.514	0.1	0	13	CTLA4
654	0	1.49612701801365	0.839	0.313	0	13	CD2
655	3.34326626470344E-281	1.41960633387184	0.767	0.281	1.12126463985624E-276	13	CD3D
656	1.26648884165274E-257	1.18783260501035	0.32	0.055	4.24755027713496E-253	13	TOX
657	3.81705780938312E-251	1.97604120591692	0.725	0.38	1.28016484811091E-246	13	NR3C1
658	1.12268055426242E-214	1.31631250930463	0.98	0.716	3.76524604288529E-210	13	SRGN
659	7.82148625863575E-200	1.25353623058691	0.328	0.072	2.62317006142126E-195	13	TNFRSF18
660	2.75545829728869E-199	1.57975747389068	0.646	0.302	9.24125603744679E-195	13	FKBP5
661	4.16210276224644E-198	1.44443180696836	0.686	0.304	1.39588602440221E-193	13	ITM2A
662	4.61029085148118E-195	1.05321670070586	0.939	0.57	1.54619934576976E-190	13	PTPRC

663	3.0357788133376E-190	1.51448565280074	0.714	0.341	1.01813949841716E-185	13	RGS1
664	4.79199591839995E-190	1.27308090547226	0.288	0.057	1.60713959111297E-185	13	KLRB1
665	2.30900890694296E-185	1.61753218760578	0.336	0.081	7.7439540721053E-181	13	CXCR6
666	7.60028137421251E-183	1.37570864422811	0.791	0.471	2.54898236728339E-178	13	RNF19A
667	9.16234043886032E-174	1.13206429970507	0.436	0.127	3.07286573638497E-169	13	TIGIT
668	4.52605013122751E-169	1.17205895769221	0.57	0.215	1.51794669301108E-164	13	CD3G
669	1.71036376903918E-168	1.30124893133804	0.532	0.196	5.73621800860361E-164	13	TRAC
670	3.34779636109536E-160	1.02426763972771	0.662	0.271	1.12278394358416E-155	13	CD3E
671	1.9742370602577E-158	1.67418905824798	0.588	0.313	6.62119625269226E-154	13	RBPJ
672	3.01131884851212E-158	1.20092953044199	0.532	0.203	1.00993611541399E-153	13	CLEC2D
673	2.41664586013572E-156	1.15514131111349	0.558	0.213	8.10494688572319E-152	13	SPOCK2
674	5.70377776920938E-156	1.29792911973722	0.677	0.34	1.91293298823744E-151	13	SAMSN1
675	2.15324810523739E-148	1.12526853139426	0.635	0.288	7.22156349534515E-144	13	SLA
676	3.19033904223225E-147	1.03181531786065	0.804	0.429	1.06997590798385E-142	13	CD52
677	6.7788354397982E-133	1.20014826407509	0.524	0.221	2.27348582979952E-128	13	ALOX5AP
678	2.06376808452987E-132	1.03332953053311	0.898	0.753	6.92146540189629E-128	13	SH3BGRL3
679	6.45205428072347E-131	1.01773201694865	0.318	0.093	2.16388996466904E-126	13	PDE3B

680	1.88011710617096E-130	1.25694990461681	0.689	0.424	6.30553675067615E-126	13	ETS1
681	1.87183032766335E-127	1.13922172995117	0.545	0.234	6.27774455291736E-123	13	TRBC2
682	2.88069603488435E-125	0.965605968204124	0.334	0.102	9.66127836179514E-121	13	IPCEF1
683	7.08812038705176E-123	1.45492544536373	0.212	0.049	2.37721381540942E-118	13	NMB
684	1.52743440815878E-121	1.35234215340933	0.428	0.163	5.12270951808291E-117	13	TRBC1
685	8.64479159144252E-121	1.37707319127638	0.537	0.272	2.89929020393799E-116	13	HMGB2
686	1.18450924125051E-116	1.03154206643888	0.656	0.353	3.97260709330597E-112	13	FYB1
687	2.00246030891873E-114	1.01734906162358	0.436	0.167	6.71585138405165E-110	13	CD96
688	2.03741898986682E-110	0.98772611329187	0.733	0.477	6.83309580821535E-106	13	STK4
689	9.10765730870358E-106	1.0284158843553	0.638	0.363	3.05452610819301E-101	13	CNOT6L
690	1.35392427865461E-104	1.10766040915823	0.127	0.021	4.54079124575183E-100	13	KRT86
691	1.97183447265824E-96	1.0366418170495	0.434	0.193	6.61313845440119E-92	13	PDE4D
692	2.51867536306967E-93	1.11805375677046	0.512	0.258	8.44713343266305E-89	13	CRYBG1
693	5.47741603620676E-85	0.976642149328807	0.394	0.173	1.83701579022302E-80	13	PIK3IP1
694	1.18918611601217E-81	0.954490731760601	0.751	0.598	3.98829239588161E-77	13	ZFP36L2
695	9.5742614934855E-81	1.20897030858772	0.473	0.274	3.21101581968517E-76	13	NAP1L4
696	7.06890500300674E-77	1.02145928564411	0.515	0.298	2.3707693599084E-72	13	EVL

697	4.40410628962843E-63	1.15574705642705	0.363	0.187	1.47704916741558E-58	13	MAF	
698	7.3610400731082E-63	1.00488487744434	0.222	0.084	2.46874561971903E-58	13	MFHAS1	
699	3.93539859077845E-51	1.10227626710774	0.246	0.11	1.31985397937528E-46	13	CHN1	
700	6.23176761096453E-44	1.04330481810294	0.366	0.218	2.09001022136528E-39	13	PRDM1	
701		0	4.08143457340585	0.975	0.094		0	14	RGS5
702		0	3.33602827783458	0.996	0.413		0	14	TIMP1
703		0	3.21782716632482	0.891	0.069		0	14	NDUFA4L2
704		0	3.03498802299989	0.969	0.21		0	14	COL4A1
705		0	3.01595795768483	0.991	0.232		0	14	BGN
706		0	3.0159162747288	0.948	0.172		0	14	THY1
707		0	2.80299466972052	0.999	0.34		0	14	SPARC
708		0	2.78114068255446	0.942	0.193		0	14	UACA
709		0	2.68466886537002	0.96	0.194		0	14	COL4A2
710		0	2.64745788198234	0.973	0.327		0	14	IGFBP7
711		0	2.61315466311859	0.969	0.186		0	14	LHFPL6
712		0	2.61293302444007	0.802	0.131		0	14	CPE
713		0	2.60108848828914	0.913	0.168		0	14	COL18A1

714	0	2.46186214432257	0.698	0.056	0	14	EGFL6
715	0	2.4408985743533	0.99	0.265	0	14	SPARCL1
716	0	2.41665790059239	0.919	0.156	0	14	PCOLCE
717	0	2.41401745890588	1	0.311	0	14	CALD1
718	0	2.38849678881742	0.204	0.004	0	14	IBSP
719	0	2.38390665006999	0.869	0.067	0	14	CCDC102B
720	0	2.34433015480934	0.919	0.149	0	14	PDGFRB
721	0	2.33124779732951	0.994	0.45	0	14	IFITM3
722	0	2.24743620772713	0.93	0.172	0	14	COL5A2
723	0	2.24615612954866	0.904	0.125	0	14	MCAM
724	0	2.18861059276215	0.96	0.212	0	14	COL6A2
725	0	2.17710425278042	0.907	0.157	0	14	TFPI
726	0	2.09237218194429	0.822	0.159	0	14	ITGA1
727	0	2.03434133188819	0.69	0.037	0	14	COX4I2
728	0	2.03426000727117	0.898	0.243	0	14	LGALS3BP
729	0	2.00514498719586	0.65	0.172	0	14	C11orf96
730	0	1.99185920850486	0.945	0.283	0	14	SERPING1

731	0	1.98070289804416	0.93	0.216	0	14	CAV1
732	0	1.9787488524355	0.749	0.048	0	14	ENPEP
733	0	1.94923308990773	0.934	0.202	0	14	COL6A1
734	0	1.94288484567484	0.868	0.175	0	14	LAMA4
735	0	1.93693027797024	0.743	0.11	0	14	MYO1B
736	0	1.91906345187428	0.921	0.227	0	14	SERPINF1
737	0	1.90927240151287	0.676	0.06	0	14	ABCC9
738	0	1.90195683508542	0.973	0.266	0	14	A2M
739	0	1.89018315915235	0.684	0.037	0	14	HIGD1B
740	0	1.88386143343529	0.843	0.105	0	14	ANGPT2
741	0	1.8807914844341	0.979	0.563	0	14	ITGB1
742	0	1.8651297660481	0.843	0.232	0	14	FILIP1L
743	0	1.84741463594299	0.837	0.23	0	14	MIR4435-2HG
744	0	1.84064093063199	0.71	0.069	0	14	PPP1R14A
745	0	1.83284686569962	0.852	0.175	0	14	EPS8
746	0	1.82316597682214	0.78	0.08	0	14	NOTCH3
747	0	1.8184239823823	0.834	0.21	0	14	ACTA2

748	0	1.80898292236443	0.989	0.263	0	14	COL3A1
749	0	1.80439519725435	0.731	0.096	0	14	TNFAIP6
750	0	1.79469298451364	0.74	0.1	0	14	FAT1
751	0	4.17131936463537	0.623	0.055	0	15	S100A9
752	0	4.13459213900366	0.514	0.019	0	15	S100A8
753	0	3.88947445544329	0.463	0.061	0	15	CXCL8
754	0	3.5562799877046	0.93	0.095	0	15	LYZ
755	0	3.01776579311192	0.654	0.01	0	15	FCN1
756	0	3.01744416254816	0.545	0.042	0	15	C15orf48
757	0	2.89383045109209	0.954	0.39	0	15	HLA-DRA
758	0	2.78324960928209	0.91	0.403	0	15	HLA-DPB1
759	0	2.72319888709394	0.908	0.087	0	15	AIF1
760	0	2.72295142515994	0.923	0.438	0	15	HLA-DPA1
761	0	2.62002566738581	0.476	0.048	0	15	G0S2
762	0	2.58041651824022	0.908	0.351	0	15	CST3
763	0	2.53442597186066	0.887	0.111	0	15	FCER1G
764	0	2.53303588129598	0.933	0.13	0	15	TYROBP

765	0	2.52608869040721	0.635	0.049	0	15	OLR1
766	0	2.5085299738629	0.921	0.339	0	15	HLA-DRB1
767	0	2.34569409161563	0.763	0.054	0	15	SERPINA1
768	0	2.33179978717694	0.73	0.085	0	15	CPVL
769	0	2.32872814001752	0.846	0.263	0	15	HLA-DRB5
770	0	2.32203166246175	0.837	0.209	0	15	HLA-DQB1
771	0	2.25409303706048	0.829	0.072	0	15	LST1
772	0	2.19096022951891	0.92	0.325	0	15	CTSS
773	0	2.09557177852119	0.684	0.048	0	15	MNDA
774	0	2.0941030278909	0.749	0.09	0	15	PLEK
775	0	2.07186389536865	0.438	0.015	0	15	EREG
776	0	2.0605417016592	0.924	0.41	0	15	NAMPT
777	0	2.03814030647849	0.563	0.063	0	15	C5AR1
778	0	2.03383500935684	0.617	0.077	0	15	BCL2A1
779	0	2.02490859772057	0.832	0.184	0	15	PLAUR
780	0	2.01163248412354	0.805	0.2	0	15	HLA-DQA1
781	0	2.01034096920756	0.81	0.205	0	15	HLA-DQA2

782	0	1.94231866081221	0.598	0.084	0	15	PPIF
783	0	1.89874673370548	0.696	0.077	0	15	MS4A6A
784	0	1.88399227836104	0.341	0.01	0	15	APOBEC3A
785	0	1.8780029703197	0.722	0.084	0	15	CYBB
786	0	1.8612099488571	0.999	0.919	0	15	FTH1
787	0	1.85358085652716	0.762	0.083	0	15	C1orf162
788	0	1.82302291060428	0.99	0.708	0	15	SAT1
789	0	1.81840341392309	0.828	0.328	0	15	TYMP
790	0	1.81043123692513	0.686	0.124	0	15	MXD1
791	0	1.8060263252898	0.455	0.028	0	15	SLC11A1
792	0	1.79491113900148	0.477	0.029	0	15	AC020656.1
793	0	1.78022575883723	0.259	0.013	0	15	IL1B
794	0	1.75573767882758	0.68	0.05	0	15	CLEC7A
795	0	1.74427865729382	0.758	0.141	0	15	FGL2
796	0	1.73400258008417	0.403	0.018	0	15	CLEC10A
797	3.72648141044051E-299	1.83707742845708	0.59	0.15	1.24978733543354E-294	15	BASP1
798	1.01513850624111E-286	2.10059066155998	0.95	0.656	3.40457152223143E-282	15	CD74

799	9.72151497425946E-136	2.24056457904873	0.219	0.042	3.26040169206714E-131	15	MT1G
800	3.95406699482677E-115	3.14956313212874	0.19	0.037	1.326114988725E-110	15	SPP1
801	0	5.07103889633011	0.76	0.141	0	16	CD24
802	0	4.18997290485962	0.789	0.039	0	16	KRT19
803	0	3.49597870670766	0.722	0.063	0	16	KRT18
804	0	3.4882990049423	0.719	0.038	0	16	KRT8
805	0	3.45307017279568	0.476	0.018	0	16	AZGP1
806	0	3.41857398965857	0.37	0.025	0	16	KRT17
807	0	3.29568929962932	0.461	0.027	0	16	SLPI
808	0	3.11805442326799	0.629	0.026	0	16	KRT7
809	0	3.04765330218293	0.435	0.018	0	16	CALML5
810	0	3.02832133733885	0.316	0.015	0	16	SAA1
811	0	2.97183296830056	0.151	0.006	0	16	SCGB1D2
812	0	2.90570558430003	0.531	0.012	0	16	CLDN4
813	0	2.88816611072864	0.301	0.021	0	16	KRT81
814	0	2.68552806445518	0.371	0.013	0	16	KRT6B
815	0	2.54953652819633	0.354	0.011	0	16	KRT14

816	0	2.53872905497669	0.624	0.099	0	16	PERP
817	0	2.52048165247598	0.531	0.011	0	16	ELF3
818	0	2.49447067088411	0.493	0.024	0	16	TACSTD2
819	0	2.36203481936835	0.38	0.009	0	16	WFDC2
820	0	2.31160833819289	0.298	0.016	0	16	MMP7
821	0	2.29303562615427	0.707	0.211	0	16	NUPR1
822	0	2.22839657801784	0.361	0.007	0	16	KRT23
823	0	2.22766101810345	0.236	0.015	0	16	TFF3
824	0	2.20293129657981	0.614	0.1	0	16	SPINT2
825	0	2.171728670824	0.559	0.009	0	16	FXYD3
826	0	2.16109377012606	0.468	0.009	0	16	S100A14
827	0	2.14370647488267	0.512	0.046	0	16	DSP
828	0	2.10864541185881	0.328	0.013	0	16	S100P
829	0	2.0256559787909	0.294	0.013	0	16	KRT5
830	0	1.96892747051701	0.51	0.047	0	16	MGST1
831	0	1.96660779007107	0.507	0.06	0	16	CRABP2
832	0	1.91010867848487	0.512	0.006	0	16	SMIM22

833	0	1.82894807780495	0.391	0.019	0	16	IRX3
834	0	1.7833062205408	0.292	0.008	0	16	CEACAM6
835	0	1.78157908997871	0.388	0.005	0	16	CLDN3
836	0	1.74106315204194	0.237	0.006	0	16	AGR2
837	0	1.706634417205	0.227	0.004	0	16	ANKRD30A
838	0	1.65901507833674	0.426	0.007	0	16	MAL2
839	0	1.65793259620747	0.573	0.122	0	16	S100A16
840	2.59866596771194E-296	1.8617211275075	0.554	0.122	8.71540592251232E-292	16	TM4SF1
841	2.35836174322364E-272	2.32593930523819	0.169	0.012	7.90947361442345E-268	16	PIP
842	7.2917397787715E-219	2.37661962109214	0.452	0.116	2.44550368700438E-214	16	CRYAB
843	2.16238395252839E-190	2.06657007542628	0.15	0.014	7.25220329998971E-186	16	LTF
844	1.06212064615851E-167	2.29237681887349	0.342	0.082	3.5621402230864E-163	16	RARRES1
845	8.99816364131386E-166	2.7142418427285	0.745	0.385	3.01780412202384E-161	16	MGP
846	3.7130002090873E-133	5.63327321241388	0.257	0.057	1.2452660101237E-128	16	SCGB2A2
847	1.90968986540778E-127	1.82575303000378	0.397	0.129	6.4047178706046E-123	16	EIF4EBP1
848	4.90161129418762E-121	4.74390678919746	0.243	0.056	1.64390239584465E-116	16	MUCL1
849	1.11179576583843E-40	2.12881408394581	0.2	0.077	3.72874063946891E-36	16	PTN

