Determining the role of CD200 in Basal Cell Carcinoma

Alex Gibbs

Thesis submitted for the award of PhD

The European Cancer Stem Cell Research Institute

School of Biosciences

Cardiff University

February 2024

Ysgoloriaethau Sgiliau Economi Gwybodaeth
Knowledge Economy Skills Scholarships

UNDEB EWROPEAIDD
EUROPEAN UNION

Llywodraeth Cymru
Welsh Government

Cronfa Gymdeithasol Ewrop
European Social Fund

Acknowledgements

I would firstly like to thank KESS2 and Aneurin Bevan Health Board for the funding of my PhD project. Without this funding, I wouldn't be where I am today. I would also like to thank the patients who kindly donated samples for this project.

I would like to give a massive thanks to Girish Patel, who took me under his wing and realised my potential some 4+ years ago where I started in the lab as a MRes student. It was here that we developed a strong working relationship and friendship. Girish has always seen the potential in me which has helped in my transition to become a bioinformatician. I would also like to thank my second supervisor, Matt Smalley, who provided valuable input and feedback throughout my PhD.

Next, I would like to thank everyone at ECSCRI. ECSCRI welcomed me during my PTY with open arms and has provided a nurturing environment for me to thrive. Special thanks go to Mark and Jo, who have been like my second parents during my time there. Both have gone above and beyond for me and have been vital for the success of my project and for the running of ECSCRI. I cannot thank them enough for everything they have done for me! I would also like to thank Neil Rodrigues, who took me on for my PTY project and exposed me to the wonderful world of bioinformatics. If it was not for Neil, I wouldn't be where I am today. I also need to give a huge thanks to the other members of ECSCRI, in no particular order, Josh, Liam, Jasmine, Non, Katie, Adam, Anna, and nanny Cleo. You have all been, and still are, the best of friends during my time there. I would also like to thank past ECSRCI members, Geraint Parfitt, Carlotta Olivero, and Juan Bautista Menendez Gonzalez, who played a massive part in my scientific career and have been excellent mentors and supporters! Another thanks goes to Gemma and Charley, past GKP lab members who have helped me get through this PhD, I don't think I could have done it without you girls! And last but not least, a massive thanks goes to my colleague, best friend and fellow coffee snob, Huw Morgan. Huw has been here since day one, where we found a bond through the love of coffee. Huw has literally been there for me at every step of the project, and I cannot thank him enough for everything he has done for me!

I also need to thank those who have also provided me with their knowledge and expertise throughout my project. Ngoc-Nga Vinh donated her time and expertise which allowed me to perform the single cell RNA sequencing aspect of my project. Robert Andrews has provided me with his knowledge and expertise in anything and everything to do with bioinformatics analyses. It was Rob who took me under his wing during my PTY project and introduced me to bioinformatics. It was from here

where we developed a strong connection, and I cannot thank him enough for his input. Giusy Tornillo performed the cell line transduction, which was vital for our labs work, including my experiments for chapter 3. Boris Shorning helped me perform the xenograft experiments and passed a great deal of knowledge on to me during these times. He also looked after the mice for me and took care of all their needs, which allowed me to continue focusing on other experiments. Lastly, I would like to thank Adam Sawko who is the rep for Novogene. Adam has worked with me since I started my project and has provided me with excellent service and advice. We have developed a great friendship as a result.

Finally, to my family, I cannot thank you enough for all the love and support you have given me since I started this journey. I first need to mention my best friend and life-long partner, Sophie. You have been my rock since day one. You have kept me grounded and have stopped me losing my mind. Thank you for putting up with me for all this time. Another shout out to my boy (dog), Marlo, who has provided me with endless hugs and kisses and loves me even if I am grumpy after a long day of writing. I want to give a big thanks to Mam, Dad, our family dog Millie, Nanna, and Daddo, who have been my biggest supporters since day one. Lastly, I want to thank my sister, Gemma, and Sophies family, Karen, Spencer, and Marilyn for supporting me. I honestly don't think I could have made it without any of you. Love you always!

Summary

Background

CD200 is an immunomodulatory protein that is expressed on the cell membrane of hair follicle bulge stem cells and basal cell carcinoma (BCC) cancer stem cells, consistent with the literature suggesting that human BCC arise from hair follicle bulge stem cells. The “driver” mutations in human BCC activate the hedgehog (Hh) signalling pathway. Similarly, growth factor pathway activation is common in cancer; notably Hh pathway activation is observed in one-quarter of all cancers.

The recent development of targeted cancer therapies, typically small-molecule drugs that block growth factor activation, represents a great advance in cancer management. Yet targeted therapies, such as smoothened antagonists for the treatment of BCC, are beset by primary and secondary resistance, suggesting the presence of additional cell survival signalling pathways.

Hypothesis

Given primary and secondary resistance is seen with smoothened antagonists in BCC, I hypothesised that cancer stem cells facilitate this resistance via additional pathway activation.

Aims

1. Determine whether CD200 is a functionally relevant marker for the isolation and function of BCC cancer stem cells.

2. Use RNA sequencing to elucidate cell signalling pathways specific to the BCC cancer stem cell-enriched population.

Methods

To assess functional relevance of CD200 to CSCs, CD200-transduced HeLa and UW-BCC1 cells were used *in vivo* to assess whether CD200 causes cancer immune cell evasion. Flow sorting followed by bulk and single-cell RNA sequencing approaches were used to define BCC cancer stem cells and signalling pathways in UW-BCC1 and primary human BCC samples. Immunohistochemistry was used to validate signalling pathway cascades identified by sequencing in primary human BCC.

Results

CD200 did not alter tumour growth rates but led to a significant reduction in tumour necrosis, immune cell infiltrate, and NK cell viability. Hence the CD200+ BCC subpopulation aids immune evasion and is therefore a functionally relevant cell surface marker to BCC CSCs.

Bulk RNA sequencing comparing flow-sorted CD200+ and CD200- CD45- primary human BCC subpopulations revealed upregulation of endothelial cell processes, including VEGF signalling. The VEGF pathway differentially expressed genes included those predicated to be expressed on both tumour and endothelial cells, confirmed by VEGFR2 (KDR) immunohistochemical labelling. Single-cell RNA sequencing showed that up to 44% of flow-sorted CD200+ CD45- primary human BCC tumour dissociate were endothelial cells. Isolating the BCC keratinocytes, we identified 17 clusters based on transcriptomic similarities, which we were able to separate into three

functional categories. All BCC keratinocytes expressed Hh pathway-regulated genes. The CD200 high expressing “basal” BCC population demonstrated active TGFβ signalling, which we validated by immunohistochemical labelling to identify phosphorylated SMAD3 nuclear translocation in BCC tissues. This suggests that BCC CD200+ CSCs upregulate the TGFβ pathway which may facilitate metastasis and treatment resistance.

Conclusions

CD200 functions to protect BCC cancer stem cells from immune attack. The BCC cancer stem cell-enriched CD200+ CD45- tumour population, which included endothelial cells, is enriched for VEGF-mediated angiogenesis. This enrichment, however, may be emanating from the endothelial cell contamination of the dataset. Specifically, BCC cancer stem cells may use TGFβ signalling for survival and intrinsic resistance to Hh pathway antagonists.

Word Count: 44,410

Table of Contents

Chapter 1---------- 1
1.1 Hallmarks of Cancer---------- 1
1.1.1 Sustaining Proliferative Signalling 5
1.1.2 Evading Growth Suppressors 5
1.1.3 Tissue Invasion and Metastasis 6
1.1.4 Limitless Replicative Potential 6
1.1.5 Sustained Angiogenesis 7
1.1.6 Evading Apoptosis 8
1.1.7 Avoiding Immune Destruction 9
1.1.7.1 The Cancer-Immunity Cycle 9
1.1.7.2 Cancer Immunoediting 13
1.1.7.2.1 Elimination 13
1.1.7.2.2 Equilibrium 13
1.1.7.2.3 Escape 14
1.1.7.2.4 Mechanisms of Immune Escape 17
1.1.7.2.4.1 Decreasing of Immune Recognition 17
1.1.7.2.4.2 Immune Resistance 18
1.1.7.2.4.3 Immunosuppressive Tumour Microenvironment 18
1.1.7.2.4.4 Negative Co-Stimulatory Pathways 19
1.3 Tumour Initiating Cells ----------20
1.3.1 Tumour Propagation Models 21
1.3.1.1 The Cancer Stem Cell Model 21
1.3.1.2 Stochastic Model 22
1.3.1.3 Alternative Models 22
1.4 Carcinoma ----------25
1.5 Skin Biology ----------25
1.6 Hair follicle ----------27

1.7 Cutaneous Basal Cell Carcinoma ----------30
1.7.1 Incidence 30
1.7.2 Clinical Presentation 31
1.7.3 Aetiology 38
1.7.3.1 Age 38
1.7.3.2 Previous History of BCC 38
1.7.3.3 Genetic Factors 38
1.7.3.3.1 Inheritance 38
1.7.3.3.2 Genetic Predisposition 39
1.7.3.3.3 Genetic Syndromes 39
1.7.3.4 Ultraviolet Radiation 42
1.7.3.5 Medications 43
1.7.3.5.1 Photosensitizing Medications 43
1.7.3.5.2 Immunosuppressive Agents 43
1.7.3.6 Ionizing Radiation 44
1.7.3.7 Other Occupational Risks 45
1.7.3.8 Arsenic Exposure 45
1.7.3.9 Skin Type and Gender 46
1.7.3.10 Pre-Existing Skin Disease 47
1.7.3.11 Vitamin D 47
1.7.4 Pathogenesis of BCC 48
1.7.4.1 Hedgehog Signalling Pathway 49
1.7.4.2 Driver DNA mutations in BCC 53
1.7.4.3 Mouse models of BCC 53
1.7.4.4 SMO antagonists in the treatment of BCC 54
1.7.4.5 DNA sequencing of BCC 55
1.7.4.6 RNA sequencing of BCC 56
1.7.4.7 Single Cell RNA Sequencing of BCC 57

1.7.4.7.1 BCC tumour cell heterogeneity 57
1.7.4.7.2 BCC tumour-associated Fibroblasts 62
1.7.4.7.3 T-cell response in BCC 63
1.7.4.8 Identification of BCC Cancer Stem Cells 63
1.8 Hypothesis 64
1.8.1 Aims 65
Chapter 2 66
2.1 Cell Culture 66
2.1.1 Established Cell Lines 66
2.1.2 Maintenance of Cell Lines 68
2.1.2.1 Thawing Cell Lines 68
2.1.2.2 Sub-Culturing Cell Lines 68
2.1.2.3 Cryopreserving Cell Lines for Storage 69
2.1.2.4 Transduction of Cell Lines 70
2.1.3 Colony Forming Cell Assays 70
2.2 Immunohistochemistry 71
2.3 Immunofluorescence 72
2.4 Flow Cytometry 74
2.5 qRT-PCR 75
2.6 Image Analysis 77
2.6.1 Colony Forming Cell Assays 77
2.6.2 Immunohistochemical Image Analysis 77
2.7 Mouse Experiments 81
2.8 Human Tissue Samples 81
2.8.1 Tissue Dissociation 82
2.8.1.1 Patel Lab BCC Dissociation Protocol 82
2.8.1.2 MACS Diccociation Protocol 83
2.8.2 Tissue Freezing, Cryosectioning, and Staining 84

2.9 Statistical Analyses ---- 84
2.10 Sequencing Sample Preparation ---- 85
2.10.1 Bulk RNA Sequencing ... 85
2.10.2 Single-Cell RNA-sequencing ... 85
2.11 Sequencing Data Processing ---- 86
2.11.1 Bulk RNA-sequencing ... 86
2.11.2 Single-Cell RNA Sequencing ... 87
2.12 Bioinformatics Analyses ---- 88
2.12.1 Bulk RNA-sequencing ... 88
2.12.2 Single-Cell RNA-sequencing ... 88
2.12.2.1 Optimisation of Analysis Protocol ... 88
2.12.2.2 Quality Control ... 93
2.12.2.3 Normalisation ... 94
2.12.2.4 Cell Cycle Scoring ... 95
2.12.2.5 Highly Variable Gene Selection ... 96
2.12.2.6 Scaling ... 97
2.12.2.7 Linear Dimensionality Reduction ... 98
2.12.2.7.1 Principal Component Analysis ... 99
2.12.2.8 Clustering ... 106
2.12.2.9 t-Stochastic Neighbour Embedding (t-SNE) ... 108
2.12.2.10 Uniform Manifold Approximation and Projection (UMAP) ... 109
2.12.2.11 Cell Type Identification ... 110
2.12.2.12 Dataset Integration ... 111
2.12.2.12.1 Seurat v3 Integration ... 112
2.12.2.12.2 scTransform Normalisation ... 114
2.12.2.12.3 Harmony Integration ... 116
2.12.2.12.4 Integration Method Testing ... 117
2.12.2.13 Downstream Analysis of Integrated Data ... 118

2.12.2.13.1 Calculating the Percentage of Cells Expressing Marker Genes in Each Cluster for Each Sample 118
2.12.2.13.2 Copy Number Variation Analysis 118
2.12.2.13.3 RNA Velocity Analysis 120
2.12.2.13.4 Single Cell GSEA 121
Chapter 3 122
3.1 Introduction 122
3.2 Optimisation of CD200 Flow Sorting Protocol 124
3.3 CD200 Expression is associated with a Reduction in Immune Cell Infiltration and Increases NK Cell Apoptosis in Transduced HeLa Xenografts 128
3.4 CD200 Expression is Associated with a Reduction in Immune Cell Infiltration and an Increase in NK Cell Apoptosis in Transduced UW-BCC1 Xenografts 134
3.5 Discussion 145
Chapter 4 147
4.1 Introduction 147
4.2 Results 148
4.2.1 RNA Sequencing of Human BCC Cell Line, UW-BCC1 148
4.2.2 Bulk RNA Sequencing of Primary Human BCC Cells 156
4.2.2.1 Optimisation of Human BCC Dissociation Protocol 156
4.2.2.2 Bulk RNA Sequencing of Primary Human BCC Cells 169
4.2.2.3 Immunohistochemical validation of VEGFR2 expression in primary human BCC tissue 208
4.3 Discussion 211
Chapter 5 215
5.1 Introduction 215
5.2 BCC single-cell RNA Sequencing Analysis Protocol Optimisation 217
5.2.1 Cell Type Identification on Individual Samples 226
5.3 Cell Type Identification Within the Integrated Dataset 239
5.4 Chromosomal Analysis of BCC Cells 243
5.5 Re-clustering of BCC Cells 249

5.6 RNA Velocity of BCC Cells ---------- 255
5.7 Stem-like Cluster Identification ---------- 262
5.8 CD200+ Cancer Stem Cell Cluster Identification ---------- 266
5.9 Characterisation of BCC Cancer Stem Cell/Stem-like Populations ---------- 268
5.10 Validation of TGFβ Signalling in BCC CD200+ Basal Cells ---------- 283
5.11 Discussion ---------- 286
Chapter 6 ---------- 290
6.1 CD200: a functionally relevant cancer stem cell-associated protein ---------- 290
6.2 CD200+ CD45- BCC subpopulation included endothelial cells and identified angiogenesis ---------- 292
6.3 Single cell sequencing: A TGFβ signature ---------- 294
6.4 Future Directions ---------- 297
References ---------- 298
Appendix ---------- 357
Appendix 2.2 | nCount RNA violin plots used for QC. 357
Appendix 2.3 | nFeature RNA violin plots used for QC. 358
Appendix 2.4 | Mitochondria gene percentage violin plots used for QC. 359
Appendix 2.5 | QC filter parameters. 360
Appendix 2.6 | Variable feature selection plots. 361
Appendix 2.7 | PCA Theory. 362
Appendix 2.8 | Clustering methods overview. 363
Appendix 2.9 | Seurat integration method overview 365
Appendix 2.10 | Harmony integration overview 367
Appendix 3.1 | CD200 expression by IHC in HeLa CD200+/- tumours 368
Appendix 3.2 | CD200 expression by flow cytometry in UW-BCC1 and UW-BCC1 CD200+/- Cell lines 369
Appendix 3.3 | CD200 expression by IHC in UW-BCC1 CD200+/- tumours 370
Appendix 4.1 | UW-BCC1 RNAseq sample information table 371
Appendix 4.2 | Human BCC RNAseq sample information table 372

Appendix 5.1 | Low resolution clustering to identify cell types 373
Appendix 5.1.1 | Writing 373
Appendix 5.1.2 | Figures 375
Appendix 5.2 | Visualisation 377
Appendix 5.3 | Markers used to identify cell types 378
Appendix 5.4 | BCC_1 PTPRC expression via (CD45) violin plot 379
Appendix 5.5 | g:Profiler results examples for BCC_1 380
Appendix 5.6 | g:Profiler results examples for BCC_3 381
Appendix 5.7 | Dataset integration 382
Appendix 5.7.1 | Writing 382
Appendix 5.7.2 | Figures 385
Appendix 5.8 | InferCNV analysis without background reduction 387
Appendix 5.9 | single cell GSEA results 388
Appendix 5.9.1 | Reactome 388
Appendix 5.9.2 | WikiPathways 389
Appendix 5.9.3 | PID 390
Appendix 5.9.4 | KEGG 391
Appendix 5.9.5 | Hallmarks 392
Appendix 5.9.6 | Biocarta 393
Appendix 5.10 | g:Profiler results of Basal_BCC cluster DEGs 394
Appendix 5.11 | EnrichR results of Basal_BCC cluster DEGs 395
Appendix 5.12 | single cell GSEA results for Basal_BCC cluster DEGs 396
Appendix 5.12.1 | Biocarta 396
Appendix 5.12.2 | Hallmark 397
Appendix 5.12.3 | KEGG 398
Appendix 5.12.4 | PID 399
Appendix 5.12.5 | Reactome 400
Appendix 5.12.6 | WikiPathways 401

Table of Figures

Figure 1.1 | The hallmarks of cancer----------3
Figure 1.2 | The cancer-immunity cycle----------11
Figure 1.3 | The caner immunoediting process----------15
Figure 1.4 | Tumour propagation methods----------24
Figure 1.5 | Diagram of the layers and components of human skin----------26
Figure 1.6 | Histomorphology of the hair follicle----------29
Figure 1.7 | Distribution of BCC on the face surface----------37
Figure 1.8 | Hedgehog growth factor signalling pathway----------51
Figure 1.9 | Cell heterogeneity within human BCC----------59
Figure 1.10 | Cellular heterogeneity within human facial BCC and normal skin----------60
Figure 2.1 | QuPath tool overview----------79
Figure 2.2 | Schematic of single cell RNA sequencing workflow----------91
Figure 2.3 | PCA and JackStraw analysis of pilot dataset----------103
Figure 3.1 | Assessment of CD200 expression in HeLa cell lines----------127
Figure 3.2 | Xenotransplantation of HeLa CD200+/- cells----------130
Figure 3.3 | Immunohistochemical analysis of HeLa CD200 +/- xenograft tumours----------132
Figure 3.4 | In vitro analysis of CD200 expression in UW-BCC1 cell line--------135
Figure 3.5 | Colony forming assays of CD200-sorted UW-BCC1 cells----------137
Figure 3.6 | Xenotransplantation of UW-BCC1 CD200+/- cells----------141
Figure 3.7 | Analysis of UW-BCC1 CD200+/- xenograft tumours----------143
Figure 4.1 | Sample correlation analyses following bulk RNA sequencing------151
Figure 4.2 | Heatmap of UW-BCC1 CD200+ and CD200- samples----------153
Figure 4.3 | Volcano plot of UW-BCC1 CD200+ vs CD200- differentially expressed genes----------155
Figure 4.4 | comparison of enzyme batches on the conventional human BCC dissociation protocol----------157
Figure 4.5 | Comparison of conventional and MACS human tumour dissociation protocols in basal cell carcinoma----------160
Figure 4.6 | Assessment of CD200 expression in BCC dissociated by two different protocols----------163

Figure 4.7 | CD200 expression in human BCCs----------165
Figure 4.8 | Immunofluorescence showing expression of BCC cancer stem cell marker, CD200, on CD200+/- CD45- BCC cells----------167
Figure 4.9 | Validation of CD200 expression in bulk RNA sequencing data from CD200+CD45- and CD200-CD45- BCC cells----------171
Figure 4.10 | Validation of expression of keratinocyte and BCC signalling pathway members in CD200+CD45- and CD200-CD45- bulk RNA sequencing data----------172
Figure 4.11 | Sample correlation analyses following bulk RNA sequencing of human BCC CD200+CD45- and CD200-CD45- populations----------175
Figure 4.12 | Heatmap of human BCC CD200+CD45- and CD200-CD45- samples----------177
Figure 4.13 | Volcano plot of human BCC CD200+CD45- vs CD200-CD45- differentially expressed genes----------179
Figure 4.14 | Gene-ontology (GO) gene-concept network (CNET) plot of BCC CD200+CD45- vs CD200-CD45- DEGs----------181
Figure 4.15 | Gene Ontology (GO) enrichment plot of BCC CD200+ CD45- vs CD200- CD45- DEGs----------182
Figure 4.16 | Wiki Pathways gene-concept network (CNET) plot of BCC CD200+ CD45- vs CD200- CD45- DEGs----------184
Figure 4.17 | Wiki Pathways enrichment plot of BCC CD200+ CD45- vs CD200- CD45- DEGs----------185
Figure 4.18 | GO and Wiki Pathways enrichment plots of BCC CD200+ CD45- vs CD200- CD45- upregulated DEGs----------187
Figure 4.19 | GO enrichment plot of BCC CD200+ CD45- vs CD200- CD45- downregulated DEGs. ----------190
Figure 4.20 | Pre-ranked Gene Set Enrichment (GSEA) analysis of Hallmark and C2- Curated gene sets in human BCC CD200+ CD45- vs CD200- CD45- DEGs---193
Figure 4.21 | Pre-ranked Gene Set Enrichment (GSEA) analysis of C5 – Ontology and C8 – Cell Type gene sets in human BCC CD200+ CD45- vs CD200- CD45- DEGs----------196
Figure 4.22 | Leading-edge analysis of Pre-ranked Gene Set Enrichment Analysis of Human BCC CD200+ CD45 vs CD200- CD45- DEGs----------199

Figure 4.23 | Ingenuity Pathway Analysis (IPA) overview of human BCC CD200+ CD45- population DEGs. ----------201
Figure 4.24 | Ingenuity Pathway Analysis results of human BCC CD200+ CD45- population DEGs. ----------204
Figure 4.25 | VEFG signalling pathway with human BCC CD200+ CD45- DEGs overlaid----------206
Figure 4.26 | Immunohistochemical staining for VEGFR2 in human BCCs------209
Figure 5.1 | t-SNE and UMAP cluster visualisations of 4 single cell samples---222
Figure 5.2 | 3-Dimensional t-SNE and UMAP cluster visualisations of 4 single cell samples. ----------224
Figure 5.3 | Cell annotation of BCC_1 sample----------228
Figure 5.4 | Cell annotation of BCC_2 sample----------231
Figure 5.5 | Cell annotation of BCC_2_NS sample----------234
Figure 5.6 | Cell annotation of BCC_3 sample----------237
Figure 5.7 | Integration, clustering, and cell annotation of 13 single cell samples----------241
Figure 5.8 | Chromosomal landscape of inferred copy number variations and identification of putative tumour cells in single cell samples----------245
Figure 5.9 | Identification of Basal Cell Carcinoma (BCC) cells. ----------247
Figure 5.10 | Identification of Basal Cell Carcinoma (BCC) cells within keratinocyte subpopulations. ----------250
Figure 5.11 | Assessment of Basal Cell Carcinoma (BCC) markers between CD200- and CD200+ sorted populations----------253
Figure 5.12 | RNA velocity of normal skin keratinocytes----------257
Figure 5.13 | RNA velocity of Basal Cell Carcinoma (BCC) keratinocytes--------260
Figure 5.14 | Visualisation of stemness-related genes in Basal Cell Carcinoma (BCC) keratinocytes----------264
Figure 5.15 | Assessment of CD200 expression in Basal Cell Carcinoma (BCC) keratinocytes----------267
Figure 5.16 | Cluster analysis of BCC keratinocytes----------270
Figure 5.17 | Basal BCC keratinocytes show functional similarity----------272
Figure 5.18 | Heatmap of commonly enriched signalling pathways in BCC basal cells. ----------277

Figure 5.19 | Assessment of TGFβ Signalling in BCC keratinocytes----------279
Figure 5.20 | SMAD2/3 signalling within CD200+ basal cell carcinoma basal keratinocytes----------281
Figure 5.21 | pSMAD3 nuclear staining of human BCC tissues----------284

Table of Tables

Table 1.1 | Clinical and histological features of BCC subtypes. ----------33
Table 1.2 | Genetic syndromes linked to increased susceptibility to BCC--------41
Table 2.1 | Cell lines used, their required media, and their passaging ratio/seeding densities----------67
Table 2.2 | Culture vessels used in the culturing of cells and their volumes of media required----------69
Table 2.3 | Primary antibodies used----------73
Table 2.4 | Secondary antibodies used----------74
Table 2.5 | Antibodies used for flow cytometry----------75
Table 2.6 | Top PCA loadings for BCC_1----------105
Table 5.1 | Single cell RNA sequencing data information table----------219

Abbreviations

AF	Alexa Fluor
AML	Acute Myeloid Leukaemia
APC	Antigen Presenting Cell
BAK	BCL-2 Homologous Antagonist Killer
BAX	BCL-2=associated X Protein
BCC	Basal Cell Carcinoma
BCL-2	B Cell Lymphoma-2
BCL-XL	B Cell Lymphoma Extra-Large
c-FLIP	FADD-like-IL-1β-converting enzyme-inhibitory protein
CCA	Canonical Correlation Analysis
CD	Cluster of Differentiation
cDNA	Complimentary DNA
CK1	Casein Kinase 1
CLL	Chronic Lymphocyte Leukaemia
CLR	Centred Log Ratio
CNV	Copy Number Variation
COSMIC	Catalogue of Somatic Mutations in Cancer
CSC	Cancer Stem Cells
cSCC	Cutaneous Squamous Cell Carcinoma
CTL	Cytotoxic T Lymphocyte
CTLA-4	Cytotoxic T-Lymphocyte-Associated Protein 4
CTS	Connective Tissue Sheath
DC	Dendritic Cell
DEG	Differentially Expressed Gene
DHh	Desert Hedgehog
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic Acid
DP	Dermal Papilla
DR5	TRIAL Death Receptor 5
FACS	Fluorescent Activated Cell Sorting
FC	Field Cancerisation
FPKM	Fragments Per Kilobase Per Million Mapped Reads

gDNA	Genomic DNA
GEM	Gel-Beads in Emulsion
GFP	Green Fluorescent Protein
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
GWAS	Genome Wide Association Study
H&E	Haematoxylin & Eosin
HF	Hair Follicle
Hh	Hedgehog
HIF	Hypoxia-Inducible Factor
HPA	Human Protein Atlas
HS	Hair Shaft
HVG	Highly Variable Gene
IF	Immunofluorescence
IFT	Intraflagellar Transport
IHC	Immunohistochemistry
IHh	Indian Hedgehog
IPA	Ingenuity Pathway Analysis
IRS	Inner Root Sheath
KSC	Keratinocyte Stem Cell
Log2FC	Log2 Fold-Change
M	Matrix
MHC	Major Histocompatibility Complex
MM	Multiple Myeloma
MNN	Mutual Nearest Neighbour
MOMP	Mitochondrial Outer Membrane Permeabilization
mRNA	Messenger Ribonucleic Acid
NES	Normalised Enrichment Score
NK	Natural Killer cell
NMSC	Non-Melanoma Skin Cancer
NS	Normal Skin
NSAID	Non-Steroidal Anti-Inflammatory Drug
OCT	Optimal Cutting Temperature
ORS	Outer Root Sheath

OTR	Organ Transplant Recipient
PAGA	Partition-based Graph Abstraction
PBS	Phosphate-buffered Saline
PC	Principal Component
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PD-1	Programmed Death Protein 1
PKA	Protein Kinase A
PSM	Photosensitising Medication
Ptch	Patched
PTCH1	Patched-1
QC	Quality Control
qPCR	Quantitative Polymerase Chain Reaction
qRT-PCR	Quantitative Reverse Transcription Polymerase Chain Reaction
RB	Retinoblastoma
RC	Relative Counts
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
RPM	Revolutions Per Minute
RT	Room Temperature
scGSEA	single cell Gene Set Enrichment Analysis
scRNAseq	single cell RNA Sequencing
SG	Sebaceous Gland
SHH	Sonic Hedgehog
SMO	Smoothened
SNN	Shared Nearest Neighbour
SNP	Single Nucleotide Polymorphism
SuFu	Suppressor of Fused
Sv	Sievert
t-SNE	t-Stochastic Neighbour Embedding
TAC	transit-amplifying cells
TD	terminal differentiation
TGFβ	Transforming Growth Factor Beta
TIC	Tumour-Initiating Cells

TME	Tumour Microenvironment
TNF	Tumour Necrosis Factor
TP53	Tumour Protein 53
UK	United Kingdom
UMAP	Uniform Manifold Approximation and Projection
UMI	Unique Molecular Identifier
UV	Ultraviolet Radiation
VDR	Vitamin D Receptor
VEGF	Vascular Endothelial Growth Factor
WGP	Wales Gene Park
WHO	World Health Organisation

Chapter 1 Introduction

1.1 Hallmarks of Cancer

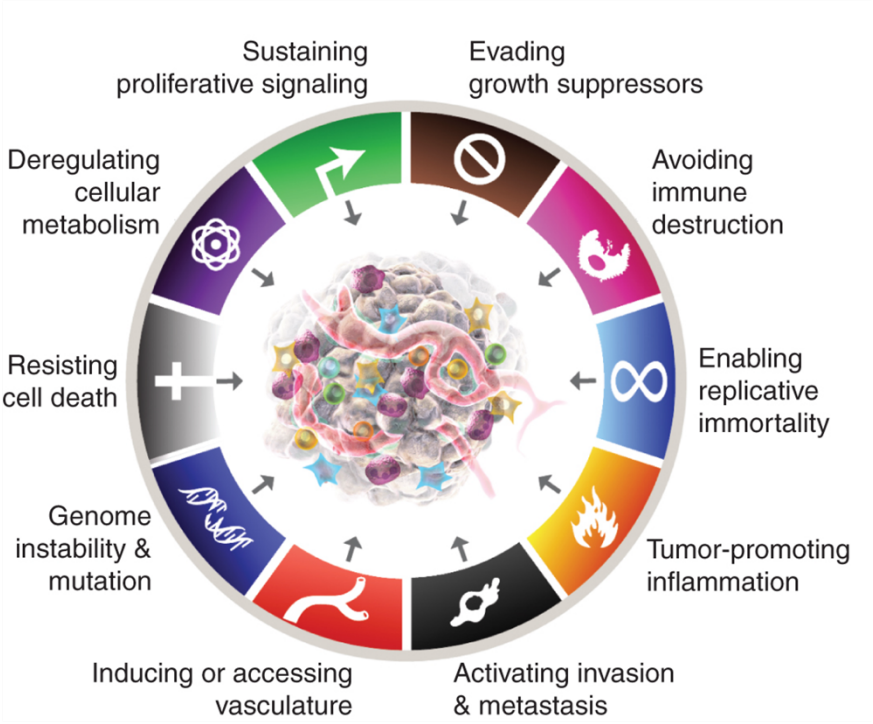
Cancer is a term given to a group of over 100 diseases that involve the disruption of cell homeostasis resulting from uncontrolled proliferation. Unchecked, these cells can form a malignant tumour which is capable of metastasising to other body areas and ultimately can cause death. For a normal cell to transform, it must accumulate several deoxyribonucleic acid (DNA) mutations within genes that regulate cell growth and differentiation.

Genome integrity is constantly under attack from endogenous or exogenous DNA-damaging agents such as reactive oxygen species (ROS), ultraviolet (UV) light, DNA replication errors, mutagenic chemicals, and ionizing radiation (Cheong and Nagel 2022; Steele et al. 2022). The DNA damage that these agents cause is usually repaired flawlessly, however, DNA-mutations do occur when these repair mechanisms fail. DNA mutations steadily accumulate over the lifetime of an individual and the majority of these are known as passenger mutations, which do not alter gene function and the fitness of the cell (Bozic et al. 2010). For a tumour to form, a driver mutation must occur resulting in deregulation of cell proliferation. Notably DNA-mutations activate genes that promote cell proliferation (called oncogenes) or block the activity of genes that are associated with cell cycle regulation (called tumour

suppressors; Chial 2008b; Chial 2008a; Stratton et al. 2009). In the UK from 2017 to 2019, cancer was responsible for over 25% of all deaths, higher in men (53% of all deaths) than women (47% of all deaths) (Cancer Research UK Annual Report 2021). In 2019, the World Health Organisation (WHO) estimated that cancer was the first or second leading cause of death for individuals <70 years in 112 of 183 countries (Bray et al. 2021). Hence Worldwide, cancer is a leading cause of death (Sung et al. 2021).

Although the transformation of normal cells relies on DNA mutations, other factors are required in order for the transformation into cancer cells. These processes have been elucidated over many decades of research and can be summarised as the 'Hallmarks of Cancer' (Figure 1.1; Hanahan and Weinberg 2000; Hanahan 2022). These ancillary processes that underpin the complex nature of cancer growth, as summarised below.

a)


Sustaining proliferative signaling
Evading growth suppressors
Deregulating cellular metabolism
Avoiding immune destruction
Resisting cell death
Enabling replicative immortality
Genome instability & mutation
Tumor-promoting inflammation
Inducing or accessing vasculature
Activating invasion & metastasis

b)

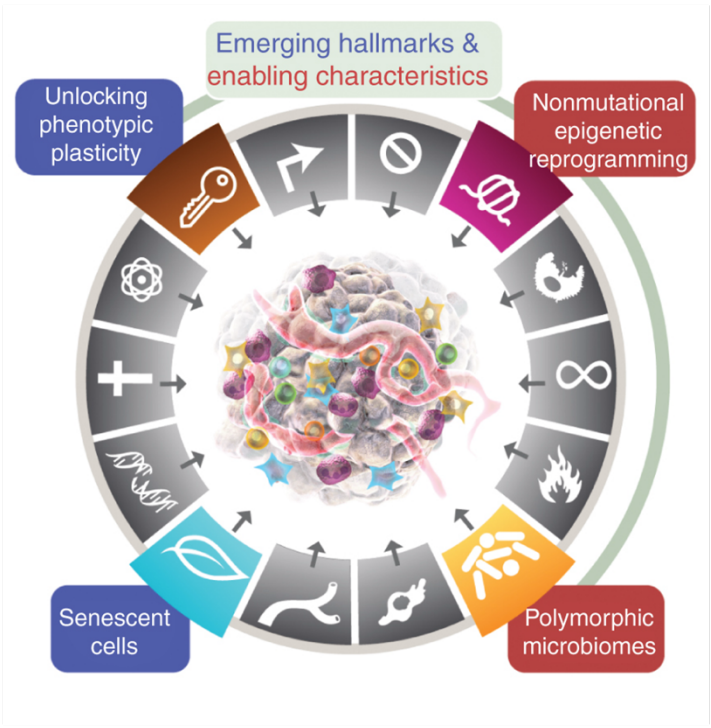

Emerging hallmarks & enabling characteristics
Unlocking phenotypic plasticity
Nonmutational epigenetic reprogramming
Senescent cells
Polymorphic microbiomes

Figure 1.1 | The hallmarks of cancer. a) *the current hallmarks of cancer embodying eight hallmarks and two enabling characteristics. These include evasion of growth suppressors, immune destruction avoidance, enabling of replicative immortality, tumour-promoting inflammation, activation of invasion and metastasis, vasculature induction or accession, genome instability and mutation, resisting of cell death, deregulation of cellular metabolism, and sustaining of proliferative signalling.* ***b)*** *two additional emerging hallmarks (blue) and two additional enabling characteristics (red) were introduced in 2011 and 2022 by the same authors. These included nonmutational epigenetic reprogramming, polymorphic microbiomes, senescent cells, and unlocking phenotypic plasticity. Figure taken from (Hanahan 2022).*

1.1.1 Sustaining Proliferative Signalling

Cancer cells are defined by sustained chronic proliferation. In normal cells, the cell cycle is carefully regulated to maintain tissue homeostasis. As mentioned earlier, cancer cells circumvent these controlling signals through driver mutations in one or more of the cell-signalling pathways (Vogelstein et al. 2013). Thus, cancer cells demonstrate cell autonomous proliferation, either in an autocrine manner (in which they produce and respond to their signalling ligands) and/or paracrine manner (in which the cancer cells signal to neighbouring normal cells (Hanahan and Weinberg 2011). There are many examples by which DNA mutations promote cell proliferation, such as overexpression through gene amplification, constitutive activation through gene mutations, aberrant signalling through protein fusions as a result of chromosomal translocation, or impairment of the receptor recycling and degradation systems (Bache et al. 2004; Gschwind et al. 2004; Lemmon and Schlessinger 2010; Fouad and Aanei 2017).

1.1.2 Evading Growth Suppressors

Despite aberrant growth signalling, tight regulation of the cell cycle controls cell division, usually through the actions of tumour suppressor proteins. In order for growth of the cancer cells, they must deregulate the cell cycle and disrupt the checkpoints (Malumbres and Barbacid 2009). There have been dozens of confirmed tumour suppressor genes that do control cell proliferation. Two key suppressors encode the retinoblastoma-associated (RB) and Tumour protein P53 (TP53) proteins, whose main functions are to monitor DNA damage and initiate cellular responses to

various stresses, and regulate the cell cycle checkpoints, respectively, which in turn halt cancer growth (Hanahan and Weinberg 2011).

1.1.3 Tissue Invasion and Metastasis

The successive steps of the metastatic process, the spread to a nearby or distant unrelated secondary site, is represented by four additional hallmarks; 1) Motility and invasion, 2) modulation of secondary sites or local microenvironments, 3) plasticity, and 4) colonisation of secondary tissues (Welch 2006; Nguyen and Massagué 2007; Chiang and Massagué 2008; Welch and Hurst 2019). This cascade involves the local invasion and dissemination of cells away from the originating tumour, followed by intravasation into the haematogenous or lymphatic systems, then extravasation of the cells into distant tissues to form secondary tumours that then colonise the site (Hanahan and Weinberg 2011; Eslami-S et al. 2020). It has since been shown that cells can also exit their residence via coelomic cavities, or along nerve fibres (Lugassy et al. 2004; Sleeman et al. 2011; Li et al. 2013; Dart 2017).

1.1.4 Limitless Replicative Potential

Under physiological conditions, cells have a limited capacity for division before they become senescent, which is the entry of a cell into an irreversible, but viable, non-proliferative state (Shay and Wright 2011). Initially, it was observed that normal cells would divide between 40 and 60 times before entering programmed cell death (Hayflick and Moorhead 1961). Cells that proliferate beyond senescence enter crisis, resulting in genomic instability (Fouad and Aanei 2017).

1.1.5 Sustained Angiogenesis

Angiogenesis describes the process of new blood vessel formation from existing vasculature. Angiogenesis is an important feature of tissue homeostasis as it ensures adequate vasculature for the delivery of nutrients to- and removal of metabolic wastes and carbon dioxide from cells. Angiogenesis is also an important aspect to tumour growth. Studies have shown that in the absence of angiogenesis, tumours cannot grow more than 1-2mm^3 in diameter without growth arrest and cell death (Folkman 1971; Fouad and Aanei 2017). To overcome these problems, tumours generate a neovascularisation network that allows the vessels to reach the distant tumour cells, ensuring adequate nutrient delivery and waste removal (Majidpoor and Mortezaee 2021). Angiogenesis is a vital process during embryogenesis, where production of new endothelial cells and their subsequent aggregation into new tubules is seen (Breier 2000). Normal vasculature becomes quiescent following early development, and is only subsequently switched on in adult tissues during wound healing and menstrual cycles (Okada et al. 2014; Raina et al. 2021).

The switching from quiescence to active angiogenesis is known as the 'angiogenic switch', and is required by tumours to escape dormancy and initiate rapid growth (Hanahan and Folkman 1996; Lugano et al. 2020). The most important trigger of the angiogenic switch is hypoxia, which is recognised by vascular endothelial cells via oxygen-sensing mechanisms which primarily interact with the hypoxia-inducible transcription factor (HIF) family that in turn regulate a plethora of genes involved in the expression of pro-angiogenic proteins (Fraisl et al. 2009; Yang et al. 2013; Lv et al. 2016; Fouad and Aanei 2017; Lugano et al. 2020). Hypoxia is a common feature of solid tumours due to their rapid growth (Emami Nejad et al. 2021), and so it is not a

surprise that HIF levels are higher in many cancers (Hashimoto and Shibasaki 2015). A key and potent regulator of angiogenesis is the vascular endothelial growth factor (VEGF) (Ferrara et al. 2003). VEGF is produced by many cells, including tumour cells, usually in response to an increase in tumour hypoxia through activation of the HIF signalling pathways (Sakurai and Kudo 2011; Kieran et al. 2012). VEGF signalling leads to endothelial cell proliferation, survival, migration, differentiation, and mediation of vascular permeability (Hicklin and Ellis 2005; Fouad and Aanei 2017).

1.1.6 Evading Apoptosis

Programmed cell death via apoptosis serves as a natural barrier to the development of cancer (Sharma et al. 2019). Apoptosis is an additional response to the transformation-associated stress that the growing tumour experiences. In cancer, the balance between cell death and proliferation is disturbed (Lopez and Tait 2015; Fouad and Aanei 2017). There are two pathways that lead to triggering apoptosis: the extrinsic and intrinsic pathways. The extrinsic pathway involves interactions between cell surface receptors and their ligands. The intrinsic pathway involves a shift in the balance between pro- and anti-apoptotic proteins as a result from internal stress signals. Mitochondrial outer membrane permeabilization (MOMP) is also seen, causing a release of activating molecules from the intermembrane space (Pfeffer and Singh 2018). Both pathways involve the activation of protein-cleaving caspases, which are vital to successfully completing apoptosis (Zaman et al. 2014).

In cancer, there is inhibition of the intrinsic apoptotic pathways. But also, other mechanisms exist that allow the cancer cells to evade apoptosis (Xu et al. 2015;

Arbiser et al. 2017; Pfeffer and Singh 2018). A classic example of apoptosis evasion in cancer cells can be seen through the upregulation of the anti-apoptotic B-cell Lymphoma-2 (BCL-2) family of regulatory proteins (Hockenbery et al. 1990; Fuchs and Steller 2011; Martinou and Youle 2011). The BCL-2 protein family includes pro-apoptotic effector proteins, pro-apoptotic BH3-only proteins, and anti-apoptotic BCL-2 proteins (Carneiro and El-Deiry 2020). Anti-apoptotic BCL-2 proteins inhibit the pro-apoptotic BCL-2 proteins, BCL-2-associated X protein (BAX) and BCL-2 homologous antagonist killer (BAK). In cancer, there is a common occurrence of increased expression of anti-apoptotic BCL-2 proteins and a decrease of pro-apoptotic proteins, BAX and BAK. This trend is observed in over fifty percent of all cancer types, indicating a prevalent mechanism of evasion from programmed cell death, irrespective of cancer type (Lopez and Tait 2015; Campbell and Tait 2018).

1.1.7 Avoiding Immune Destruction

The immune system is able to detect cells expressing aberrant proteins resulting from driver mutations in cancer. The mechanism of destruction of cancer cells is known as the cancer-immunity cycle. Interestingly, the immune system has also shown to constrain and/or promote cancer growth via cancer immunoediting.

1.1.7.1 The Cancer-Immunity Cycle

The immune system has long been known to play a fundamental role in virtually all types and stages of cancer (Mellman et al. 2023). The cancer-immunity cycle is a process in which the body eliminates cancer cells. Introduced in 2013, the process explains a cyclic process which is initiated by cancer cells that release antigens, and

concludes with the destruction of the cancer cells (Chen and Mellman 2013). The cycle illustrates that T cells do not respond or work on their own, but instead they exist in a series of steps that lead to the accumulation of immune-stimulatory factors which increase and broaden their responses (Figure 1.2). The cycle is split into 7 major steps. The first step sees the release of antigens by the cancer cells as a result of cell death or destruction. This is followed by the capture of these antigens by antigen-presenting cells, dendritic cells (DCs). Here, the DCs process the antigens and produce peptides that bind to major histocompatibility complex (MHC) on cytotoxic T lymphocytes (CTLs). In step 3, we see the presentation of the MHC molecules by DCs to CTLs, which in turn primes and activates them. The activated CTLs then migrate toward the tumour via the blood vessels (step 4) and infiltrate the tumour bed (step 5). Here, the CTLs bind to cancer cells via their T cell receptor and MHC antigens (step 6), which ultimately results in the killing of the cancer cell (Step 7) and completion of the cycle. As the cancer cell dies, we see a release of antigens by the cell, which restarts the cycle again.

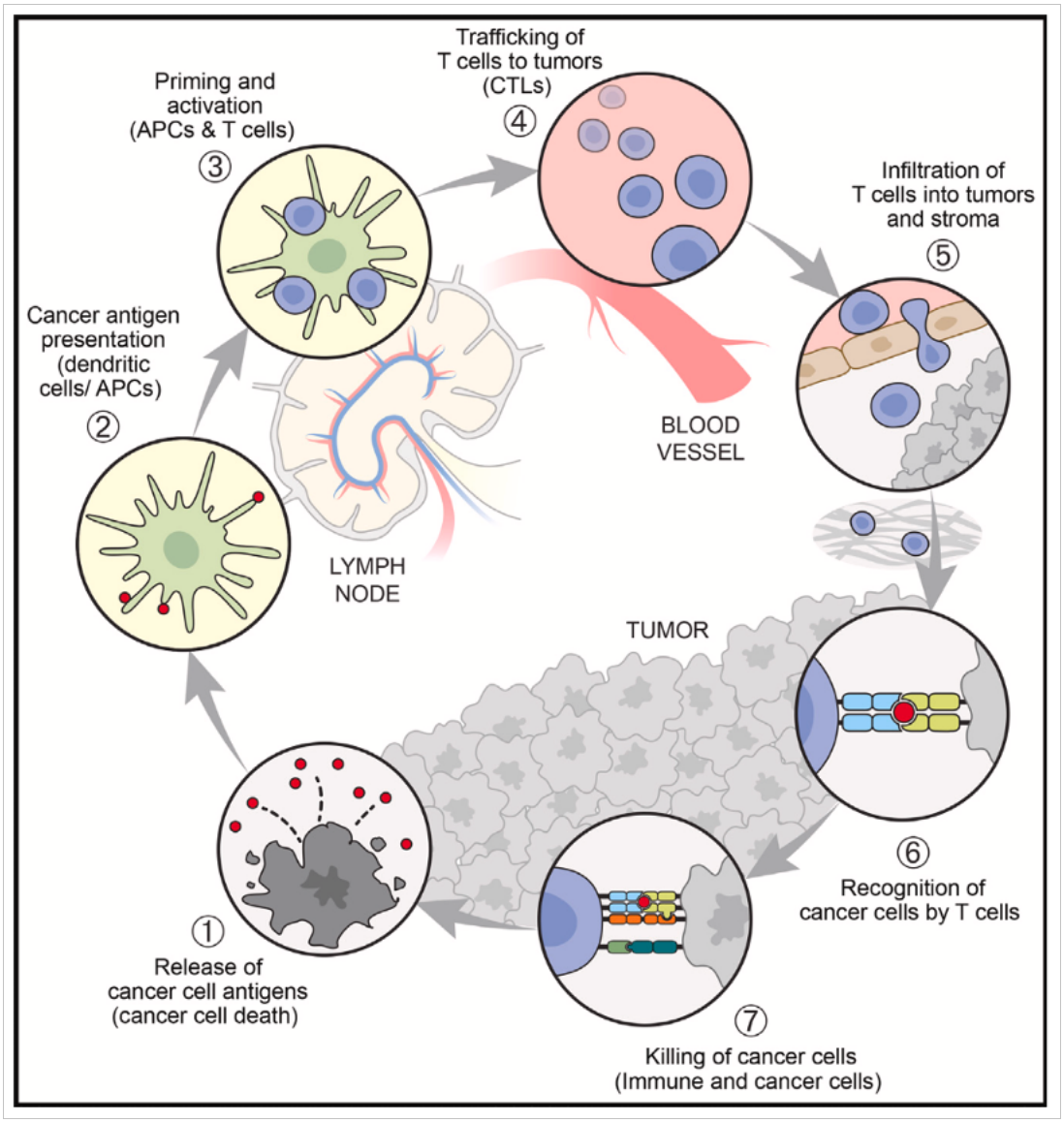


Figure 1.2 | The cancer immunity cycle. *The generation of immunity to cancer is a cyclic process that can be self-propagating, leading to an accumulation of immune-stimulatory factors that in principle should amplify and broaden T cell responses. The cycle is also characterized by inhibitory factors that lead to immune regulatory feedback mechanisms, which can halt the development or limit the immunity. This*

cycle can be divided into seven major steps, starting with the release of antigens from the cancer cell and ending with the killing of cancer cells. In step one, cancer cells release cancer-specific antigens. In step two, dendritic cells process these antigens and produce peptides that bind to major histocompatibility complex (MHC) and present them on their surface where CD4+ T cells then recognise them. In step three, T cells become primed and activated and migrate towards the cancer cells in step four. In step five, the T cells infiltrate the tumour bed and bind to the cancer cells in step six via their T cell receptor and MHC antigens. This ultimately leads to the destruction of the cancer cells, which then cause further release of cancer-specific antigens, restarting the process. Each step is described above and, in the figure, with the primary cell types involved and the anatomic location of the activity listed. Abbreviations are as follows: APCs, antigen presenting cells; CTLs, cytotoxic T lymphocytes. Figure and Taken from Chen and Mellman 2013.

1.1.7.2 Cancer Immunoediting

Cancer immunoediting was introduced as a framework to understand how the immune system can constrain and/or promote the cancer through three phases: Elimination, Equilibrium, and Escape (Gubin and Vesely 2022). From these phases, we see the adaptive immune system editing the tumour immunogenicity, and these edited tumours are then able to escape the host immune system, resulting in unrestrained growth (Figure 1.3).

1.1.7.2.1 Elimination

The elimination phase can be described as a classic immunosurveillance concept, in which both adaptive and innate immune cells detect transformed cells that have escaped the intrinsic tumour suppression mechanisms and destroy these cells before they become clinically visible. Here, we see normal cells transform into cancer cells by genotoxic insults and carcinogens and escape the classic tumour suppressor mechanisms such as p53. These cells express stress-induced molecules such as tumour antigens in the context of MHC class I molecules and NKG2D ligands which are recognised by CD8+ T cells and natural killer (NK) cells, respectively. From here, we see the activation of an effective cancer cell killing response triggered by the action of DCs (Mittal et al. 2014).

1.1.7.2.2 Equilibrium

The equilibrium phase can be described as the molecular mechanisms that result in triggering of immune-mediated tumour dormancy. In this phase, cancer cells that

were not destroyed during the elimination phase enter a state of dynamic dormancy. Here, we see the immune system maintains the tumour mass but does not kill the cells. This phase involves only the adaptive immune system, unlike the elimination phase which involves both adaptive and innate immunities. Tumours in this phase can exist for many years, but given that they are under constant immune pressures, develop many cancer cell clones that contain genetic and epigenetic changes. This can also lead to the dominance of certain subclones which are able to evade immune recognition and subsequently evade the immune destruction, which in turn can lead to the development of clinically relevant disease (Rojas-Domínguez et al. 2022).

1.1.7.2.3 Escape

For cancer cells to reach the escape phase, they must first have gained beneficial genetic and epigenetic changes which enabled them to avoid immune detection and elimination. This phase is also now identified as a hallmark of cancer (Fouad and Aanei 2017), and here tumours display unrestricted growth with high likelihoods of evading the innate and adaptive immune systems (Dunn et al. 2004). Tumour cells can escape the immune system in one of three ways: Loss or downregulation of tumour antigens and/or antigen presenting mechanisms combined with the reduction of co-stimulatory molecules that results in a reduced immune recognition and activation; induction of a tumour microenvironment (TME) that has immunosuppressive qualities via upregulation of suppressive immune cells and production of anti-inflammatory cytokines; and upregulation of anti-cytotoxic mechanisms (Mittal et al. 2014).

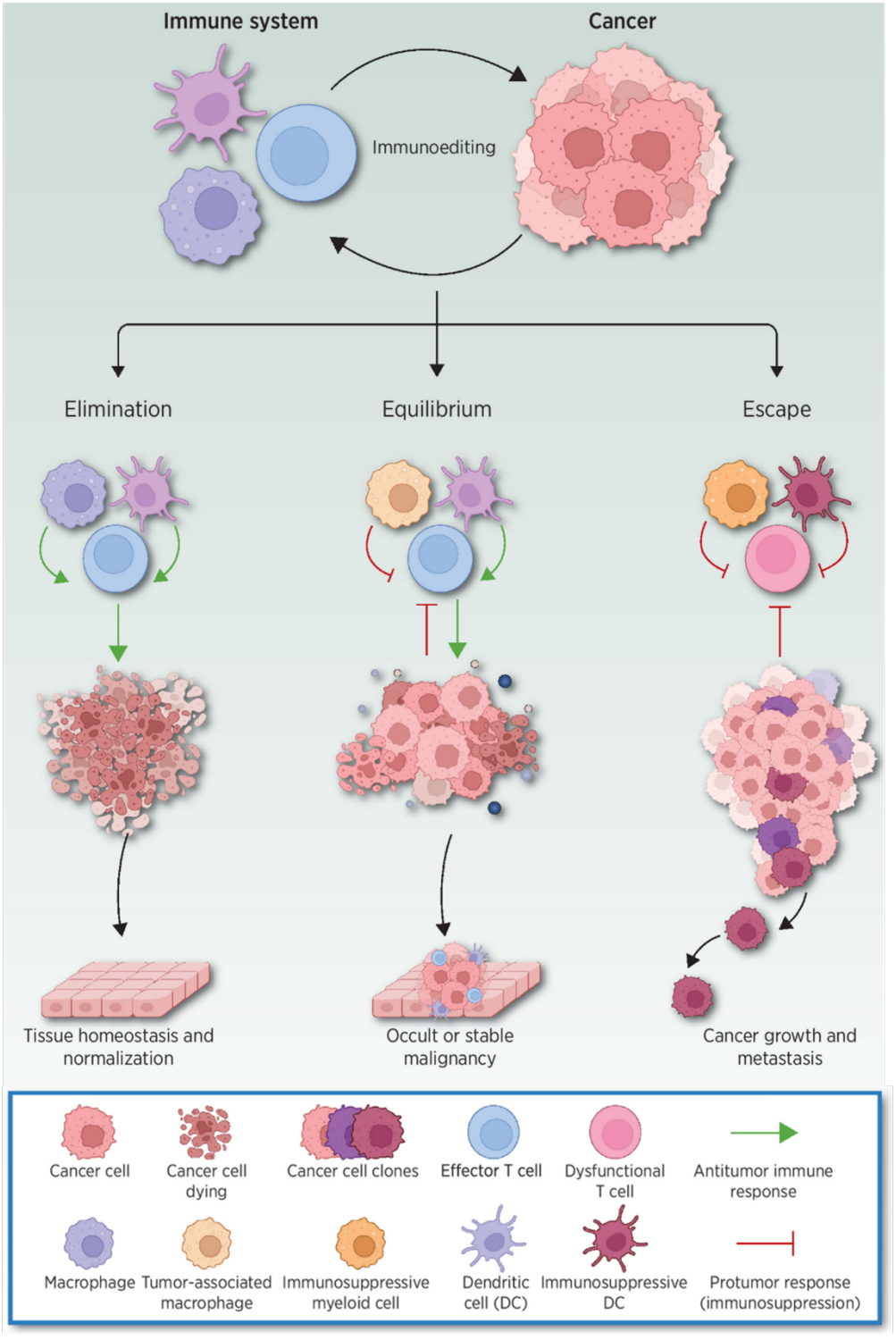

Immune system
Cancer
Immunoediting
Elimination
Equilibrium
Escape
Tissue homeostasis and normalization
Occult or stable malignancy
Cancer growth and metastasis
Cancer cell
Cancer cell dying
Cancer cell clones
Effector T cell
Dysfunctional T cell
Antitumor immune response
Macrophage
Tumor-associated macrophage
Immunosuppressive myeloid cell
Dendritic cell (DC)
Immunosuppressive DC
Protumor response (immunosuppression)

Figure 1.3 | The cancer immunoediting process. *The cancer immunoediting hypothesis offers a framework for understanding the intricate and dynamic interactions between immune cells and cancer cells within the tumour microenvironment (TME). These interactions can either be beneficial in fighting against tumours, or supportive of tumour growth, illustrating the dual role of immune cells in tumour suppression and promotion. Cancer immunoediting involves three main phases: elimination, equilibrium, and escape. During the elimination phase, the innate and adaptive immune systems work together to eradicate emerging tumours, resulting in tumour destruction and restoration of tissue homeostasis. During the equilibrium phase, the immune system fails to completely eradicate the tumour, and the cancer enters a phase of dormancy, where the immune system maintains the mass but does not eradicate the cancer. During the escape phase, continued immune pressures cause several genetic and epigenetic changes in the tumour cells which enable their selective advantage and enable them to escape the immune system and then develop into clinically relevant disease. Figure and taken from Gubin and Vesely 2022.*

1.1.7.2.4 Mechanisms of Immune Escape

There are a number of mechanisms that allow cancer cells to evade the immune system. These mechanisms allow evasion of both innate and adaptive immune systems, which in turn enables unrestricted tumour development and growth. We also see that cancer cells are able to recruit and induce immunosuppressive cytokines and cells within the TME, which in turn hinders the anti-cancer immune responses (Kim and Cho 2022).

1.1.7.2.4.1 Decreasing of Immune Recognition

The immune system recognises normal and malignant cells through antigen presentation, namely MHC-antigen complexes expressed at the surface of these cells. One of the most common mechanisms to evade immune recognition is for the cancer cell to downregulate the MHC class I antigens. Downregulation of the MHC I antigens has been shown to have no detrimental effect on the cells ability to grow, survive, or metastasize (Cornel et al. 2020). This downregulation is an independent prognostic factor in many cancer types, has been linked to possible immunotherapy resistance, and is seen in 40-90% of human cancers (Taylor and Balko 2022). According to the 'missing self' hypothesis, the loss of MHC I should result in an increase in NK cell infiltration and NK-mediated killing of these cells (Cruz-Muñoz et al. 2019). However, in reality, there have been no reports of MHC I negative tumours displaying an increase in NK cell infiltrate when compared to those with normal MHC I expression, implying a utilisation of other immune evasion methods (Taylor and Balko 2022).

1.1.7.2.4.2 Immune Resistance

In addition to the downregulation of cell surface antigens to avoid immune cell detection and killing, tumour cells can additionally develop further resistance mechanisms to avoid the cytotoxic effects of the immune cells. Resistance mechanisms include the expression of anti-apoptotic molecules such as cellular FLICE (FADD-like IL-1β-converting enzyme)-inhibitory protein (c-FLIP) and B cell lymphoma extra-large (BCL-XL), as well as the constitutive activation of pro-oncogenic transcription factors like STAT3 (Wang et al. 2018). Furthermore, tumour cells can resist apoptosis induced by cytotoxic T lymphocytes (CTLs) and NK cells by downregulating and mutating death receptors such as TRAIL death receptor 5 (DR5) and Fas (Vesely et al. 2011).

1.1.7.2.4.3 Immunosuppressive Tumour Microenvironment

The TME is a complex ecosystem of cells, signalling molecules, and extracellular matrix that surrounds a tumour. The interactions between the tumour and the TME is essential for tumour survival, growth, and invasion. A key characteristic that the tumour needs for immune evasion is the development of an immunosuppressive TME. Immunosuppressive TMEs have shown to have a significant impact on patient survival and response to treatment (O'Donnell et al. 2018). The balance between a pro- or anti-tumour TME is dependent on the presence and activation status of immune cells that infiltrate the TME and the cytokines and chemokines that they produce (Labani-Motlagh et al. 2020). To create an immunosuppressive TME, the tumour must secrete regulatory cytokines, growth factors, and proteases in order to alter the phenotype of the infiltrating immune cells. Coupled with an increased

recruitment of immunosuppressive regulatory cells at the tumour site, the tumour is able to create the immunosuppressive environment that is capable of reducing the effectiveness of the anti-tumour immune response (Baghban et al. 2020). A key mediator of this activity is transforming growth factor beta (TGFβ), which is able to inhibit multiple stages of the anti-tumour immunity. This includes the inhibition of DC migration and activation, inhibition of NK and T cell functions, and promotion of Th2 and Treg cell development through differentiation of CD4+ T cells (Labani-Motlagh et al. 2020). Another important mediator of note is VEGF. A critical component of tumour angiogenesis, VEGF enables the neovascularisation around the tumour, which can act as a barrier and inhibit the trafficking of effector immune cells to the tumour (Geindreau et al. 2021). VEGF has also shown to inhibit the function of T cells and DCs, reduce the cytotoxic effects of NK cells, and prevention of sufficient antigen uptake and presentation (Ribatti 2022). Likewise, a number of inflammatory mediators residing within the TME have shown to contribute to the development of cancer and metastasis. These mediators include tumour necrosis factor alpha (TNFα), IL-6, IL-10, and IL-12, to name a few. Additionally, these mediators also contribute to immune evasion through key immune cell suppression, such as macrophages and T cells. Subsequently, these have become attractive targets for cancer treatments (Morris et al. 2022).

1.1.7.2.4.4 Negative Co-Stimulatory Pathways

To avoid an overactive immune response, immune cell activation is tightly regulated by the balance between co-stimulatory and co-inhibitory signals. The activation of an immune cell requires this balance to tip in the favour of the stimulatory signals. In cancer however, we see this balance is often favoured towards inhibitory signals

which in turn results in the downregulation and subsequent inhibition of the anti-tumour response. Regulation of the immune response is mediated through immune cells such as macrophages and T cells, regulatory cytokines such as TGFβ, and immune checkpoints such as cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) and programmed death protein 1 (PD-1; He and Xu 2020). Under physiological conditions, we see these mechanisms successfully regulate and maintain homeostasis, but in cancer, we see these mechanisms are enhanced which allow the cancer to subsequently evade the immune system (Marin-Acevedo et al. 2018). Immune checkpoints are known to be negative co-stimulatory pathways and are activated by their respective ligand-receptor interaction. This causes T cell activation and function at numerous points throughout the immune response. Several immune checkpoints have been studied and identified to have a role in cancer development. The most notable, and most studied, are CTLA-4 and PD-1. Both have involvement at different stages of the immune response. CTLA-4 ligand-receptor binding occurs during the T cell priming stage and results in limited T cell activation. PD-1 ligand-receptor binding occurs during the effector phase and results in regulating the activity of activated T cells (Dyck and Mills 2017).

1.3 Tumour Initiating Cells

Tumour initiating cells (TICs), also known as cancer stem cells (CSCs), are cancer cells that harbour the ability to self-renew and initiate tumour growth. TICs are believed to recapitulate the original tumour through stem cell processes of self-renewal and differentiation (Qureshi-Baig et al. 2017). TICs form a small, distinct population within the tumour and have been shown to be the cause for relapse and metastases

(Suraneni and Badeaux 2013). As such, TICs have long since been a hot topic of study and therapeutic intervention.

1.3.1 Tumour Propagation Models

The method of tumour propagation has long been debated. Within different tumour types, we see functional heterogeneity of certain cells. This functional heterogeneity exists in various ways, such as differences in proliferative capacity, metastatic potential, and response to therapy. In breast cancer, certain subpopulations of cells may exhibit enhanced migratory abilities, facilitating metastasis to distant organs (Kong et al. 2020). Similarly in glioblastoma, subsets of cells with stem-like properties can resist conventional treatments and drive tumour recurrence (Mattei et al. 2021). We also see that tumours are formed from cells that display various proliferative and differentiation potentials. This heterogeneity has led to the development of two propagation models to account for this: the cancer stem cell model and the stochastic model.