850	2.17288314845776E-36	1.70984941007748	0.617	0.455	7.28741550329763E-32	16	MIF
851	0	3.84930431013594	0.99	0.114	0	17	GZMB
852	0	3.51449494182625	0.951	0.076	0	17	IRF4
853	0	2.92444217094545	0.882	0.138	0	17	LDLRAD4
854	0	2.77916395004085	0.935	0.193	0	17	IRF7
855	0	2.77153120735784	0.905	0.046	0	17	IL3RA
856	0	2.7419106579092	0.827	0.04	0	17	EGLN3
857	0	2.66459039248216	0.843	0.12	0	17	GPR65
858	0	2.6091212254165	0.85	0.043	0	17	PLAC8
859	0	2.51334536490321	0.807	0.075	0	17	POLB
860	0	2.42357663856417	0.833	0.004	0	17	LILRA4
861	0	2.39755873870714	0.778	0.082	0	17	IRF8
862	0	2.3831031692537	0.794	0.015	0	17	PLD4
863	0	2.21418694330794	0.801	0.121	0	17	C12orf75
864	0	2.10326156360958	0.719	0.069	0	17	TSPAN13
865	0	2.06460101664617	0.444	0.004	0	17	TCL1A
866	0	1.99570307293444	0.83	0.096	0	17	CYBB

867	0	1.99482049021158	0.657	0.037	0	17	CLIC3
868	0	1.92964734907746	0.693	0.051	0	17	SLC15A4
869	0	1.92550209588633	0.745	0.058	0	17	SEL1L3
870	0	1.89737410718091	0.68	0.006	0	17	SMPD3
871	0	1.88328303414543	0.758	0.093	0	17	THEMIS2
872	0	1.87885019323255	0.363	0.017	0	17	KRT5
873	0	1.84339559992563	0.601	0.017	0	17	P2RY6
874	0	1.82456007073694	0.634	0.019	0	17	SPIB
875	0	1.80296712354396	0.696	0.051	0	17	RUBCN
876	6.03172177605431E-301	2.18005035495885	0.719	0.101	2.02291884925309E-296	17	GRASP
877	7.95858196101838E-298	4.42801683368466	0.98	0.264	2.66914921808634E-293	17	GPR183
878	1.99306623010505E-283	3.6694981600322	0.542	0.058	6.68434552252631E-279	17	PTGDS
879	3.52949956364828E-279	1.81938169219165	0.775	0.124	1.18372356365636E-274	17	FCHSD2
880	4.77991319356782E-274	3.34002425625207	0.873	0.174	1.60308728685878E-269	17	AREG
881	2.67619816879648E-234	3.4143952913781	0.899	0.24	8.97543341850963E-230	17	SOX4
882	1.24601511003947E-232	2.68513033390403	0.908	0.237	4.17888547605037E-228	17	NR4A3
883	4.39807941144457E-231	1.85597517073818	0.814	0.157	1.47502787301028E-226	17	SLC7A5

884	3.76583197119423E-226	2.86773859227139	0.922	0.262	1.26298472649912E-221	17	RGS2
885	1.23636030249869E-222	1.84966255140465	0.866	0.189	4.14650518252011E-218	17	ITM2C
886	1.87828639074884E-220	1.94760232773458	0.931	0.225	6.29939689729345E-216	17	ALOX5AP
887	2.47888839901386E-215	2.45116116215964	0.84	0.218	8.3136959126127E-211	17	PPP1R14B
888	6.28932311599494E-214	2.03160040591348	0.546	0.076	2.10931318664238E-209	17	CHML
889	9.96581502648118E-204	2.01160277460435	0.752	0.157	3.34233504358126E-199	17	DUSP5
890	8.92445906604974E-183	2.03367084531104	0.941	0.313	2.99308508157176E-178	17	TCF4
891	1.92628729396688E-181	1.86171832055726	0.82	0.197	6.46038232650611E-177	17	RAB11FIP1
892	5.67893685032745E-169	2.20816975553452	0.951	0.533	1.90460184086282E-164	17	SEC61B
893	6.87225795161289E-164	2.4618681572443	0.935	0.387	2.30481787181193E-159	17	NR3C1
894	5.26325942081659E-162	1.80469310567606	0.768	0.198	1.76519194455347E-157	17	PTPRE
895	8.80040303055726E-131	1.85963341380656	0.722	0.186	2.95147916838829E-126	17	CD83
896	2.8061243161186E-125	2.01863169056561	0.948	0.584	9.41117973139857E-121	17	HERPUD1
897	1.68722810906202E-124	1.94253557157188	0.863	0.334	5.65862563217221E-120	17	RANBP2
898	2.88880771239854E-114	1.95239845884602	0.977	0.638	9.68848330584221E-110	17	HSP90B1
899	2.46137371997449E-108	2.04812895323875	0.908	0.5	8.25495518205044E-104	17	TXN
900	3.19916689716556E-107	1.80605712175595	0.866	0.404	1.07293659397139E-102	17	ZC3HAV1

901	0	3.01910846295272	0.881	0.017	0	18	MKI67
902	0	3.01284037794777	0.857	0.069	0	18	CENPF
903	0	2.86057682330401	0.623	0.011	0	18	HIST1H1B
904	0	2.81123086511458	0.77	0.022	0	18	TOP2A
905	0	2.3708645899739	0.841	0.052	0	18	NUSAP1
906	0	2.30197757164499	0.631	0.009	0	18	ASPM
907	0	1.92679392882595	0.667	0.028	0	18	TPX2
908	0	1.86017606161774	0.702	0.026	0	18	CLSPN
909	0	1.8027147452827	0.698	0.067	0	18	ATAD2
910	0	1.76297051853278	0.563	0.021	0	18	TYMS
911	0	1.75922996509624	0.528	0.012	0	18	PCLAF
912	0	1.70898504042919	0.615	0.007	0	18	UBE2C
913	0	1.65521472201144	0.595	0.015	0	18	KNL1
914	0	1.60750770189769	0.571	0.014	0	18	TK1
915	0	1.60172251514312	0.54	0.029	0	18	CENPE
916	0	1.53010807112942	0.635	0.008	0	18	BIRC5
917	0	1.42696262147337	0.615	0.016	0	18	CDK1

918	0	1.42164747225081	0.298	0.004	0	18	HIST1H3B
919	2.54191749905985E-286	1.63474963728806	0.702	0.081	8.52508290834691E-282	18	SMC2
920	2.84488291359381E-274	3.40168831829775	0.976	0.215	9.54116831561093E-270	18	STMN1
921	9.00129270626651E-256	1.8213122942346	0.75	0.098	3.01885354782766E-251	18	PTTG1
922	5.28314841214196E-244	1.46261265767365	0.667	0.084	1.77186231446417E-239	18	CKS1B
923	1.03440803688497E-227	1.9773879496357	0.583	0.065	3.46919767410482E-223	18	HIST1H1D
924	1.96822149383793E-207	2.27032000506649	0.69	0.105	6.60102124603365E-203	18	HIST1H1C
925	1.0994381928373E-198	1.51863099671224	0.31	0.021	3.68729581113775E-194	18	HIST1H2AG
926	5.97774882522325E-188	2.72245140874031	0.948	0.276	2.00481740100337E-183	18	HMGB2
927	1.02383724169225E-185	1.32946002492297	0.611	0.085	3.43374534118745E-181	18	KIF20B
928	2.18201429094029E-177	1.33841958561236	0.524	0.068	7.31803952895554E-173	18	MCM7
929	9.39268088266789E-174	1.81104802778759	0.825	0.184	3.15011731442916E-169	18	SMC4
930	1.33261051086824E-164	1.74002061195315	0.639	0.112	4.4693091313499E-160	18	PCNA
931	4.30135561648955E-162	1.65824315515079	0.774	0.167	1.44258864665827E-157	18	TMPO
932	7.1114520546826E-157	2.63902947332168	0.984	0.468	2.38503879009945E-152	18	HMGN2
933	8.81297377072798E-155	1.94362208756181	0.607	0.102	2.95569514322675E-150	18	HIST1H1E
934	2.93691806158295E-145	2.71002819650146	0.968	0.416	9.84983579493689E-141	18	TUBB

935	1.79489791954012E-138	1.76932938192482	0.877	0.252	6.01972864255365E-134	18	ANP32E
936	1.95863629417128E-135	2.57238602322161	0.992	0.633	6.56887440339164E-131	18	TUBA1B
937	1.45915844674933E-128	1.88781475864768	0.988	0.816	4.8937255987079E-124	18	HMGB1
938	1.28639768655024E-126	1.94116380959393	0.762	0.222	4.31432056115221E-122	18	DUT
939	1.37310675954606E-120	3.21158347598526	0.913	0.404	4.60512545016557E-116	18	HIST1H4C
940	2.49155972478746E-117	1.40806818002362	0.468	0.078	8.35619300499217E-113	18	HIST2H2AC
941	3.87031392462406E-114	1.42618753241151	0.821	0.246	1.29802588404042E-109	18	SMC1A
942	2.56375470154834E-111	1.74127354832358	0.976	0.591	8.59832051805282E-107	18	H2AFZ
943	9.29647132140042E-104	1.3400308554611	0.726	0.203	3.11785055177127E-99	18	DNMT1
944	1.17585014904468E-87	1.59037896212823	0.865	0.399	3.94356622986605E-83	18	DEK
945	1.77765116603214E-87	1.33408532621725	0.833	0.304	5.96188648063861E-83	18	CALM3
946	1.72431729697521E-85	1.47716657640215	0.762	0.283	5.78301535059546E-81	18	H2AFV
947	1.29670494282609E-76	1.50153399371621	0.984	0.724	4.34888903725015E-72	18	PFN1
948	7.39741068132385E-71	1.40697057377696	0.992	0.876	2.48094359430239E-66	18	GAPDH
949	1.23550819341052E-53	1.48075011864768	0.73	0.353	4.14364737906022E-49	18	CORO1A
950	3.83743823584551E-42	1.49508991001724	0.996	0.914	1.28700003553787E-37	18	ACTB
951	0	5.28766447894644	0.982	0.084	0	19	FABP4

952		0	5.06059299839712	0.818	0.02		0	19	CCL21	
953		0	3.76922526406088	0.915	0.017		0	19	TFF3	
954		0	3.48824770645013	0.873	0.017		0	19	MMRN1	
955		0	3.43195946042583	0.964	0.076		0	19	RAMP2	
956		0	3.34337511768166	0.909	0.068		0	19	CAVIN2	
957		0	2.7415351417598	0.848	0.041		0	19	TM4SF18	
958		0	2.70786757624885	0.558	0.028		0	19	EDN1	
959		0	2.49700940148238	0.709	0.014		0	19	PROX1	
960		0	2.47709554795607	0.818	0.067		0	19	PDPN	
961		0	2.38846420679543	0.776	0.06		0	19	MRC1	
962		0	2.33325661365123	0.667	0.032		0	19	CLDN5	
963		0	2.1962501083531	0.752	0.052		0	19	SMAD1	
964		0	2.1217591810177	0.764	0.058		0	19	EGFL7	
965		0	2.00761754762062	0.739	0.038		0	19	ROBO4	
966		0	1.8509329235483	0.624	0.022		0	19	SLC24A1	
967		0	1.77284795199951	0.661	0.041		0	19	CXorf36	
968	8.37911071887384E-294		1.78498321271382	0.727	0.055	2.81018615289591E-289		19	CDH5	

969	1.12013627996946E-292	2.78266571984073	0.83	0.078	3.75671305576159E-288	19	CD34
970	1.89025972875982E-270	3.2153016876691	0.836	0.083	6.33955307831467E-266	19	EFEMP1
971	1.44503715081003E-236	2.23928870469434	0.879	0.103	4.84636559638666E-232	19	CALCRL
972	1.24870856143323E-226	2.51871633291842	0.952	0.13	4.18791877333476E-222	19	TM4SF1
973	3.08291236736376E-223	3.76862379196124	0.982	0.174	1.03394714976646E-218	19	GNG11
974	8.43770945834133E-215	3.45974620756339	0.988	0.177	2.82983899813851E-210	19	TFPI
975	1.12868430454534E-211	3.0835268632527	0.879	0.125	3.78538142058418E-207	19	ANGPT2
976	4.82062510408324E-200	1.8351755988141	0.673	0.073	1.61674124740744E-195	19	NT5E
977	1.78135165457462E-176	3.10295559151395	0.964	0.21	5.97429717911235E-172	19	CD9
978	5.39609456153613E-172	2.42706491831587	0.891	0.169	1.80974219404799E-167	19	PPFIBP1
979	6.26369158689078E-154	3.03752667653115	0.867	0.161	2.10071688441143E-149	19	RGS16
980	5.06042595913258E-138	2.98375995177133	0.933	0.235	1.69716565817388E-133	19	CAV1
981	9.96708668147817E-138	2.27312912978449	0.933	0.218	3.34276153123415E-133	19	STMN1
982	7.98971821277554E-132	2.48964942176737	0.77	0.146	2.67959169420066E-127	19	ID1
983	2.67154268095182E-127	2.84097506618338	0.945	0.262	8.95981984337622E-123	19	FABP5
984	5.91684356378311E-126	2.0190734402107	0.867	0.194	1.98439099442158E-121	19	LAMA4
985	3.29237299934298E-122	2.8328738249807	0.764	0.155	1.10419605651965E-117	19	AKAP12

986	8.22330601591931E-116	1.78926080929569	0.485	0.063	2.75793237161902E-111	19	HOMER3
987	3.08028340228549E-112	1.97690334111609	0.727	0.147	1.03306544745851E-107	19	CCND1
988	3.98319973440671E-112	2.253546268984	0.897	0.243	1.33588552692532E-107	19	SOX4
989	3.29202746921138E-99	1.9980053428826	0.885	0.263	1.10408017262411E-94	19	SPTBN1
990	1.94588374535401E-98	2.20047262730024	0.927	0.316	6.52610490516827E-94	19	TCF4
991	8.81290888746621E-94	1.76735103226992	0.879	0.254	2.95567338267842E-89	19	CNN3
992	3.00426452476968E-85	2.2951253090153	0.903	0.319	1.00757023631725E-80	19	ID3
993	2.33975323272444E-84	2.22394871389	0.915	0.371	7.84706439191122E-80	19	GYPC
994	5.50755880705204E-83	1.93899676973947	0.933	0.393	1.84712507270911E-78	19	HSPB1
995	4.56045702728127E-78	1.79212111828015	0.994	0.549	1.52948607780959E-73	19	ANXA2
996	8.78415985943439E-75	1.96489122330014	0.636	0.151	2.94603153365711E-70	19	SPRY1
997	3.0983976741447E-72	1.83039070304682	1	0.656	1.03914061195465E-67	19	S100A10
998	8.15607634260399E-69	1.91633780546215	0.885	0.376	2.73538488378253E-64	19	FKBP1A
999	5.27237700197624E-66	1.93402064526177	0.909	0.345	1.76824979892279E-61	19	IGFBP7
1000	6.29425430699256E-63	1.85588358338085	0.909	0.466	2.11096700947917E-58	19	TUBA1A
1001	0	6.23366690664738	0.951	0.007	0	20	TPSAB1
1002	0	6.17796334743313	0.993	0.003	0	20	CPA3

1003	0	6.11310792567959	0.938	0.009	0	20	TPSB2
1004	0	4.13082889866044	0.792	0.001	0	20	MS4A2
1005	0	4.05566241831384	0.736	0.018	0	20	HPGD
1006	0	3.87864274787177	0.854	0.008	0	20	HPGDS
1007	0	3.82902129480647	0.819	0.007	0	20	IL1RL1
1008	0	3.57364563984605	0.465	0.003	0	20	TPSD1
1009	0	3.41603248219947	0.694	0.009	0	20	KIT
1010	0	3.03179418708268	0.625	0.001	0	20	SLC18A2
1011	0	3.02455805823425	0.319	0.001	0	20	CTSG
1012	0	3.01774038757653	0.396	0.001	0	20	ADCYAP1
1013	0	3.00257936332397	0.625	0.024	0	20	VWA5A
1014	0	2.9613157131827	0.549	0.001	0	20	HDC
1015	0	2.31388228256927	0.444	0.007	0	20	RHEX
1016	0	1.9991849879671	0.368	0.002	0	20	CALB2
1017	0	1.83067227171023	0.368	0	0	20	AL157895.1
1018	1.68090761394616E-296	2.69850399085723	0.458	0.019	5.63742795565263E-292	20	GATA2
1019	1.01853878850242E-227	1.96432268020381	0.368	0.016	3.41597538887942E-223	20	RGS13

1020	6.4188727930059E-207	1.8966788338584	0.375	0.018	2.15276155731832E-202	20	MLPH
1021	1.48048218159964E-182	2.60368340668469	0.507	0.04	4.96524114064887E-178	20	LTC4S
1022	4.1839743547859E-144	2.9432418340179	0.847	0.153	1.4032213191081E-139	20	CLU
1023	3.0951860210373E-108	2.06323588082415	0.757	0.132	1.03806348773549E-103	20	FCER1G
1024	2.40771463729688E-89	3.19316777905919	0.75	0.178	8.07499335056629E-85	20	AREG
1025	4.86184477974316E-86	1.70875308313862	0.278	0.024	1.63056550223026E-81	20	PTGS1
1026	6.78306968108909E-85	2.46038284242608	0.375	0.045	2.27490590964366E-80	20	PTGS2
1027	6.30387129021256E-83	2.20533605160454	0.764	0.199	2.11419235331149E-78	20	CAPG
1028	7.93464957214468E-83	1.85350734803694	0.382	0.048	2.66112277350588E-78	20	BTK
1029	1.80025956356293E-77	2.02264091135743	0.396	0.055	6.03771052427734E-73	20	ALOX5
1030	1.29516310993369E-75	2.73699349528472	0.507	0.092	4.34371803809561E-71	20	BCL2A1
1031	8.67470357729395E-74	2.35205222898205	0.569	0.12	2.90932208575284E-69	20	DUSP6
1032	4.02195750262585E-65	1.87461126440952	0.486	0.095	1.34888410723066E-60	20	ARHGEF6
1033	1.28825404814215E-57	1.83492308875456	0.729	0.23	4.32054642665916E-53	20	ALOX5AP
1034	3.10974179255752E-56	2.32771369155131	0.514	0.125	1.04294520238794E-51	20	LEO1
1035	1.92466829654724E-51	2.27639503608551	0.562	0.169	6.45495253296012E-47	20	ACSL4
1036	5.30565754529878E-47	1.95697922984312	0.611	0.206	1.7794114275423E-42	20	ARHGAP18

1037	1.84117596458132E-44	1.72903845397466	0.417	0.098	6.17493595001282E-40	20	PAK1
1038	1.94979364584816E-43	2.42898346699776	0.812	0.493	6.53921792944556E-39	20	NFKBIA
1039	2.20204127829935E-42	1.8439302095187	0.917	0.613	7.38520603916036E-38	20	ANXA1
1040	2.4613877001828E-37	2.00347686128639	0.646	0.267	8.25500206887307E-33	20	RGS2
1041	7.65863837774688E-37	1.80265903347575	0.451	0.126	2.56855413912875E-32	20	GPR65
1042	8.3104052283757E-37	2.1113166785907	0.667	0.296	2.78714370549264E-32	20	GLUL
1043	9.22675238706362E-33	1.87104133745633	0.535	0.19	3.0944682155734E-28	20	CD83
1044	2.11023859427376E-30	1.92959505598088	0.681	0.366	7.07731819747535E-26	20	ID2
1045	1.52894119345253E-28	1.77326790183438	0.674	0.342	5.12776297460109E-24	20	BIRC3
1046	1.83046866916757E-28	1.82125585992918	0.389	0.122	6.1390258226542E-24	20	RBBP8
1047	6.70355220921595E-24	1.71493269784714	0.5	0.225	2.24823733992685E-19	20	NFKBIZ
1048	3.30971359296416E-20	1.71016617431417	0.562	0.319	1.11001174480832E-15	20	KDM6B
1049	1.3047810401404E-13	2.04725176091125	0.438	0.217	4.37597465242289E-09	20	CD69
1050	3.4231968234513E-12	1.84058459272606	0.465	0.277	1.1480717506491E-07	20	TSC22D1

D.12 TCR Clones

Table D.5: Top Ten TCR Clones in BRO20.