1.3.1.1 The Cancer Stem Cell Model

The cancer stem cell model, also known as the hierarchical model, formulates that tumours are hierarchically organised, with the tumorigenic cancer stem cell residing at the top which can self-renew indefinitely and differentiate to form the tumour progeny (Walcher et al. 2020; Figure 1.4a). These progenies are non-tumorigenic but are essential to the contribution of the tumour growth. This model infers that only certain cancer stem cell subpopulations can drive the cancer progression, and so implies that specific characteristics can be identified and therapeutically targeted,

which eliminates the cancer stem cells and in turn prevents further generation of the tumour bulk, subsequently long-term killing the cancer (Ayob and Ramasamy 2018). This therapeutic intervention can further be supplemented with therapies that can target the rest of the tumour bulk, such as chemotherapy to target the proliferating cells.

1.3.1.2 Stochastic Model

The Stochastic model, also known as the clonal evolution model, works on the theory that any cell within the tumour mass has the potential to re-propagate the tumour. Here, the model describes that all cells within a tumour have the ability to gain genetic and epigenetic mutations that enable cancer stem cell properties of self-renewal and differentiation (Rich 2016; Mondal et al. 2018; Figure 1.4b). These mutations could further accumulate in cells and enhance the cells fitness and resistance, which allows greater cell survival and cell selection to drive tumour progression and heterogeneity (Melzer et al. 2017).

1.3.1.3 Alternative Models

In 2010, the Stemness Phenotype Model (SPM) was also proposed, which introduced a non-hierarchical model in which there are no specific CSC populations within tumours (Kaushik et al. 2021). Instead, the model explains that cancer cells are highly plastic in stemness terms, and CSCs and non-CSCs can interconvert into each other depending on the microenvironment. The model predicts that there exists cancer cell populations ranging from pure CSC phenotype to a non-CSC phenotype, and survival of a single cell can originate a new tumour. In 2014, there was a study that stated the

gap between the CSC and stochastic models could be bridged with an alternative model. This model experimentally and mathematically showed the existence of both distinctive CSC subpopulations and the stochastic transitions of non-CSCs to CSCs (Wang et al. 2014).

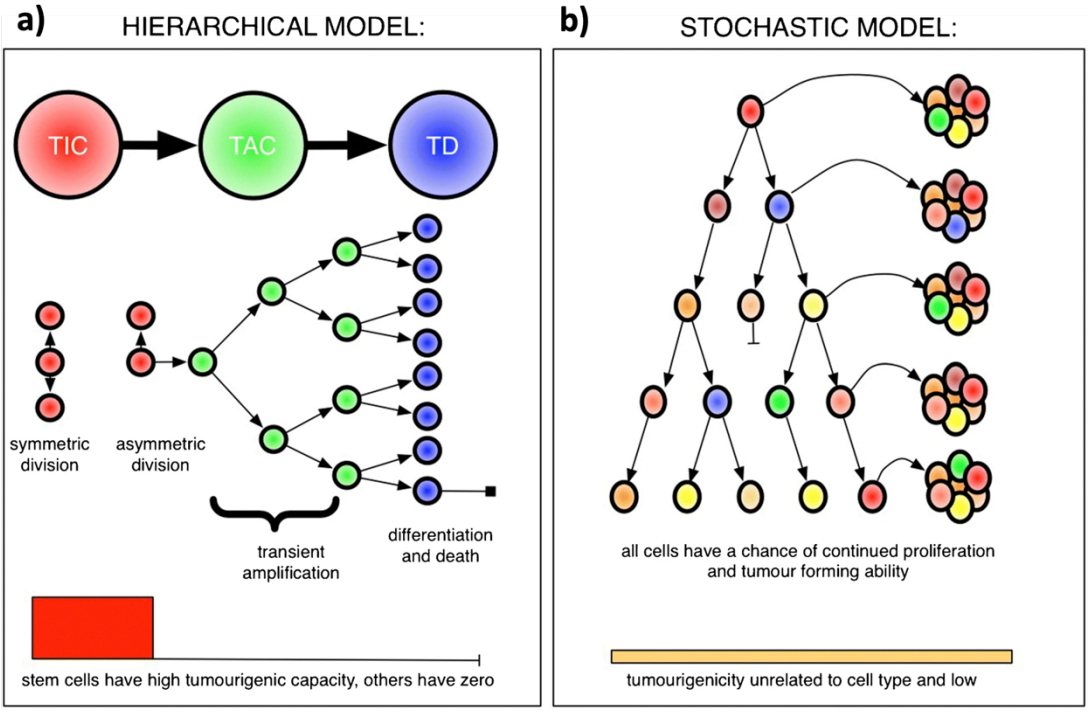


Figure 1.4 | Tumour propagation models. a) *The hierarchical (cancer stem cell) model in which only a subset of cells, the putative cancer stem cells, or tumour-initiating cells (TICs), have the ability to proliferate indefinitely and can recapitulate the entire tumour, while all others are doomed to differentiation into transit-amplifying cells (TACs) and eventual terminal differentiation (TD) and then cell death.* ***b)*** *Stochastic model, in which each tumour cell has the same ability, with low probability, to form tumours. Clonogenicity is represented on the bottom of each panel where, in a), the red box means that the TICs have high probability and all others have none, and in b), that all cells have equal, low clonogenic potential. Figure and legend taken and adapted from Scott et al. 2019.*

1.4 Carcinoma

Epithelial tissues, the environmental barriers of our bodies, are constantly exposed to cancer-causing agents. As such, carcinomas, the cancer of epithelial tissues, are the most common form of cancer accounting for 85% of all cancers and 78% of all cancer associated deaths (CR-UK database analysis; Ye and Weinberg 2015; Sung et al. 2021). Carcinomas frequently arise from a pre-cancerous transformation, known as intraepithelial neoplasia or field cancerisation (FC), within which multiple carcinomas can develop. The term FC describes the spread of premalignant changes across a tissue or organ that can lead to the development of multiple, independent tumours or areas of dysplasia within the same region (Willenbrink et al. 2020).

1.5 Skin Biology

The skin is the largest organ of the body and performs a plethora of functions. One of its main functions is to provide a barrier to the constant environmental insults that we encounter. The skin is formed of two layers, the outermost epidermis, and the innermost dermis (Figure 1.5). Both layers play an important part in skin homeostasis and differ in their structures and functions. In some areas of the body, we see differing skin thicknesses depending on the body location. We see a thicker layer of skin on hairless skin of the palms of hands and soles of feet as there is an additional epidermal layer known as the stratum lucidum, which is a layer of dead skin cells (Yousef et al. 2022). To mitigate against the environment, the epidermis undergoes a process of terminal differentiation leading to a water-tight barrier. The constant turnover of epidermal keratinocytes prevents infection and the acquisition of lasting

DNA mutations. At the basal layer, just above the interface of the epidermis and dermis, reside long-lasting keratinocyte stem cells (KSCs). KSCs undergo self-renewal, and through asymmetric division, give rise to progeny that are destined for terminal differentiation. Also interspersed in the basal layer are melanocytes, which provide pigmentation to the keratinocytes to protect against the DNA-damaging effects of ultraviolet light (Yousef et al. 2022).

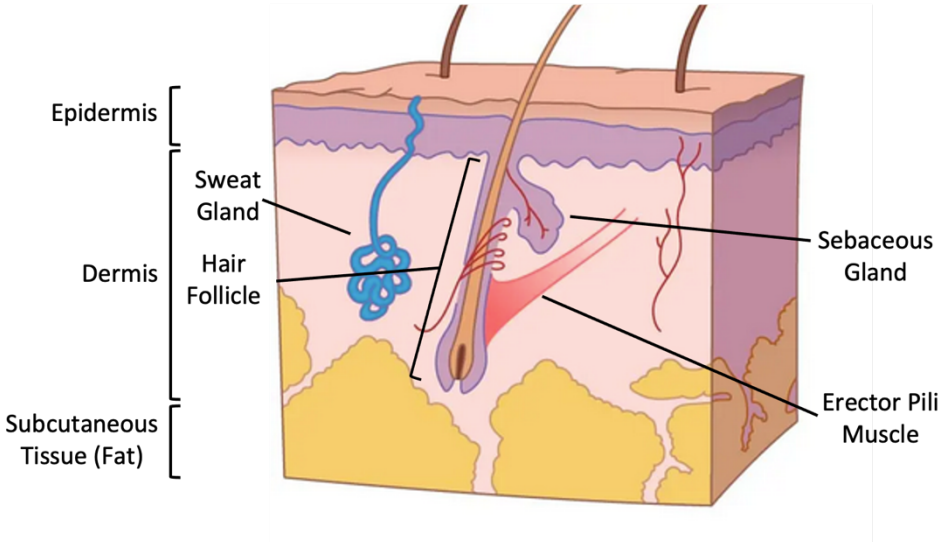


Figure 1.5 | Diagram of the layers and components of human skin. *The skin is separated into two main sections, the outermost epidermis, and innermost dermis. Below the dermis is the subcutaneous tissue layer. Within the epidermis we see the hair follicle that protrudes into the dermis. In the dermis we see the erector pili muscle that attaches to the hair follicle, and the sweat glands that protrude through the epidermis and to the surface of the skin. Figure has been adapted from Cobb and Watson 2018.*

1.6 Hair follicle

The functional anatomy of the mature hair follicle (HF) in anagen phase has been extensively studied and can be divided into two regions longitudinally. The first section is the upper HF, which is permanent and does not visibly cycle. This region is composed of the infundibulum which is the opening of the hair canal to the skin and isthmus. The second section, the lower HF, undergoes continuous remodelling during every round of the hair cycle and represents the hair factory, the anagen bulb (Figure 1.6a; Schneider et al. 2009).

At the proximal end of the infundibulum, we see the joining to the isthmus region of the outer root sheath (ORS). This is also the site at which the erector pili muscle attaches. Within the isthmus, we see a population of melanocytic and epithelial HF stem cells that is known as the bulge region (Figure 1.6b). This region marks the end of the permanent, non-cycling region of the hair follicle. Separating the bulge from the lowest region of the HF, the anagen bulb, is the suprabulbular HF epithelial region. Within the anagen bulb, we see a matrix of keratinocytes and a HF pigmentary unit (de Groot et al. 2021). The activation of these matrix keratinocytes causes them to rapidly proliferate, and their number determines the size and diameter of the bulb and hair shaft, respectively. Upon termination of matrix keratinocyte proliferation, differentiation occurs which leads to the generation of the various cell lineages of the hair shaft and inner root sheath (IRS). The ORS however, is derived from a separate population of progenitor cells (Hsu et al. 2011; Figure 1.6c, d).

As we can see from Figure 1.6d, the bulb region contains around 8 concentric layers, each with a different population that contains a unique keratin pattern. From the outside to the inside, we have the connective tissue sheath (CTS), ORS, the four layers that consist of the IRS, companion layer, Henle's layer, Huxley's layer, and the IRS cuticle, then finally the hair cuticle, cortex, and medulla (that make up the hair shaft). The ORS contains epithelial HF stem cells that are located in the bulge region. Numerous studies have been able to isolate and study the various populations within the HF through cell surface protein expression. These HF stem cells have been postulated to be the cells of origin for terminally differentiated cells in adult tissues and are identified through expression of CD200 and KRT15 (Rompolas and Greco 2014).

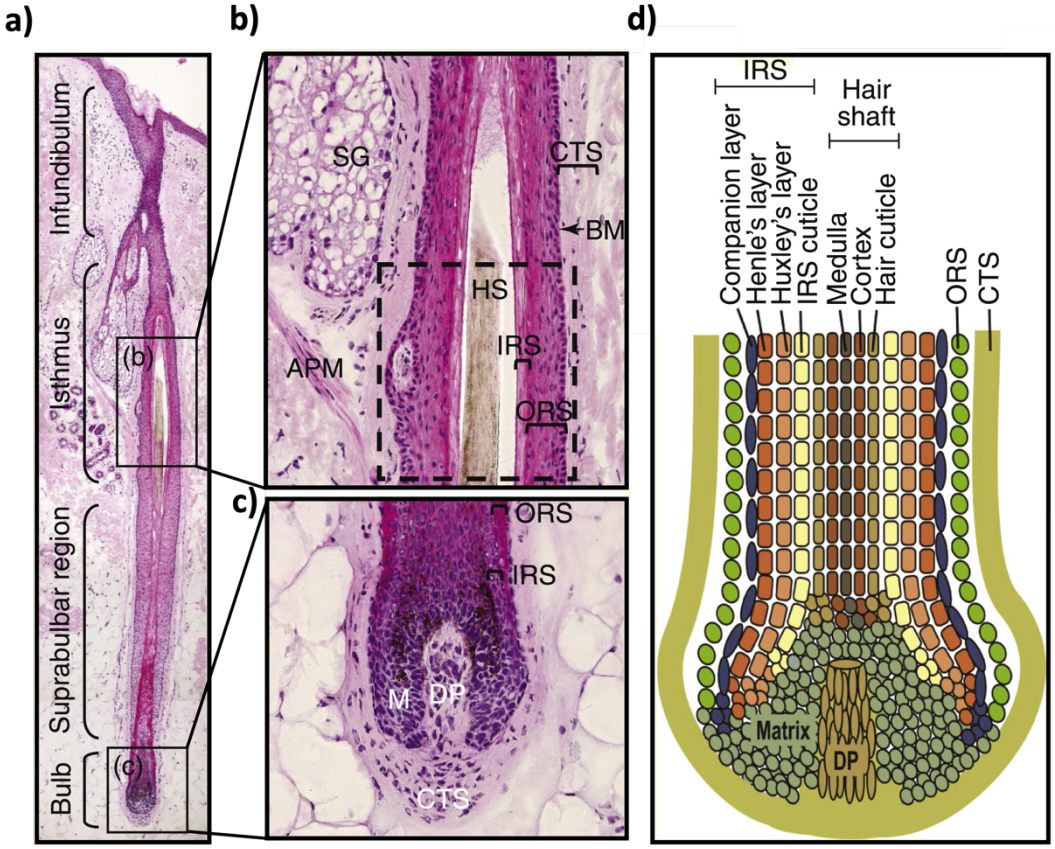


Figure 1.6 | Histomorphology of the hair follicle. a) *Sagittal section through a human scalp hair follicle (anagen VI) showing the permanent (infundibulum, isthmus) and anagen associated (suprabulbar and bulbar area) components of the hair follicle.* ***b)*** *High magnification image of the isthmus. The dashed square indicates the approximate location of the bulge.* ***c)*** *High magnification image of the bulb.* ***d)*** *Schematic drawing illustrating the concentric layers of the outer root sheath (ORS), inner root sheath (IRS) and shaft in the bulb. (BM: basal membrane; APM: arrector pili muscle; CTS: connective tissue sheath; DP: dermal papilla; M: matrix; HS: hair shaft, IRS: inner root sheath; ORS: outer root sheath; SG: sebaceous gland). Figure adapted from Schneider et al. 2009.*

1.7 Cutaneous Basal Cell Carcinoma

Human cutaneous Basal Cell Carcinoma (BCC) together with cutaneous Squamous Cell Carcinoma (cSCC) arise from skin keratinocytes, collectively referred to as keratinocyte carcinoma (previously called non-melanoma skin cancers); combined, they are the most diagnosed cancer in the world. Previously a disease of the elderly, the incidence of both keratinocyte carcinomas has continued to increase annually over the last three decades, and now increasingly affects younger adults. Early detection and surgical excision are often curative, as is common practice in the U.K., therefore the metastatic rate remains low with approximately 250 deaths per annum. Once surgery is completed, patients are followed up 4-6 monthly for up to 5-years.

1.7.1 Incidence

BCC is the most common skin cancer, with previous rates from roughly 3.6 million new cases diagnosed and 3000 deaths annually in the United States (Rogers et al. 2015; Cameron et al. 2019a; Lashway et al. 2023) to an estimated 5.4 million new cases (Naik and Desai 2022).

The metastatic and mortality rates for non-melanoma skin cancers (NMSCs) are very low, with the incidence of metastases in BCC and SCC ranging between 0.00281-0.05% and 0.5-16%, respectively, and the age-adjusted mortality rate estimated to be 0.12/100,000 person years for BCC and 0.3/100,000 person years for SCC (Rogers et al. 2015; Apalla et al. 2017). Despite the relatively low malignant potential of both

NMSCs, they are however associated with remarkable morbidity and a substantial cost to the healthcare system (Hay et al. 2014; Perera et al. 2015).

These numbers are, in reality, likely to be substantially different due to the fact that NMSCs are often not reported in global total cancer cases. The availability and quality of epidemiological data for BCC in the U.K has historically been of poor quality due to the high volume and multiplicity of cases. Coupled with the fact that most BCCs are under-diagnosed and usually treated in primary care facilities, this has led to the data frequently being omitted from national cancer registries and statistics (Venables et al. 2019).

1.7.2 Clinical Presentation

BCC is derived from epidermal cells at basal layer of either interfollicular or hair follicles. BCC has been classified by the World Health Organization (WHO) into a number of subtypes. The most common subtypes seen in the clinic are nodular, superficial, and morpheaform BCCs. The less common BCC subtypes are infundibulocystic, fibroepithelial, infiltrative, micronodular, and basosquamous BCCs (Reiter et al. 2021). Table 1.1 shows the clinical and histological features of the various BCC subtypes. In addition to these subtypes, the WHO have also classified BCCs into risk categories, which is the risk for incomplete resection of free margin, recurrence, and locally aggressive capability. These risk categories can be beneficial for therapeutic intervention. High risk BCC subtypes are morpheaform, infiltrative, micronodular, and basosquamous BCCs, whilst the low risk BCCs are nodular and superficial (Puig and Berrocal 2015; Castrisos et al. 2021; Ciążyńska et al. 2021). An

additional aggressive histological BCC feature was also added by Cameron and colleagues in 2019, which describes perineural invasion that is found in both BCC and SCC but is predominantly observed in SCC and accounts for 2-6% of all NMSCs (Cameron et al. 2019).

Table 1.1 | Clinical and histological features of BCC subtypes. *Table shows the clinical and histological features of low-risk (nodular, superficial, infundibulocystic, and fibroepithelial) and high-risk (morpheaform, infiltrative, micronodular, and basosquamous) BCC subtypes.*

BCC subtypes	Clinical features	Histological features
Low Risk		
Nodular (50-80%)	-Shiny, pearly papule or nodule with a smooth surface, rolled borders, and arborizing telangiectasias -Sharp border (a diagnostic clue) -predilection for the head and neck -If large and ulcerated, called "Rodent ulcer" (ulcus rodens) -If melanin-containing, called pigmented BCC -Originated from follicular bulge	-Discrete nests of malignant basaloid cells in the dermis -peripheral palisading -mucoid/mucinous stroma containing plump spindle cells -If pigmented, presence of increasing melanocytes, numerous melanophages and slight melanin-containing
Superficial (10-30%)	-Well-circumscribed and erythematous thin plaque or patch with central clearing, and thin rolled borders -Usually with scale -Leave atrophic or hypopigmented scar when regression -May present with irresponsive eczematous patch -mainly on the trunk, legs - If melanin-containing, called pigmented BCC -Originated from interfollicular epithelium	-Multiple lobules of basaloid cells with a myxoid stroma attached superficially to the epidermis -Prominent peripheral palisading -band-like lichenoid infiltrate
Infundibulocystic	-Well-circumscribed pearly papule -usually on the elderly head and neck	-Well-circumscribed, anastomosing strands of basaloid cells and scattered infundibulum -like cystic structures
Fibroepithelial	-an indolent skin-colored or erythematous sessile plaque or pedunculated papulo -nodule -mainly on the trunk (esp. lower back)	-Multiple collections of delicate strands of epidermal basaloid keratinocytes arranged in a reticular pattern within a spindle cell stroma and abundant collagen
High Risk		
Morphoeaform (>10%)	-Ivory-white, shiny, smooth Infiltrated plaque with poorly defined edges -often spread subclinically -commonly found on the head and neck	-Thin cords of basaloid cells surrounded by a sclerotic collagenous stroma -mostly absent peripheral palisading and stromal retraction -positive staining of tumor stroma with smooth muscle alpha-actin
Infiltrative	-Poorly defined, indurated, flat or depressed plaque with white, yellow, or pale pink color – may have overlying crusts, erosions, ulcerations, or papules	-Thin cords with angulated ends of few atypical basaloid keratinocytes, varying in sizes and shapes surrounded with a typical stroma
Micronodular (15%)	-Erythematous macule or thin papule/plaque	-Multiple small aggregates of basaloid cells (<15 µm) within the dermis, with minimal peripheral palisading, myxoid stroma and retraction artifact
Basosquamous (<2%)	- Well-defined nodular or superficial BCC with infiltrative features -Majority found on the head and neck	-both BCC and SCC histologic features presented in a continuous fashion -Atypical basaloid cells with variable eosinophilic features, prominent mitotic function and numerous apoptotic cells -Presence of transitional zone between BCC and SCC (squamous intercellular bridge formation and cytoplasmic keratinization)

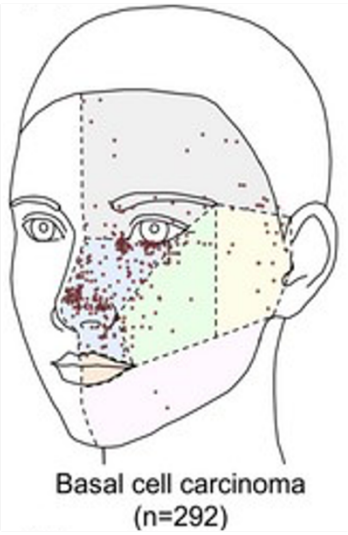
Nodular is the most common BCC subtype in both males and females, followed by superficial and infiltrative BCC subtypes (Castrisos et al. 2021; Ciążyńska et al. 2021). We see that the superficial BCC is the most common subtype founder in younger patients (67.7 +/- 11.5 years VS other subtypes: 69.3 +/- 12.7 years, $p<0.01$), has a female predominance, and is most commonly found on the trunk. Other BCC subtypes appear to have a predominance to arise in the head and neck region, but show no preferences between age and sex (Ciążyńska et al. 2021). This confirmed that variable causal relationship between sun exposure and superficial BCC subtype is related to intermittent intense UV radiation in earlier ages, whereas other subtypes align with chronic consistent UV exposure.

There is also a link between anatomical site and BCC subtype. Most BCCs occur on the face and account for 81% of high risk BCCs and 57% of low risk BCCs. The second most common anatomical region is the trunk, accounting for 10% of high risk BCCs and 26% low risk BCCs (Ciążyńska et al. 2021). The face is favored by the nodular and infiltrative BCC subtypes, whilst the trunk is favored by the superficial BCC subtypes (Castrisos et al. 2021). In the facial region, BCC occurs most commonly on the nose (29.7%), followed by eye (16.2%), cheek (15.8%), temple (12.0%), forehead (11.4%), earlobe (9.1%), lips (3.2%), chin (1.4%), nasolabial folds (0.62%), and the jawline (0.62%).

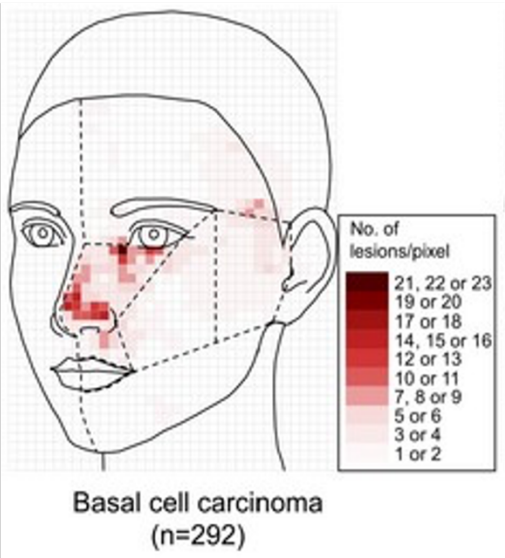
In 2021, Omodaka and colleagues performed a retrospective study on UV-related skin cancer distribution on the face (Omodaka et al. 2021). The authors collected facial BCC samples from 292 Japanese patients (139 men, 153 women, mean age = 75 years), SCC samples from 118 patients (56 men, 62 women, mean age = 78), and

melanoma samples from 51 patients (23 men, 28 women, mean age = 78) from 2000 to 2017 and found that although these skin cancers are UV-induced and are distributed on the median face, which is a UV sensitive location, their dominant areas were different. BCC arose on the central area of the median face, whilst SCC and melanoma showed occupation on the external and internal areas of the median face, respectively (Figure 1.7). However, this was a single centre study reporting on a relatively small population of patients but implies the disease-specific aetiological factors apart from UV exposure.

a)


Basal cell carcinoma
(n=292)

b)


No. of
lesions/pixel
21, 22 or 23
19 or 20
17 or 18
14, 15 or 16
12 or 13
10 or 11
7, 8 or 9
5 or 6
3 or 4
1 or 2
Basal cell carcinoma
(n=292)

c)

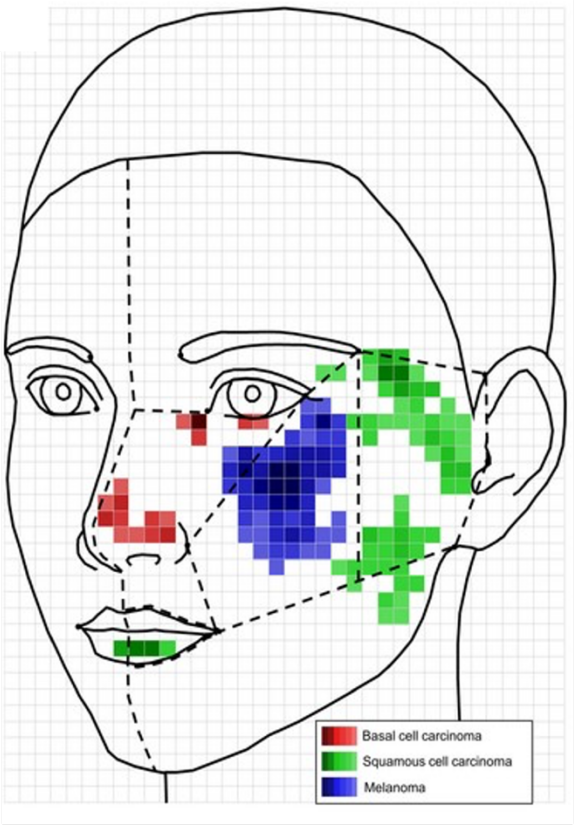

Basal cell carcinoma
Squamous cell carcinoma
Melanoma

Figure 1.7 | Distribution of BCC on the face surface. a) *The distribution of lesional centers is shown. The six areas of the face surface are arbitrarily divided by straight lines between key anatomical structures.* ***b)*** *The density of each lesion is shown. The number of lesions per pixel is shown on a colour scale representing 10 levels. The lesional areas beyond the midline were not consolidated onto one side to avoid the artefact of aggregation of lesion-positive pixels near the midline.* ***c)*** *The three patterns in* ***b)*** *are superimposed including those from SCC and melanoma. Pixels with a low number of lesions/pixel [i.e. the four lightest hues in* ***b)****] are not colored. Figure and legend adapted from Omodaka et al. 2021.*

1.7.3 Aetiology

1.7.3.1 Age

The ratio of BCC to SCC is greater in younger individuals 18-39 but decreases with age, with the ratio being 2.92 in individuals 40-64 and 1.33 in individuals 65> (Lukowiak et al. 2020). As patients get older, the odds of a NMSC becoming a BCC is decreased in both sexes. Nonetheless, the mean onset of a BCC is in patients aged 70.1 years, based on a study of 11,548 patients ranging from 1999-2019 (Ciążyńska et al. 2021). The hypothesised reason for this is that a reduced ability to repair UV-induced DNA damage is seen with age (Verkouteren et al. 2017).

1.7.3.2 Previous History of BCC

Having a history of previous BCCs, unsurprisingly, is a strong risk factor for the development of further BCCs, with a 10-times greater risk than normal populations (Fania et al. 2020). Studies have shown that patients with BCC have a 30-50% increased risk of developing an additional BCC within 5 years (Marcil and Stern 2000). Additionally, a strong risk factor is seen in patients who have a family history in skin cancers, with an estimated odds ratio of 2.2 (Vitasa et al. 1990).

1.7.3.3 Genetic Factors

1.7.3.3.1 Inheritance

A predisposition to the development of keratinocyte carcinomas has been seen by the inheritance of genes. As aforementioned, studies have shown that patients who

have a positive family history of skin cancers are at an increased risk of BCC development. A more recent study showed that a positive family history of skin cancer was seen in 17% of BCC patients (Choquet et al. 2020).

1.7.3.3.2 Genetic Predisposition

With the advent of genome wide association studies (GWAS), numerous genetic determinants of BCC have been identified (Choquet et al. 2020). A number of single nucleotide polymorphisms (SNPs) have been shown to contribute to allelic variations which subsequently leads to the tendency of BCC tumorigenesis. A table of BCC-associated SNPs that have been identified in up to 11% of inherited BCCs can be found in (Chahal et al. 2016).

1.7.3.3.3 Genetic Syndromes

There are 19 genetic syndromes that have been linked to an increased susceptibility of BCC (Table 1.2; Kilgour et al. 2021). Studies on Gorlins syndrome in the 1990's paved the way to the understanding of BCC pathogenesis. The hallmark of Gorlins syndrome is the presence of multiple BCCs, with the median number being 8, but some patients can have >1000 (Jaju et al. 2016). Investigative studies have found that patients inherit a defective PTCH1 allele, which leads to the overactivation of the hedgehog signalling pathway (Hahn et al. 1996; Johnson et al. 1996). Basex-Dupre-Christol syndrome, a rare X-linked recessive genetic syndrome, includes BCC as one of the defining features during diagnosis of this syndrome, which sees the BCC development in 50% of cases before the age of 20 (AlSabbagh and Baqi 2018). Likewise, Rombo syndrome, a rare autosomal dominant disorder, also sees patients

develop multiple BCCs (Van Steensel et al. 2001). Xeroderma Pigmentosa, an autosomal recessive disorder, renders patients with UV hypersensitivity and impaired DNA repair mechanisms and an extremely high risk of skin cancer development. Some studies have shown that patients have up to a thousand times higher risk of BCC development (Cameron et al. 2019; Choquet et al. 2020; Brambullo et al. 2022).

Syndrome	Gene(s)	Gene Function(s)
Gorlin syndrome	PTCH1, SUFU, PTCH2	HH pathway members
Bazex-Dupré-Christol syndrome	UBE2A, ACTRT1	DNA repair, Regulation of cell cycle, HH pathway
Rombo syndrome	Unknown	Unknown
Generalized follicular basaloid hamartoma syndrome	Unknown	Unknown
Happle-Tinschert syndrome	Unknown	Unknown
Muir-Torre syndrome	MSH2, MLH1, MSH6, PMS2	DNA mismatch repair
Brooke-Spiegler syndrome	CYLD	NF-κβ and EGFR pathways regulator
Cowden syndrome	PTEN	PI3K-AKT signaling pathway
Cartilage-hair hypoplasia	RMRP	Immune response
Schimmelpenning syndrome	Unknown	Unknown
Phacomatosis pigmentokeratotica	Unknown	Unknown
Xeroderma pigmentosum	XPA-XPG, XPV, POLH	Nucleotide excision repair
Bloom syndrome	BLM (REC0L3)	Chromosomal stability
Werner syndrome	WRN (REC0L2), LMNA	Chromosomal stability
Rothmund-Thomson syndrome	REC0L4, C160rf57	Chromosomal stability
Schopf-Schulz-Passarge syndrome	WNT10A	WNT/β-catenin signaling pathway, cell proliferation and migration
Epidermodysplasia verruciformis	TMC6 (EVER1), TMC8 (EVER2)	Immune response and signal transduction in the endoplasmic reticulum
Oculocutaneous albinism	TYR, OCA2, TYRP1, SLC45A2 (MATP), SLC24A5, C10orf11, 4q24	Melanin synthesis
Hermansky-Pudlak syndrome	HPS1-HPS8	Melanin synthesis

Table 1.2 | Genetic syndromes linked to increased susceptibility to BCC. *Table shows the genetic syndromes and the genes and their associated functions that burden the individual with increased susceptibility to BCC development.*

1.7.3.4 Ultraviolet Radiation

The most important risk factor for the development of BCC is cumulative ultraviolet (UV) radiation from the sun and/or tanning beds. BCC is the most mutated cancer, with a tumour mutational burden of up to 65 mutations per megabase (Sharpe et al. 2015). A direct association between BCC incidence and UV radiation has been reported, where it has been described as a primary risk factor for the occurrence and development of BCC (Olsen et al. 2015). DNA does not absorb UVA, but instead causes its damaging effects through the production of reactive oxygen species (ROS) and ROS-dependent DNA damage (Gloster and Brodland 1996). UVB is strongly absorbed by aromatic rings within DNA bases and the adjacent pyrimidines are subsequently susceptible to dimerization. This in turn leads to the formation of cyclobutene pyrimidine dimers, and if improperly repaired, can cause C>T, and CC>TT base transitions (Rodust et al. 2009). These mutations accumulate at the dipyrimidine sites and are known as a UVB signature mutations (Nishigori 2006). Increased incidence in BCC is usually seen within the aging population, where cumulative long-term exposure is seen. However, studies have also shown that intense intermittent exposure to UV can also increase the risk of BCC (Kricker et al. 1995; Rosso et al. 1996; Teng et al. 2021). Population-based studies have shown that skin cancer rates, including BCC, are higher in areas that are in close proximity to the equator, where higher levels of UV are seen (Chuang et al. 1990; Reizner et al. 1993). Furthermore, with the gaining popularity of indoor tanning and beauty products such as tanning lamps and nail polish dryers amongst the younger population, we also have seen an increase in skin cancer incidence (Green et al. 2007; Schulman and Fisher 2009; Zhivagui et al. 2023). All studies show that UV radiation, whether obtained from the sun or from an artificial source, plays a massive part in the incidence of BCC.

1.7.3.5 Medications

1.7.3.5.1 Photosensitizing Medications

Photosensitizing medications (PSMs) are frequently prescribed to the elderly to treat chronic conditions, who are also simultaneously at a higher risk of BCC development (George et al. 2021). Studies have shown that drug metabolites from PSMs may react with UV radiation and result in ROS production. Multiple meta-analyses have shown a slight increase in the risk of BCC, for example, the risk factor of using some diuretics and development of BCC is 1.17 (Bendinelli et al. 2019). Similarly, studies on beta-blockers have shown a marginal increase BCC risk of 1.09 (Tang et al. 2018). Interestingly, studies on non-steroidal anti-inflammatory drugs (NSAIDs) suggest a chemo preventative role, with a BCC risk of 0.94 (Ma et al. 2020b).

1.7.3.5.2 Immunosuppressive Agents

Patients who have had an organ transplant receive immunosuppressive medication to prevent organ rejection, such as Tacrolimus which is used to inhibit calcineurin activity and subsequently suppresses the activity of T cells and organ rejection in patients who have received transplanted organs such as kidney, heart, liver, and lung (First 2004). There have been numerous studies linking organ transplant recipients (OTRs) to an increased risk of skin cancer. When compared to the general population, OTRs are at an overall 2-fold risk of cancer incidence. When looking specifically at NMSCs, SCC has a risk of 65- to 250-fold increased risk, and BCC has a 10- to 16-fold increased risk (Collett et al. 2010; Mittal and Colegio 2017).

1.7.3.6 Ionizing Radiation

There have been a plethora of studies evaluating the mortality and incidence of cancers in occupationally exposed people that work in the healthcare and nuclear industries (Gilbert 2009). These workers are at risk of exposure to ionising radiation, such as those who work with X-rays or those who work with nuclear material. One such study found the risk of BCC and SCC in patients who had a history of radiotherapy (non-diagnostic therapeutic ionizing radiation) was 5- to 6-fold (Lichter et al. 2000). There is an estimated 7 million healthcare workers that are professionally exposed to ionising radiation worldwide (Russo et al. 2012). Cardiologists are believed to be at an estimated 3 times higher exposure rate than conventional radiologists as they work closer to the X-ray source. A study on 27,011 radiologists and technicians in China (working with X-rays) revealed a 4.1 relative risk of them developing all cancers compared to physicians working at the same hospitals, whereas those working for more than 15 years showed the highest risks of developing skin cancer (Wang et al. 1990; Wang et al. 2002). A more recent study on 65,719 Caucasian radiology staff also confirmed an increased risk of BCC development if the dosages were received before 1960 or before the age of 30 (Lee et al. 2015). Several studies have assessed cancer incidences in nuclear-related workers. One study looking at skin cancer incidence in workers from a nuclear power plant in Russia found a significant increase in NMSC incidence in workers who were exposed to ionising radiation at cumulative doses > 2.0 Sv (Sievert; a unit representing the stochastic health risk of ionising radiation) when compared to a reference dose of 0 – 0.5 Sv (Azizova et al. 2018). Furthermore, risk of BCC was significantly associated with cumulative external gamma dose with ERR/Sv 0.49 (excess relative risk per Sv).

1.7.3.7 Other Occupational Risks

Certain occupations also come with an increased risk in skin cancer. Those who work in the outdoors, such as gardeners and lifeguards are certainly at risk of developing skin cancers due to their chronic exposure to UV radiation. However, those working in occupations such as tanning salons, nuclear, healthcare, and ground works are also at risk. Tanning salon workers are believed to be at an increased risk of developing skin cancers due to their working with, and use of sun beds (Dessinioti and Stratigos 2022). In healthcare, those who work with radiating sources, such as X-rays, are also at an increase of developing skin cancer. Interestingly, an increase in skin cancer risk has been seen in welders. It is estimated that there are 11 million people working as welders worldwide. These welders are exposed to the full spectrum of UV radiation whilst working, and often experience burns or cutaneous erythema (Falcone and Zeidler-Erdely 2019). The most recent study demonstrated a significantly increased risk of BCC in the head and neck of these workers when compared to the control cohort (Heltoft et al. 2017).

1.7.3.8 Arsenic Exposure

Arsenic is a well-known carcinogen and has been attributed to multiple cancers, including skin cancer (Martinez et al. 2011). Arsenic can enter the body via ingestion, respiration, or via the skin (Huang et al. 2019). The most common source of arsenic exposure is through contaminated water. Other sources of exposure include ingestion of food that has been grown with contaminated water, smoking cigarettes, industrial manufacturing and wood preservation, and cosmetics (Lu 1990; Lim et al. 2018; Huang et al. 2019). The recommended threshold levels of arsenic

concentration in drinking water are ≤10 µg/L, but chronic exposure exceeding this limit has been reported worldwide. Most recent figures show that there are more than 100 million people that are exposed to concentrations of > 50 µg/L (Moon et al. 2012). Arsenic produces profound effects in multiple organs, but the skin is considered to be the most susceptible organ and is usually the first manifestation of toxicity (Rahman et al. 2009; Mayer and Goldman 2016). Studies have indicated the role of arsenic, in certain areas, to the formation of NMSCs. One such study of a Bangladesh population observed a dose-response relationship between arsenic and skin toxicity and following up with the same patients 6 years later revealed 1.6% of those patients developed BCC (11.2/700 patients). This translated to approximately 266 new cases/100,000 patient years in this population (Kibriya et al. 2022).

1.7.3.9 Skin Type and Gender

The BCC incidence rate was greater in men (360/100,000) than women (292.9/100,000) in a study based in Minnesota ranging from 2000-2010 (Muzic et al. 2017). In the UK, the incidence rate of BCC was higher in males than females from 2013-2015 (352/100,00 vs 219/100,000, respectively; Venables et al. 2019).

In younger individuals (<40), the incidence of BCC is seen more in women than men possibly due to the sun-seeking behaviour and a higher use of indoor tanning beds (Gangan 2022). The incidence rate of NMSCs is low in Asians, with a BCC incidence rate ranging from 1.5-15.57/100,000 people (Razi et al. 2015; Al-Qarqaz et al. 2018). The hypothesised reason for this is due to increased melanin in these individuals which asserts its protective function against UV damage. Interestingly, the incidence of SCC is higher than BCC in darker skinned populations (Talvalkar 1970; Fleming et al. 1975). Studies have also shown that individuals with light skin, light coloured eyes,

and red hair are at an increased risk of BCC (Khalesi et al. 2013; Harken and Fazio 2022; Lashway et al. 2023).

1.7.3.10 Pre-Existing Skin Disease

There have been many studies investigating the effect of inflammatory conditions on the incidence of skin cancers. Chronic inflammation is responsible for the development and progression of approximately 20% of all cancers (Tang and Wang 2016). As aforementioned, UV radiation can cause DNA damage and ROS generation, which can lead to an inflammatory response in the skin. This can also lead to the various cell populations of the skin to secrete pro-inflammatory cytokines to further add to insult (Sollberger et al. 2015). There are two types of inflammation stages, chronic and acute, and studies have linked chronic inflammation to tumorigenesis in the skin (Mantovani et al. 2008; Neagu et al. 2015; Neagu et al. 2019). A more recent study has reviewed the risk of BCC development in a range of chronic inflammatory conditions. The authors found that multiple studies confirmed that patients with atopic dermatitis, scleroderma have a significantly increased risk of BCC development whilst patients with psoriasis, hidradenitis suppurativa, prurigo nodularis had a nonsignificant increased risk (Ju et al. 2023).

1.7.3.11 Vitamin D

Vitamin D is a lipophilic vitamin produced by UV radiation from the sun and is also found in certain foods. It is known to have a role in a plethora of beneficial functions such as bone health, anti-proliferation, anti-angiogenic, and immune system modulation (Abdelwahab et al. 2022). It also has a protective role against several

cancer types. However, most recent studies have indicated vitamin D with an increased skin cancer risk. Mahamat-Saleh and colleagues found during a large meta-analysis that circulating levels of 25(OH)D, a biomarker of vitamin D status which reflects both intake and synthesis in response to direct sun exposure, was associated with higher risks of keratinocyte carcinomas, with the relative risk of BCC being 1.41 (Mahamat-Saleh et al. 2020). An additional area of study is on the genes that encode the vitamin D receptor (VDR). VDR is expressed on a plethora of cells, including epidermal basal keratinocytes (Vornicescu et al. 2020). Upon binding to the VDR, we see expression of VDR-responsive genes which exert the effects of the vitamin D. Interestingly, the VDR is highly polymorphic, with a > 618 variants. Most of these variants have been studied and are linked to increased skin cancer risk, particularly the FokI, ApaI, BsmI, and TaqI variants (Lesiak et al. 2011; Gnagnarella et al. 2014).

1.7.4 Pathogenesis of BCC

A hallmark of BCC is the dysregulation of the Sonic Hedgehog (SHH) growth factor signalling pathway (Epstein 2008). As aforementioned, the pathways' role in BCC formation was first identified during studies of Gorlins syndrome, an autosomal-dominant genetic disease that results in BCC (Johnson et al. 1996; Sinx et al. 2018). Studies showed that germ-line mutations in the Patched-1 (PTCH1) receptor led to multiple BCC's. Further studies into the SHH pathway in BCC have shown that constitutive activation of downstream signalling members also results in BCC formation, confirming a pivotal role of SHH in BCC (Fan et al. 1997a; Oro et al. 1997; Xie et al. 1998; Grachtchouk et al. 2000).

1.7.4.1 Hedgehog Signalling Pathway

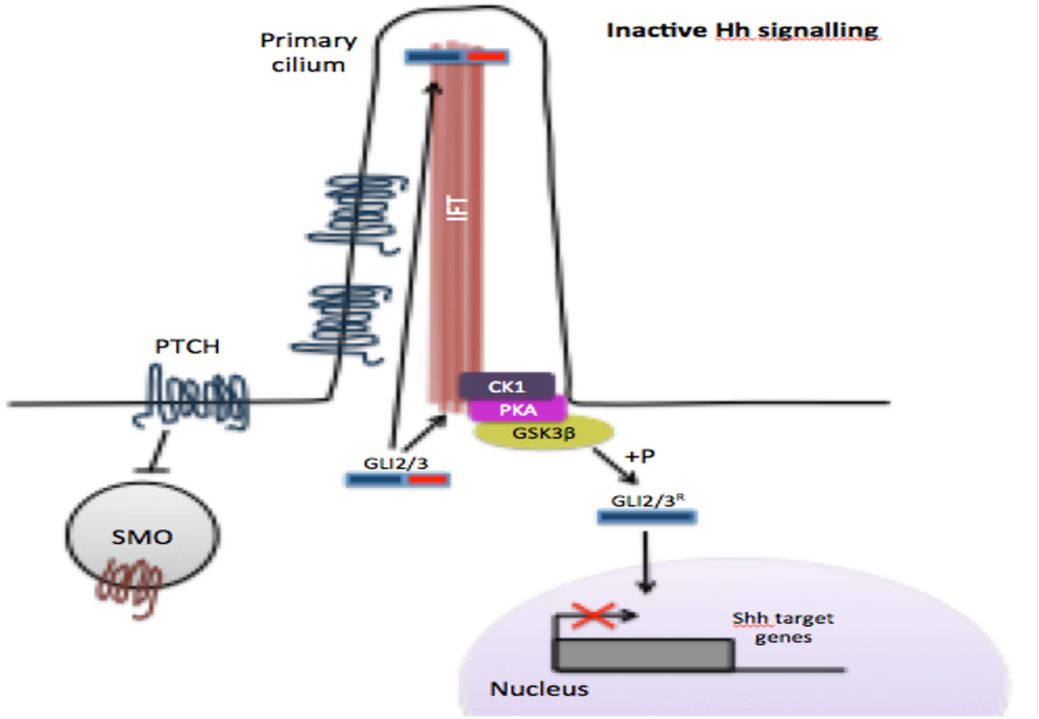
The Hedgehog (Hh) growth factor cell signalling pathway was initially identified as a 'segment-polarity' gene which is evolutionary conserved and essential during development (Figure 1.8; Nüsslein-Volhard and Weischaus 1980; Ingham and McMahon 2001; Weedon et al. 2008; Zhao et al. 2009). In vertebrates, the pathway is activated by three secreted hedgehog ligands: Sonic Hedgehog (SHh), Indian Hedgehog (IHh) and Desert Hedgehog (DHh); of which SHh is the key morphogenetic factor (Mann and Beachy 2004). All three ligands bind a single twelve-transmembrane receptor protein, Patched (Ptch), on the recipient cell (Hooper and Scott 1989). In the absence of ligand binding, Ptch acts as a negative regulator of a seven-transmembrane G-protein coupled receptor smoothened (Smo; Alcedo et al. 1996; Van den Heuvel and Ingham 1996; Stone et al. 1996). However, upon binding of Hh, Ptch relinquishes inhibition of Smo leading to Smo phosphorylation by CK1 and GRK2, which is aided by the co-receptor Cdo (Denef et al. 2000). Smo is regulated through multi-site phosphorylation in a dose-dependent manner, which regulates both the subcellular localization and conformation (Chen et al. 2011). The kinases CK1 and GRK2 are both required for the phosphorylation and activation of Smo at six Serine/Threonine clusters found in the Smo C-terminus (Sheng et al. 2004; Meloni et al. 2006; Chen et al. 2011). This phosphorylation induces the ciliary localization of Smo through the recruitment of β-arrestins that link Smo to the anterograde kinesin-II motor (Kovacs et al. 2008; Chen et al. 2011). CK1/GRK2-mediated phosphorylation of the C-terminus that results in dimerization/oligomerisation of the Smo C-tail induces a conformational switch (Zhao et al. 2007). In the absence of ligand, Smo adopts an inactive conformation, which masks the kinase binding sites from CK1/GRK-2 (Chen et al. 2011). Furthermore, CK1 is accumulated in the primary cilia

in response to Hh ligand and therefore enhances the phosphorylation of Smo (Chen et al. 2011).

Suppressor of fused (Sufu), a negative regulator, forms complexes with Gli proteins to inhibit their ability to translocate into the nucleus and thus drive gene transcription (Ding et al. 1999; Cheng and Michael Bishop 2002; Merchant et al. 2004; Barnfield et al. 2005). However, upon Hh ligand binding this complex dissociates and thereby relinquishes the full-length Gli proteins (Humke et al. 2010; Tukachinsky et al. 2010). Several kinases have been implicated in the regulation of Gli activity, including Cdc211, which binds Sufu to block inhibition of the pathway (Evangelista et al. 2008).

Thus, the Hh signal transduction pathway leads to the activation of the Gli zinc-finger transcription factors. In vertebrates, there are three Gli proteins. Gli2 and 3 are activated by SMO, while Gli1 is transcriptionally induced by pathway activation and therefore acts in a positive feedback loop to reinforce Gli activity (Hui and Angers 2011). In the absence of ligand, Gli proteins undergo proteolytic processing into a truncated form (GliR), which also functions as a transcriptional repressor (Figure 1.8a). Hedgehog signalling inhibits Gli processing and converts the full length GliF into the active form GliA, which then acts as a transcription factor inducing downstream target genes (e.g. GLI 1, Gli2, K17, PDGFRα, PTCH1, BCL2) (Figure 1.8b).

a)


Primary cilium
Inactive Hh signalling
IFT
PTCH
CK1
PKA
GSK3β
GLI2/3
+P
GLI2/3R
SMO
Shh target genes
Nucleus

b)

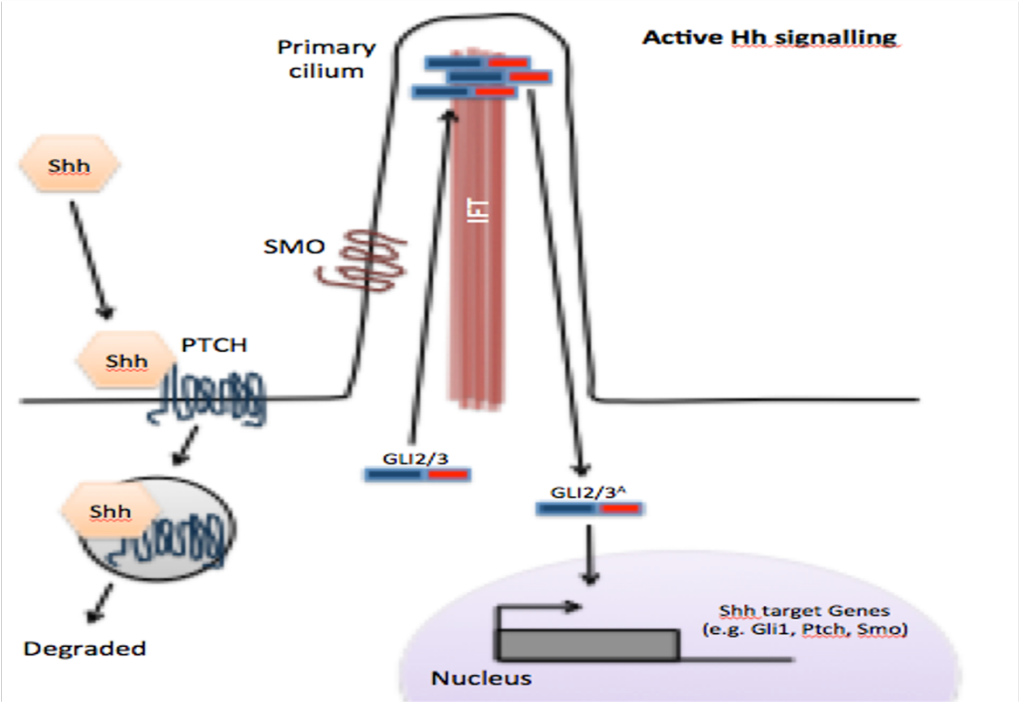

Primary cilium
Active Hh signalling
Shh
IFT
SMO
Shh
PTCH
GLI2/3
GLI2/3A
Shh
Shh target Genes (e.g. Gli1, Ptch, Smo)
Degraded
Nucleus

Figure 1.8 | Hedgehog growth factor signalling pathway. a) *In the absence of the Hh ligand, Patched (Ptch) localises to the primary cilia and inhibits Smoothened (SMO). SMO is then sequestered into endocytic vesicles. Intraflagellar transport (IFT) machinery shuttle GLI2/3 in and out of the primary cilium. Also, at the base of the IFT, GLI2//3 is phosphorylated by PKA, CK1, and GSK3β which results in cleavage of GLI2/3 activator domain (red) and generates GLI2/3R, a transcriptionally repressive form. Additionally (not shown in figure), suppressor of fused (SuFu) can form complexes with, and sequester GLI2/3, which stops their phosphorylation into the active form, repressing transcription.* ***b)*** *Binding of Hh to Ptch causes Ptch to be transported from the cilium and marked for degradation, subsequently relinquishing the inhibition of SMO. SMO is then phosphorylated by CK1 and GRK-2, causing ciliary localisation via the IFT machinery. Activated SMO then inhibits PKA. GLI2/3 is then accumulated in the cilium and phosphorylated by CK1 to form a transcriptionally active form, GLI2/3A, which can then enter the nucleus and cause transcription of Hh target genes. Additionally (not shown in figure), PKA inhibition by SMO is insufficient for the phosphorylation of the SuFu/GLI complex. Unphosphorylated SuFu/GLI complex is transported to the cilia tip via the IFT machinery where it marked for ubiquitinylation, causing the release of GLI2/3, which can then be phosphorylated by CK1 to form GLI2/3A. Adapted from Petrova and Joyner 2014.*

1.7.4.2 Driver DNA mutations in BCC

Similarly to Gorlins syndrome, mutations of PTCH1, SMO, and SUFU have also been found in sporadic BCC (Epstein 2008; Kool et al. 2008). All cases of sporadic BCC show evidence of hedgehog pathway activity, with 90% associated with a loss of PTCH1 (Gailani et al. 1996; Kim et al. 2002) and 10% with activating mutations in SMO (Xie et al. 1998; Reifenberger et al. 2005). Consistent with the allelic loss on chromosome 9q22 in BCC formation previously observed (Gailani et al. 1996). Therefore, the mutational basis for BCC and therefore the key driver mutations reside within the Hh pathway, of which the most frequently mutated gene is PTCH1.

1.7.4.3 Mouse models of BCC

Mouse models also support the role of the Hh pathway in BCC pathogenesis. Wild-type mice do not develop BCC, even after exposure to UV or ionizing radiation. However, transgenic mice in which one of the Ptch gene alleles has been knocked out, Ptch+/-, upon exposure to UV radiation frequently develop BCC (Aszterbaum et al. 1999). In this model, UV exposure is necessary to mutate the remaining Ptch allele (Mancuso et al. 2004). Furthermore, overexpression of the constitutively activated Hh signalling pathway members (SMO, GLI1 and GLI2) in transgenic mouse models or in human skin keratinocytes grafted onto mice, spontaneously develop BCC (Fan et al. 1997a; Fan et al. 1997b; Grachtchouk et al. 2000). Thus, mouse models support the fundamental role of the Hh pathway in the formation of BCC.

1.7.4.4 SMO antagonists in the treatment of BCC

BCC rarely metastasise in part due to their growth on the skin and so early detection is common. Surgical excision is the gold standard treatment. Providing surgery achieves clear histological margins, there is a <2% recurrence rate after 5 years (Griffiths et al. 2005; Walker and Hill 2006). For BCC that have extended beyond the ability to surgically excise them, or for those rare BCC that have metastasised, SMO antagonists have been developed. A phase 1 clinical trial of 33 patients with locally advanced or metastatic BCC treated with the Smo antagonist vismodegib, 2 patients experienced complete response and 18 had a partial response (Von Hoff et al. 2009). A subsequent nonrandomized study on 104 patients (71 locally advanced BCC and 33 metastatic BCC), found that vismodegib led to at least a partial response in 43% of locally advanced and 30% of metastatic BCC (Sekulic et al. 2012). A randomised, double-blind, placebo-controlled trial in Gorlins syndrome demonstrated reduced tumour burden (Tang et al. 2012). Similarly, Sonidegib (Novartis), a potent SMO inhibitor, showed a partial response in the treatment of locally advanced, metastatic, or Gorlins syndrome BCC (Skvara et al. 2011; Migden et al. 2015). In addition to the lack of primary response, continued treatment with SMO antagonists is also associated with secondary resistance (relapse); and occasionally with transformation into aggressive SCC (Chang and Oro 2012; Zhu et al. 2014; Saintes et al. 2015).

Secondary resistance to SMO antagonists have been attributed to insensitive protein variants, genomic bypass signalling (such as loss of SUFU or amplification of Gli2), and activation of alternative growth factor pathways (for example RAS/MAPK, PI3K, or aPKC) (Buonamici et al. 2010; Dijkgraaf et al. 2011; Kool et al. 2014; Atwood et al.

2015; Zhao et al. 2015). Thus, BCC growth appears not to be solely dependent on Hh pathway driver mutations.

1.7.4.5 DNA sequencing of BCC

In contrast to targeted DNA sequencing, whole exome sequencing (WES) has been undertaken on 6 BCCs from a single patient (Xie et al. 2022). These BCCs were taken from an 83-year-old Caucasian male who had an extensive history of sun exposure, including 3 superficial BCCs on his legs and 3 nodular BCCs on his head and neck region. A total of 41,946 somatic cancer-associated mutations when comparing all 6 tumours to the normal skin were found, with the Catalogue of Somatic Mutations in Cancer (COSMIC) UV radiation signature being the most dominant (27-62% of mutations). PTCH1 and TP53 were the most frequently mutated genes in 5/6 and 4/6 BCCs, respectively. CSMD1 mutations were observed in all nodular BCC, while CSMD3 and TPTE/PTEN2 mutations were in all superficial BCCs.

Atwood and colleagues undertook Whole Exome Sequencing to determine the genetic basis for BCC resistance to SMO antagonists (Atwood et al. 2015). As expected, mutations were identified in PTCH1 (8/14) and TP53 (4/14), however, there was no greater frequency associated with BCC resistance. Instead, the authors found mutations within the remainder of the Hh pathway in 15/29 (in 85% of samples), in SMO (6/14), cAMP/PKC (4/14), and GLI2 amplification. Direct sequencing determined that SMO mutations were seen in 77% of resistant BCCs (23/30) and 33% of sporadic-untreated BCCs (12/36).

1.7.4.6 RNA sequencing of BCC

To understand the biology of BCC growth, beyond the driver mutations, researchers have undertaken several different approaches, including transcriptomic analysis. In a study comparing adjacent normal skin with BCC, RNA sequencing identified upregulation of Hh signalling-associated target genes (Mackeh et al. 2018; Wan et al. 2021). In addition, differentially expressed genes demonstrated enrichment of hair follicle and epidermal stem cell gene sets.

Litvinov and colleagues undertook an analysis of three publicly available datasets, totalling 75 BCCs and 34 normal skin samples (Litvinov et al. 2021). The authors found KEGG gene set enrichment for: BCC signalling, Hh signalling, Wnt Signalling and IL-17 signalling. PANTHER gene set enrichment analysis also identified Wnt, integrin, and cadherin pathways. By generating differentially expressed genes between normal and locally advanced BCC, the authors determined that normal BCCs were enriched for target genes of the Hh pathway [PTCH1 (15/15 BCCs), SMO (15/15 BCCs), and PTCH2 (12/15 BCCs)], Wnt pathway [LEF1 (14/15 BCCs), FZD8 (14/15 BCCs), and DACT1 (14/15 BCCs)], IL-17 pathway [CXCL9 (14/15 BCCs)] and MMPs [MMP1 (13/15 BCCs) and MMP10 (10/15 BCCs)]. Advanced BCC was associated with additional gene set enrichment of KEGG pathways including PI3K-Akt signalling, TLR signalling, and proteoglycans in cancer.

1.7.4.7 Single Cell RNA Sequencing of BCC

Single-cell RNA sequencing has been used to understand the BCC tumour heterogeneity and importantly interaction between tumour cells and the microenvironment. There have been a number of single-cell RNA sequencing studies that have been published since 2020. These studies have addressed the tumour cell heterogeneity, tumour-associated fibroblasts, and the immune cell response.

1.7.4.7.1 BCC tumour cell heterogeneity

An analysis of 4 human BCCs, resulting in 45,656 cells, elucidated the cell heterogeneity within BCC, comprising of: keratinocyte clusters (characterised by keratin 14 expression and involucrin for differentiating keratinocytes), an endothelial cell cluster (characterised by PECAM1 expression), an immune cell population (characterised by PTPRC expression) and a fibroblast cluster (characterised by vimentin expression) (Figure 1.9; Yao et al. 2020). The isolated keratinocyte clusters were used to define cell surface proteins that could be used for the further isolation of vismodegib resistant cells characterised by MRTF expression. The most recent single cell study comes from Magnus Lynch's group, who employed multiscale spatial mapping of cell populations across normal skin and BCC tissues (Ganier et al. 2024). The study used 9 BCCs and 11 normal skin samples from the face, and 15 normal skin samples from the rest of the body, which generated a scRNAseq dataset consisting of 233,379 cells, though only 155,401 cells passed quality control thresholds. They subsequently identified 30 skin cell clusters which consisted of 16 cell types (Figure 1.10a). Their findings showed that in the face, there are 20 conserved skin cell populations, and found that there were differences in cell abundances. T cells and

NK cells were more abundant in BCC when compared to normal body skin (Figure 1.10b).