	Clones	Proportion	CDR3.nt	CDR3.aa	V.name	D.name	J.name	Sequence	chain
1	26	0.013478486262312	TGTGCTGTGAGAGACGAGTACGAC CAGGCAGGAACTGCTCTGATCTTT; TGTGCCAGCAGCTTGGCATTAGC GGGAGGGCGTCTTGAGTACTTC	CAVRDEYDQAGTALIF; CASSLALAGGRLEYF	TRAV3;TRBV5-6	NA;TRBD2	TRAJ15; TRBJ2-5	TGTGCTGTGAGAGACGAGTACGAC CAGGCAGGAACTGCTCTGATCTTT; TGTGCCAGCAGCTTGGCATTAGC GGGAGGGCGTCTTGAGTACTTC	TRA;TRB
2	25	0.012960082944531	TGTGCCAGCAGCTCCCCGGGCCGGGGG CTTCGAGGGGCCAACGTCCTGACTTTC	CASSSPGRG LRGANVLTF	TRBV12-3	TRBD1	TRBJ2-6	TGTGCCAGCAGCTCCCCGGGCCGGGGG CTTCGAGGGGCCAACGTCCTGACTTTC	TRB
3	19	0.009849663037843	TGTGCTCTCTTCACGGGAG GAGGAAACAAACTCACCTTT; TGTGCTCTGAGTGAGGCTG GAGATGACAAGATCATCTTT; TGTGCCAGCAGCCTTAATCCGGGACTA GCGGGAGGAGGGGATACGCAGTATTTT	CALFTGGGNKLTF; CALSEAGDDKIIF; CASSLNPGLAGGGDTQYF	TRAV16;TRAV19; TRBV4-2	NA;NA;TRBD2	TRAJ10;TRAJ30; TRBJ2-3	TGTGCTCTCTTCACGGGAGG AGGAAACAAACTCACCTTT; TGTGCTCTGAGTGAGGCTG GAGATGACAAGATCATCTTT; TGTGCCAGCAGCCTTAATCCGGGACTA GCGGGAGGAGGGGATACGCAGTATTTT	TRA;TRA;TRB
4	14	0.007257646448937	TGTGCCAGCAGCTTAGCCA GGAACACTGAAGCTTTCTTT	CASSLARNTEAFF	TRBV7-9	NA	TRBJ1-1	TGTGCCAGCAGCTTAGCCA GGAACACTGAAGCTTTCTTT	TRB
5	13	0.006739243131156	TGTGCTCTCTTCACGGGAGG AGGAAACAAACTCACCTTT; TGTGCCAGCAGCCTTAATC CGGGACTAGCGGGAGGAG GGGATACGCAGTATTTT	CALFTGGGNKLTF; CASSLNPGLAGGGDTQYF	TRAV16;TRBV4-2	NA;TRBD2	TRAJ10; TRBJ2-3	TGTGCTCTCTTCACGGGAGG AGGAAACAAACTCACCTTT; TGTGCCAGCAGCCTTAATCCGGGACTA GCGGGAGGAGGGGATACGCAGTATTTT	TRA;TRB
6	12	0.006220839813375	TGTGCAGCTGGAAACACCA ATGCAGGCAAATCAACCTTT; TGTGCCATCAGCCCCACA AACTATGGCTACACCTTC	CAAGNTNAGKSTF; CAISPTNYGYTF	TRAV29/DV5; TRBV10-3	NA;NA	TRAJ27; TRBJ1-2	TGTGCAGCTGGAAACACCA ATGCAGGCAAATCAACCTTT; TGTGCCATCAGCCCCACA AACTATGGCTACACCTTC	TRA;TRB
7	12	0.006220839813375	TGTGCTCTGAGTGAAACC TCCTACGACAAGGTGATATTT; TGTGCCAGCAGCCTGTTTCGGG GGTACTATGGCTACACCTTC	CALSETSYDKVIF; CASSLFRGYYGYTF	TRAV19;TRBV4-2	NA;NA	TRAJ50;TRBJ1-2	TGTGCTCTGAGTGAAACCT CCTACGACAAGGTGATATTT; TGTGCCAGCAGCCTGTTTCGG GGGTACTATGGCTACACCTTC	TRA;TRB
8	12	0.006220839813375	TGTGCTTATAGGCGGACGAAC ACCGGTAACCAGTTCTATTTT; TGTGCCAGCAGCCCGGGCGGGAC GGTATACAATGAGCAGTTCTTC	CAYRRTNTGNQFYF; CASSPGGTVYNEQFF	TRAV38-2/DV8; TRBV11-3	NA;NA	TRAJ49;TRBJ2-1	TGTGCTTATAGGCGGACGA ACACCGGTAACCAGTTCTATTTT; TGTGCCAGCAGCCCGGGCGGGACG GTATACAATGAGCAGTTCTTC	TRA;TRB
9	11	0.005702436495594	TGTGCTCTGAGTGAGGCTG GAGATGACAAGATCATCTTT; TGTGCTCTCTTCACGGGAGGAG GAAACAAACTCACCTTT; TGTGCCAGCAGCCTTAATCCGGGACT AGCGGGAGGAGGGGATACGCAGTATTTT	CALSEAGDDKIIF; CALFTGGGNKLTF; CASSLNPGLAGGGDTQYF	TRAV19;TRAV16; TRBV4-2	NA;NA;TRBD2	TRAJ30;TRAJ10; TRBJ2-3	TGTGCTCTGAGTGAGGCTG GAGATGACAAGATCATCTTT; TGTGCTCTCTTCACGGGA GGAGGAAACAAACTCACCTTT; TGTGCCAGCAGCCTTAATCCG GGACTAGCGGGAGGAGG GGATACGCAGTATTTT	TRA;TRA;TRB
10	11	0.005702436495594	TGTGCTCTGCTGGAAT ATGGAAACAAGCTGGTCTTT; TGTGCTAGTGGTTTGGGGGG AGGCGTGCCCGCTTTCTTT	CALLEYGNKLVF; CASGLGGGVPAFF	TRAV9-2;TRBV12-5	NA;NA	TRAJ47;TRBJ1-1	TGTGCTCTGCTGGAATATG GAAACAAGCTGGTCTTT; TGTGCTAGTGGTTTGGG GGGAGGCGTGCCCGCTTTCTTT	TRA;TRB

Table D.6: Top Ten TCR Clones in BRO23.

	Clones	Proportion	CDR3.nt	CDR3.aa	V.name	D.name	J.name	Sequence	chain
1	81	0.022816901408451	TGTGCAATGAGGGTCCCGAGGGATGCTG GTGGTACTAGCTATGGAAAGCTGACATTT; TGTGCCAGTAGCTCCGGACA GGCTCGCGAAGCTTTCTTT	CAMRVPRDAGGTSYGKLTF; CASSSGQAREAFF	TRAV12-3;TRBV19	NA;NA	TRAJ52;TRBJ1-1	TGTGCAATGAGGGTCCCGAGGGATGCTG GTGGTACTAGCTATGGAAAGCTGACATTT; TGTGCCAGTAGCTCCGGACA GGCTCGCGAAGCTTTCTTT	TRA;TRB
2	47	0.013239436619718	TGTGCTGTCCGCGGAACTTTAGCG AGAGGTGCTGACGGACTCACCTTT; TGTGCCAGCAGCTTAGGGA CTCCAGGGATTCAGTACTTC	CAVRGTLARGADGLTF; CASSLGTPGIQYF	TRAV41;TRBV13	NA;NA	TRAJ45;TRBJ2-4	TGTGCTGTCCGCGGAACTTTAGCG AGAGGTGCTGACGGACTCACCTTT; TGTGCCAGCAGCTTAGGGA CTCCAGGGATTCAGTACTTC	TRA;TRB
3	43	0.012112676056338	TGTGTGGTGAACATCGGTCGAGATTCA GGAGGAGGTGCTGACGGACTCACCTTT; TGTGCCAGCCGACAGGGTTT TCTTAATGAGCAGTTCTTC	CVVNIGRDSGGGADGLTF; CASRQGFLNEQFF	TRAV12-1;TRBV19	NA;NA	TRAJ45;TRBJ2-1	TGTGTGGTGAACATCGGTCGAGATTCA GGAGGAGGTGCTGACGGACTCACCTTT; TGTGCCAGCCGACAGGGTT TTCTTAATGAGCAGTTCTTC	TRA;TRB
4	40	0.011267605633803	TGTGCAGGATACGATAACT ATGGTCAGAATTTTGTCTTT; TGTGCCAGCAGCTTAGATCGAGGGA ATCTTTTCACCGGGGAGCTGTTTTTT	CAGYDNYGQNFVF; CASSLDRGNLFTGELFF	TRAV27;TRBV11-2	NA;NA	TRAJ26;TRBJ2-2	TGTGCAGGATACGATAACTA TGGTCAGAATTTTGTCTTT; TGTGCCAGCAGCTTAGATCGAGGGA ATCTTTTCACCGGGGAGCTGTTTTTT	TRA;TRB
5	40	0.011267605633803	TGTGCCGTGATTTCTGG TGACAAGATCATCTTT; TGCGCCAGCAGCCCTGAG GAAGTTACCCAGTACTTC	CAVISGDKIIF; CASSPEEVTQYF	TRAV12-2;TRBV4-1	NA;NA	TRAJ30;TRBJ2-5	TGTGCCGTGATTTCTGG TGACAAGATCATCTTT; TGCGCCAGCAGCCCTGAG GAAGTTACCCAGTACTTC	TRA;TRB
6	32	0.009014084507042	TGTGCTGTGAGTGATAAGAGGG ATAGCAGCTATAAATTGATCTTC; TGCGCCAGCAGCCAGAGCGGGG AGGAAGGAGAGACCCAGTACTTC	CAVSDKRDSSYKLIF; CASSQSGEEGETQYF	TRAV8-6;TRBV4-3	NA;NA	TRAJ12;TRBJ2-5	TGTGCTGTGAGTGATAAGAGGG ATAGCAGCTATAAATTGATCTTC; TGCGCCAGCAGCCAGAGCGGGG AGGAAGGAGAGACCCAGTACTTC	TRA;TRB
7	31	0.008732394366197	TGTGCAGAGAGGGGAAGATA CAGCACCCTCACCTTT; TGTGCCAGCAGCTTAGGCTCCGGGAC AGGGGCTACCGGGGAGCTGTTTTTT	CAERGRYSTLTF; CASSLGSGTGATGELFF	TRAV5;TRBV7-8	NA;TRBD1	TRAJ11;TRBJ2-2	TGTGCAGAGAGGGGAAGA TACAGCACCCTCACCTTT; TGTGCCAGCAGCTTAGGCTCCGGGAC AGGGGCTACCGGGGAGCTGTTTTTT	TRA;TRB
8	29	0.008169014084507	TGTGCTGTGAGAGATCTCAACT TTGGAAATGAGAAATTAACCTTT	CAVRDLNFGNEKLTF	TRAV1-2	NA	TRAJ48	TGTGCTGTGAGAGATCTCAACT TTGGAAATGAGAAATTAACCTTT	TRA
9	29	0.008169014084507	TGTGTGGTATATGGA AACAAGCTGGTCTTT; TGTGCCAGCAGCGCCAGTCCAGGGTC CAATAACTATGGCTACACCTTC	CVVYGNKLVF; CASSASPGSNNYGYTF	TRAV12-1;TRBV9	NA;NA	TRAJ47;TRBJ1-2	TGTGTGGTATATG GAAACAAGCTGGTCTTT; TGTGCCAGCAGCGCCAGTCCAGGG TCCAATAACTATGGCTACACCTTC	TRA;TRB
10	27	0.007605633802817	TGCGCCAGCAGCCAGAGCGGGG AGGAAGGAGAGACCCAGTACTTC	CASSQSGEEGETQYF	TRBV4-3	NA	TRBJ2-5	TGCGCCAGCAGCCAGAGCGGGGA GGAAGGAGAGACCCAGTACTTC	TRB

D.13 Pseudotime Analysis

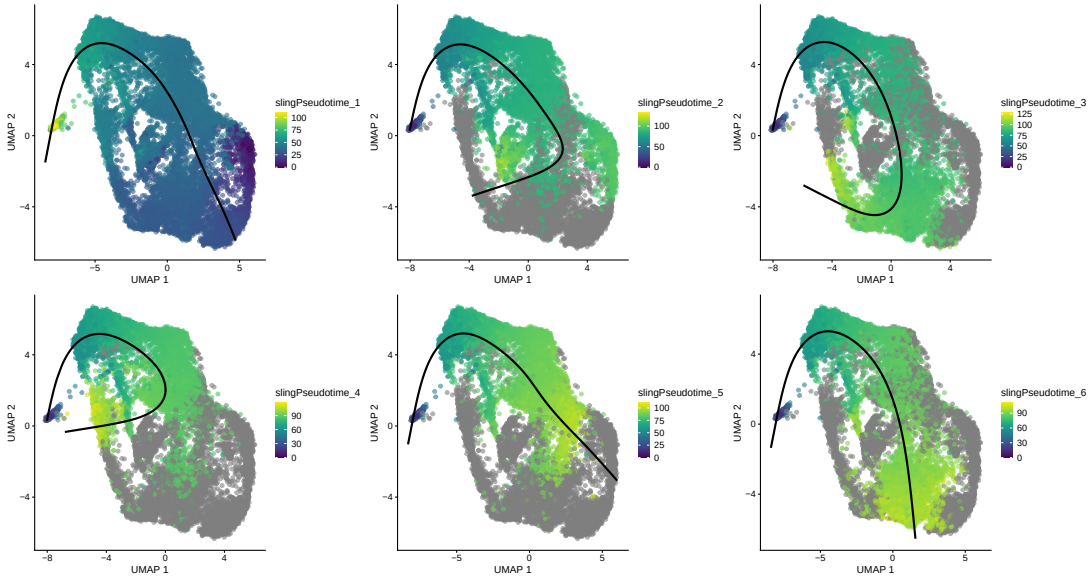


Figure D.1: T Cell Trajectories.
Principal curves representing each of six predicted lineage paths through T cell clusters with pseudotime values obtained by orthogonal projection onto the curves.

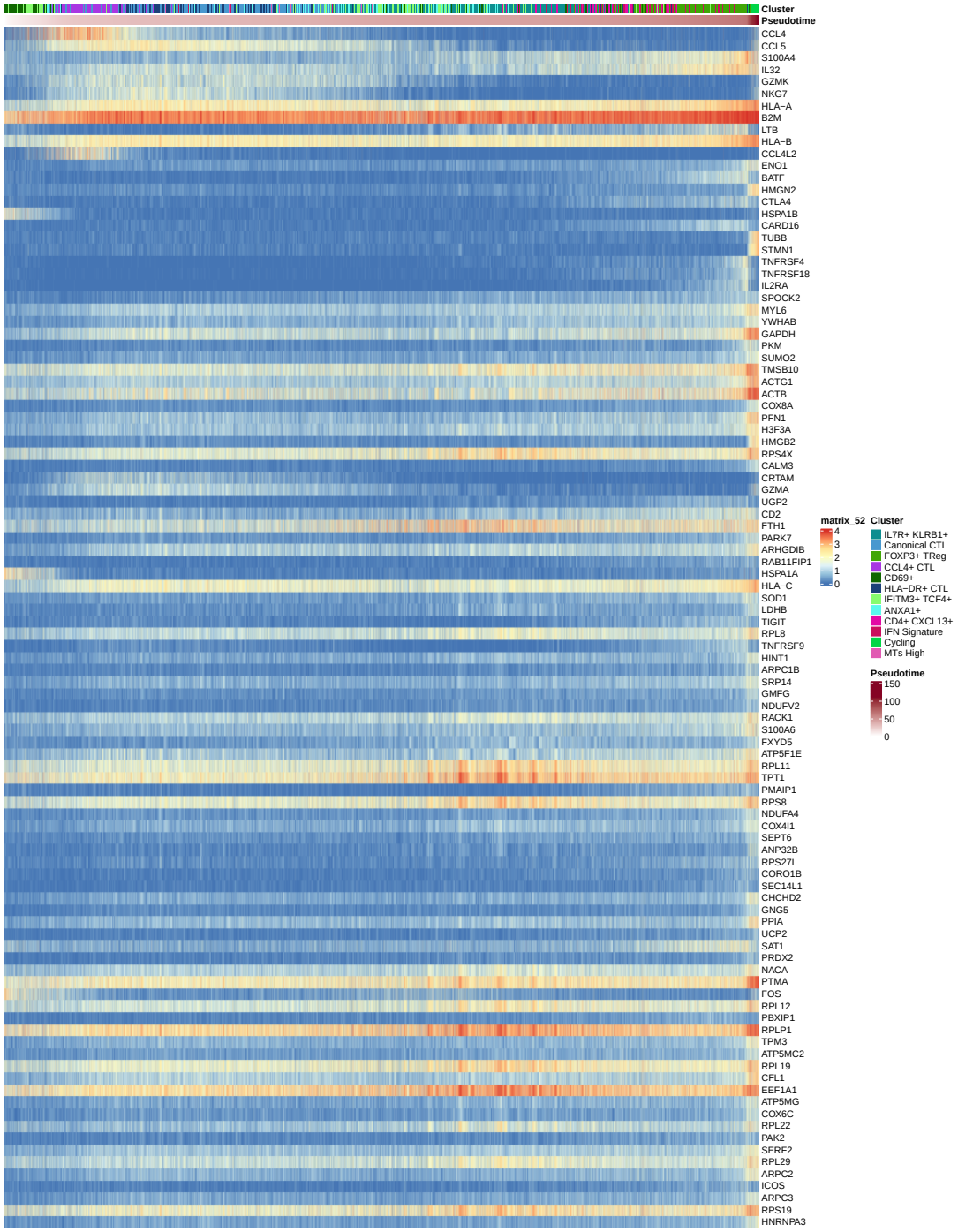


Figure D.2: Genes Varying Along Pseudotime.
Heatmap showing top 100 temporally expressed genes identified by fitting a generalised additive model using pseudotime as the independent variable. The RNA counts were log1p transformed.

D.14 Cellular Interactions

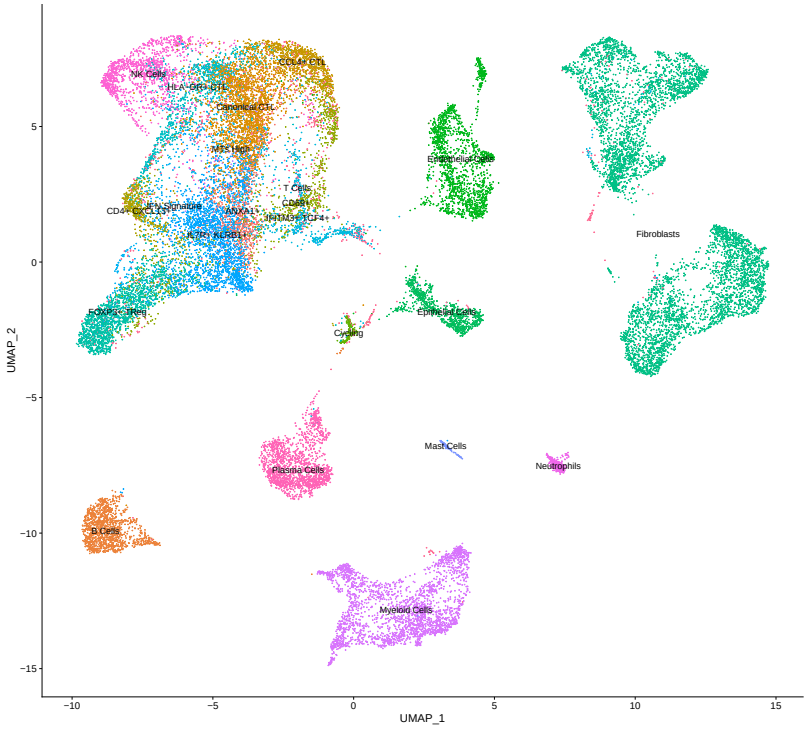
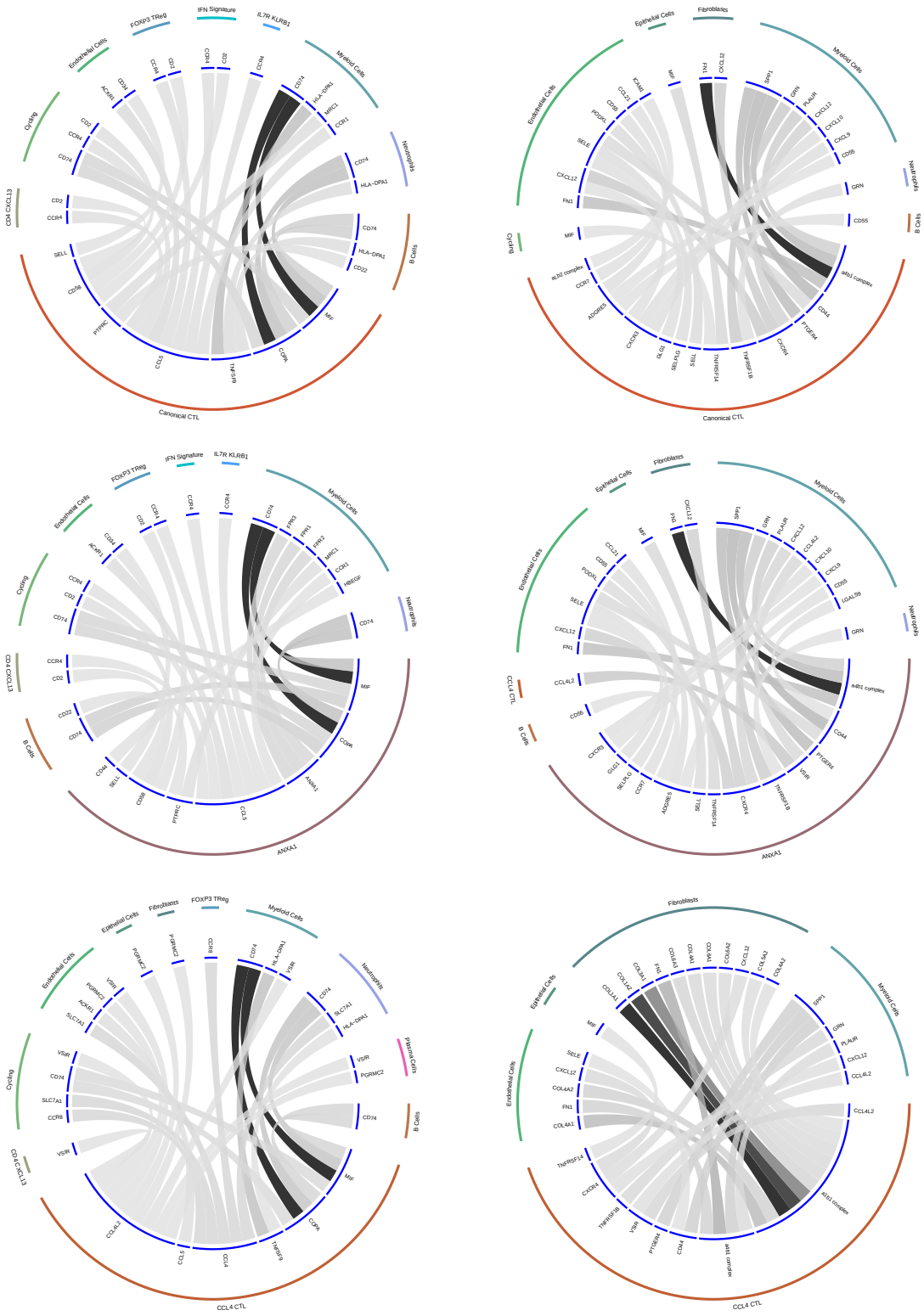
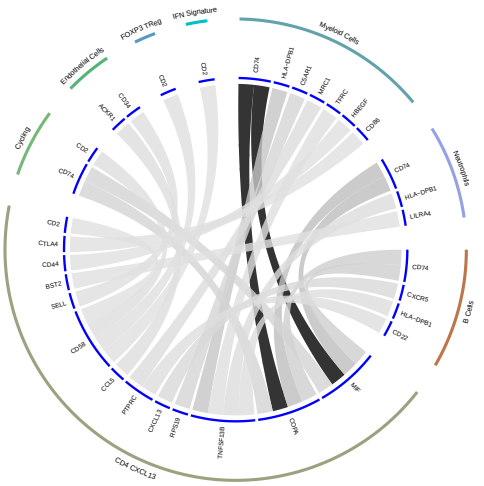
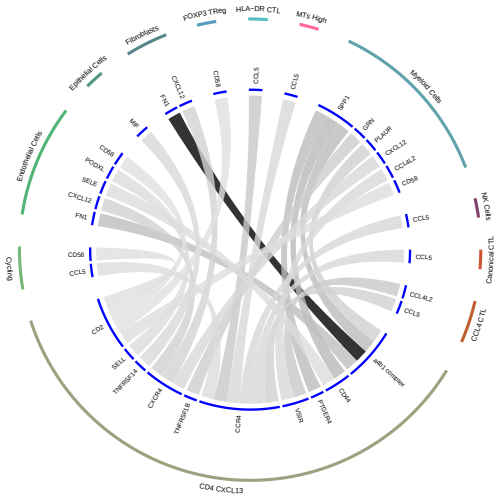
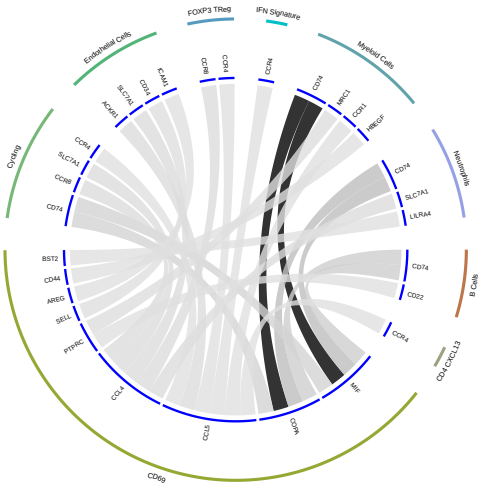


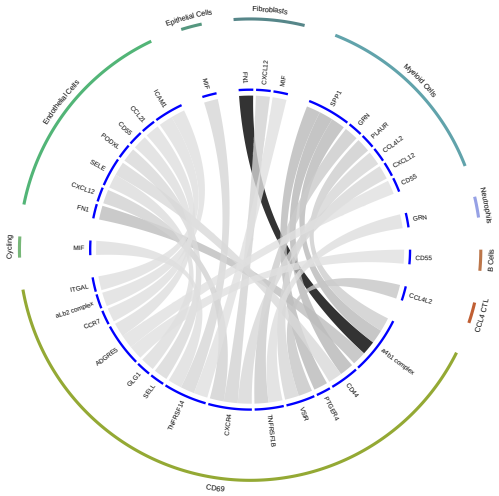
Figure D.3: T Cell Subpopulations and Major Cell Types.
UMAP plot of the integrated datasets representing cellular clusters coloured by cell type along with T cell subclusters.

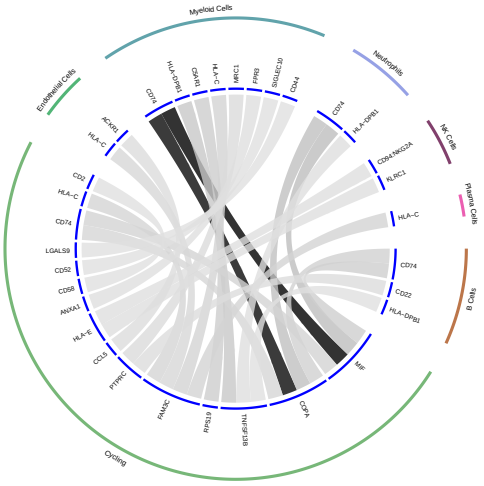

IFN Signature
IL7R KLRB1
Myeloid Cells
Neutrophils
B Cells
Canonical CTL
FOXP3 TReg
Epithelial Cells
Endothelial Cells
Cycling
CD4 CXCL13
Fibroblasts
Plasma Cells
CCL4 CTL
ANXA1

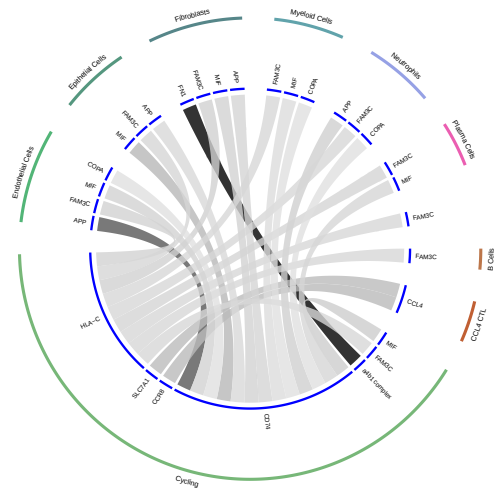

Myeloid Cells
Neutrophils
B Cells
CD4 CXCL13
Cycling
Endothelial Cells
FOXP3 TReg
IFN Signature
CD74
HLA-DPB1
CIITA
MRC1
PTPRC
HBEGF
CD96
CD7
CD5
CD34
ACKR1
CD2
CTLA4
CD44
BST2
SELL
CD28
CCL5
CXCL13
RPS19
TNFSF10B
CD4
CXCR5
CD22
LILRA4
Mif

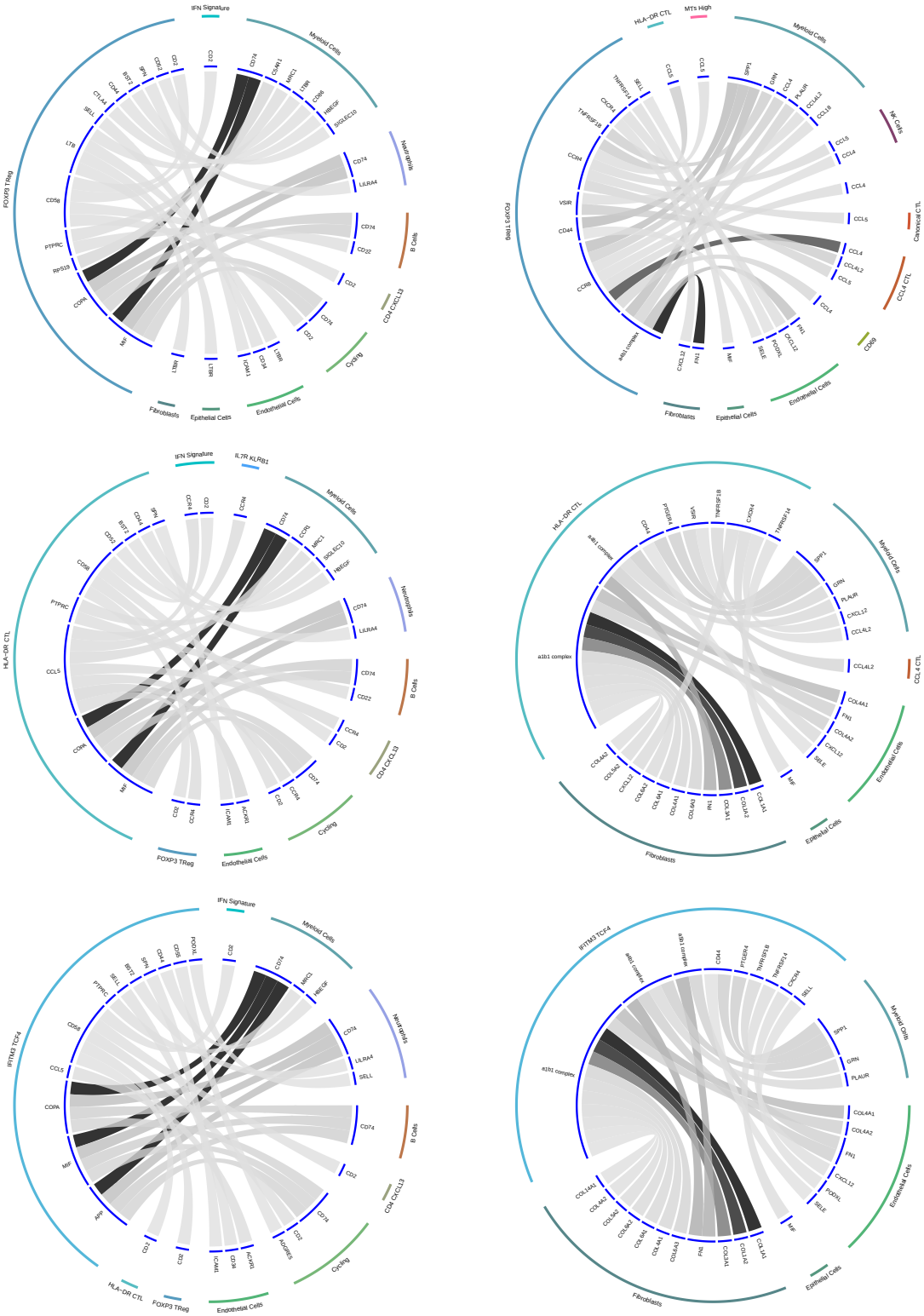

Myeloid Cells
NK Cells
Cytotoxic TLS
CCL4 CTL
CD4 CXCL13
Cycling
Endothelial Cells
Fibroblasts
FOXP3 TReg
HLA−DR CTL
MTs High
Spp1
GRN
PLAUR
CCL4L2
CXCL12
CD58
CCL5
FN1
MIF
CD55
PODXL
SELE
CD2
SELL
TNFRSF14
CXCR4
TNFRSF18
CCR4
LGR
PTGER4
CD4
adh1 complex