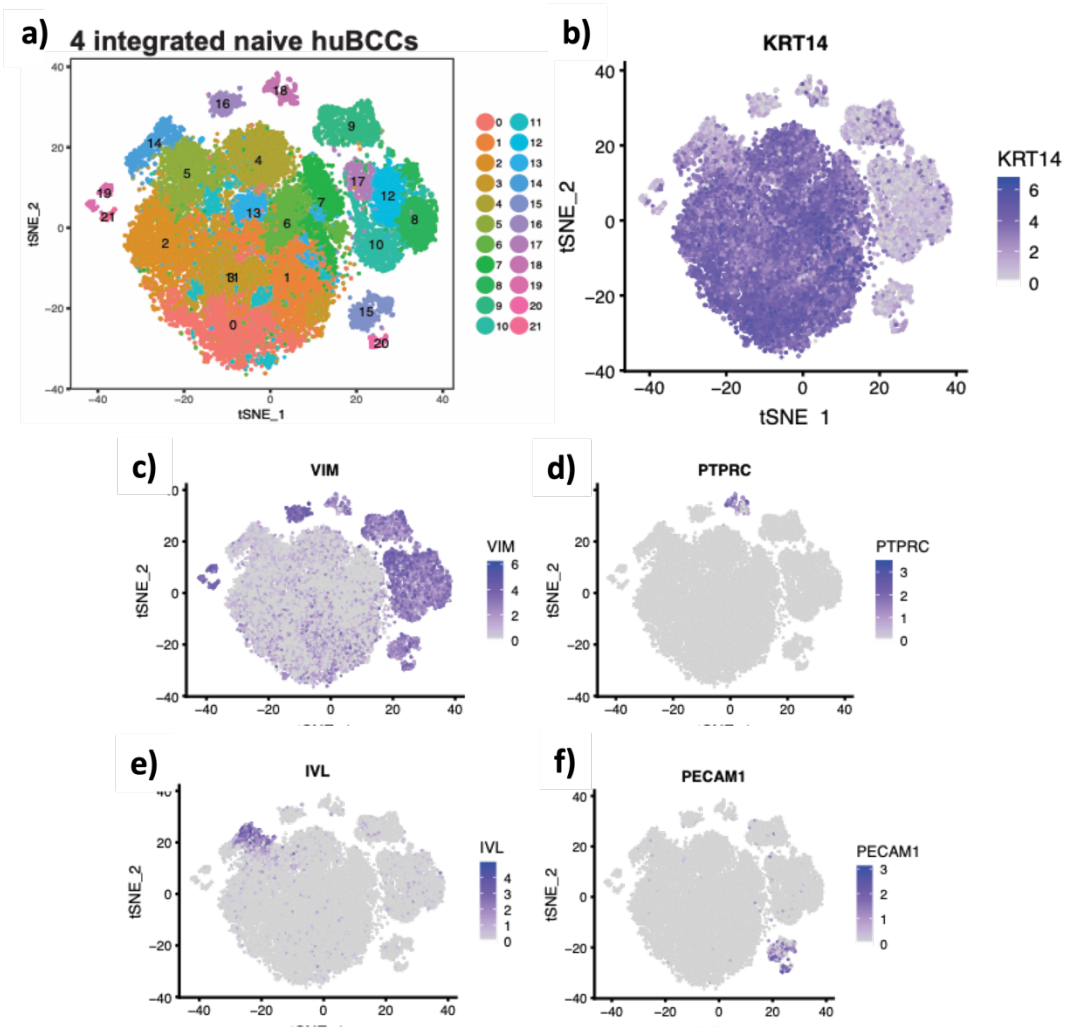
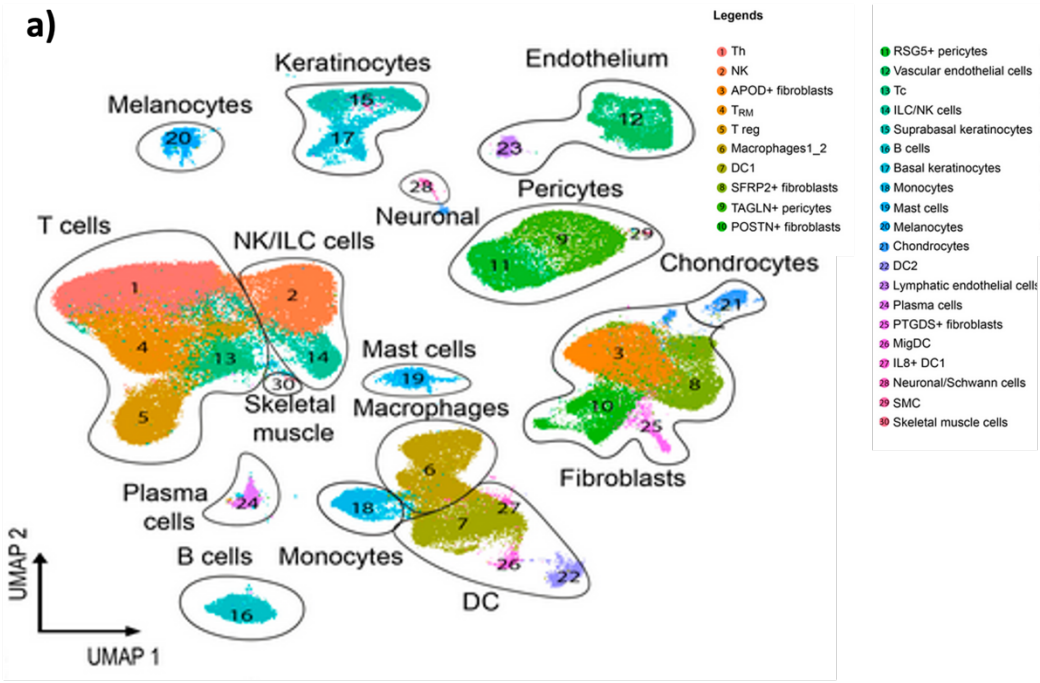


Figure 1.9 | Cell heterogeneity within human BCC. *Yao and colleagues analysed 4 human BCCs resulting in 45,656 cells.* ***a)*** *Total clusters generated from the dataset.* ***b-f)*** *UMAPs showing expression of KRT14, VIM, PTPRC, IVL, and PECAM1 which highlighted keratinocytes, fibroblasts, immune cells, differentiating keratinocytes, and endothelial cells, respectively. The authors used the sequencing to identify three prognostic surface markers which correlated to their previously identified SMO resistance mechanisms in BCC. Figure adapted from Yao et al. 2020.*


a)
Legends
1 Th
2 NK
3 APOD+ fibroblasts
4 TRM
5 T reg
6 Macrophages1_2
7 DC1
8 SFRP2+ fibroblasts
9 TAGLN+ pericytes
10 POSTN+ fibroblasts
11 RSG5+ pericytes
12 Vascular endothelial cells
13 Tc
14 ILC/NK cells
15 Suprabasal keratinocytes
16 B cells
17 Basal keratinocytes
18 Monocytes
19 Mast cells
20 Melanocytes
21 Chondrocytes
22 DC2
23 Lymphatic endothelial cells
24 Plasma cells
25 PTGDS+ fibroblasts
26 MigDC
27 IL8+ DC1
28 Neuronal/Schwann cells
29 SMC
30 Skeletal muscle cells
Keratinocytes
Endothelium
Melanocytes
Pericytes
Neuronal
T cells
NK/ILC cells
Chondrocytes
Mast cells
Skeletal muscle
Macrophages
Fibroblasts
Plasma cells
B cells
Monocytes
DC
UMAP 2
UMAP 1

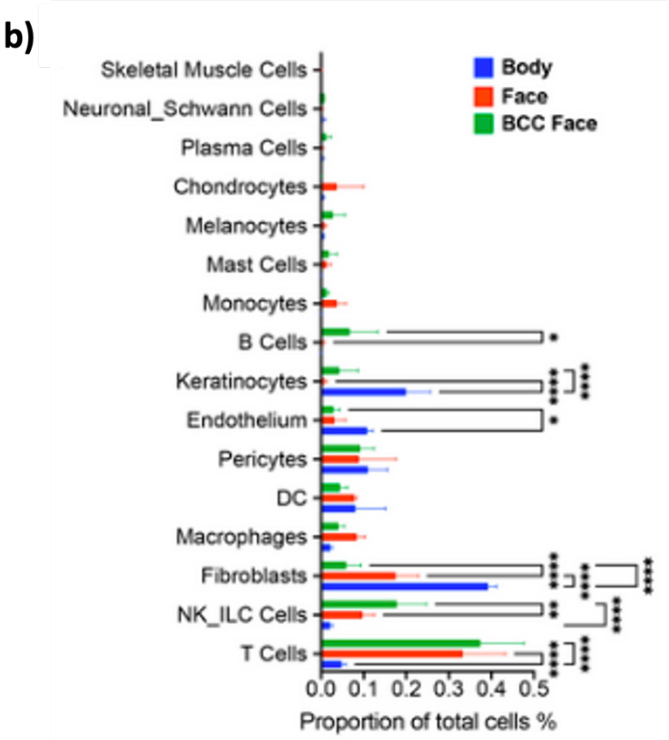

b)
Body
Face
BCC Face
Skeletal Muscle Cells
Neuronal_Schwann Cells
Plasma Cells
Chondrocytes
Melanocytes
Mast Cells
Monocytes
B Cells
Keratinocytes
Endothelium
Pericytes
DC
Macrophages
Fibroblasts
NK_ILC Cells
T Cells
0.0 0.1 0.2 0.3 0.4 0.5
Proportion of total cells %

Figure 1.10 | Cellular heterogeneity within human facial BCC and normal skin. *Ganier and colleagues used 9 BCCs and 11 normal skin samples from the face, and 15 normal skin samples from the rest of the body which generated a scRNAseq dataset consisting of 155,401 cells that passed quality control thresholds. The authors also used spatial transcriptomics and in vivo optical coherence tomography to generate a multiscale spatial atlas of healthy skin and basal cell carcinoma.* ***a)*** *UMAP showing identification of 30 skin cell clusters and 16 cell types. Skin cell populations are represented by different colours and cell types are encircled and annotated.* ***b)*** *The abundance of cell types was assessed across different body areas. The bar graph shows the differences in cell type abundance. Figure adapted from Ganier et al. 2024.*

1.7.4.7.2 BCC tumour-associated Fibroblasts

Ganier and colleagues showed that there were 4 major fibroblasts populations in the face and body in both normal and BCC skin. They found the same populations were present across each sample location (normal and BCC face skin and normal body skin) but observed differences in population abundances and distributions. BCC skin showed a significant increase in POSTN+ fibroblasts when compared to normal skin, and when investigated further, were found to be located around the HFs and considerably increased around the BCC tumour nodules (Ganier et al. 2024). Another study by Yerly and colleagues investigated the invasive niche of BCC and found that tumour cells displayed a collective migratory phenotype which was characterised by cell-cell junction complexes. They also found that cancer-associated fibroblasts were present at this niche and displayed features of extracellular matrix remodelling. Further investigation into the BCC cells at this niche revealed that they strongly expressed cytokine Activin A and subsequently increased Activin A-induced gene signatures in the adjacent cancer-associated fibroblasts (Yerly et al. 2022). Another study by Guerro-Juarez and colleagues investigated the interaction between the BCC and the TME (Guerrero-juarez et al. 2022). Using a range of downstream analyses such as pseudotime, RNA velocity, and cellular entropy, they revealed a cancer-specific rewiring of fibroblasts in which STAT1, TGFβ, and other inflammatory signals induced a non-canonical WNT5A program that maintained the inflammatory stromal state. They also found that tumour cells rapidly react to inflammatory signals released by the fibroblasts through the production of heat shock proteins, indicating that inflammatory cells act as an inflammatory hub that regulate BCC growth via these proteins.

1.7.4.7.3 T-cell response in BCC

By combining single-cell sequencing and T-cell receptor sequencing, Yost and colleagues showed in advanced BCC that after immune checkpoint inhibitor therapy targeting PD-1, novel T-cell clones infiltrated the BCC and replaced their pre-existing exhausted counterparts rather than invigorating them (Yost et al. 2019; Srivastava et al. 2023). Using trajectory analysis, the authors assessed T cell evolution in response to this treatment and revealed a bifurcated trajectory and becoming either terminally activated or exhausted.

1.7.4.8 Identification of BCC Cancer Stem Cells

Our lab has been studying the heterogeneity to further elucidate the biology of BCC. BCC was found to demonstrate hair follicle differentiation patterns, in keeping with BCC arising from hair-bearing skin (Morgan et al. 2020a). Intriguingly, the pattern of hair follicle differentiation is stalled because of the lack of dermal papilla signals (Morgan et al. 2020b). Restoration of the dermal§ papilla signals was able to restore the ability of BCC cells to complete the pattern of differentiation.

The presence of differentiation within the hair follicle raised the possible presence of a tumour-initiating cell population that would maintain BCC tumour growth similar to normal keratinocyte stem cells within the hair follicle. CD200, which identifies keratinocyte stem cells of the hair follicle bulge, was found to be expressed in small clusters of BCC tumour cells located in the basal and immediately adjacent suprabasal layers of tumour nodules. Using flow cytometry, we saw that all BCC samples contained a CD200+ tumour cell population of 1.63% $+/-$ 1.11% (Colmont

et al. 2013a). When 10^5 flow-sorted BCC cells from 5 independent BCCs were plated into a colony-forming assay, CD200+ cells gave rise to three-fold more colonies when compared to their CD200- counterparts. The CD200- cells gave rise to fewer colonies than the unsorted BCC cells. Further experiments showed that this CD200+ population when xenografted into mice, was able to reproducibly form tumours that resembled the original BCC with as little as 1 x 10^4 cells, whilst the CD200- counterpart failed to produce tumours. Limiting dilution assays revealed the tumour-initiating cell frequency of this population to be 1 in 822. We therefore concluded that the CD45- CD200+ BCC subpopulation is enriched in BCC cancer stem cells.

Consistent with BCC chemotherapy resistance, a proportion of CD200+ CD45- BCC expressed the multidrug resistance protein ABCB1 (Colmont et al. 2014). The CD45 cell surface marker was used to eliminate all haematopoietic cells. We showed that both the ABCB1+ CD200+ CD45- BCC sub-population and the ABCB1- CD200+ CD45- BCC sub-population could recreate BCC growth in vivo, wherein in vivo growth induced ABCB1 expression in the ABCB1- CD200+ CD45- BCC sub-population (Colmont et al. 2014) .

1.8 Hypothesis

BCC growth after the acquisition of driver mutations follows differentiation patterns that are consistent with normal hair follicle tissue pattern of differentiation and also consistent with the presence of BCC CSCs. For BCC to persist and grow after SMO antagonist therapy, BCC CSCs must have additional cell survival pathways that allow the circumvention of the Hh pathway inhibition.

1.8.1 Aims

To elucidate the cell survival pathways within BCC CSCs, I have two aims:

1. Determine whether CD200 is a functionally relevant marker for the study of BCC CSCs.
2. Use RNA sequencing to elucidate cell signalling pathways specific to the BCC CD200+ CSC-enriched population. I will be investigating the signalling pathways related to cell survival post-Hh inhibition.

Chapter 2: Methods

2.1 Cell Culture

2.1.1 Established Cell Lines

The cell lines that have been used in this project have been summarised in Table 2.1. Both cell lines were already present at the Patel laboratory upon my arrival. The long-term human BCC cell line, UWBCC1, is well characterised and will serve as a biologically relevant cell line. The cell line was kindly gifted by Vladimir Spiegelman (Pennsylvania State University, Hershey, Pennsylvania, USA). The human cervical keratinocyte cancer cell line, HeLa (ATCC ref: CCL-2), has been previously transduced by an ex-PhD student into a HeLa CD200+ and a HeLa CD200- cell line and will subsequently serve as biologically relevant positive and negative controls, respectively (Morgan et al. 2022).

Cell Line	Media	Passaging Ratio/Seeding Density
HeLa	RPMI (Gibco #11875093) containing 10% fetal bovine serum (FBS; Gibco #A5670701), 100IU/mL Penicillin (Gibco #15070063), 100IU/mL Streptomycin (Gibco #15070063), 20µg/mL Gentamycin (Gibco #15710064), 2mM L-glutamine (Gibco #A2916801)	1:10
UWBCC1	DMEM containing 10% (Gibco #11965092) fetal bovine serum (FBS; Gibco #A5670701), 100IU/mL Penicillin (Gibco #15070063), 100IU/mL Streptomycin (Gibco #15070063), 2mM L-glutamine (Gibco #A2916801)	1:10

Table 2.1 | Cell lines used, their required media, and their passaging ratio/seeding densities. *All mediums and reagents were purchased from Thermofisher Scientific (UK). Medias were made under sterile conditions using a laminar flow class 2 tissue culture hood.*

2.1.2 Maintenance of Cell Lines

2.1.2.1 Thawing Cell Lines

All cell lines were thawed at 37°C in a bead bath and then transferred to a 15mL Falcon tube (Corning). 9 volumes of pre-warmed media were added drop-by-drop and then cells were pelleted at 300xg for 5 minutes at RT and supernatant discarded. Cell pellets were then resuspended in 15mL previously mentioned media and transferred to a T75 culture flask and incubated at 37°C (5% CO2). Media was changed every 2-3 days and cells were sub-cultured at 80% to 90% confluence. Culture medias used to maintain cell lines are summarised in Table 2.1.

2.1.2.2 Sub-Culturing Cell Lines

For both cell lines, media was aspirated, and cells were washed with 5mL pre-warmed PBS (Gibco, UK). Cells were then incubated at 37°C for 10 minutes with 0.05% Trypsin-EDTA or TrypLe (Gibco, UK), or until cells detached from the culture flask. Four volumes of media containing 10% FBS was then added to inhibit the Trypsin activity (serum-free media was also used to dilute-out the TrypLe). Cells were then transferred to a 15mL Falcon and centrifuged for 300xg for 5 minutes at RT. Supernatant was removed and the cells were resuspended in media and counted using a haemocytometer. Cells were then seeded into the appropriate culture vessel (Table 2.2). Cells were then incubated at 37°C until 80% to 90% confluent.

Culture Vessel	Relative Surface Area (cm^2)	Volume (µL)
T-25	25	5,000
T-75	75	15,000
T150	150	30,000
6-well	9.6	2,000/well

Table 2.2 | Culture vessels used in the culturing of cells and their volumes of media required. *Culture vessels were purchased from Sigma-Aldrich (UK). Vessels were opened under sterile conditions inside a laminar flow class 2 tissue culture hood.*

2.1.2.3 Cryopreserving Cell Lines for Storage

All cell lines were prepared in the same manner as sub-culturing, except for the resuspension of the cells in freezing medium following centrifugation. All cell lines were frozen at a cell concentration of $1.5x10^6$ cells/mL. All cell lines were frozen in media containing 10% DMSO and 90% FBS. Cells prepared for cryopreservation were placed in 2mL cryovials and stored at -80◦C overnight in a cryopreservation container, and then logged and transferred to liquid nitrogen the next day.

2.1.2.4 Transduction of Cell Lines

The UWBCC1 was transduced to overexpress CD200 (Morgan et al. 2022). This was kindly done by Giusy Tornillo (ECSCRI, Cardiff University). Briefly, complimentary DNA for CD200 was sub-cloned into the PINCO retroviral expression vector which co-expresses the green fluorescent protein (GFP). CD200-expressing PINCO-GFP and PINCO-GFP control vectors were then transfected into Phoenix packaging cells. Phoenix cells were then cultured, and virus was harvested. CD200-expressing PINCO-GFP and PINCO-GFP control retroviral supernatants were then added to RetroNectin-coated 24-well plates and centrifuged at 4,000rpm for 90 minutes. Viruses were then aspirated, and 2.5 x 10^5 UWBCC1 cells were then added per well and incubated for 24 hours. Cells were then trypsinised and CD200 expression levels were confirmed in the UWBCC1 CD200pos and CD200neg cell lines through flow cytometry (Appendix 3.3).

2.1.3 Colony Forming Cell Assays

Colony forming cell assays were performed on UW-BCC1 cells. This assay was used to assess the ability of the cells to survive, proliferate, and expand to form small colonies. All experiments were performed in triplicate, using 3 biological replicates and 3 technical replicates per biological replicate. Cells were first antibody labelled and sorted into CD200+ and CD200- populations. 6-well plates were then taken and labelled into CD200-, CD200+, and unsorted plates. One plate was used for each condition. 2mL of media was added to each well (Table 2.2). 100 and 500 sorted and unsorted cells were then added to their respective wells and plates were gently agitated to disperse the cells. Plates were then incubated for 7 days. On day 8, media

was gently aspirated, and wells were washed twice with PBS. 400 µl Crystal Violet was added to each well and incubated on a rocker plate for 20 minutes. Crystal Violet was then aspirated, and plates were gently rinsed with tap water. Plates were then left to air dry at room temperature. Colonies were then imaged on a GelCount plate reader (Oxford Optronix) and quantified using QuPath software.

2.2 Immunohistochemistry

IHC was performed with an additional no primary antibody control either on a separate tissue section, or on a spot on the same tissue section where possible. Table 2.3 shows antibodies used. Paraffin embedded tissue sections were baked for 30 min at 60˚c and dewaxed and rehydrated using the following method: Xylene (10 min x2), 100% ethanol (5 min x 2), 95% ethanol (5 min x 1), 70% ethanol (5 min x 1) then phosphate-buffered saline (PBS; 5 min x 1). Sections were incubated in citrate buffer (8 mM sodium citrate, pH 6.0) in a pressure cooker heated in a microwave at full power (750 W, 10 min) for antigen retrieval. After cooling specimens were washed in PBS (5 min x 1), a hydrophobic barrier was drawn around the tissue using an ImmEdge PAP pen (Vector Laboratories; #H-4000) and endogenous peroxidase activity blocked with DAKO Dual Endogenous Enzyme-Block (Agilent Technologies; #K4065) for 10 min at room temperature (RT) and blocked for 1 hour at RT in BlockAid (Thermofisher; #B10710). Tissue sections were washed (PBS; 5min x 4) and incubated overnight with primary antibodies diluted in BlockAid. Sections were washed in PBS-Tween20 (0.05%) (5 min x 4) and incubated for 1 hour at RT with appropriate biotinylated secondary antibody diluted in BlockAid. Protein labelling was detected using Vectastain avidin-biotin enzyme complex (ABC-DAB; Vector Laboratories) or

ABC-alkaline phosphatase (ABC-AP; Vector Laboratories) with diaminobenzidine (DAB; DAKO, Agilent) or Immpact Vector Red (Vector Laboratories) as the chromagen, respectively. After washing (PBS; 5 min x 4) sections were incubated with ABC or AP for 30 min at RT and washed again in PBS (5 min x 4). Chromogen was then added and observed under a light microscope for appropriate colour development. Sections were then counterstained with haematoxylin (Atom Scientific) and dehydrated using the following method: 70% ethanol (20 sec x 1), 95% Ethanol (20 sec x 1), 100% ethanol (2 min x 3), Xylene (2 min x 2). Slides were mounted and images acquired using a light microscope.

2.3 Immunofluorescence

Immunofluorescence was performed on cells that were sorted into drops/spots on a Poly-L-lysine coated microscope slide (Thermofisher: #160005) at a rate of 50 cells/drop/spot. Sorted cells were firstly allowed to attach to the microscope slide for 10 minutes. Areas were then circled with an ImmEdge PAP pen (Vector Laboratories; #H-4000) and then fixed in 4% formalin for 10 minutes at RT followed by a 5-minute PBS wash. Fixed cells were then permeabilised in 0.25% Triton/PBS for 10 mins at RT followed by another 5-minute PBS wash. Permeabilised cells were then blocked for 1 hour at RT in BlockAid. Samples were washed in PBS (4 x 5 mins) and incubated overnight at 4°C with primary antibodies in BlockAid. A summary of primary antibodies used can be found in Table 2.3. Sections were washed in PBS-Tween20 (0.05%) (5 minutes x4) and incubated at RT for 1 hour with secondary antibodies (Table. 2.4) diluted in PBS with 1µL of 20µg/mL DAPI for nuclear staining. Specimens were washed in PBS-Tween20 (0.05%) (4 x 5 mins) and a coverslip mounted over the

areas using Vectashield mounting medium (Vector Laboratories). Images were acquired using a DM6000B upright fluorescence microscope (Leica Microsystems). Image analysis was performed using QuPath.

Target (Dilution)	Reactivity	Host	Cat. #	Company
anti-CD200 (1:400)	Human, Mouse, Rat	Mouse	AF2724	R&D Systems
anti-Pan-Cytokeratin (1:200)	Human, Mouse, Rat	Mouse	ab7753	Abcam
anti-α-Tubulin (1:400)	Human, Mouse	Mouse	MABT205	Sigma
anti-NK1.1 (1:400)	Human, Mouse	Mouse	MA1-70100	ThermoFisher Scientific
anti-Cleaved Caspase 3 (1:200)	Human, Mouse	Rabbit	9661	Cell Signalling Technology
anti-CD200R (1:400)	Human	Rabbit	ab198010	Abcam
anti-VEGFR2 (1:400)	Human, Mouse	Rabbit	2479	Cell Signalling Technology

Table 2.3 \| Primary antibodies used. *All primary antibodies were diluted in BlockAid and incubated on samples overnight at 4 °C.*

Target (Dilution)	Fluorophore	Cat. #	Company
Donkey anti-Mouse (1:1000)	Alexa Flour 488	A-21202	Life Technologies
Donkey anti-Mouse (1:1000)	Alexa Fluor 647	A-31571	Life Technologies
DAPI		4083	Cell Signalling Technology

Table 2.4 | Secondary antibodies used. *All secondary antibodies were diluted in PBS. 1 µL of DAPI (20 µg/mL) was also added to the PBS. Samples were then incubated with the secondary antibodies for 1 hour at room temperature.*

2.4 Flow Cytometry

Single cell suspensions were taken, antibody labelled, and then analysed or flow sorted for further analyses. All antibody incubations were carried out for 1 hour at 4°C in FACS buffer (PBS containing 10% FBS and 0.1% sodium azide) in FACS tubes. All centrifugations were performed at 1400rpm for 5 minutes at 4°C. Single cell suspensions up to $1x10^6$ cells/mL were stained with anti-CD45-FITC, anti-CD200-AF-

647, and the live-dead stain, DAPI. Table 2.5 summarises the antibodies used. Cells were analysed on an LSR Fortessa (BD Biosciences) or sorted on a FACS Aria (Becton Dickenson, Oxford, UK). Cells were sorted into sterile 1.5mL Eppendorf tubes maintained at 4°C. Single stained samples and Isotype controls were used as compensation controls and to assess background staining, respectively. Doublets, dead cells (DAPI+) and pan-leukocytes (CD45+) were excluded. FlowJo software was used to analyse the resulting data and create figures.

Antibody	Dilution	Cat. #	Source
CD200 (AF647)	1:1000	FAB27241R	R&D Systems
CD45 (FITC)	1:1000	341071	BD Biosciences
DAPI	1µL/1x10^5 cells	564907	BD Biosciences

Table 2.5 \| Antibodies used for flow cytometry. *Single cell suspensions were antibody labelled and then analysed or flow sorted on an LSR Fortessa or FACS Aria, respectively.*

2.5 qRT-PCR

RNA was extracted from cells using an RNeasy Mini Kit (Qiagen). Cells were resuspended in RLT Plus buffer, transferred to a gDNA eliminator spin column with 2mL collection tube and centrifuged for 30s at 8,000xg. One volume of 70% ethanol

was added to the flow through and transferred to an RNeasy spin column and centrifuged for 30s at 8,000xg. The membrane was subsequently washed with Buffer RW1 and centrifuged for 30s at 8,000xg and then washed twice in Buffer RPE and centrifuged for 2 mins at 8,000xg. The RNA was then eluted by adding 30-50µl of RNase-free water and centrifuging for 1 minute at 8,000xg. RNA quality and concentration was then measured on a Nanodrop 2000 (Thermo Scientific). cDNA synthesis was performed using the Quantitect Reverse Transcription Kit (Qiagen). gDNA was removed from the sample by adding 500ng of RNA to 2µl of gDNA wipeout and made up to a total volume of 14µl with RNase-free water. A 2-minute incubation at 42°C was performed in a thermocycler. 1µl of reverse transcriptase and primer mix, and 4µl of RT buffer was then added to each sample and incubated for a further 15 minutes at 42°C followed by a 3-minute incubation at 95°C in a thermocycler. Resulting cDNA was then used as input for qRT-PCR analyses. briefly, TaqMan probes were purchased from Thermofisher Scientific (CD200; Hs01033303_m1 and β-Actin; Hs01060665_g1). All analyses were run on the QuantStudio 7 Flex Real-Time PCR system (Applied Biosystems) using the Quantstudio software (v1.7.2). All reactions were run in triplicates, and housekeeping gene β-Actin was used as the reference gene for each plate. Reactions were run using the TaqMan Universal Master Mix II, with UNG (Applied Biosystems) according to manufacturer's guidelines. Reactions were assembled in 96-well plates using a final reaction volume of 15µl and final concentration of 20ng of RNA/cDNA. Reactions were run using the following conditions: 50°C for 2 mins, 95°C for 10 mins, followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 min. Samples with reproducible cycle times (CT) values were analysed using the 2-delta delta Ct method. The Ct value of the reference sample was used to calculate absolute units.

2.6 Image Analysis

Image analysis was performed using QuPath (v0.5.0). A new QuPath project was created for each experiment and each image was imported into that project.

2.6.1 Colony Forming Cell Assays

Each well image was opened, and the wand tool was used to select each colony. The benefit of using the wand tool is that it also recognises the size of the colony, and subsequently highlights it. Once all colonies were selected, the points tool was selected, which enumerated the number of selected colonies. Additionally, given that the wand tool highlighted the size and shape of each colony, the colony areas were measured. This was done by clicking on the measurements table button and then clicking the annotation measurements option.

2.6.2 Immunohistochemical Image Analysis

H&E images were used to count total tumour cells and necrotic areas. To count total tumour cells, the brush tool was selected and used to draw around the whole tissue (Figure 2.1a). The cell detection tool was then used to count total number of cells in the tumour. The parameters of the cell detection tool, specifically the requested pixel size, were adjusted as to only detect larger cells (Figure 2.1b). Doing this enabled the detection of tumour cells and not detection of smaller immune cells. The number of detected cells were then obtained by clicking on the show annotations measurements tool (Figure 2.1c). To enumerate the number of immune cells present within each tumour section, the H&E sections were first inspected to identify areas

of immune cell infiltrate. Immune cells are small and are densely packed, forming darker patches within the tissue. These areas were highlighted with the brush tool and the cell detection tool was again used to count the cells. Necrotic areas were defined as areas with high immune cell infiltrate that have no cells present on the inside. These areas appear as dark patches with light pink areas in the centre. To calculate the ratio of necrotic areas to tumour cell areas, the brush tool was first used to select all necrotic areas and then the total number of cells present in these areas were calculated using the cell detection and annotations measurement tool. Then the brush tool was used to select the whole tissue and cell detection and annotations measurement tools were used to count all cells present in the tissue. The necrotic ratio was then calculated by dividing the total number of necrotic cells by the total number of cells. For enumeration of marker-positive cells, the tissue was first selected using the brush tool, and then the positive cell detection tool was used with the optical density sum option selected (Figure 2.1d). The nucleus: DAB OD mean threshold was adjusted to ensure proper positive cell detection (Figure 2.1e, f). Again, the annotation measurements tool was used to obtain the number of positive and negative cells. To enumerate the percentage of NK cells expressing Cleaved Caspase 3 and CD200R, I used serial tissue sections. I firstly identified clusters of positive NK1.1 staining and then highlighted these areas with the brush tool. I then stained the serial sections with CD200R and Cleaved Caspase 3 and cross referenced those highlighted areas. I again used the brush tool to select those areas on the respective slides and used the positive cell detection tool to count positive cell staining.

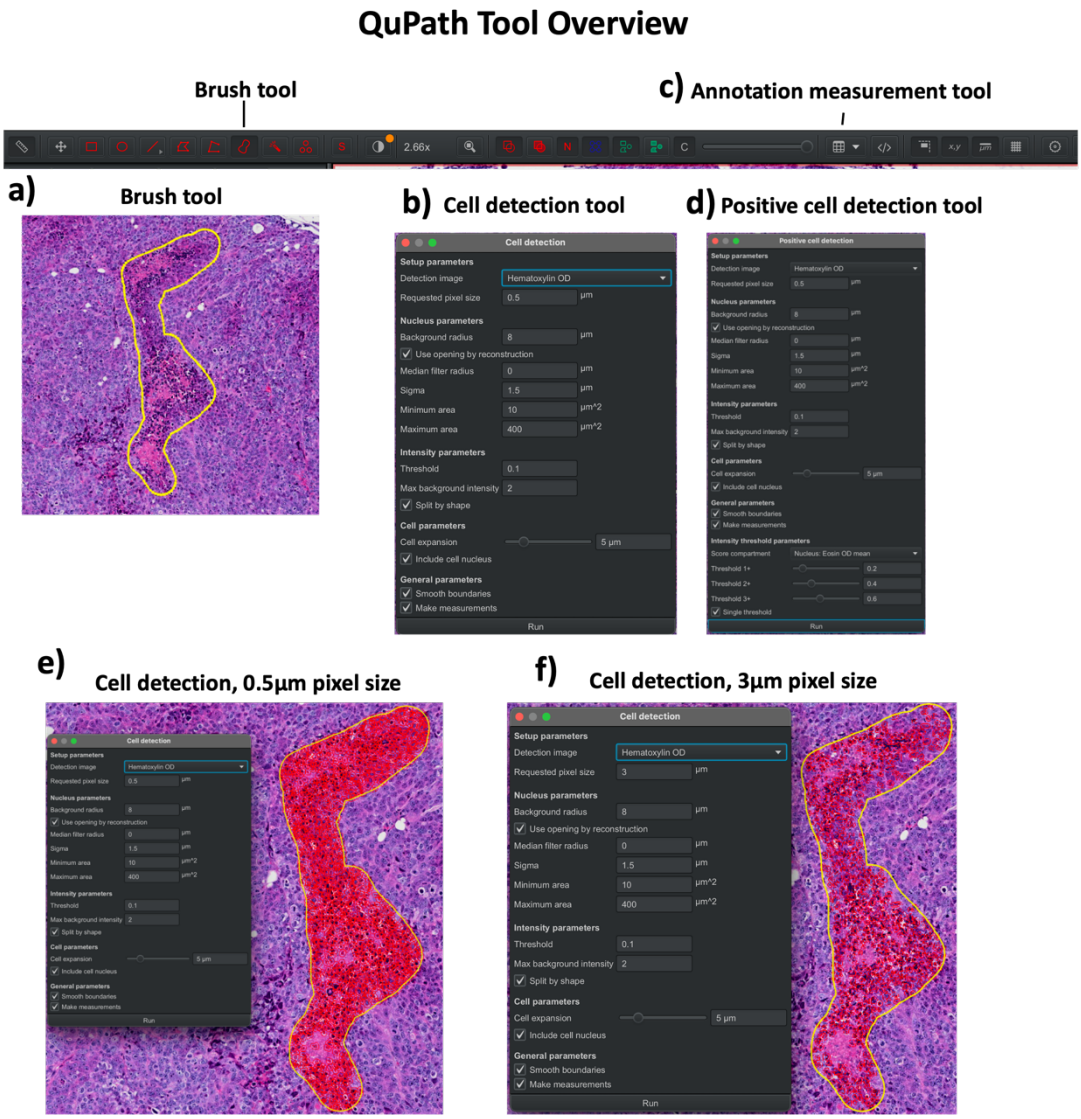


Figure 2.1 | QuPath tool overview. *QuPath tool was downloaded (version 0.5.0) and used to analyse and quantify imaging data.* ***a)*** *The brush tool is used to highlight areas on the image.* ***b)*** *The cell detection tool is used to detect cells within highlighted areas.* ***c)*** *The annotation measurement tool is used to quantify the number of detected cells from the cell detection or positive cell detection tool.* ***d)*** *The positive cell detection*

tool is used to detect positive cells on immunohistochemistry slides. ***e, f)*** *An example of the number of cells detected when changing the pixel size whilst using the cell detection tool at 0.5µm and 3µm, respectively.*

2.7 Mouse Experiments

All mouse experiments were performed in accordance with a UK Home Office Licence (Project Licence 30/3382). Nude mice (NU(NCr)-*Foxn1*^{nu}, 6- to 8-week-old females; Charles River) were used to study the impact of CD200 expression on in vivo tumour growth. Mouse husbandry was performed by our laboratories research technician, Boris Shorning, and xenografting was performed by myself under the supervision of Boris. Both Boris and I have personal licences (My PIL #I19564663) to perform these experiments. Mice were anesthetised with isoflurane (induced at 5% Isoflurane concentration and maintained at 2%) and then injected with 1 x 10^6 UWBCC1 CD200+ and UWBCC1 CD200- (n=5 each) that had been mixed with 100 µl of Matrigel matrix (Corning) into the left and right flank of the mouse. Tumours were left to establish over 5 days, and then mice were culled by cervical dislocation and tumours harvested. Tumours were first measured for total area using callipers, and then sent to Cardiff Universities Histology Hub for paraffin embedding, sectioning, and H&E staining. H&E slides were then assessed under a microscope for tumour cellularity and CD200 expression (Figure 3.1 and Appendix 3.3; Morgan et al. 2022).