Myeloid Cells
Neutrophils
B Cells
CD4 CXCL13
CD69
Endothelial Cells
FOXP3 TReg
IFN Signature
CCR6
CD3
CD74
HLA-DPB1
CD34
ACKR1
SLC7A1
CD96
CD7
CD2
BST2
CD44
AREG
SELL
PTPRC
CCL5
CCL3
CD4
CD22
LILRA4
Mif


Myeloid Cells
Neutrophils
B Cells
CCL4 CTL
CD69
Epithelial Cells
Endothelial Cells
Fibroblasts
Cycling
Spp1
GRN
PLAUR
CCL4L2
CXCL12
CD58
CD55
ITGAL
itgb2 complex
CCR7
ADGRE5
SLC1
SELL
TNFRSF14
CXCR4
TNFRSF18
LGR
PTGER4
CD4
adh1 complex
FN1
CXCL12
Mif
ICAM1
CCL5
CD96
PODXL
SELE
CXCL12
FN1
MIF


Myeloid Cells
Neutrophils
NK Cells
Plasma Cells
B Cells
Cycling
Endothelial Cells
CD74
HLA-DPB1
CD4
SIGLEC10
CBLB
HLA-C
MRC1
PPIB
CD4-NKG2A
KLRC1
HLA-C
CD74
CD22
HLA-DPB1
Mif
CD74
HLA-DPB1
ACKR1
HLA-C
CD2
HLA-C
CD74
LGALS9
CD52
CD58
ANXA1
HLA-E
CCL5
PTPRC
FAM3C
RPS19
TNFSF10B
CD4


Fibroblasts
Myeloid Cells
Neutrophils
Plasma Cells
B Cells
CCL4 CTL
CD1
Cycling
Epithelial Cells
Endothelial Cells
FN1
FAM3C
MIF
APP
CD9A
FAM3C
APP
FAM3C
MIF
CD9A
MIF
FAM3C
APP
FAM3C
MIF
FAM3C
CCL4
MIF
FAM3C
adh1 complex
HLA−C
SLC7A1
CCR6


IFN Signature
Myeloid Cells
Neutrophils
B Cells
CD4 CXCL13
Cycling
Endothelial Cells
Epithelial Cells
Fibroblasts
FOXP3 TReg
HLA−DR CTL
MTs High
NK CD56
Cancercell CTL
CCL4 CTL
CD69
IL7R KLRB1
IFITM3 TCF4

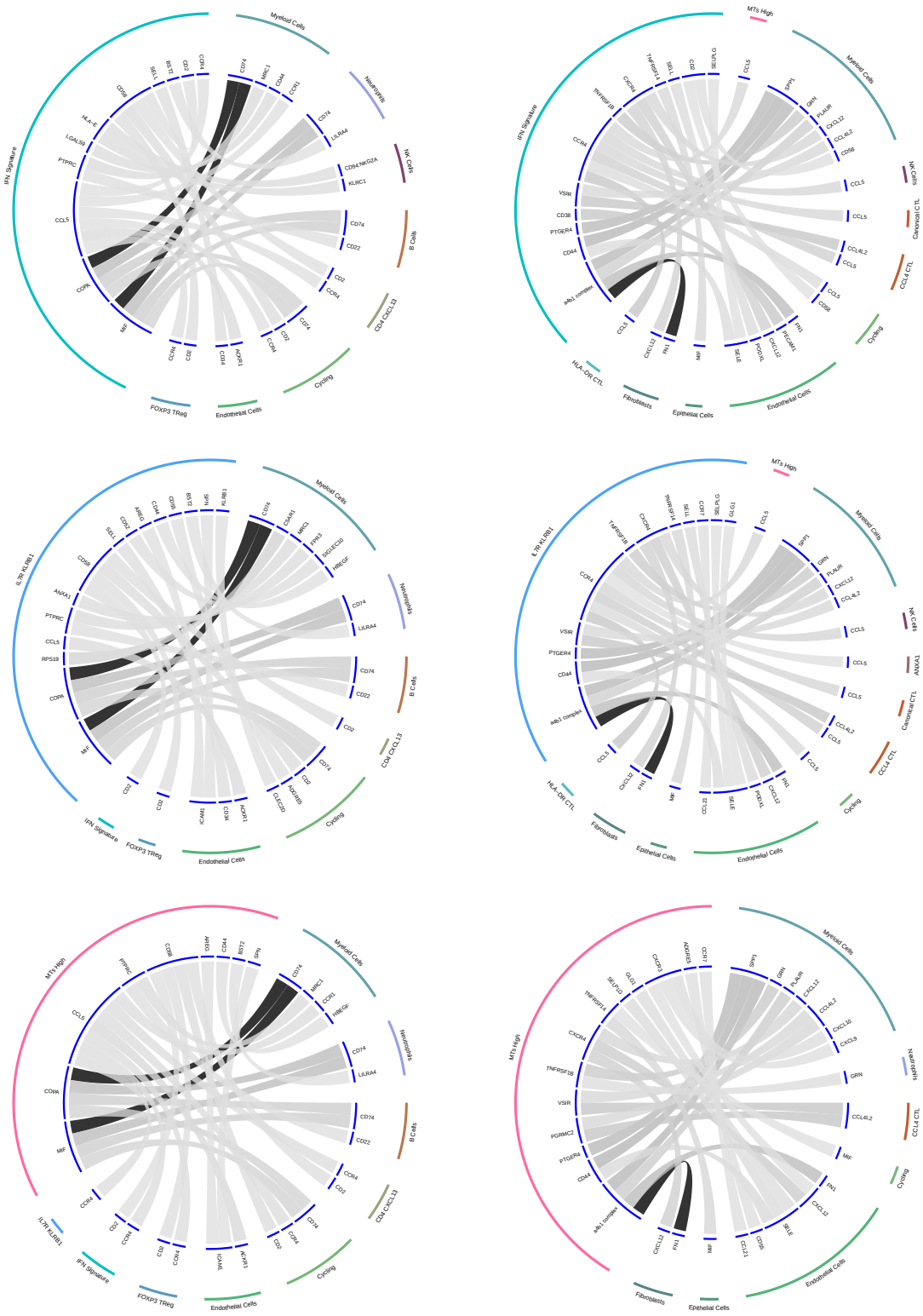


Figure D.3: Circle Plots of T Cell Interactions.
Top 25 outgoing (left) and incoming (right) interactions for each T cell cluster.

D.15 Slide-Seq Pucks

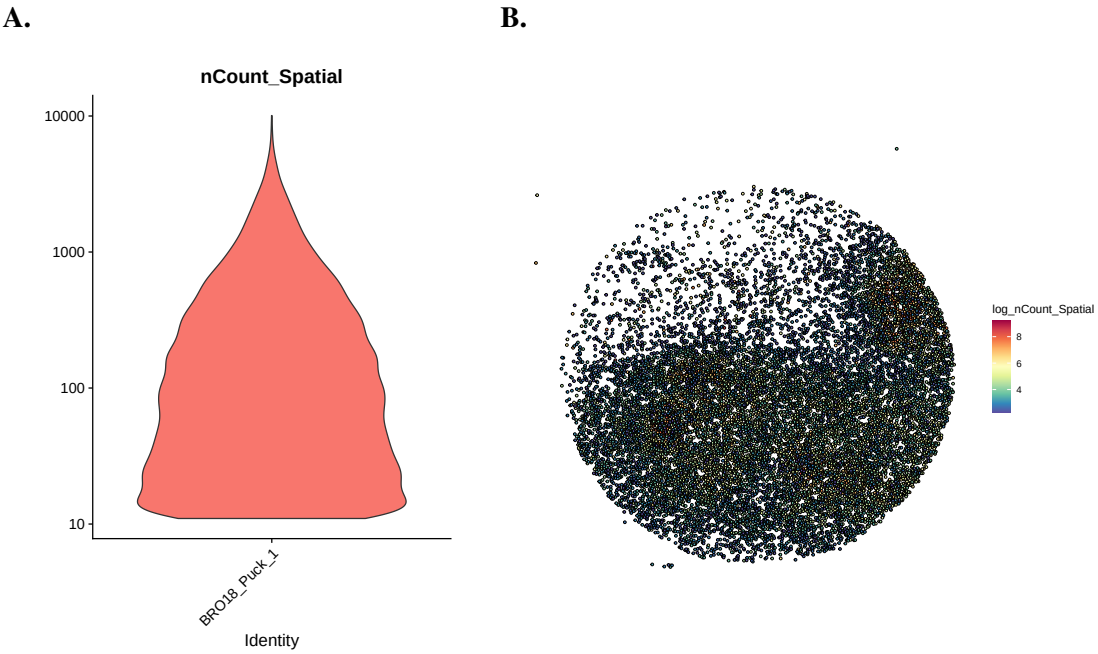


Figure D.4: Total Counts per Bead for Puck 1.

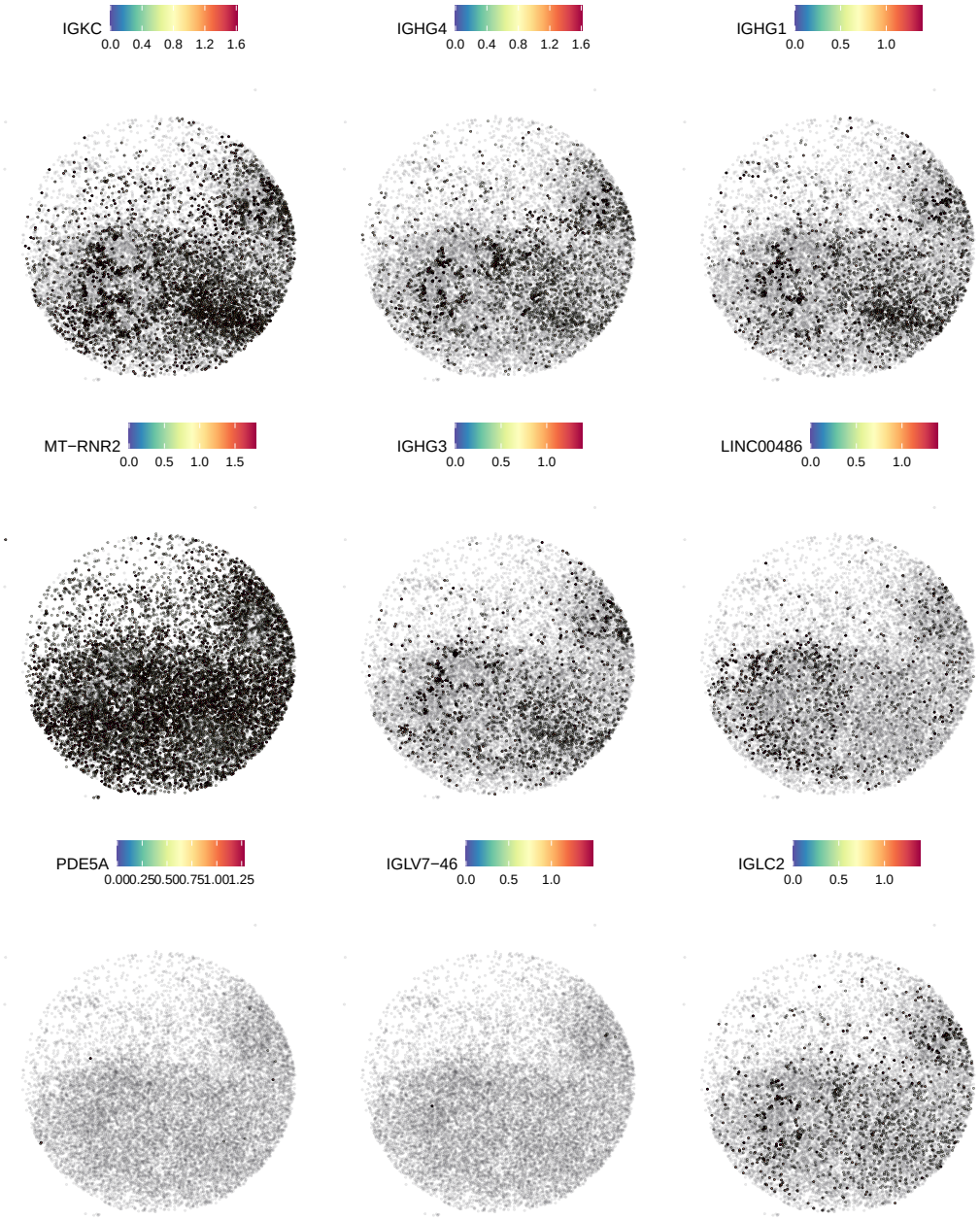
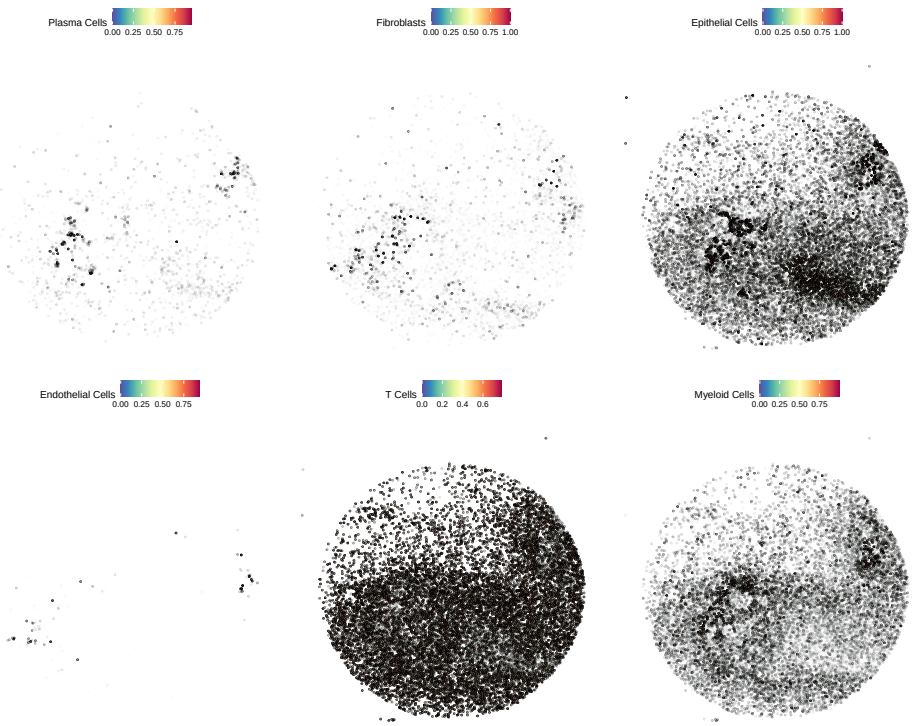


Figure D.5: Top Spatially Variable Features for Puck 1.

A.



B.

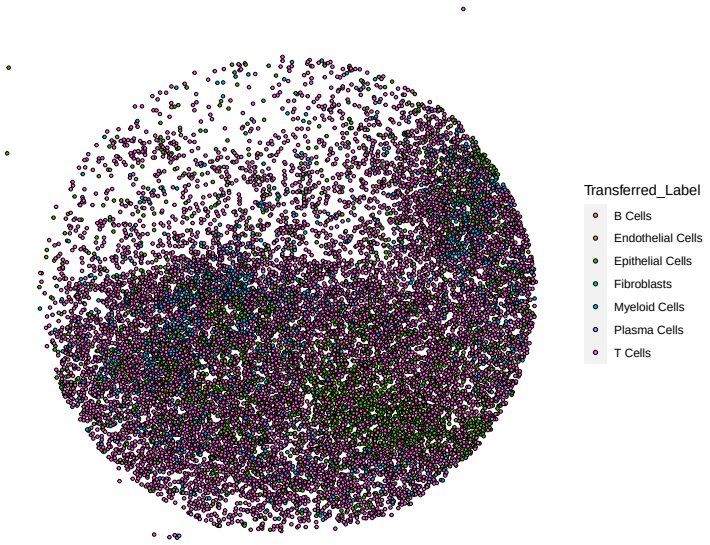


Figure D.6: Transferring scRNA-Seq Cell-Type Annotations Onto Puck 1.

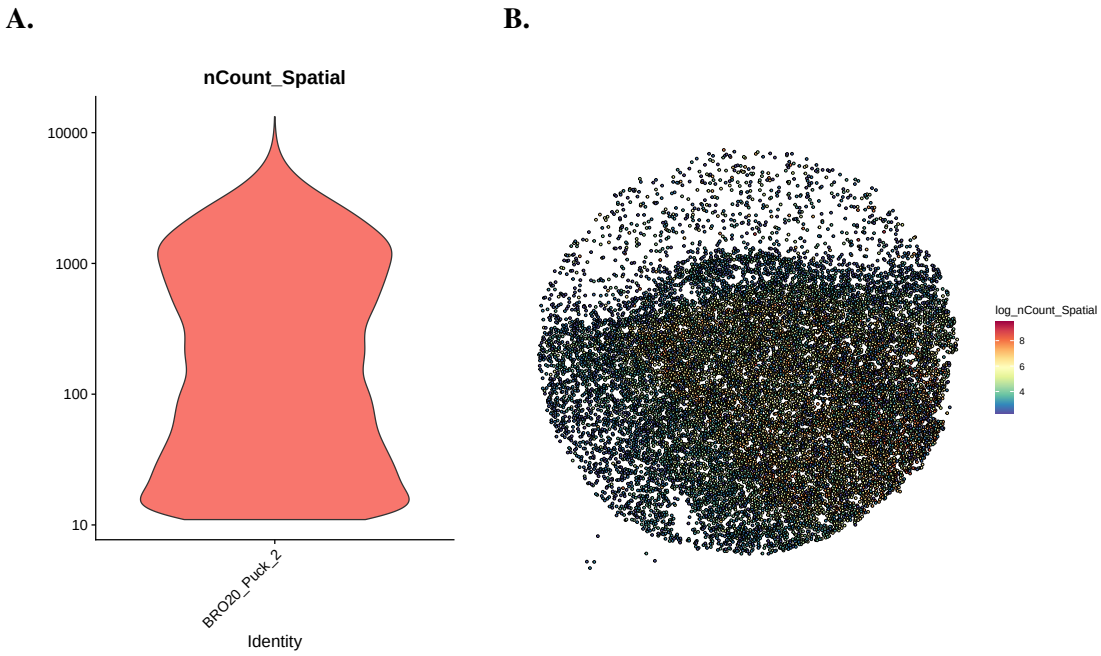


Figure D.7: Total Counts per Bead for Puck 2.

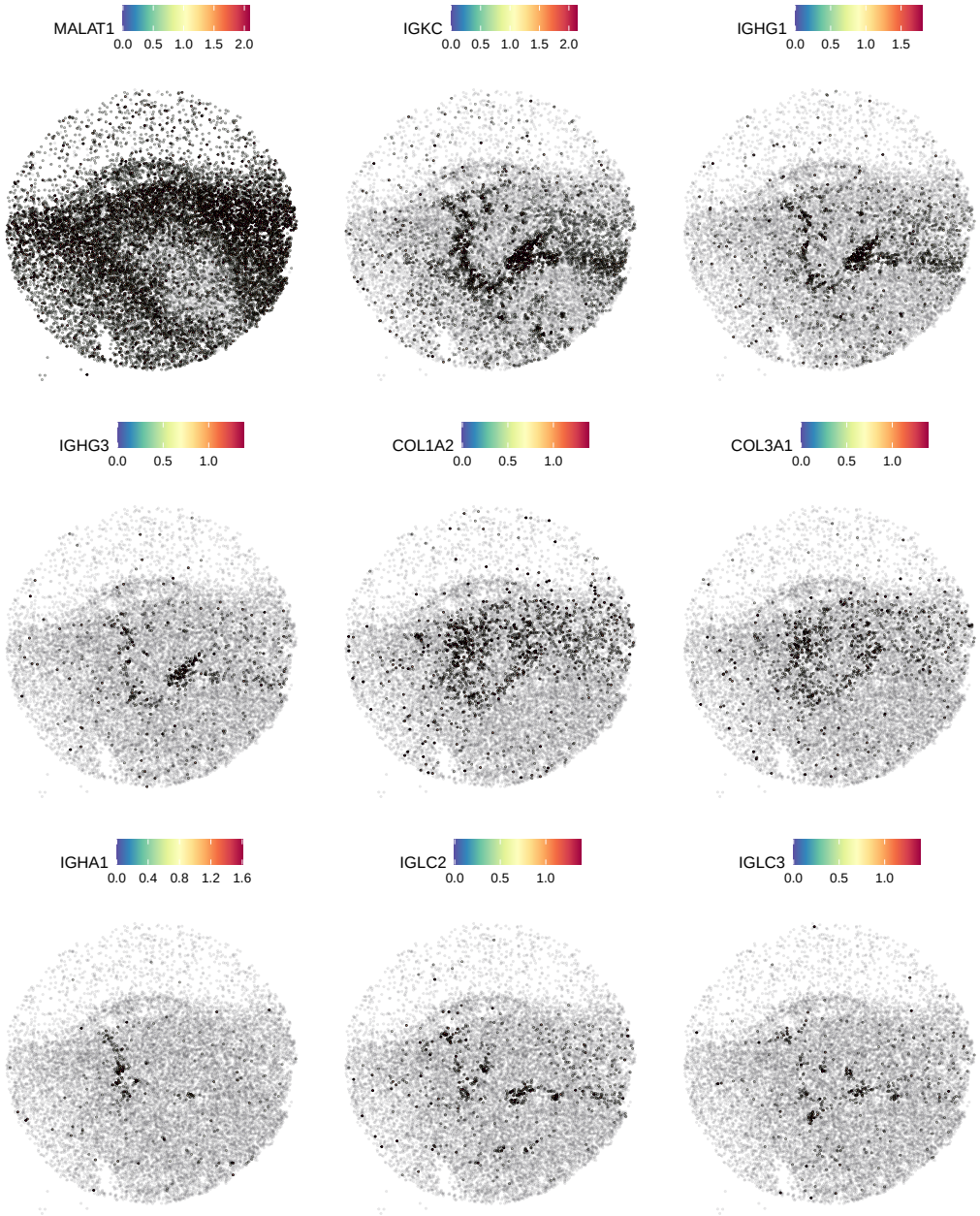
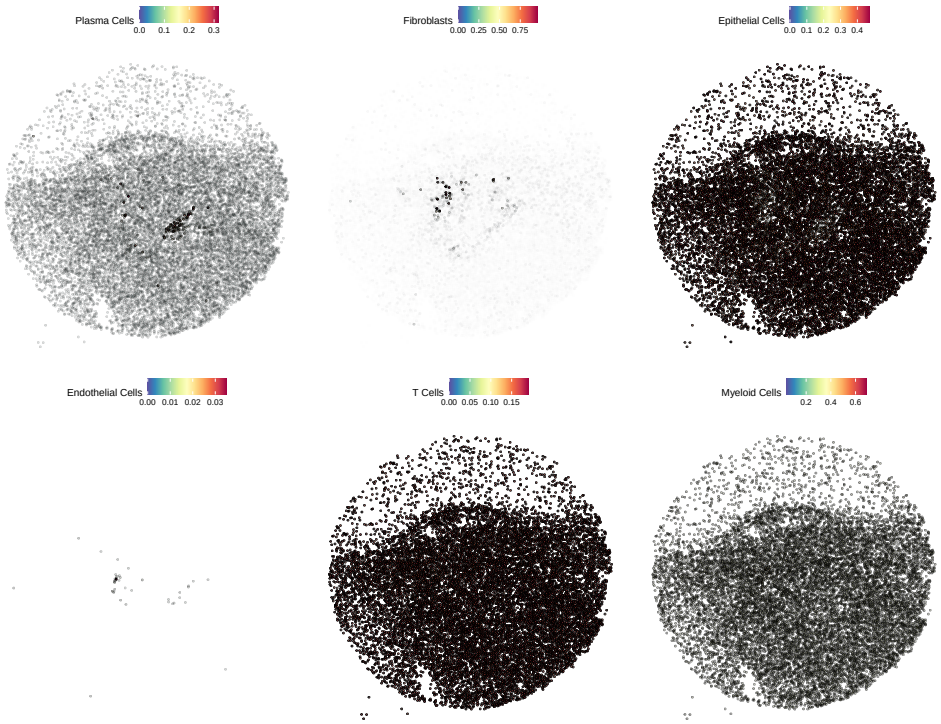


Figure D.8: Top Spatially Variable Features for Puck 2.

A.



B.

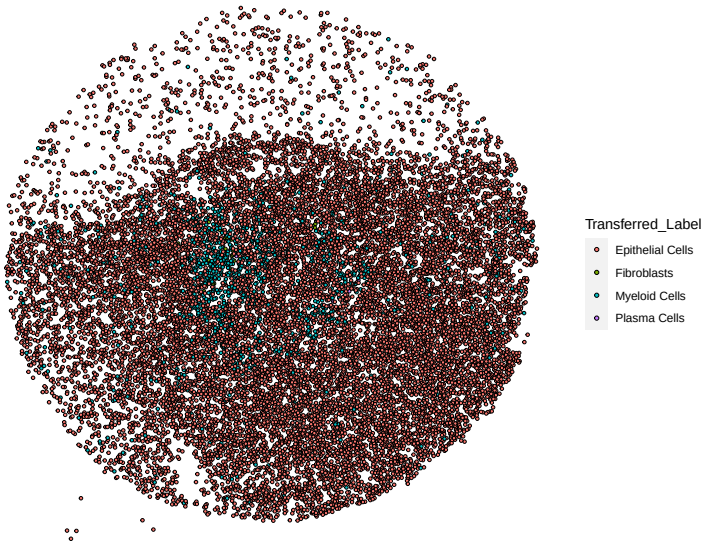


Figure D.9: Transferring scRNA-Seq Cell-Type Annotations Onto Puck 2.

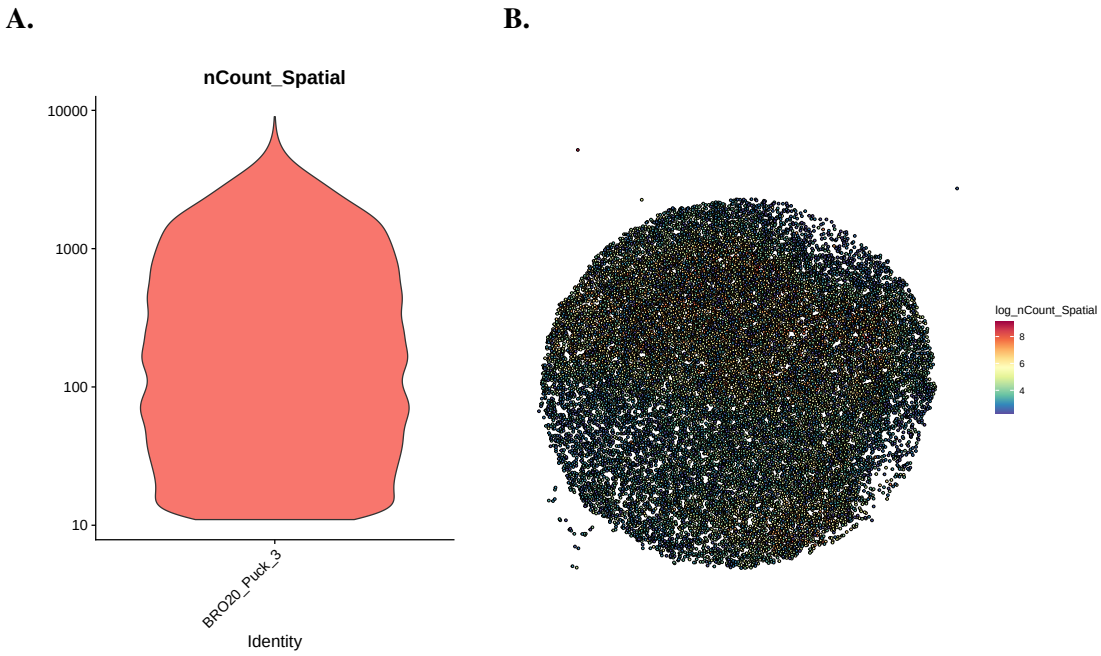


Figure D.10: Total Counts per Bead for Puck 3.

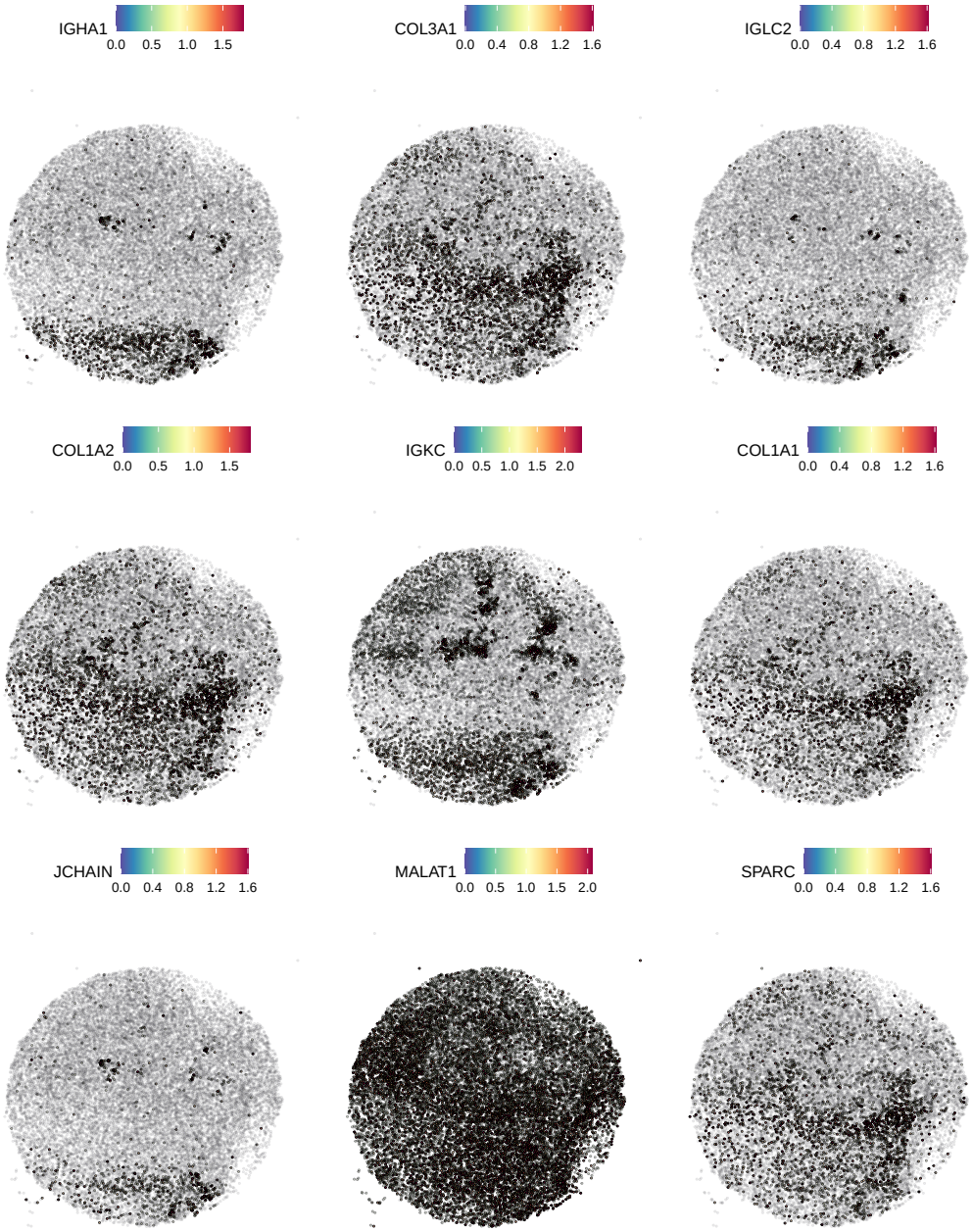
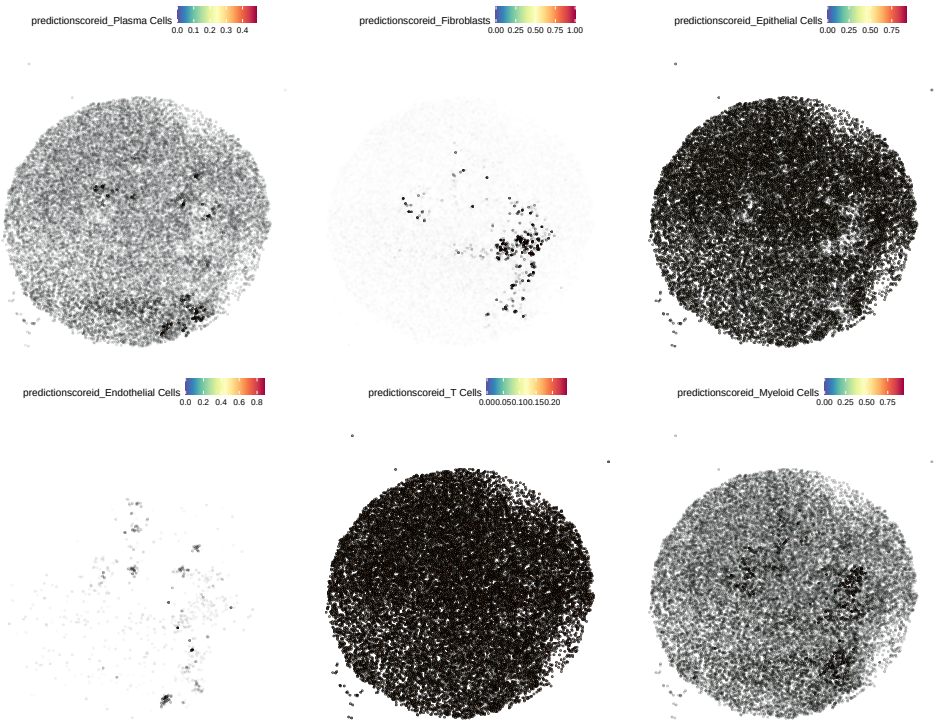


Figure D.11: Top Spatially Variable Features for Puck 3.