2.8 Human Tissue Samples

Human BCC tissue was obtained following appropriate NHS and ethics approval (09-WSE-02-1) with informed and written consent from the patients. After excision, tumours were placed in an appropriately labelled 50mL Falcon tube containing Cryo-SFM (PromoCell) and immediately stored at 4°C until I came to collect them. All samples entering the laboratory were logged in accordance with the Human Tissue

Act and were processed within 24 hours. Undissociated tissue was disposed of in the appropriate human waste bins. All incoming samples were anonymised by the dermatologist and then re-labelled by myself before proceeding.

2.8.1 Tissue Dissociation

Tumour tissues were dissociated into a single-cell suspension in a laminar flow class 2 tissue culture hood for cell culture. Briefly, 1mL of Medium 199 (conventional protocol) or RPMI 1640 (MACS protocol) was pipetted into a sterile 10cm^2 petri dish. Using sterile forceps, the tumour tissue was dipped in PBS to rinse and then placed into the petri dish. Then, using sterile iris scissors, the tumour was minced for 15 to 20 minutes. The minced tumour was then transferred to either a 6 well plate or a MACS C-tube. Following enzymatic digestion, red cell lysis was performed by resuspending cell pellets in 1mL Ammonium Chloride Solution (Stemcell Technologies; #07800) for 2 minutes and then 9mL PBS was added to dilute and inactivate it. Samples were centrifuged again and resuspended in FACS buffer.

2.8.1.1 Patel Lab BCC Dissociation Protocol

This method has been described previously (Morgan et al. 2019; Olivero et al. 2019). Briefly, minced tumour tissue is divided into two wells of a 6 well plate, and each well was filled with 2mL Media 199 supplemented with 25U Neutral Protease (Worthington). Tumour tissue was then agitated using a P1000 and 1mL pipette tip before being incubated for 90 minutes at 37°C. Tumour tissue was agitated every 30 minutes. 200U of Collagenase III (Worthington) was then added to each well and agitated again before a 30-minute incubation at 37°C. Tissue was disrupted every 15

minutes. Tissue was transferred to two 50mL Falcon tubes (1 well/Falcon) and 25mL Media 199 was added to each tube. Samples were then centrifuged for 7 minutes at 1300rpm. Supernatant was discarded and samples were resuspended in 5mL TrypLe (GibCo, UK), vortexed, and placed in a bead bath at 37°C for 15 minutes. Samples were vortexed every 5 minutes. Following TrypLe digestion, 25U DNase I (Thermofisher; #18047019) was added to each sample (According to manufacturer instructions), vortexed, and incubated in a bead bath at 37°C for 5 minutes. Samples were then filtered through a 70µm filter into one Falcon tube containing 25mL Media 199. The filter was rinsed with 5mL Media 199, and the sample was centrifuged for 7 minutes at 1300rpm.

2.8.1.2 MACS Diccociation Protocol

Minced tumour tissue was placed into a MACS C-tube along with 4.7mL RPMI 1640 or DMEM, 200µL Enzyme H, 100µL Enzyme R and 25µL Enzyme A. The lid of the MACS tube was tightly shut, and the tube was placed upside down into a gentleMACS Dissociator. The h_tumour_01 protocol was run, and the tube was then transferred to a MACSMix Tube Rotator and incubated at 37◦C for 30 minutes. Following incubation, the C-tube was again placed in the gentleMACS Dissociator and the H_Tumour_ protocol was run (the number of the protocol depended on the amount of digestion of the tissue). Another 30-minute incubation on the MACSMix Tube Rotator was performed. The H_Tumour_ protocol was performed one more time and the C-tube was placed in a centrifuge and quickly spun for 5-10 seconds. 4mL supernatant was taken from the C-tube and filtered through a 70µm filter into a 50mL Falcon using a 70µm strainer. The remaining 1mL 'pellet' was then gently manipulated to the bottom of the C-tube and placed onto the gentleMACS

Dissocator. The M_imp_Tumour_01 protocol was run and then the C-tube was quickly spun again for 5 – 10 seconds. The supernatant was taken and filtered into a 50mL Falcon. The C-tube was then rinsed with 5mL RPMI 1640 or DMEM and filtered into the 50mL Falcon. The filter was then washed with 15mL RPMI 1640 or DMEM. The 50mL Falcon tube was then centrifuged at 300xg for 7 minutes.

2.8.2 Tissue Freezing, Cryosectioning, and Staining

Residual tumour tissue was collected from 70µm filters and embedded in OCT (Tissue-Tek) compound on dry ice. Frozen samples were then cryosectioned using the Leica Cryostat CM1860 UV (Leica Systems, England). The cryostat chamber was cleaned with 70% ethanol and samples were placed in the chamber to acclimatise for 10 minutes. Samples were then removed out of the plastic mould and attached to the metal chuck using OCT compound. Samples were then sectioned to a thickness of 15µm onto glass slides and left to air-dry at RT. Glass slides were then stained with Toluidine Blue (Sigma) for 10 seconds and then washed under a running tap face-down for 1 minute. Samples were then left to air-dry at RT. Stained samples were then imaged under a light microscope.

2.9 Statistical Analyses

Statistical analyses were performed in Graphpad Prism 10 software. Normality tests of data sets were tested using a Shapiro-Wilks test. Normally distributed data between 2 groups was tested using an unpaired t-test, non-normally distributed data were compared using a Mann-Whitney U test. To compare the means of >2 non-

normal distributed groups, a Brown-Forsythe and Welch analysis of variance (ANOVA) was used. Bar graphs were then auto generated in Graphpad. Statistically significant differences were marked as * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$, and **** = $p \leq 0.0001$.

2.10 Sequencing Sample Preparation

2.10.1 Bulk RNA Sequencing

RNA sequencing was performed on CD200+ CD45- and CD200- CD45- cancer stem cells from primary human BCCs. RNA was subsequently extracted from each cell population as described in Chapter 2.5 with the exception of using the RNeasy Micro Kit (Qiagen) and eluting RNA into 14µL of RNase-free water. RNA was then sent the Wales Gene Park (WGP) for sequencing. Whilst at WGP, cDNA libraries were made from each RNA sample and NEBNext sequencing adaptors were added along with sample barcodes. Samples were then paired-end sequenced using a Novaseq S1.

2.10.2 Single-Cell RNA-sequencing

Freshly sorted cells were firstly centrifuged at 1400rpm for 3 minutes at 4°C. the supernatant was then aspirated, and cells were resuspended in cold Media 199 at an optimal cell concentration according to the target cell recovery table provided by 10x Genomics. Prepared cells were then handed over to Ngoc-Nga Vinh (Dementia Research Institute, Cardiff University), who processed the cells according to manufacturer's Single Cell 3' Reagent Kit (v3.1) protocol. Briefly, cells were loaded

onto the 10x Chromium controller along with the GEM kit (v3.1) to generate Gel Beads-In-Emulsion (GEMs). Reverse transcription is then performed on the GEMs, and cDNA is purified, amplified, and again purified. Using the Single Cell Library Kit (v3.1), cDNA libraries were then made through multiple rounds of PCR fragmentation, end repair, poly-A-tailing, size selection purification, and indexing. Finally, libraries were quality checked using a TapeStation 4200, and library yields were assessed via qPCR. Libraries were then frozen at -80°C until ready for sequencing. Libraries were sent via 24-hour tracked courier on 5-10kg dry-ice to Novogene, where they were sequenced on a Novaseq S4 (PE150) at a depth of 200 million reads per sample. Post-sequencing, data underwent an in-house quality check pipeline before being returned to me.

2.11 Sequencing Data Processing

2.11.1 Bulk RNA-sequencing

FastP was used on all sequencing reads to provide a quality control check and trim the adaptors. Quality checked reads were then mapped against the human reference genome (GRCH38 using Spliced Transcripts Alignment to a Reference (STAR) software in 2-pass mode. Genomic features were then counted from the mapped sequence reads using the reads summarisation program, featureCounts, which generated a normalised count matrix for each sample against the GENCODE v27 gene model. A normalised count matrix consists of a table containing the amount of sequence fragments that have been assigned to each gene. The normalised count matrices were then used to generate differentially expressed genes (DEGs) through the use of

the SARTools package (wrapper to DESeq2). Normalised count matrices were then used as input data for PCA, heatmap and correlation analyses (ggfortify, heatmap.2, and corrplot packages in R, respectively. All packages were used with default parameters). DEGs were used as input for IPA, GSEA, enrichment analyses and to generate a volcano plot (EnhancedVolcano package in R). Enrichment analyses were performed using the clusterprofiler package in R.

2.11.2 Single-Cell RNA Sequencing

Sample fastq files were first downloaded from Novogene and checksums were generated using the '*md5sum'* program in Linux and cross-referenced with the checksums provided by Novogene to ensure full downloads and no data loss. Fastq files were then uploaded onto Cardiff Universities HAWK supercomputer and were again checked for complete upload through regeneration of checksums. Next, the human reference genome (GRCh38) dataset was downloaded as a gzipped tarball directly from the 10x Genomics website into the working directory on HAWK using the *'curl'* command. The gzipped tarball was unpackaged using the 'tar' command.
To generate count matrices for each sample, the *'cellranger'* package was loaded, and the *'count'* function was run on each sample whilst using 10,000 as the expected number of cells, which mapped and quantified genes for each sample. Results were then either downloaded from HAWK and used as input to R for downstream processing or kept on HAWK where they were then used as input for copy number variation analysis (Section 2.11.2.13.2). The 'cellranger count' pipeline also generated a webpage report for each sample, which serves as an initial QC.

2.12 Bioinformatics Analyses

2.12.1 Bulk RNA-sequencing

Data was analysed using R (version 4.3.1). The normalised count matrices were imported into R and the standard SARTools pipeline (wrapper to DESeq2) was run to generate normalised count matrices and differentially expressed genes. Normalised count matrices were then used as input data for PCA, heatmap and correlation analyses (ggfortify, heatmap.2, and corrplot packages in R, respectively. All packages were used with default parameters). DEGs were used as input for IPA, GSEA, enrichment analyses and to generate a volcano plot (EnhancedVolcano package in R). For GSEA heatmaps, the top normalised enrichment scores (NES) were taken and input into Morpheus (Broad Institute). Venn diagrams were generated by using the normalised read counts as input into the ggVennDiagram package in R. Enrichment analyses were performed using the clusterprofiler package in R.

2.12.2 Single-Cell RNA-sequencing

2.12.2.1 Optimisation of Analysis Protocol

Over the course of the PhD project, a total of 13 samples were successfully processed for scRNAseq, these consisted of ten BCCs and three NS samples (Table 5.1). Samples were dissociated, labelled and FACS sorted as previously described in Chapter 2.4. The first batch consisted of three BCCs and one NS (BCC_1-3 and BCC_2_NS) and was generated via the four-reaction chromium kit immediately after the first COVID lockdown. This batch was sent for sequencing as soon as the samples were generated

to obtain the sequencing data before the inevitable second lockdown. For the second batch, incoming samples were immediately processed into libraries and frozen back until a total of thirteen libraries were generated using the sixteen-reaction chromium kit. This second batch consisted of seven BCCs, and two NS (BCC_4-8, and BCC_NS_1 and BCC_7_NS). The remaining libraries consisted of murine samples from our lab's other projects.

The first stage of the scRNAseq project was to optimise the analysis protocol. Since its development, there has been an explosion of tools and methods to perform and analyse scRNAseq data, which makes it difficult to select the best method for your data (Hie et al. 2020; Schneider et al. 2021). However, the sequencing data from the first batch of samples that were sent to Novogene was used as a 'pilot' dataset, which enabled the generation and optimisation of an analysis script, which allowed me to test various tools and methods, and was later used to process the full dataset. Samples were pre-processed individually, and clustering was assessed for putative cell types that were identified using canonical markers. Samples were then integrated into one dataset using a selection of integration methods. Clustering was performed on the integrated dataset to assess how well each method performed at separating the putative cell types. A schematic workflow for the scRNAseq analysis can be seen in Figure 2.2.

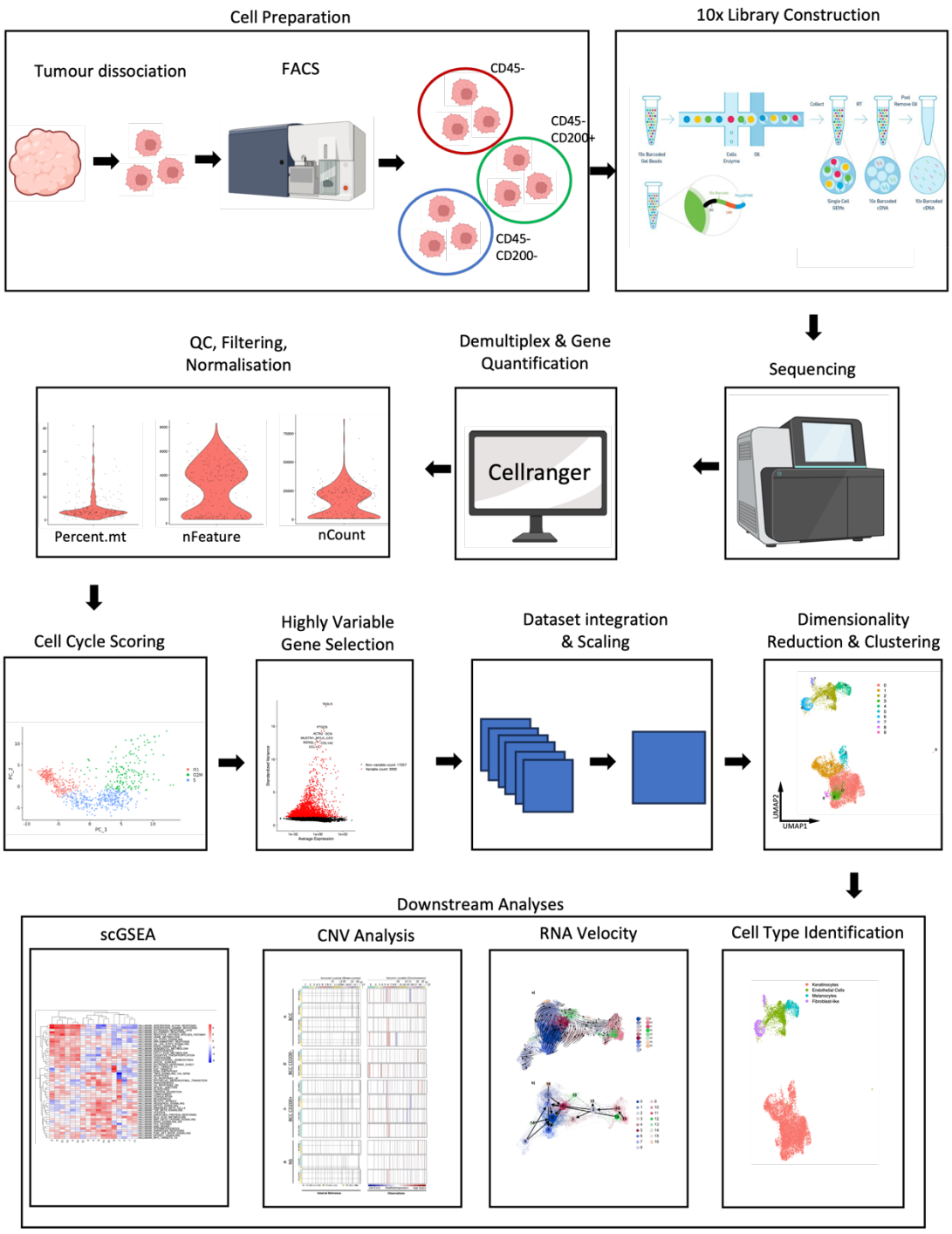

Cell Preparation
Tumour dissociation
FACS
CD45-
CD45-
CD200+
CD45-
CD200-
10x Library Construction
Sequencing
Demultiplex & Gene
Quantification
Cellranger
QC, Filtering,
Normalisation
Percent.mt
nFeature
nCount
Cell Cycle Scoring
Highly Variable
Gene Selection
Dataset integration
& Scaling
Dimensionality
Reduction & Clustering
UMAP2
UMAP1
Downstream Analyses
scGSEA
CNV Analysis
RNA Velocity
Cell Type Identification

Figure 2.2 | Schematic of single cell RNA sequencing workflow. *Briefly, tumours were taken from patients and immediately dissociated into single cell suspensions. Cells were then antibody labelled for CD45, CD200, and DAPI, and then flow sorted into 3 populations. These populations were then used as input into 10X Genomics single cell pipeline to generate single cell libraries. Libraries were then sequenced at Novogene (Cambridge, UK). Sequenced samples were then input into 10X Genomics Cellranger pipeline where they were firstly demultiplexed and subsequently genes were counted. Putative doublets, empty cells, cells with high mitochondrial reads, and bad quality cells were then removed during the quality control step, which was then followed by normalisation. Cell cycle stages were then calculated, and highly variable genes were selected. All samples were then integrated into one dataset. Dimensionality reduction was then performed in the form of PCA, followed by Seurat clustering using the PCAs, followed by UMAP. Downstream analyses were then performed: Cell type identification, RNA Velocity, CNV analysis, and scGSEA.*

A total of 936,290,456 raw sequencing reads were generated across 6,224 raw single cells from all four samples (BCC = 693,860,583 raw reads across 5,326 raw cells and NS = 242,429,873 raw sequencing reads across 888 raw cells). On average, there were 3,160 median genes and 10,428 median Unique Molecular Identifiers (UMI)'s per cell for the BCC samples and 3,172 median genes and 14,014 median UMI's for the NS sample. Each sequencing fragment is tagged with a Unique Molecular Identifier (UMI), which allows for identification of PCR duplicates and unique molecules during the library preparation stage. There was an average of 90.45% reads confidently mapped to the human genome (BCC = 90.3% average and NS = 90.9%) and 52.2% reads confidently mapped to the human transcriptome (BCC = 49.9% and NS = 59.1%). Cell Ranger also generates a sequencing saturation statistic, which provides an insight into library complexity and sequencing depth. The statistic looks at the number of reads originating with an already-observed UMI. Taking the inverse of one minus sequencing saturation indicates the number of additional reads it would take to detect a new transcript. There will be differing amounts of RNA present in different cell types and subsequently results in differing amounts of total transcripts in the sequencing library, which is referred to as library complexity. As sequencing depth increases so does the number of genes that are detected, but this reaches saturation at certain depths depending on cell type. In essence, the sequencing saturation gives an indication into how complex your libraries are and how deep they have been sequenced. On average, sequencing saturation was 79.95% across all samples (BCC = 77.2% average saturation and NS = 88.1% saturation). Raw sequencing statistics can be found in Table 5.1.

2.12.2.2 Quality Control

Whilst at Novogene, sample libraries are pooled together to increase sequencing efficiency. This is known as multiplexed sequencing. Following sequencing of the libraries, one sequencing file is produced which contains the sequencing reads from all samples. This file is then demultiplexed, which uses the samples UMI's to generate sequencing files for each sample. Demultiplexing was performed by Novogene and the resulting files were returned to me. Demultiplexed samples were then input into R (version 4.2.1) and processed individually using Seurat (version 4.3.0). Samples were initially read into R via the *'Read10x'* function and made into Seurat objects using the *'CreateSeuratObject'* function with the minimum number of cells = 3 and minimum number of features = 200. QC was then performed in order to remove doublets and low-quality cells. To do this, violin plots were generated for the following metrics: nCount_RNA, nFeature_RNA, and percent.MT (Appendix 2.2, 2.3, and 2.4, respectively) nCount_RNA is a measure of the total number of molecules detected within a cell, nFeature_RNA is a measure of how many genes are detected within a cell, and percent.MT is a measure of the percentage of reads that map to the mitochondrial genome. Given the nature of the library preparation method, it is possible to obtain more than one cell (doublets or multiplets) per GEM bead and obtain empty GEM beads. The former can generate aberrantly higher gene count per 'cell' and the latter can contain very few gene counts. A Gel Bead in Emulsion (GEM) is a key component to the 10x Genomics sequencing method. Each GEM contains a single bead with a unique barcode plus the reagents (including UMI's) required for RNA amplification. During the library generation process, each single cell is encapsulated within one GEM and the resulting protocol generates reads that have the same barcode, allowing identification of those reads to that one cell. A low

nFeature_RNA count for a cell indicates that it may be dead or dying, or that it was an empty droplet. Likewise, a high nFeature_RNA and/or nCount_RNA count indicates that the 'cell' is a doublet/multiplet. Low-quality and/or dying cells also demonstrate extensive mitochondrial contamination, which increases the percent.MT metric. Cells were first filtered using nFeature_RNA and percent.MT and then nCount_RNA was assessed. In all samples, the former filtering metrics were deemed appropriate at filtering out these cells. Filter metrics are displayed in Appendix 2.5. Following QC and filtering, a total of 5,020 viable cells remained, BCC = 4288 (BCC_1 = 2119, BCC_2 = 1773, and BCC_3 = 396) and NS =732.

2.12.2.3 Normalisation

The next step in the analysis protocol is to normalise each sample. Post-QC and filtering, scRNAseq datasets can also contain noise and variability that can be attributed to cell type, size, RNA content, cell cycle stages, gender differences, and sequencing depth (Lafzi et al. 2018). These technical variances need to be corrected through a process called normalisation. Normalisation methods correct for these variations so that gene expression profiles are comparable across cells and samples.

There are a number of normalisation methods that have been developed and benchmarked specifically for scRNAseq, which range from variations of bulk-RNAseq methods to the development of completely new methods specific to scRNAseq (Lytal et al. 2020). However, none of them have come out universally superior during benchmark studies. The most common approach to processing and analysing scRNAseq data is via the Seurat package in R (Satija et al. 2015a). Seurat has a built-in function, 'NormalizeData', which ensures expression values across cells are on a

comparable scale. There are three options for normalising the data through the 'NormalizeData' function: log transformation (LogNormalize), centred log ratio transformation (CLR), and relative counts (RC). Given the number of parameters available at each step of the analysis protocol, the 'LogNormalize' argument was pre-selected as the preferred method for 'NormalizeData', as studies have shown that CLR and RC perform similarly, generating fewer clusters downstream, having similarity scores >90%, and generally underperforming when compared to 'LogNormalize (Schneider et al. 2021). The 'LogNormalize' argument divides the counts for each gene by the total counts in the cell, multiplies that value by the scale factor, which by default is 10^4, and then log transforms the results. This method aims to normalise the biases in sequencing coverage between cells and then log-transforming the data ensures downstream analyses are comparable on a relative, rather than absolute, scale (Lun 2018). The scale factor is used to ensure that one normalised count corresponds to roughly one raw count that is observed in one 'average-sized' single cell. By default, this is 10^4 as this is the usual number of UMIs per cell in a typical 10x Genomics experiment. However, in deeper sequencing methods, such as Smart-seq2, the number of UMIs per cell can range from 10^5-10^6, and so the scaling factor would need to be changed accordingly. For my samples, the average UMI per cell was ~1.4×10^4 for both the pilot dataset and the full dataset, so the default scaling factor was chosen.

2.12.2.4 Cell Cycle Scoring

Following normalisation, cell cycle stage was calculated for each cell in each sample. Stages of the cell cycle can introduce unwanted sources of heterogeneity within scRNAseq datasets and can affect downstream clustering. Seurat's 'CellCycleScoring'

function assigns cells a score based on expression of S and G2/M phase markers and then predicts the phase of the cell as either G2M, S or G1 phase. The cell cycle score for each cell can then be added to the Seurat object as a new metadata column and used to regress out the cell cycle effects before performing clustering. Regressing out the cell cycle phase removes all signals associated with the cell cycle. This could, however, potentially negatively impact downstream analysis, especially in differentiating processes where stem cells are quiescent and differentiating cells are proliferating. Regressing out the cell cycle in this instance can impede the distinction between stem and progenitor cells. Instead, Seurat recommends that the difference between the G2M and S phases should be regressed. The resulting regression would allow signals that separate non-cycling/cycling cells to remain and differences in phase among often uninteresting proliferating cells to be removed.

2.12.2.5 Highly Variable Gene Selection

The dataset still currently consists of noisy data which comprises mostly lowly expressed genes that have similar expression across cells and samples. Using all genes for downstream clustering would be unwise as many of them provide no biological insight. Instead, a selection of highly 'informative' genes is made and used for downstream clustering. In doing so, we are simultaneously removing irrelevant genes, background noise, and computation time. Assuming that biological processes are driven as a result of changes in gene expression, we can exploit this by selecting genes that show high variance across cells, genes that are highly expressed in some cells, but lowly expressed in others (Brennecke et al. 2013; Satija et al. 2015b). Selecting these highly variable genes (HVGs) highlights biological signals and vastly improves downstream clustering. However, scRNAseq data includes many zero

counts that can inflate gene variance, and so directly using gene variance is not advised as it contains both biological and technical variance. Seurat's 'FindVariableFeatures' function circumvents this problem by fitting the gene variance against the average expression and then calculates a standard variance for each gene, which represents the genes variance from the fitted value. The selected HVGs then represent genes that are more variable than genes which have similar average expression. This method is known as variance-stabilizing transformation and is the default correction method in the method (selection.method = 'vst') (Mayer et al. 2018; Hafemeister and Satija 2019a). 3000 HVGs were selected for each sample in the pilot dataset for downstream clustering (Appendix 2.6).

2.12.2.6 Scaling

The next step of the protocol is to scale the data. Scaling the data linearly transforms the data, which shifts the expression of each gene so that the mean expression is zero and the variance across the cells is one. This ensures that highly expressed genes do not dominate during downstream analyses. Scaling also allows for the regression of 'uninteresting' signals. In Seurat, the 'ScaleData' function models the relationship between gene expression and the features that we want to be regressed. Scaled residuals of the model then represent a corrected expression matrix which can then be used for downstream analysis. Here, the cell cycle difference, nCount_RNA (total number of molecules detected within a cell), and percent.MT (percentage of reads that map to the mitochondrial genome) were regressed from the data.

2.12.2.7 Linear Dimensionality Reduction

In order to visually assess putative cell populations within each sample, dimensionality reduction was performed. Each sample is highly dimensional, consisting of hundreds to thousands of cells with thousands of expressed genes. Humans are unaccustomed to perceiving greater than three dimensions, so dimensionality reduction is an important process for the visualisation and identification of cell clusters. Reducing the dimensionality of the data also reduces the noise of the data (Ilicic et al. 2016; Tsuyuzaki et al. 2020), avoids the curse of dimensionality (Altman and Krzywinski 2018) in which a higher number of dimensions leads to an increased volume in which that data can occupy, resulting in extremely sparse data and makes it difficult to distinguish differences between two points due to the distances between them becoming too similar. Dimensionality reduction also compresses the data, making downstream analyses less compute intensive. An increase in similarity between dimensions is also seen as the number of dimensions increases, which can make downstream analyses difficult due to the high multicollinearity.

A reduction has already been made in the gene space in the form of HVG selection, but the resulting datasets are still highly dimensional. Consideration was made as to which reduction method was chosen, as some methods aim to preserve global structure, whereas others aim to preserve local structures (Nguyen and Holmes 2019).

2.12.2.7.1 Principal Component Analysis

The most common dimensionality reduction method is principal component analysis (PCA). PCA is a multivariate linear statistical dimensional reduction technique that transforms a multidimensional dataset into a smaller set of summative indices, known as principal components (PCs) (Karl Pearson 1901; Hotelling 1933; Lever et al. 2017). For scRNAseq, the technique allows for each cell, which currently contains 3000 HVGs, to be condensed into one point along each PC, enabling its visualisation in a one- or two-dimensional plot. Running PCA on a dataset X containing N rows (observations/genes) and K columns (variables/cells) will produce a K-dimensional variable space. Each variable (K) in the dataset (X) represents a coordinate axis (Appendix 2.7a, using three variable axes as example). The length of each variable is first standardised by scaling the unit variance to ensure equal contribution to the analysis. Each observation (N) is then projected onto the K-dimensional variable space and the variable averages are then subtracted from the data to form a vector of averages and represents one point in the K-dimensional space (Appendix 2.7b and c, respectively). Variable average, red point). This average point is then mean-centred (repositioned to the origin of the K-dimensional space), which re-positions the coordinate system (Appendix 2.7d). The first summary index, PC1, is then fitted onto the K-dimensional space so that it passes through the average point and contains most of the variance. Each observation is then projected onto the PC1 line and the coordinates from the average point (origin) are recorded (Appendix 2.7e). This reduces the multidimensional data into one-dimension coordinates. A second summary index, PC2, is then calculated by plotting it orthogonal to the PC1 line (Appendix 2.7f). Again, the PC2 line will pass through the average point (origin) in a direction that contains most of the variance, observations are projected onto PC2,

and coordinates taken. PCA generates K-PCs for the dataset. The first PC contains the maximum information, the second PC contains the remaining maximum information and so on. The PCs are orthogonal vectors, meaning that the variance captured by each PC is independent to each other and the explained variances do not overlap. This method is an unsupervised approach to dimensionality reduction.

PCA was performed on all 4 samples using the 'RunPCA' function, which by default, generates 50 PCs. DimPlots were first generated on the PC1 and PC2 to assess for any primary sources of heterogeneity (Figure 2.3a, d, g, j). When plotted against each other, PC1 and PC2 appear to separate out 3 general populations of cells across all samples. Probing further into the BCC_1 sample, Table 2.6 shows the top 5 sources of heterogeneity in PCs 1-5. Amongst the top positive genes showing heterogeneity in PC1 are Keratins 5, 14, and 17, which are all common markers of epithelial keratinocytes (Wang et al. 2020; Zhang et al. 2022), suggesting the larger cluster at around coordinates 0, 0 (PC1, PC2) are keratinocytes. Along PC2, there are two general populations that reside on either end of the coordinates, with the prospective keratinocyte population residing in the middle. The top positive genes showing heterogeneity in PC2 are PECAM1, PLVAP, EGFL7, RAMP2, and ADGRL4, which are expressed by endothelial cells (Fitch et al. 2004; Malta et al. 2018; Favara et al. 2019; Goncharov et al. 2020; Lai et al. 2021; Denzer et al. 2023). Together, PC1 and PC2 appear to define keratinocytes, fibroblasts, and endothelial cells. Subsequent PCs identify a possible melanocyte population (PMEL, MLANA, CDH19, TRPM1, and CAPN3, defined by PC4 Weeraratna et al. 2004; Guo et al. 2012; Moody et al. 2016; Belote et al. 2021) and contain considerable variance.