A.



B.

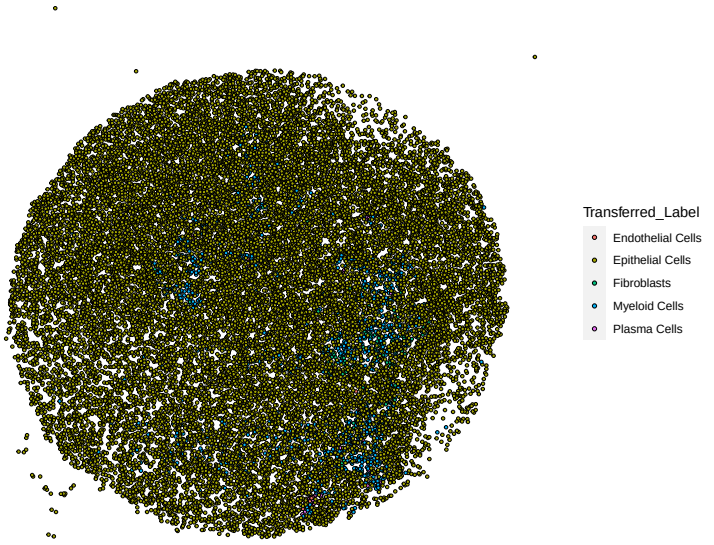


Figure D.12: Transferring scRNA-Seq Cell-Type Annotations Onto Puck 3.

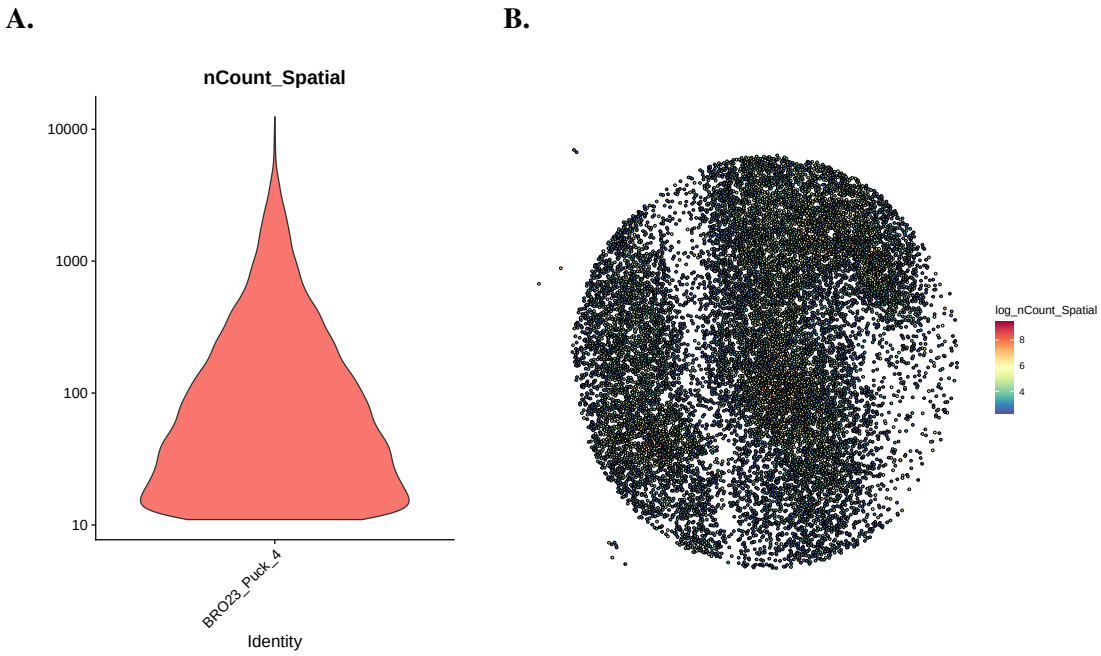


Figure D.13: Total Counts per Bead for Puck 4.

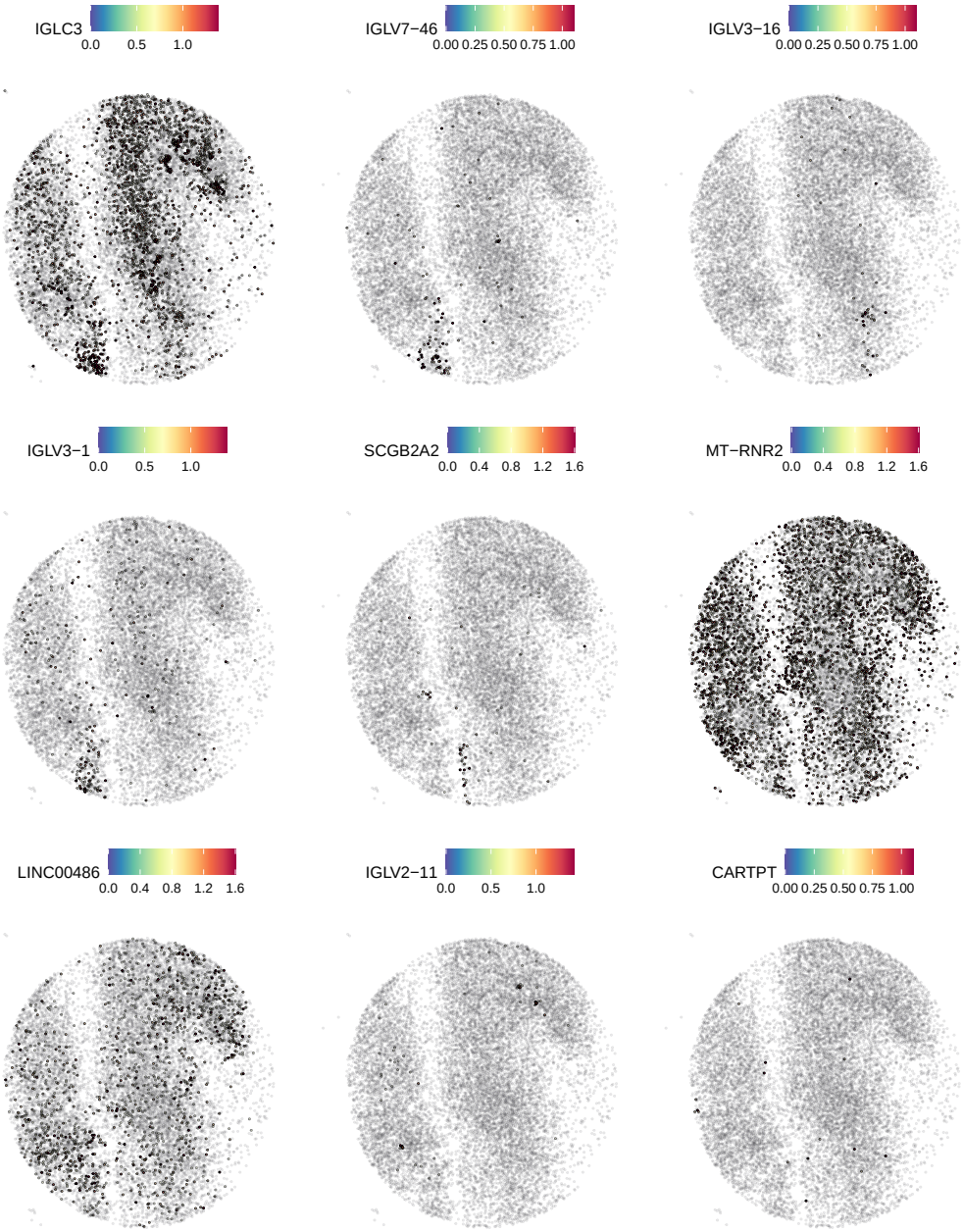
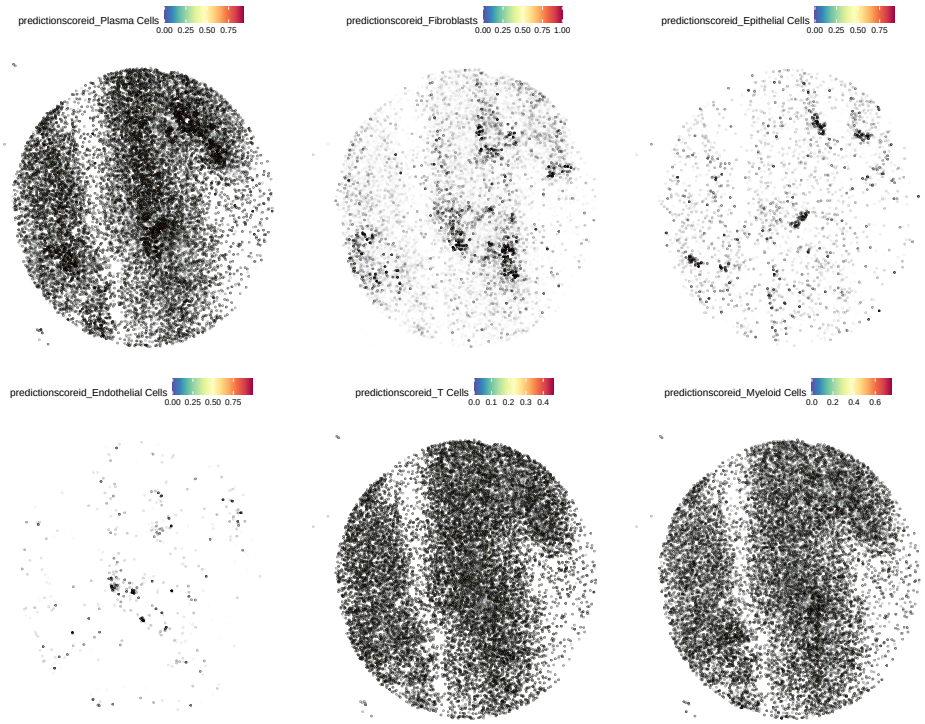


Figure D.14: Top Spatially Variable Features for Puck 4.

A.



B.

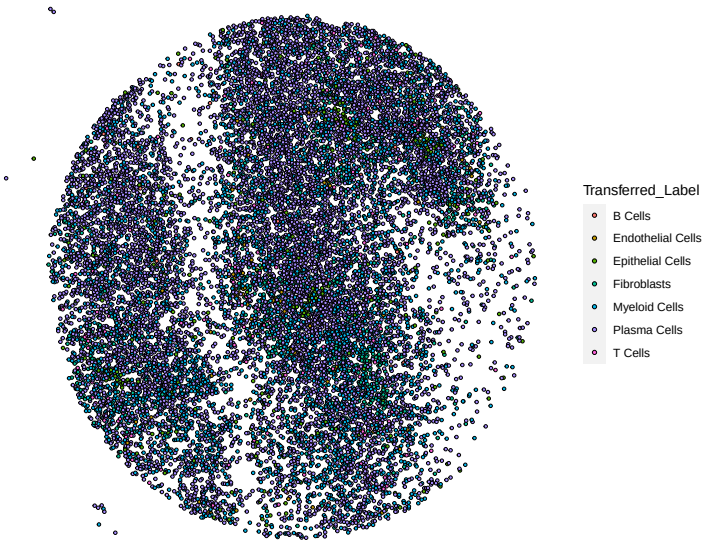


Figure D.15: Transferring scRNA-Seq Cell-Type Annotations Onto Puck 4.