The difficulty here is selecting which PCs best represent the data, as using just PC1 and PC2 are not enough to define all expected cell types within the samples. The PCs also are unable to subcluster any populations. Unlike bulk-RNAseq, where PC1 usually defines most of the variance, PC1 in scRNAseq usually defines <10% of the variance due to the complexity of the data (multiple cell types and random noise, for example). Multiple PCs are therefore needed to correctly represent the majority of variance in the datasets. To do this, two methods are commonly used: Jack Straw Plot (Macosko et al. 2015) and Elbow Rule (Nguyen and Holmes 2019). The Jack Straw Plot is a resampling test that randomly permutes the dataset, reruns PCA and constructs a 'null distribution' of feature scores and repeats the process. The method allows for the identification of significant PCs, those who have strong enrichment of low p-value features. Significant PCs should show a left skew in p-value distributions (coloured curves) when compared to the null distribution (dashed line). All samples, except for BCC_3, showed a left skew (Figure 2.3c, f, i , l). BCC_3 sample showed left skew for around 35/40PCs. For BCC_3 and BCC_2_NS samples however, p-value significance started to lose significance after 30PCs.

Another method of PC selection is the Elbow Rule, where the number of PCs is selected when an 'elbow' in the plotted variances is detected. The variances of each PC are plotted using a scree plot and ranked in accordance to variance explained. The variance explained will tail off towards latter PCs and form an 'elbow' in the data, meaning that the majority of the variance has been better explained by previous PCs and that subsequent PCs will explain roughly the same variance. Again, taking Mix_BCC_CD45 sample as example, the scree plot shows that an elbow is present at PC9/10 (Figure 2.3b). However, the variance starts to plateau after PC30, therefore

PCs 1-30 were selected for clustering. Similar patterns of variance were seen with the other three samples, with variance plateauing after PC30 (Figure 2.3e, h, k). Taking both methods into consideration, 30 PCs were therefore selected for clustering.

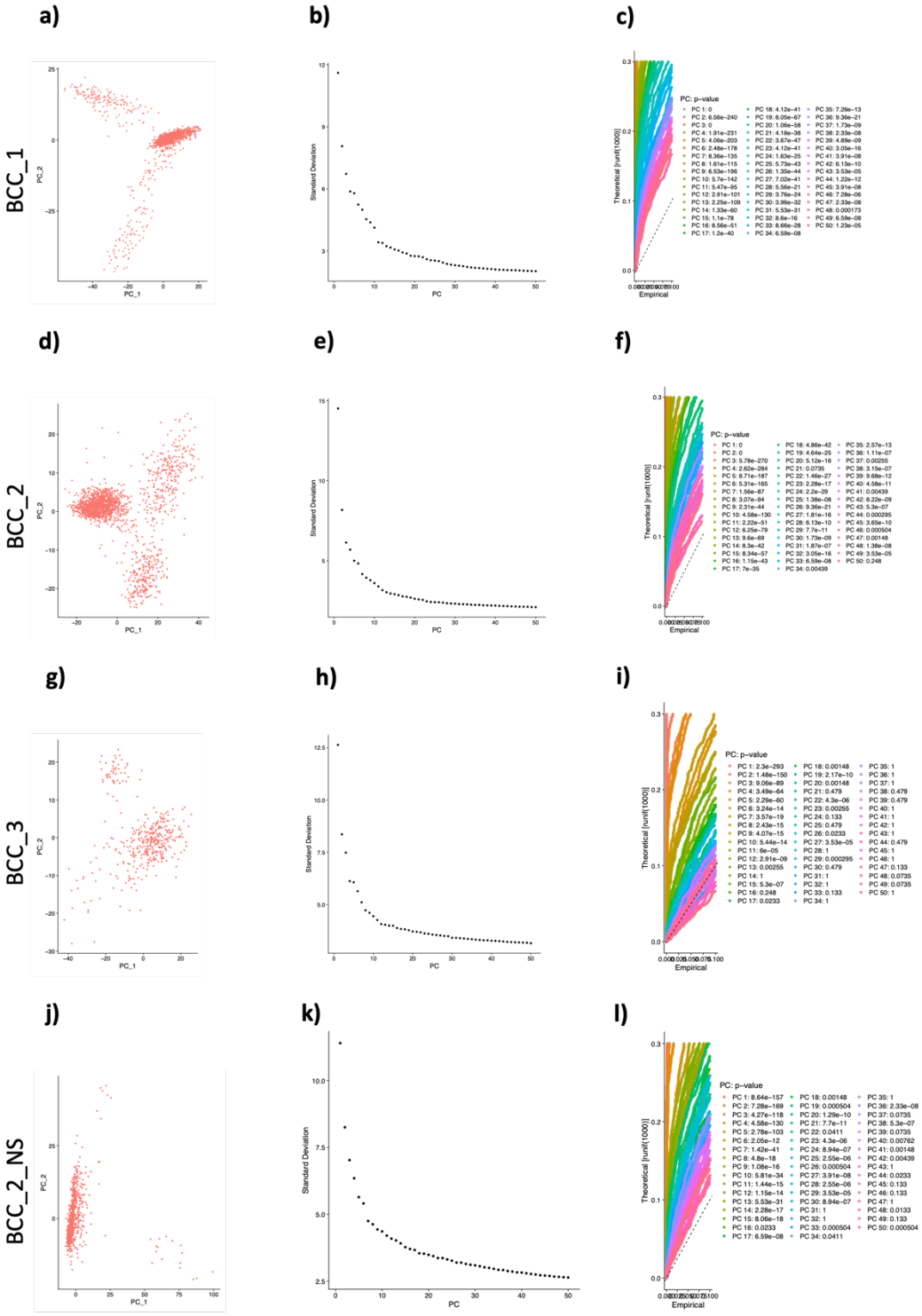

a)
b)
c)
BCC_1
d)
e)
f)
BCC_2
g)
h)
i)
BCC_3
j)
k)
l)
BCC_2_NS
PC_1
PC_2
Standard Deviation
PC
Theoretical [runif(1000)]
Empirical
PC: p−value

Figure 2.3 | PCA and JackStraw analysis of pilot dataset. *PCA was performed on each sample. JackStraw analysis was also performed, which resamples the dataset and randomly permutates it, reruns PCA and constructs a 'null distribution' of feature scores and repeats the process. Each PC is represented as a coloured line on the plot. The method allows for identification of significant PCs, those who have strong enrichment of low p-value features. Significant PCs show a left skew in p-value distributions when compared to the null distribution (dashed line). The elbow rule was also used to select PCs. Variances of each PC are plotted using a scree plot and are ranked in accordance to the variance explained. The number of PCs selected is based on where an 'elbow' occurs in the plot, which means that most of the variance has been better explained by the previous PCs and that subsequent PCs will explain roughly the same amount of variance.* ***a, d, g, j)*** *Biplot of PC's 1 and 2 for each sample. These plots show the spread of the cells along both axes.* ***b, e, h, k)*** *Scree plot for each sample. These plots were used for the elbow rule.* ***c, f, I, l)*** *JackStraw Plot for each sample. dotted line depicts null distribution.*

Principle Component	Positive	Negative
PC_ 1	KRT17	GNG11
	KRT5	A2M
	KRT14	VIM
	DSP	SPARCL1
	CEBPB	CALCRL
PC_ 2	PECAM1	PRRX1
	PLVAP	COL3A1
	EGFL7	PDGFRB
	RAMP2	TWIST1
	ADGRL4	COL6A3
PC_ 3	LGALS7B	ARC
	KRT6A	BGN
	CALML5	ID3
	MT1X	EGR2
	S100A7	ID1
PC_ 4	PMEL	TMSB4X
	MLANA	KRT6B
	CDH19	MGP
	TRPM1	DSC2
	CAPN3	S100A2
PC_ 5	HSPH1	GADD45G
	REL	LMO3
	BAZ1A	TMCO3
	MAFF	RASD1
	HSPD1	DDIT4

Table 2.6 | Top PCA loadings for BCC_1. *Principle Component Analysis (PCA) was performed on the BCC_1 sample. Table shows the top 5 sources of heterogeneity for the first 5 PCs.*

2.12.2.8 Clustering

Having chosen the dimensionality of the samples, clustering can now be performed. Clustering is an important step in the analysis, as it allows for the separation of the cells into unique clusters based on similarity (Appendix 2.8). Appropriate clustering enables cell type separation as well as separation of cell subtypes. The most common approach to clustering is to use an unsupervised hypothesis-free method. Many of these methods exist, including hierarchical clustering, *k*-means, and graph-based (Lafzi et al. 2018).

Hierarchical clustering can either combine individual cells into larger clusters, known as agglomerative clustering, or divide larger groups into smaller ones, known as divisive clustering (Appendix 2.8c; Kiselev et al. 2019). Hierarchical methods provide consistent results without a predefined number of clusters, however, can be extremely compute intensive with increasing dataset size.

K-means clustering methods use *k* centroid estimation (centres of each cluster), assigns each cell to its nearest centroid, recomputes *k* centroids on the basis of the means of the cells in each centroid clusters, and then reiterates the steps until convergence (when the assignments no longer change; Appendix 2.8b; Lafzi et al. 2018). K-means clustering is a fast approach method to clustering given the simplicity of the algorithm. However, the method has shown to favour spherical clusters, which can lead to incorrect clustering on irregular size and shaped datasets (containing multiple cell types, for example). The number of *k* clusters must also be specified, which can lead to under- or over-clustering of the dataset, and the method is completely dependent on the randomly chosen starting coordinates, meaning the

method must be run multiple times to ensure the clustering is stable (Hicks et al. 2021).

Graph based clustering has become a popular approach since the development of the Seurat pipeline(s). The graph-based method embeds cells into a highly dimensional graph-like structure using the PCA space as input, where each node represents a cell, and each cell is connected to its nearest neighbour. Edges are then drawn between cells with similar gene expression patterns, with more similar cells having more weighted edges and less similar cells having less weighted edges. The graph is then partitioned into 'quasi-cliques' or highly connected 'communities' through the use of a Louvian algorithm (Appendix 2.8d, e) (Blondel et al. 2008; Satija et al. 2015a). The graph construction avoids making assumptions about cluster shape and number of cells per cluster and is scalable to larger datasets. The method also accepts a range of data input types, including processed data such as PCs (Lafzi et al. 2018) which has shown to generate better cluster separation. The user also is not required to provide a number of output clusters, instead a resolution parameter is needed. The resolution parameter determines how many nearest neighbours each cell has when constructing the graph, a higher resolution generates more tightly connected clusters (i.e. more, and potentially smaller clusters) and a lower resolution generates less-connected clusters (i.e. less, and larger clusters). A low resolution can be useful for identifying general cell populations within a heterogeneous dataset, whereas a higher resolution may be better at finding rare subpopulations.

The main disadvantage to the method is that the algorithms don't work well with smaller datasets (< 1000 cells), as they rely on many connections between nodes. Information on clustering parameters can be found in Chapter 5.2 and Appendix 5.1

2.12.2.9 t-Stochastic Neighbour Embedding (t-SNE)

T-SNE models the probability distribution of neighbours (set of points) around each point in the dataset. Higher probabilities are given to points (cells) that are more similar, and lower probabilities are given to points that are more dissimilar. Euclidean distances are first measured for each point between all points. Probability (which represents similarity) is then measured, and a Gaussian centred probability distribution is created. The same method is then performed on the points in the two- or three-dimensional space (dimensions are provided by the user) with exception to a t-distribution used to measure the probability between the low dimensional points. The distance between the two distributions, in relation to the coordinates of the points, is then minimised using Kullback-Leibler divergence model and the data is successfully reduced. The overall goal of t-SNE is to create a two- or three-dimensional mapping that maximises similarities and minimises differences between the distributions over all points in the dataset. t-SNE efficiently highlights local structure, however, suffers at computing inter-cluster relationships and is faced with slow compute times when scaled to large-scale datasets (Becht et al. 2018). Users also must provide a 'perplexity' value when performing t-SNE. Perplexity is a fitting parameter which is coarsely equivalent to the number of nearest neighbours to use when running the algorithm. A low perplexity value aims to preserve local interactions, whereas higher values aim to preserve a more global structure. T-SNE is also a stochastic method, which chooses random starting points each time it is run.

This starting point is also referred to as a seed value. The method is also sensitive to perplexity values, and so users should therefore save the seed which was first used and run the method multiple times to choose the best perplexity value (Andrews and Hemberg 2018). T-SNE parameters can be found in Chapter 5.2 and Appendix 5.2

2.12.2.10 Uniform Manifold Approximation and Projection (UMAP)

To overcome some of t-SNE's limitations, UMAP was developed by McInnes and colleagues in 2018. The authors claim this method preserves just as much local structure and more global structure, coupled with significantly faster compute times. The method is comparable to t-SNE, where it aims to preserve the relationships between the datapoints in a high-dimensional space and project into a two- or three-dimensional space.

The method first constructs a 'fuzzy simplical complex'. A weighted graph is constructed using a complex mathematical formula, involving the generation of a Čech complex and a Vietoris-Rips complex. A Čech complex is a method of representing a topology, which uses building blocks known as simplices (a k-dimensional object which is formed through the connection of k + 1 points). A 0-simplex is a point, a 1-simplex is a line, and a 2-simplex is a triangle. The topology of the dataset can be accurately captured and represented through the use of simplices, and a Čech complex can be formed through the combination of these simplices. Each datapoint in the dataset represents a 0-simplex and extending a radius, the size of which is determined by the distance to its kth-nearest neighbour, from each 0-simplex will construct a set of 1-, 2-, and higher-simplices. When two radii intersect, a 1-simplex is generated. As the radii continue to expand, increasing simplices are

formed. The graph then becomes ‘fuzzy’. Each connection is made into a probability of similarity, the further away cells are from each other, the lower the probability and the more ‘fuzzier’ the graph. Therefore, as the radius grows, the less likely cells are to be related and the ‘fuzzier’ the graph becomes. Most of the dataset can be represented by 0- and 1-simplices, which compose the Vietoris-Rips complex. Computationally, this complex is very easy to work with, and is what gives UMAP its superior speeds. Local structure is furthermore preserved by requiring that each point must be connected to its nearest neighbour. This generates a highly dimensional space which is then used to construct and optimise a two- or three-dimensional space in a similar manner to t-SNE. This projection into a lower dimensional space is significantly quicker than t-SNE due to the mathematical approaches used during the construction of the fuzzy simplical complex.

UMAP also has two hyperparameters that can be adjusted: n_neighbours and min_dist. N_neighbours is the number of approximate nearest neighbours used to construct the initial graph. This parameter is used to balance local versus global structure. Min_dist is the minimum distance between the points in the low dimensional space and controls how tightly packed the points are together on the UMAP plot. These parameters are easier to understand than the t-SNE parameters and help to generate more meaningful representations of the dataset. UMAP parameters can be found in Chapter 5.2 and Appendix 5.2.

2.12.2.11 Cell Type Identification

Clusters were next identified and appropriately annotated using marker visualisation and DEG analysis. Cell type markers were obtained from the Human Protein Atlas

(HPA) and various journal articles (Ji et al. 2020; Solé-Boldo et al. 2020; Wang et al. 2020; Guerrero-Juarez et al. 2022; Yerly et al. 2022; Quah et al. 2023; Ren et al. 2023; The Human Protein Atlas. 2023). Markers used can be found in Appendix 5.3. UMAP and Violin Plots were used to visualise each marker, and Seurat's 'FindAllMarkers' function was used to generate DEGs for each cluster. Top cluster DEGs were then visually assessed for common cell type markers, and also used as input into g:Profiler and the g:GOSt functional profiler analysis was performed. The HPA result section was scrutinised the most, as results such as 'skin1: melanocytes' worked in conjunction with the marker visualisation. Once clusters were identified, identities were assigned using the 'RenameIdents' function.

2.12.2.12 Dataset Integration

The method addresses common problems associated with the investigation of multiple samples, such as the number and type of cells, and technical properties such as sample processing, library preparation, and sequencing depths. There are different integration methods, each encompassing different algorithms and assumptions on the data (Argelaguet et al. 2021; Forcato et al. 2021). All methods of integration use an 'anchor', a feature which is used to link all datasets. There are three types of anchors that can be used, and therefore three types of integration: Horizontal integration (features as anchors), vertical integration (cells as anchors), and diagonal integration (no anchor). Horizontal integration is used for same-modality experiments where cells have been profiled from the same tissues across independent samples. This method uses the common gene set as the integration anchor. Vertical integration is used for multi-modality experiments where different assays have been performed on the same cells (matched assays). For this method,

the cells would be used as the integration anchors. Finally, diagonal integration is used where both the cells and the genomic features differ between experiments. An example of this is where scRNAseq and single cell assay for transposase-accessible chromatin using sequencing (scATACseq) are applied to separate groups of cells. All of my experiments involved the transcriptomic profiling of the same tissues across independent samples, so horizontal integration was performed. I selected and tested three of what I considered to be the most used horizontal integration methods at the time: Seurat v3 (Stuart et al. 2019), scTransform (Hafemeister and Satija 2019a), and Harmony (Korsunsky et al. 2019). These three methods, along with others, are classically referred to as batch correction tools, as they address the problems faced with different batches of experiments (both biological and technical variation). It is important to note that whilst the integration techniques group cells based on similar gene expression, this does not necessarily mean that those grouped cells belong to the same cluster. These integration techniques are used primarily as a batch correction method.

2.12.2.12.1 Seurat v3 Integration

Samples are first processed individually by log normalising and selecting HVGs. An initial dimensional reduction is made in the form of a Canonical Correlation Analysis (CCA; Butler et al. 2018), which finds linear combinations of features across cells in the datasets that are maximally correlated, therefore identifying shared correlation structures across those datasets (also known as diagonal-CCA). The CCA returns vectors, similar to PCA, but whose gene-level projections are correlated between the datasets, but not necessarily aligned. Similar to Seurat's clustering method, each CCA vector represents a metagene, and each metagene is linearly transformed, and the

mapping between each metagene is stretched or compressed through 'dynamic time warping', which corrects for changes in population density. This results in a single, aligned, low-dimensional space that represents all the datasets. The key difference between CCA and PCA here is whilst CCA identifies correlated gene modules that are present in all datasets that define shared biological states, PCA will only identify sources of variation in individual datasets whether said variation is present across all samples or not. However, the authors note in a later publication that the comparison of CCA vectors across datasets can still be problematic (Stuart et al. 2019). CCA, unlike PCA, does not account for global differences in scale between each dataset (such as differences in normalisation). A new version of Seurat's integration was presented (v3), which addressed this issue and also built upon the algorithm to further improve the integration of datasets. The initial CCA dimensional reduction is firstly improved through L2-normalisation of the CCA vectors across all samples, which scales the expression of genes across the datasets, so they are on a comparable scale. Post-dimensional reduction, K-Nearest Neighbours (KNNs) are calculated for each cell within each paired dataset based on the L2-normalised CCA vectors. KNNs is an algorithm that calculates the k (k is a user defined hyperparameter) most similar points to a new data point, and then assigns the new data point to the same class (or cluster) as the majority of its neighbours. KNN firstly calculates the distances between each data point and all other data points, and the k data points that have the shortest distances are then considered to be the nearest neighbours of the new data point. Finally, the KNNs are then used to calculate the Mutual Nearest Neighbours (MNNs) for each cell across different datasets that are in the same cluster. These pairwise incidences are referred to as 'anchors'. This method is essentially identifying clusters of cells that have similar gene expression for each

dataset first, and then overlaying clusters between datasets that have similar gene expression to create one integrated dataset (Appendix 2.9). An interesting point to make about this method is that when performing integration, the user can use a 'reference' dataset. This reference dataset can be used to guide the algorithm (as cluster identities are already known) when integrating 'query' datasets. The method can also be run without using a reference. The anchor identification is key for the integration of the datasets. In some incidences, there will be a population of cells that are present in one dataset, but not present in the other and should therefore not create any anchor pairs. However, in practice, this does happen in low frequencies. To combat this, anchor filtering and scoring is performed before the final integration occurs. Firstly, the nearest neighbours of each anchor query cell are checked in the reference dataset. If the anchor reference cell is found within the first k.filter neighbours (again, user defined), then the anchor is retained. Secondly, to further minimise the likelihood of incorrectly identified anchors, a shared nearest neighbours (SNN) graph is generated for each reference anchor cell. This shows which cells are neighbours of the anchor cell in both the reference and the query datasets. The shared neighbour overlap is then computed between anchor and query cells and is assigned as an anchor score. Higher scoring anchors are more likely to be accurate than lower scoring anchors. This anchor score is then implemented into the algorithm during final integration. After Seurat integration, the dataset is then processed as previously described (scaled, PCA, Clustering, UMAP).

2.12.2.12.2 scTransform Normalisation

scTransform, whilst not an integration method, is a more recently endorsed normalisation strategy that aims to avoid some of the pitfalls associated with the

standard normalisation processes. The authors (Satija lab – Seurat developers) have included this normalisation as an improvement to their standard integration pipeline at the bottom of their manual page, hence why I have included it as another integration method in this section. The publication came at the end of 2019, shortly after their Seurat v3 integration update (Hafemeister and Satija 2019a).

The authors observed, for the standard log normalisation approach, did not affect all genes equally. Genes with low-medium abundances were typically normalised more effectively than genes with high abundances. This is because the variances of low abundance genes are amplified with log normalisation, but very little effect is seen on variances of high abundance genes. They also observed a significant imbalance of variances in the data. Cells with low total UMI counts tend to have higher variances for higher abundance genes, this is due to the amplification of variance in all genes by log transformation, but the variance is already high in cells with low total UMI counts, leading to a situation where the variance of high abundance genes is disproportionately high, making identification of differentially expressed genes difficult. To combat this, the authors solution was to use Pearson's residuals to transform the data. Briefly, the Pearson's residuals from a 'regularised negative binomial regression', in which cell sequencing depth is utilised as a covariate in a generalised linear model, can successfully remove the technical noise from downstream analyses whilst simultaneously preserving biological heterogeneity. Rather than applying a global scaling factor (as is the case in log normalisation), this model multiplies gene expression by a gene-specific weight which is governed by evidence that a gene is non-uniformly expressed across cells. Genes with a more uniform expression across cells will have a lower weight, whereas genes with a more

non-uniform distribution of expression will have a higher weight. The benefit to this approach is that the model considers both the expression of the gene and the distribution of its expression across the cells, allowing for the effective identification of rare cell types.

2.12.2.12.3 Harmony Integration

Harmony integration was first introduced by Korsunsky and colleagues in late 2019 (Korsunsky et al. 2019) as a 'superior' approach to integrating of datasets. The authors claimed this approach could be performed on up to one million cells on a personal computer. The method first starts with a low-dimensional space generated through PCA analysis, and then applies L2-normalisation to each cell to generate a cosine distance. This cosine difference is a more robust measurement of cell-to-cell similarities. This projects the cells into a unit hypersphere, and so the next step of the algorithm uses k-means clustering (Chapter 2.11.2.8) with 10 random restarts to find initial locations of the cluster centroids. These cluster centroids are then used as input into another algorithm that assigns cluster membership to each cell. This algorithm calculates the distance of each cell to each cluster centroid and assigns a value from 0 to 1. The initial cluster diversity is then calculated by aggregating the number of cells from each dataset assigned to each cluster. This initialised Harmony object (initial cluster assignments, scaled PC embeddings, and cluster diversity statistics) is then used to perform Harmony clustering (or maximum-diversity clustering), which performs iterative (user defined) centroid estimation and cluster assignment. Cells are then re-assigned to the cluster centroids using the cluster diversity to maximise diversity within the clusters, and cluster centroids are then re-positioned to be closer to their assigned cells. The result here is the identification of

shared groups of cells between datasets. A correction model is then applied to the data, which, for each cell, estimates how much its dataset identity contributes to its PCA scores and subtracts this contribution from that cells PCA score. Harmony then reiterates all steps above until cell cluster assignments are stable (Appendix 2.10).

2.12.2.12.4 Integration Method Testing

Default parameters were used for all three methods, with exception for Seurat v3 integration, where no reference dataset was used. For Seurat v3 integration, samples were processed individually (QC), normalised, cell cycle scored, and HVGs were selected (as described in Chapter 2.11.2.2-2.11.2.5). Samples were then integrated using Seurat's 'SelectIntegrationFeatures', 'FindIntegrationAnchors', and 'IntegrateData' functions. The integrated dataset was then scaled, and PCA, clustering, and UMAP was performed (as described in Chapter 2.11.2.6-2.11.2.8, and Chapter 2.11.2.10, respectively). For scTransform, samples were processed individually, and cell cycle scoring was performed. Samples were then normalised, scaled and HVGs were selected using the 'SCTransform' function. Integration was then performed using 'SelectIntegrationFeatures', 'PrepSCTIntegration', and 'FindIntegrationAnchors' and 'IntegrateData' using "SCT" as the normalisation method. The integrated dataset was then processed in the same manner as the Seurat v3 method. For Harmony integration, samples were first processed individually and then merged into one dataset using the 'merge' function. The merged dataset is then normalised, cell cycle scored, HVGs selected, and PCA performed. 'RunHarmony' function is then used to perform Harmony's integration on the merged dataset. The integrated dataset is then clustered and UMAP is performed.

2.12.2.13 Downstream Analysis of Integrated Data

2.12.2.13.1 Calculating the Percentage of Cells Expressing Marker Genes in Each Cluster for Each Sample

To account for these differences, a duplicate dataset was made for each marker gene and an expression filter of > 1 (log of normalised reads + 1) was used. The following calculation was then used to identify the percentage of cells expressing the marker gene > 1 in each cluster for each sample group:

$$\frac{\boldsymbol{Number\ of\ cells\ with\ expression > 1\ in\ cluster}}{\boldsymbol{Total\ number\ of\ cells\ in\ cluster}} \times \mathbf{100}$$

2.12.2.13.2 Copy Number Variation Analysis

inferCNV was used to identify putative tumour cells in the samples. inferCNV is a Python package that can also be run in R via the Reticulate package (version 1.34.0) which provides a Python interface. For the inferCNV package (version 1.3.3) to operate functionally, the following packages were additionally installed: igraph (version 1.5.1), Seurat-disk (version 0.0.0.9015), Seurat-data (version 0.1.0), DECIPHER (version 2.30.0), and Phylogram (version 2.1.0). A conda environment was created and selected using the '*reticulate::use_condaenv*' function, and the following Python packages were installed using the '*reticulate::py_install*' function: NumPy (version 1.22.3), Python-igraph (version 0.11.2), Pandas (version 2.0.3), UMAP-learn (version 0.5.4), and Leidenalg (version 0.10.1). The integrated dataset was next subsetted into individual samples via the '*subset*' function using the 'orig.ident' metadata slot to select each sample. subsetting of the integrated dataset enabled

each individual sample to retain the clustering and cell type information. To run inferCNV, count matrices were generated for each sample by pulling the counts for each cell from the RNA assay for each sample. Sample annotations were also generated by pulling the Seurat cluster information from the metadata and reformatting it into a data frame. An InferCNV object was then created using the '*CreateInfercnvObject*' function, using the generated counts matrices, sample annotations, and a downloaded human Gencode gene coordinate file (version 19) as input, and Melanocytes, Fibroblast-like, and Endothelial cells as the references. inferCNV was then run on the inferCNV object via the '*infercnv::run*' function with analysis mode = 'samples', 'ward.D2' as the hierarchical clustering method, and a value of 0.1 as the cut-off which is vital for 10x Genomics data. Following UMAP and cell annotation of the main dataset, keratinocyte populations were subsetted via the '*subset*' function using the cell identities as the subsetting factor. PCA, clustering, and UMAP was rerun on the keratinocyte dataset, with a higher resolution parameter to generate smaller, more defined clusters. Keratinocyte populations were then identified through marker expression and gene set expression. Gene set expression was assessed through the '*AddModuleScore*' function, which takes a list of genes as input and calculates a 'module score' for that gene set. Gene sets were obtained from the Molecular Signatures Database (mSigDB), CellMarker 2.0 database, and journal articles. Keratinocyte populations were annotated as previously described, and populations of interest were manually selected via the '*CellSelector*' function and selected cells were annotated and added to the cell identities in the respective metadata slot. DEGs were then generated between the various populations using the '*FindMarkers*' function, using the 'ident.1' and 'ident.2' as the comparative entities.

2.12.2.13.3 RNA Velocity Analysis

A minor pitfall to scRNAseq experiments is that they provide only a snapshot of cellular dynamics at the time the cell was processed. Whilst providing an unbiased and high-resolution transcriptomic landscape of captured cells, it does not conventionally offer a look into the dynamical transitions between cellular states (Kolodziejczyk et al. 2015; Li et al. 2023). This naturally led to the development of trajectory inference algorithms which constructed potential branching trajectories which was based on the similarity of the transcriptomic profiles of the cells (Bendall et al. 2014; Haghverdi et al. 2016; Setty et al. 2019). The main challenge of these algorithms was determining the direction of the velocities and the root and terminal cell states. One such method to incorporate all this is RNA velocity analysis, which counts nascent (unspliced) mRNA's and mature (spliced) mRNA's and computes their correlation (La Manno et al. 2018). RNA velocity analysis was performed using the python package, scVelo (Bergen et al. 2020). The Seurat object was first exported from R into an anndata object and then the standard scVelo pipeline was performed using the stochastic model. RNA velocity was performed using the scVelo Python package (version 0.2.5). Three datasets were analysed: BCC & NS Keratinocytes, BCC keratinocytes, and NS keratinocytes. The BCC and NS keratinocyte datasets were generated by subsetting the BCC & NS keratinocyte dataset with the '*subset*' function. Firstly, metadata tables were generated for each dataset by extracting the following: cell IDs, UMAP embeddings, and cell barcodes. Expression matrices were generated by extracting the count data from the RNA data slot via the '*GetAssayData*' function. Dimensionality matrices were generated by extracting the PCA embeddings and finally, gene names were also extracted. Sample loom files were generated using the HAWK supercomputer. Briefly, the velocyto (version 0.17) and samtools (version

1.10) modules were loaded and velocyto was run on each sample using each samples barcode and binary alignment map (BAM) files as input along with a human gene reference file (Gencode GRCh38). In Python, the following packages were then loaded: ScanPy (version 1.9.4), anndata (version 0.10.0), NumPy (version 1.22.3), Pandas (version 2.0.3), cellrank (version 1.5.1), and scVelo (version 0.2.5). Metadata tables were read in along with the loom file to make an anndata object for each sample. Sample anndata objects were then concatenated into one object and pre-processed using the '*scv.pp.filter_and_normalize'* and '*scv.pp.moments'* functions. RNA velocity was then estimated using the '*scv.tl.velocity'* function and embedding fields were projected using the '*scv.pl.velocity_embedding_stream'* function. Trajectory inferences were also generated by using PAGA graph abstraction. PAGA velocity graphs were generated using the '*scv.tl.paga'* and '*scv.pl.paga'* functions.

2.12.2.13.4 Single Cell GSEA

Single cell GSEA (scGSEA) was performed on the clusters to understand the enriched pathways and gene sets between them. The hierarchical clustering from the scGSEA analyses was also used to investigate cluster similarity. scGSEA was performed using the singleseqgset package in R. Firstly, the gene sets of interest were downloaded and loaded into R: Hallmarks, C2 Biocarta, C2 KEGG, C2 PID, C2 Reactome, and C2 WikiPathways. The log fold change was then calculated between clusters for all genes using the logFC function. scGSEA was then performed on each cluster for each gene set using the Wilcoxon Mann Whitney correlation corrected GSEA function (wmw_gsea). The statistics and p value columns were then extracted and corrected for multiple comparisons. Heatmaps were then generated for each gene set database using the heatmap3 package and scaling the rows.

Chapter 3: The Role of CD200 on BCC Cancer Stem Cells

3.1 Introduction

Our previous work demonstrates that CD200 is expressed by BCC cancer stem cells and is essential for the formation of the BCC tumour (Colmont et al. 2012; Colmont et al. 2013b). However, we were unable to definitively conclude whether CD200 is a functionally relevant marker of these cancer stem cells, or if it is merely just a protein that happens to be located on the surface of these cells. Given that only 1 in 822 CD45- CD200+ BCC cells are able to recapitulate the tumour, the latter statement is likely to be true. However, CD200 is an immune modulator and therefore may be essential for BCC cancer stem cell survival.