D.16 Reference for Spatial Transcriptomics

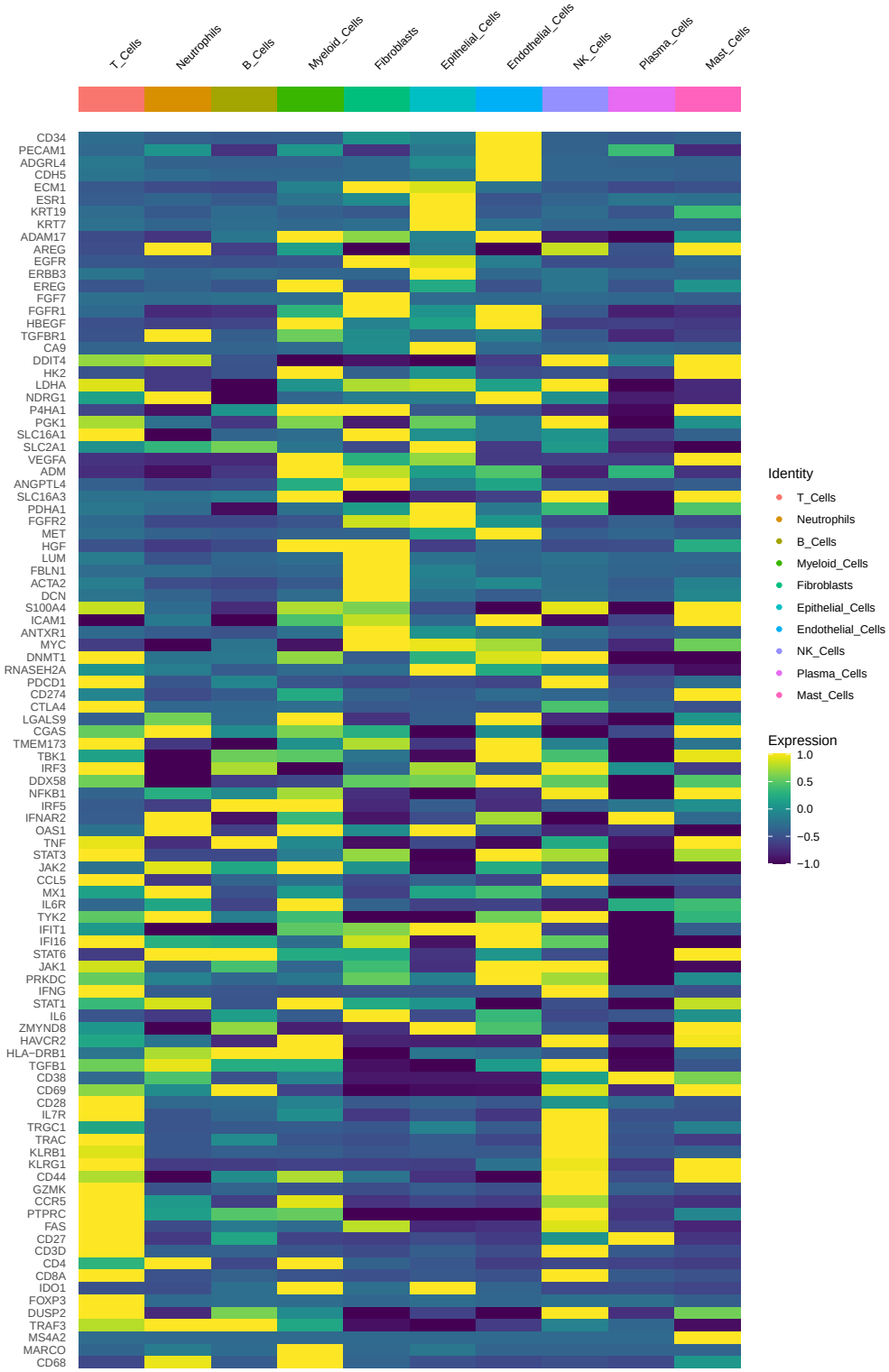
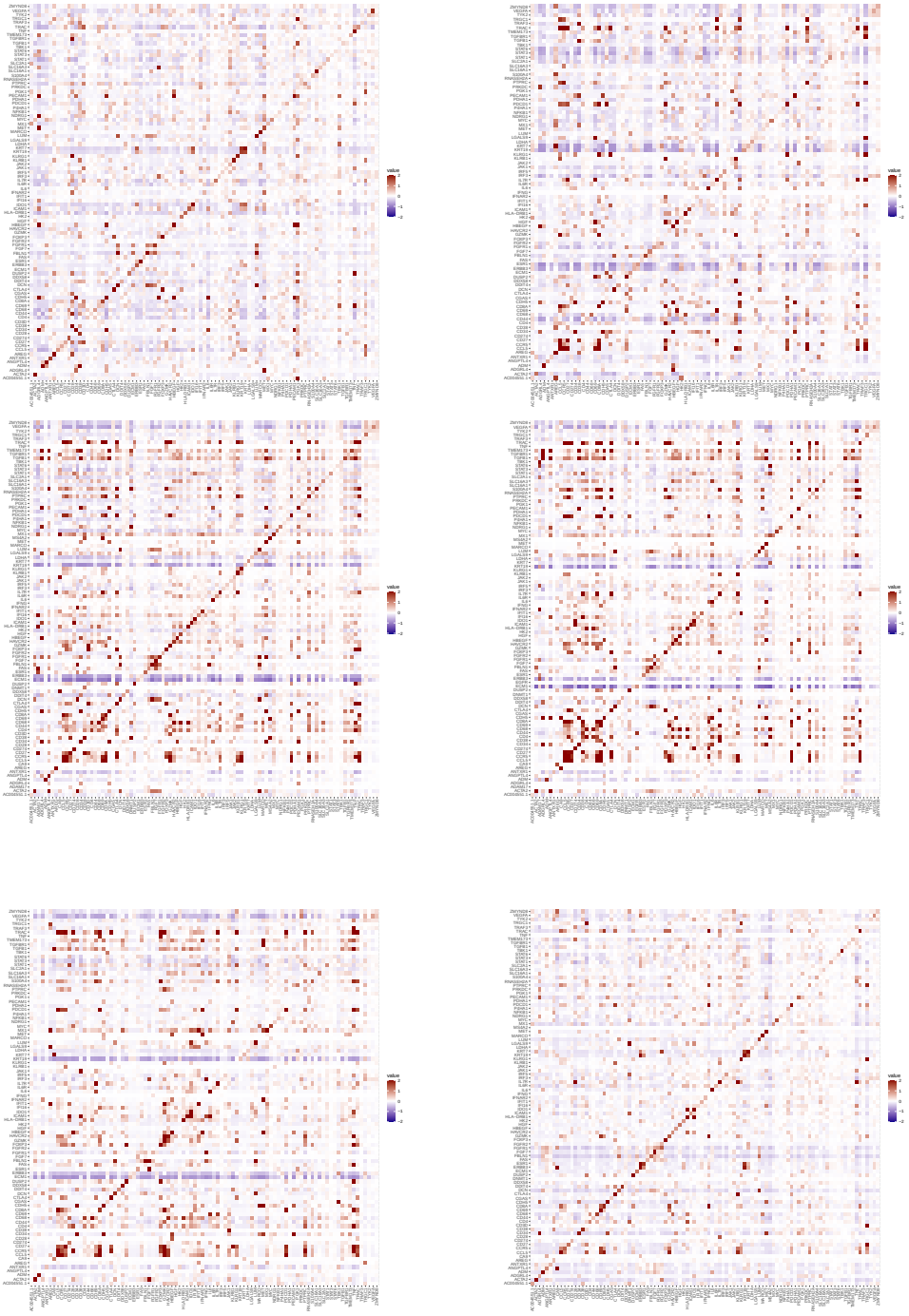
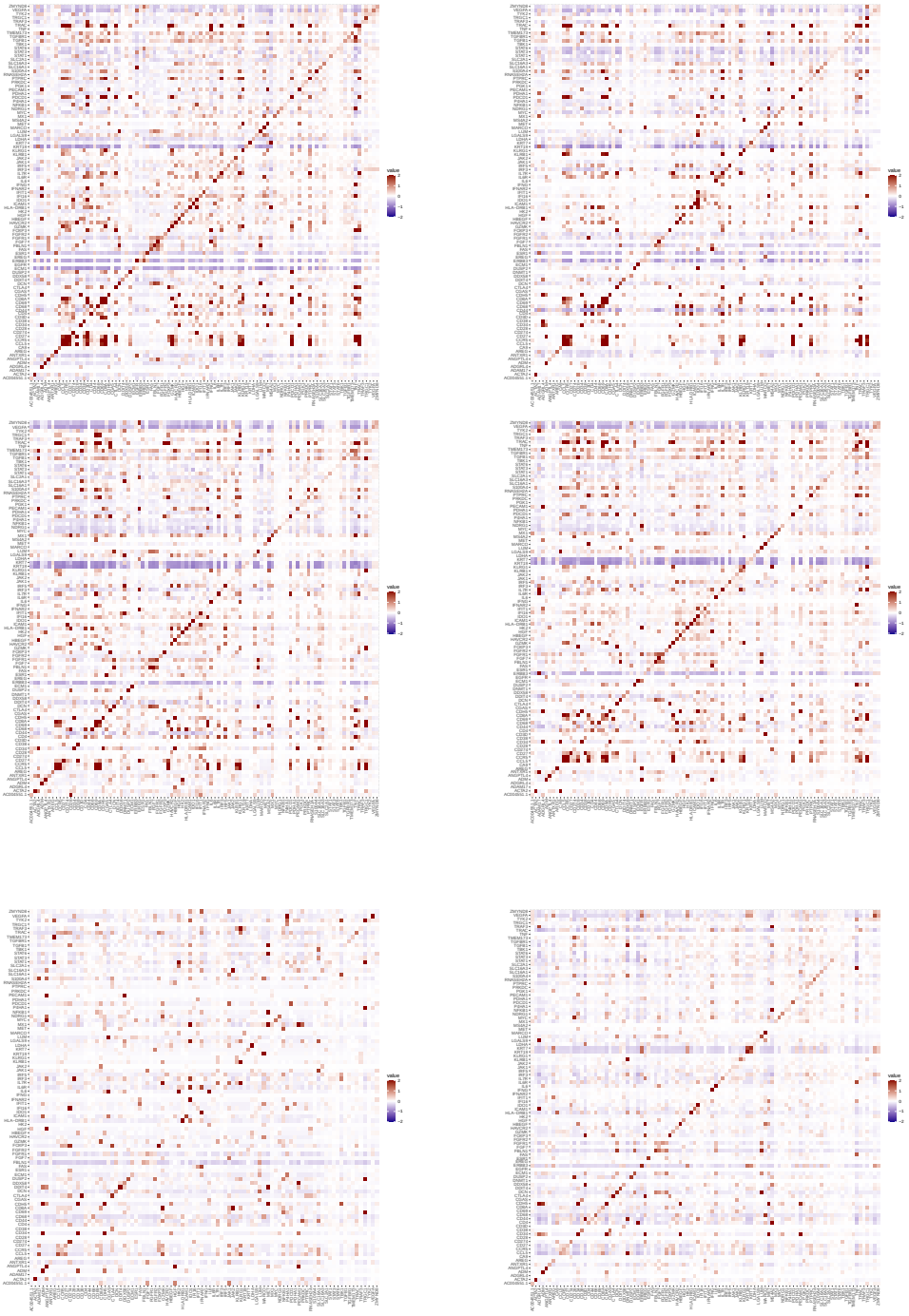


Figure D.16: Heatmap Displaying Average Expression of Spatial Probe Targets in scRNA-Seq

D.17 Colocalisation Matrices





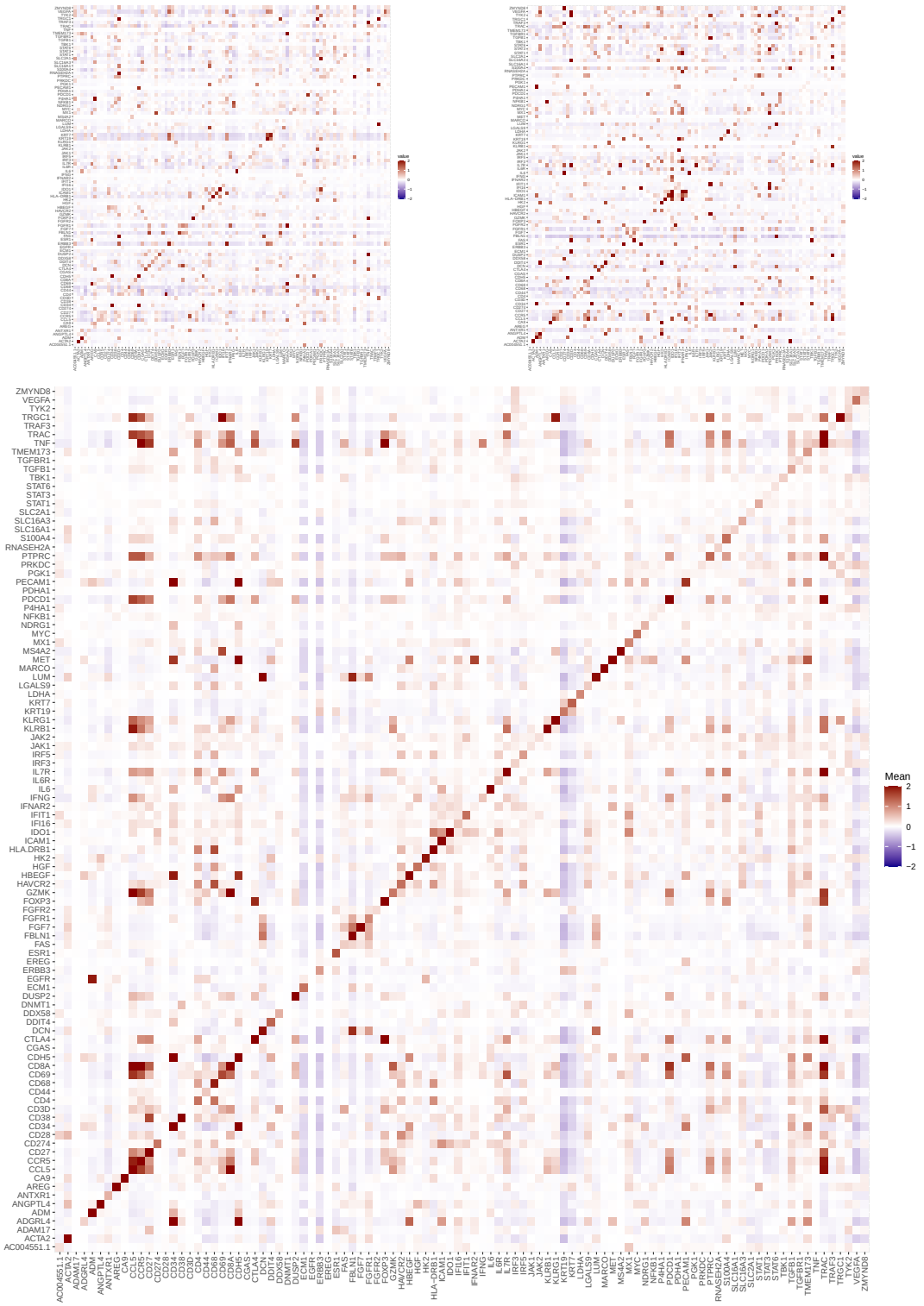


Figure D.16: Molecular Cartography Colocalisation Matrices.
A. A1-2. **B.** A1-3. **C.** A2-1. **D.** A2-2. **E.** A2-3. **F.** B1-1. **G.** B2-1. **H.** C2-1. **I.** D1-1. **J.** D1-2. **K.** D2-1. **L.** D2-2. **M.** D2-3. **N.** D2-4. **O.** Mean.

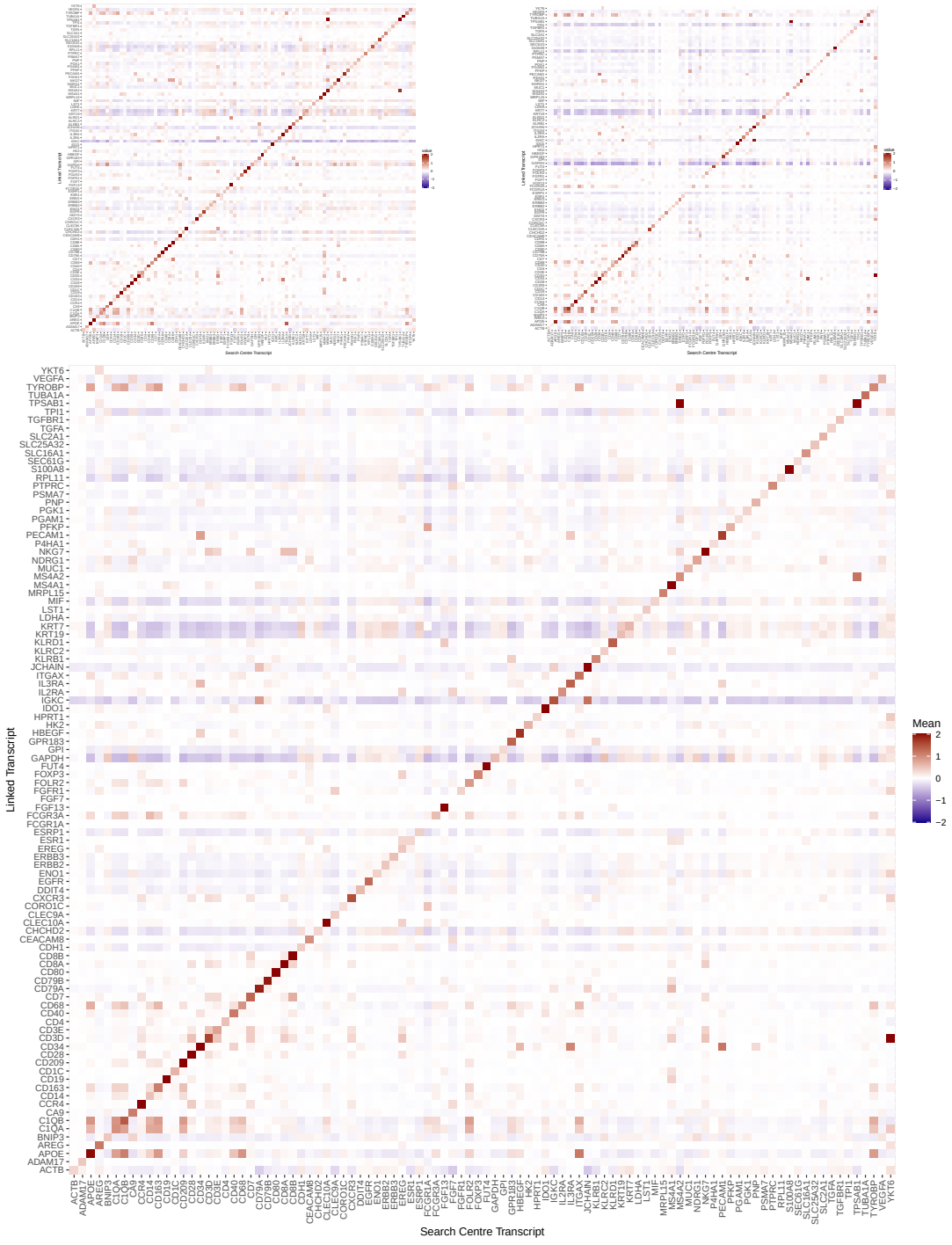


Figure D.16: CARTANA Colocalisation Matrices.
A. BBP1467. **B.** BBP1557. **C.** Mean.

APPENDIX E

Supplementary Material for Chapter 5

E.1 Multi-Agent System

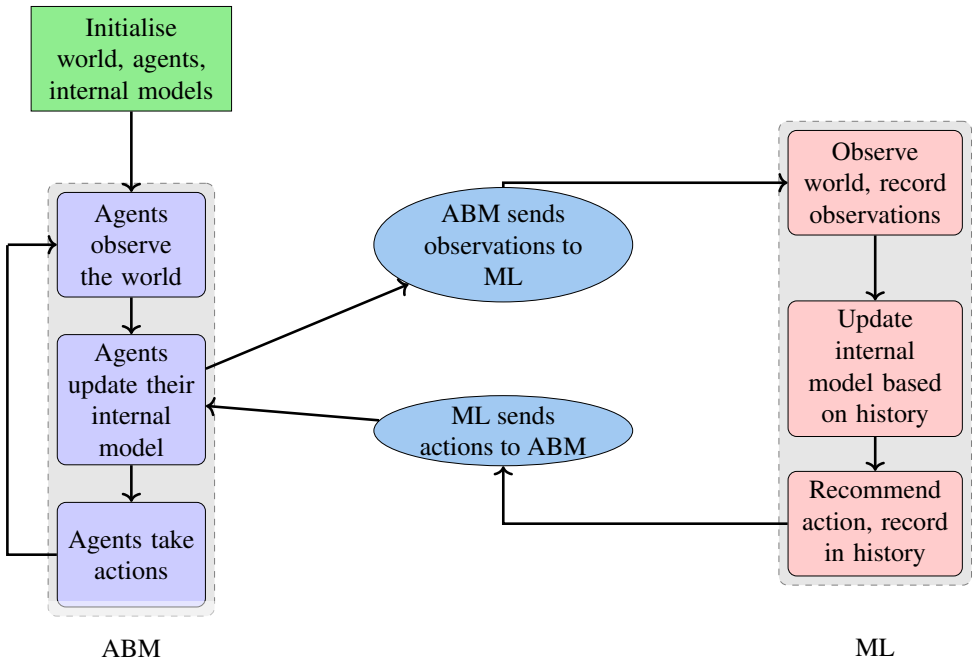


Figure E.1: Multi-Agent System Flow Chart.
High-level description of integrating machine learning (ML) algorithms into agent-based systems. Adapted from Rand (2006).

APPENDIX F

Source Code

Chapter 3 — https://github.com/CBigOxf/Identifying-T-Cells-scRNA

Chapter 4 — https://github.com/CBigOxf/TME-TNBC-T-IFN-Analysis

Chapter 5 — https://github.com/CBigOxf/Tumour-CAR-T-Cells

Appendix A — https://github.com/MBoemo/nanopredict

Appendix B — https://github.com/CBigOxf/Viral-Cancer-T-IFN-Analysis