There have been multiple lines of evidence to suggest the ability of the host immune system is able to eradicate BCC. In the immunosuppressed setting, such as organ transplant recipients, the incidence of BCC is higher. Cessation of the immunosuppression has shown that the risks associated with immunosuppressive therapies are diminished (Lobo et al. 1992; Euvrard et al. 1995; Otley et al. 2001). Similarly, an increased risk in BCC incidence is also seen in HIV patients (Lobo et al. 1992; Crum-Cianflone et al. 2009). Histological analyses have shown that 20% of BCCs demonstrate regression that is associated with an invading inflammatory cell

infiltrate (Go et al. 2019). Multiple studies have also investigated the effects of various therapies, such as direct stimulation of the immune system by injection of interferons which resulted in BCC resolution, and peri-lesional injection of interleukins which resulted in BCC killing (RC et al. 1990; Lkić et al. 1991; Alpsoy et al. 1996; Kaplan and Moy 2000). The BCC Tumour Microenvironment (TME) contains a large amount of tumour infiltrating CD45+ inflammatory cells, representing ~13% +/- 11% of all cells. The BCC immune response consists mainly of T cells, with a ratio of 2:1 (CD4+/CD8+), and compared to normal surrounding skin, BCC tumours contain significantly greater numbers of Tregs (>25% of tumour associated CD4+ T cells; Kaporis et al. 2007). Attenuation of dendritic cells (DCs), T cell effector responses, and NK cell recruitment is also seen in BCC TME due to the release of IL-2, TGF-β, IL-2, and IL-12 by the BCC keratinocytes (Curson and Weedon 1979; Wong et al. 2000; Stary et al. 2007; Vivier et al. 2011).

CD200 is an immunomodulatory protein that binds to its receptor, CD200R, which in turn downregulates immune cell activity (Barclay et al. 2002). CD200 is widely expressed in a range of cell types including B and T cells, dendritic cells, endothelial cells, and kidney cells, to name a few (Wright et al. 2001). CD200:CD200R interaction has been shown to modulate immune cell activity via inhibition of proinflammatory cytokine secretion and NK and basophil cell activities, enhancement of anti-inflammatory cytokine secretion, and promotion of regulatory T cell induction (Choe and Choi 2023). Overexpression of CD200 has been seen in a number of solid cancers and haematological malignancies and its increased expression has been linked to poor prognosis (Coles et al. 2011). The expression of CD200 is also thought to provide cells, in particular stem cell populations, with an immune-protective function which

in turn protects from immune damage (Gorczynski et al. 2001). Our lab has previously shown that CD200 is also expressed by BCC cancer stem cells and have hypothesised that its expression provides an immune-evasive role (Colmont et al. 2013b; Morgan et al. 2019).

Although the CD200+ CD45- BCC cell subpopulation has a tumour initiating cell (TIC) frequency of 1 in 822, expression of CD200 may still be functionally relevant. Therefore, I first sought to determine whether CD200 expression in BCC, in particular BCC cancer stem cells, was functionally important. To do this, I first assessed whether CD200 in a non-skin cancer cell line was able to cause immune evasion. I then xenografted CD200+ cells into mice and assessed the same thing. Lastly, I then used a well characterised BCC cell line and performed the same experiments and additionally assessed colony forming ability of the sorted cells to determine the functional relevance of CD200. The majority of the work from this chapter was included in a recent publication where we assessed CD200's ability to cause immune evasion (Morgan et al. 2022).

3.2 Optimisation of CD200 Flow Sorting Protocol

I first wanted to determine the specificity of the CD200 antibody and the ability of it to detect positive, and not negative, cells. I used a cervical keratinocyte cancer cell line, HeLa, which does not normally express CD200, that had been transduced in our lab with CD200+-GFP or vector control-GFP (Morgan et al. 2022). Flow cytometric analysis of HeLa CD200+ and CD200- cell lines reproducibly showed that the CD200 antibody specifically identified CD200+ cells and could distinguish CD200- cells (n=5;

Figure 3.1a, HeLa wt; Figure 3.1b, HeLa-GFP+-CD200-; Figure3.1c, HeLa-GFP+-CD200+). Both cell lines also demonstrated consistent protein expression through flow cytometry (Figure 3.1d; HeLa CD200+, 93.90% +/- 0.21 SEM. HeLa CD200-, 0.31% +/- 0.11 SEM) and CD200 RNA expression was confirmed by RT-PCR when flow-sorting the HeLa CD200+ cell line into CD200+/CD200- cells (Figure 3.1e; HeLa CD200+, 2090 AU. HeLa CD200-, 0 AU. n=3).

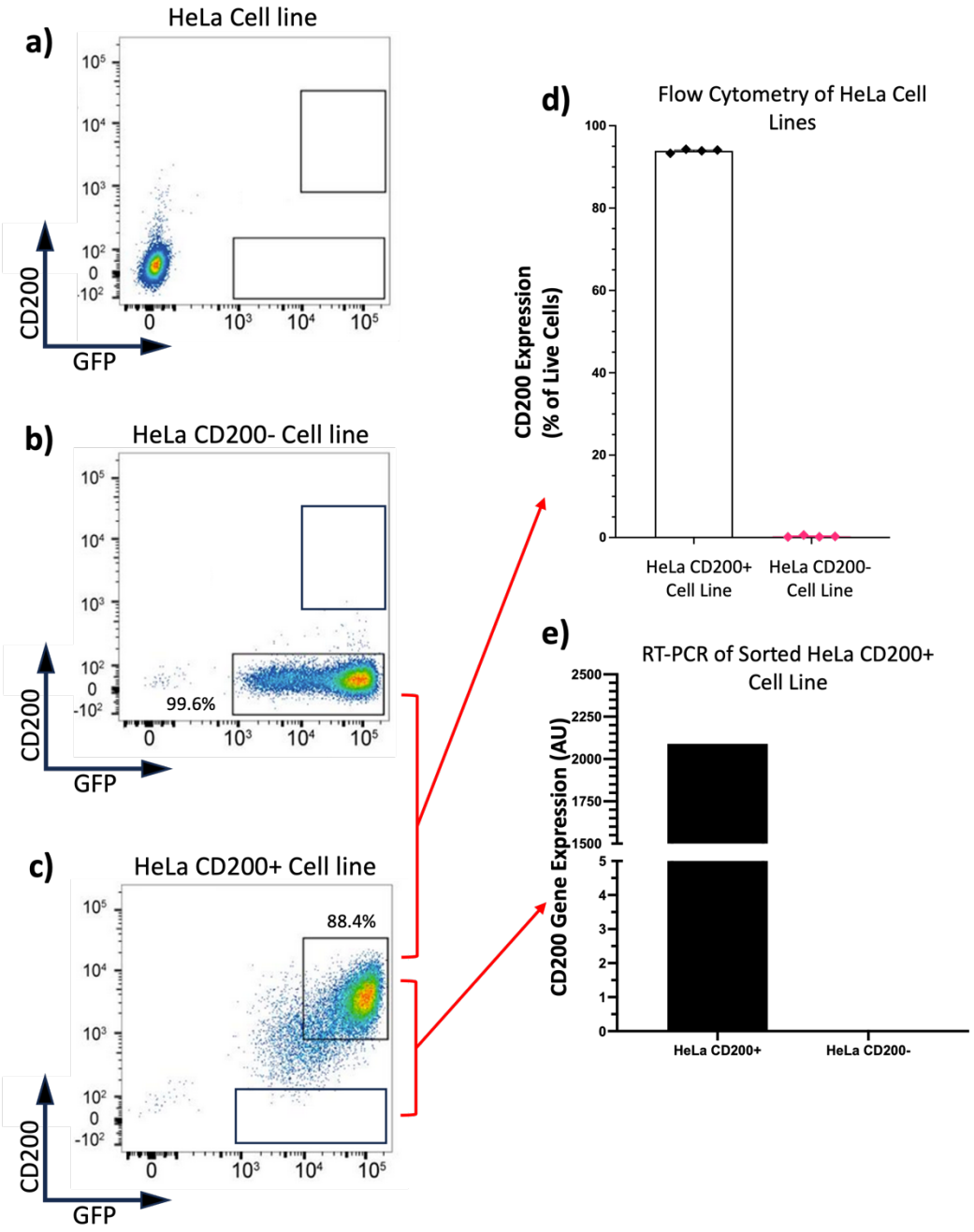

a)
HeLa Cell line
CD200
GFP
b)
HeLa CD200- Cell line
99.6%
CD200
GFP
c)
HeLa CD200+ Cell line
88.4%
CD200
GFP
d)
Flow Cytometry of HeLa Cell Lines
CD200 Expression (% of Live Cells)
HeLa CD200+ Cell Line
HeLa CD200- Cell Line
e)
RT-PCR of Sorted HeLa CD200+ Cell Line
CD200 Gene Expression (AU)
HeLa CD200+
HeLa CD200-

Figure 3.1 | Assessment of CD200 expression in HeLa cell lines. *HeLa cells were transduced with CD200+-GFP or vector control-GFP and run through a flow cytometer to assess GFP and CD200 expression.* ***a)*** *HeLa wild type cell line,* ***b)*** *HeLa vector control (CD200-) cell line,* ***c)*** *HeLa CD200+ cell line.* ***d)*** *Percentage of live cells expressing CD200 in HeLa CD200+ and HeLa CD200- cell lines, n = 4 biological replicates per cell line. Error bar shows mean +/- SEM* ***e)*** *HeLa CD200+ cell line was flow sorted into CD200+ and CD200- cells, and RT-PCR was performed to assess CD200 RNA expression, n=3 technical replicates.*

3.3 CD200 Expression is associated with a Reduction in Immune Cell Infiltration and Increases NK Cell Apoptosis in Transduced HeLa Xenografts

Our in vitro data, which is now published (Morgan et al. 2022), had determined that BCC demonstrated an absence of NK cells within the tumour microenvironment due to CD200-mediated apoptosis. As CD200 is a cell surface marker used to enrich BCC cancer stem cells, I next wanted to determine whether CD200 had a functional role that may substantiate its role as a cancer stem cell associated protein. To address this, I performed xenotransplantation of HeLa CD200+ and HeLa CD200- cells into athymic nude mice (Figure 3.2a). Having determined that 10^6 HeLa cells could generate a 1cm tumour within mice (Liu et al. 2012), 1 x 10^6 CD200+ and CD200- cells were injected into the left and right flanks of 6- to 8-week old nude mice (both n = 5). After 5 days, tumours reached size thresholds and mice were subsequently culled. HeLa CD200+ tumours demonstrated ubiquitous CD200 expression when compared to HeLa CD200- tumours (Appendix 3.1). Initial measures of the tumours showed no significant differences in volume (CD200+, 8.26 +/- 5.26 mm^3 and CD200-, 11.42 +/- 11.52 mm^3, mean +/- SEM; Figure 3.2c). H&E-stained sections (Figure 3.2b) comparing HeLa CD200+ to HeLa CD200- tumours determined significantly greater number of tumour cells (1,952 +/- 114.80 vs 1,240 +/- 90.99 per mm^2 of tissue, mean +/- SEM; $P < 0.001$; Figure 3.2d) and lower necrotic areas (Figure 3.2c, necrotic areas marked with asterisk, and Figure 3.2e; 0.14 +/- 0.04 vs 0.36 +/- 0.07, mean +/- SEM; $P < 0.05$). Importantly, CD200+ tumours demonstrated a significant decrease in immune cell infiltrate (1,180 +/- 77.39 vs 1,551 +/- 128.60 per mm^2 of tissue, mean +/- SEM; $P < 0.05$; Figure 3.2f). HeLa CD200+ tumours also demonstrated fewer NK cells within the infiltrate when compared to HeLa CD200- tumours (NK1.1; 1,024 +/-

239.8 vs 2,085 +/- 251.9 NK cells per mm^2 of tumour tissue, respectively; Figure 3.3a, c and Figure 3.3bi). Thus, CD200 expression on HeLa cells led to immune evasion.

To determine whether CD200 led to NK cell apoptosis, I next investigated the number of NK cells that were expressing CD200R and Cleaved Caspase 3. To address this, I used serial sectioning and staining, and identified the same region of NK cells across the sections. This allowed me to firstly quantify the number of NK cells (Figure 3.3bi and Figure 3.3c) and then quantify the percentage of these NK cells that express CD200R (Figure 3.3bii and Figure 3.3d) and Cleaved Caspase 3 (Figure 3.3biii and Figure 3.3e). I found that HeLa CD200+ tumours had a significantly greater number of apoptosing NK cells when compared to HeLa CD200- tumours (25.45 +/- 5.18% vs 9.88 +/- 2.70%; mean +/- SEM; Figure 3.3biii and Figure 3.3e). There was no significant difference in CD200R expressing NK cells (11.76 +/- 3.00% vs 19.05 +/- 7.31%; mean +/- SEM; Figure 3.3bii and Figure 3.3d). Together, these findings supported our in vitro findings such that BCC CD200 ectodomain shedding leads to an absence of NK cells within the TME, thus suggesting that its functionally relevant to BCC cancer stem cells (Morgan et al. 2022).

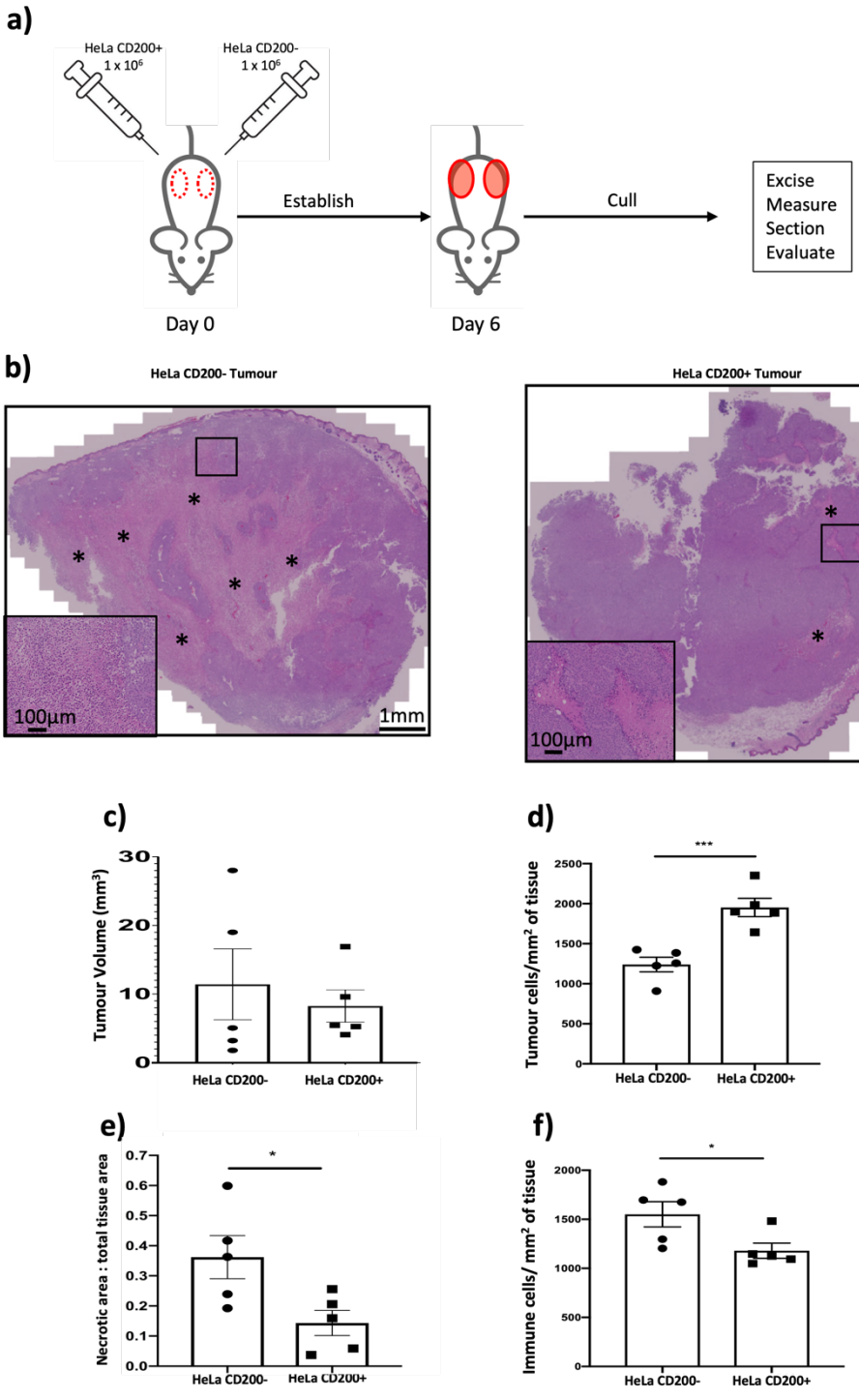

a)
HeLa CD200+
1 x 10⁶
HeLa CD200-
1 x 10⁶
Establish
Cull
Excise
Measure
Section
Evaluate
Day 0
Day 6
b)
HeLa CD200- Tumour
HeLa CD200+ Tumour
100µm
1mm
100µm
800µm
c)
Tumour Volume (mm³)
HeLa CD200-
HeLa CD200+
d)
***
Tumour cells/mm² of tissue
HeLa CD200-
HeLa CD200+
e)
*
Necrotic area : total tissue area
HeLa CD200-
HeLa CD200+
f)
*
Immune cells/ mm² of tissue
HeLa CD200-
HeLa CD200+

Figure 3.2 | Xenotransplantation of HeLa CD200+/- cells. *HeLa cells were transduced with CD200+-GFP or vector control-GFP and injected into the left and right flanks of athymic nude mice (both n=5) and left to establish for 5 days. Mice were then culled, and tumours harvested. Dissected tumours were fixed, paraffin embedded, and sectioned.* ***a)*** *Experiment schematic.* ***b)*** *Haematoxylin & Eosin (H&E) staining of both tumours. Asterisks denote necrotic areas. Zoomed panel shows enlarged necrotic areas.* ***c)*** *Tumours were measured post-dissection and volumes calculated in* mm^3. ***d)*** *Total tumour cells were counted and displayed as cells per* mm^2 *of tissue.* ***e)*** *Ratio of necrotic area to total tissue was calculated.* ***f)*** *Total immune cells were counted and displayed as cells per* mm^2 *of tissue. Error bars = mean +/- SEM; *P<0.05; ***P<0.001 by Mann Whitney test.*

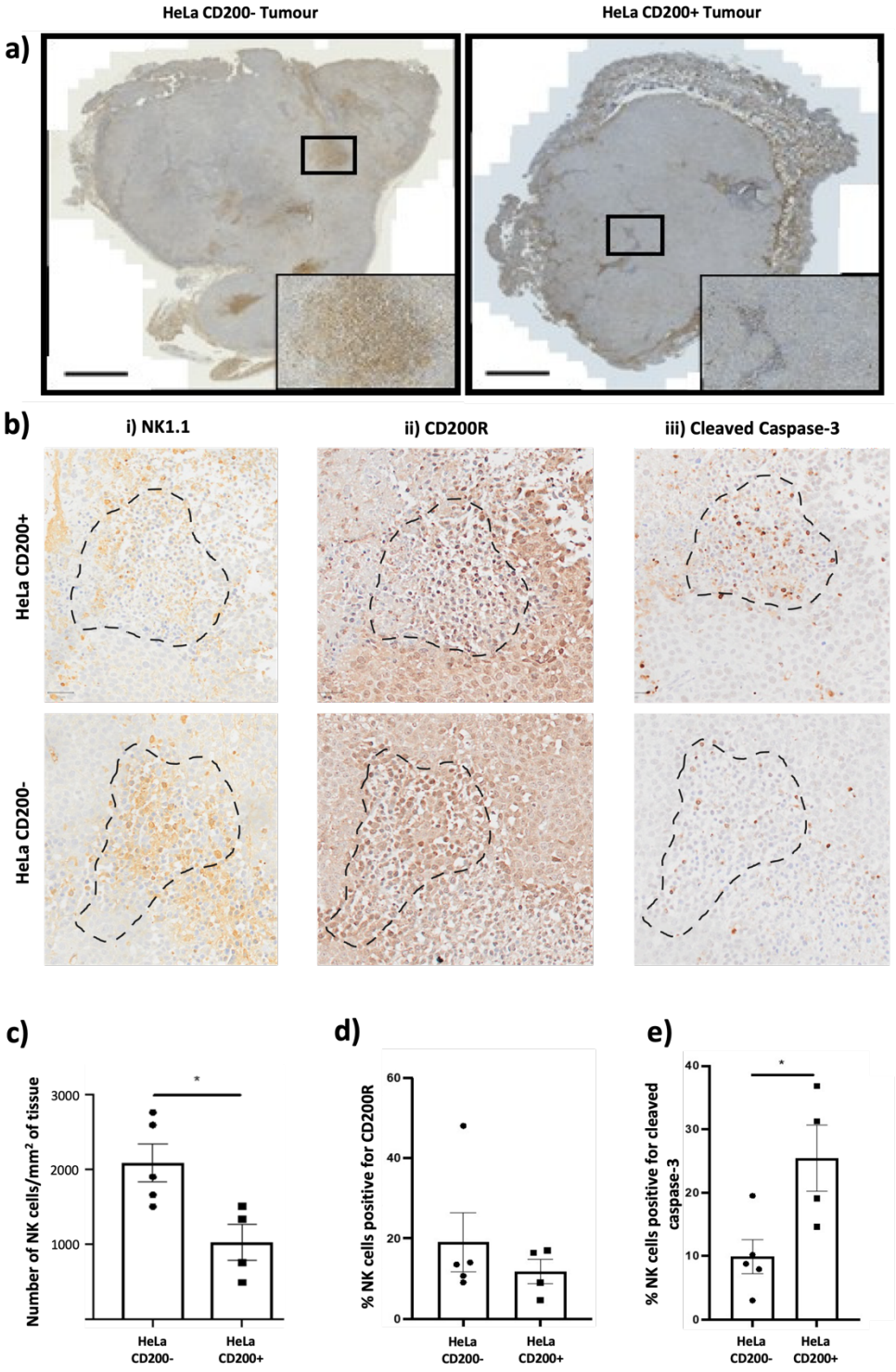

a)
HeLa CD200- Tumour
HeLa CD200+ Tumour
b)
i) NK1.1
ii) CD200R
iii) Cleaved Caspase-3
HeLa CD200+
HeLa CD200-
c)
Number of NK cells/mm² of tissue
3000
2000
1000
*
HeLa CD200-
HeLa CD200+
d)
% NK cells positive for CD200R
60
40
20
0
HeLa CD200-
HeLa CD200+
e)
% NK cells positive for cleaved caspase-3
40
30
20
10
0
*
HeLa CD200-
HeLa CD200+

Figure 3.3 | Immunohistochemical analysis of HeLa CD200 +/- xenograft tumours. *HeLa CD200+/- cells were transplanted into athymic nude mice and left to establish. Tumours were harvested 5 days later, and tumours sectioned.* ***a)*** *Immunohistochemical (IHC) staining for NK1.1 denotes NK cells. Zoomed panel shows enlarged area with NK1.1 staining.* ***b)*** *IHC staining for NK1.1* ***(i)****, CD200R* ***(ii)****, and cleaved caspase-3* ***(iii)****. Dotted line depicts the same area of NK cells (NK1.1-positive) across the three histological sections.* ***c)*** *Quantification of the NK1.1 staining and depicted as number of NK cells per* mm^2 *of tissue.* ***d and e)*** *Percentage of NK cells (NK1.1-positive) that express CD200R and cleaved caspase-, respectively. P < 0.05 (*). P < 0.001 (***). QuPath was used to analyse images. Graphs and statistical analyses were generated and performed in Graphpad. Error bars = mean +/- SEM; *P<0.05 by Mann Whitney test.*

3.4 CD200 Expression is Associated with a Reduction in Immune Cell Infiltration and an Increase in NK Cell Apoptosis in Transduced UW-BCC1 Xenografts

I next wanted to assess if CD200 could similarly invoke immune evasion in a BCC cell line, UW-BCC1. I first assessed the expression of CD200 in the cell line by flow cytometry which revealed a distinct population of CD200+ cells consisting of 3.9% +/- 0.75 SEM of the total population (n=8, Figure 3.4a, b). I next flow sorted the CD200+ and CD200- populations and assessed CD200 RNA quantity and performed colony forming assays. Both populations contained extremely low amounts of CD200 RNA with no significant difference between them (1.76 AU vs 1.58 AU, respectively; Figure 3.4c). I undertook three separate experiments, plating 100 and 500 cells per well in colony forming cell assays, to compare CD200+ to CD200- flow sorted cells. In all experiments, the colony forming efficiency of unsorted cells was higher than the sorted populations (Figure 3.5). There was significantly greater number of colonies in the unsorted population when 500 cells were plated in a colony forming cell assay (Figure 3.5c, $P < 0.01$). There were no significant differences in colony number between CD200+ and CD200- (26.33 +/- 2.53 vs 28.50 +/-1.97 and 77.57 +/- 5.52 vs 67.63 +/- 6.06, mean number of colonies +/- SEM, Figure 3.5a, c, respectively) or area (2147 +/- 229.10 vs 2256 +/- 182.30 and 1974 +/- 288.70 vs 2167 +/- 259.40, mean colony area in mm^2, Figure 3.5b, d, respectively). Hence there was no increase in colony forming efficiency in the CD200 population.

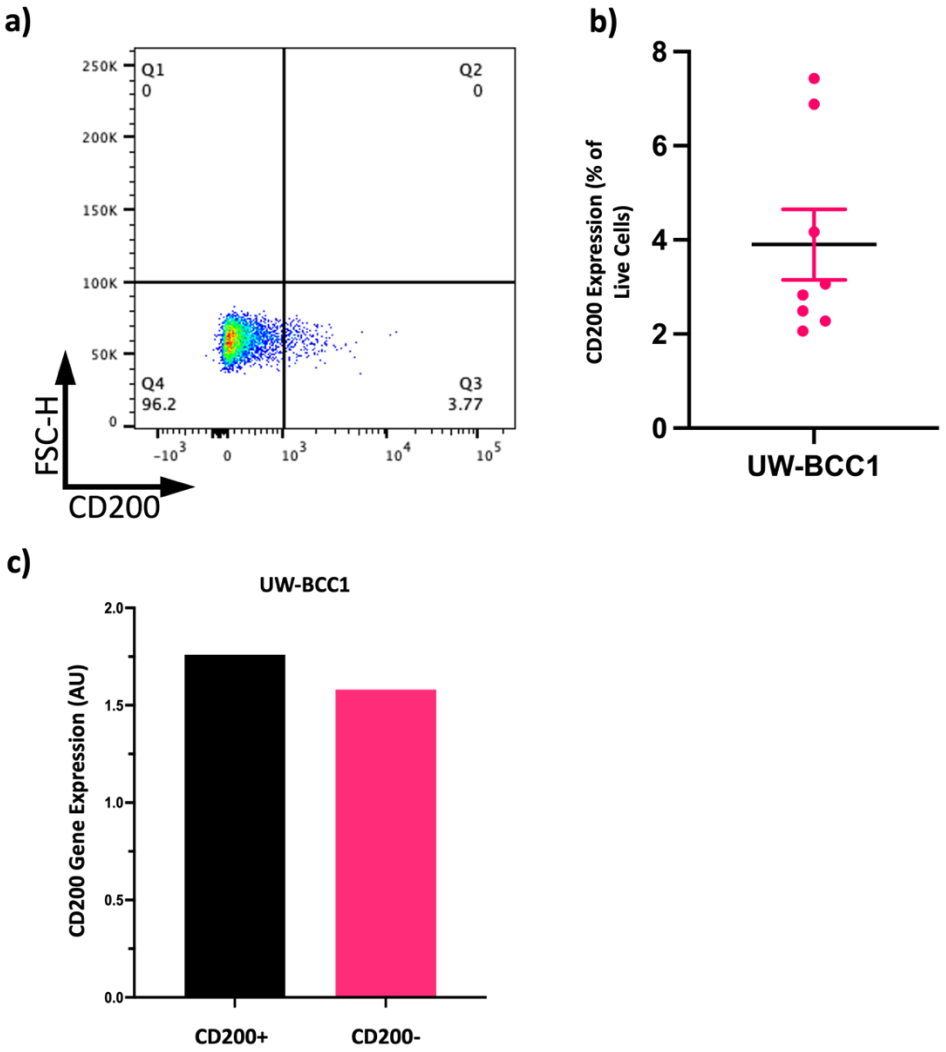


Figure 3.4 | In vitro analysis of CD200 expression in UW-BCC1 cell line. a) *UW-BCC1 was stained for CD200 antibody and expression was analysed on a flow cytometer. plot shows CD200 expression across live cells (DAPI-).* ***b)*** *graph shows CD200-positive cells as a percentage of live cells (DAPI-), n=8. Mean +/- SEM.* ***c)*** *UW-BCC1 cell line was sorted for CD200+ and CD200-. RNA was extracted from both populations and RT-PCR was performed with 3 technical replicates per population. Graph shows the*

CD200 arbitrary units (AU) calculated by the 2-delta delta Ct method. Error bar = mean +/- SEM.

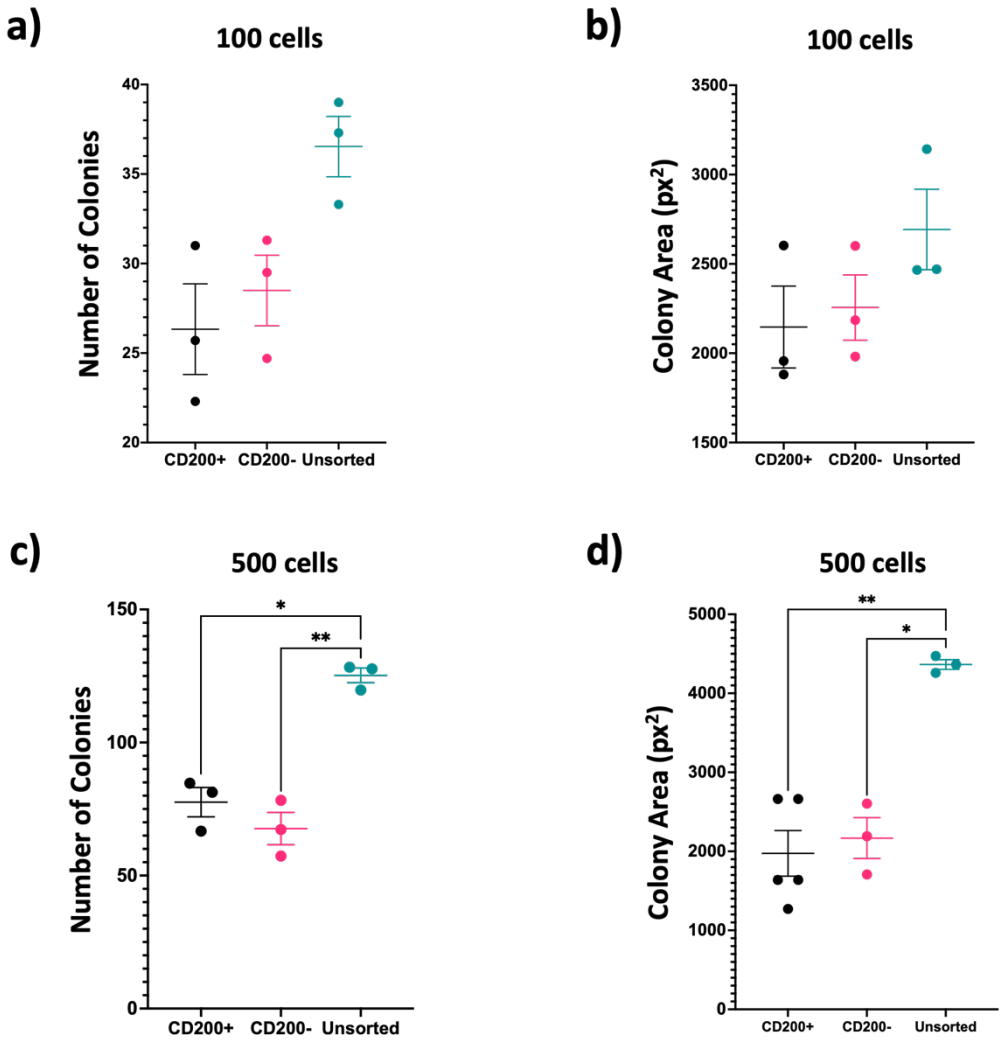


Figure 3.5 | Colony forming assays of CD200-sorted UW-BCC1 cells. *UW-BCC1 cells were sorted for CD200+ and CD200-. 100* ***(a, b)*** *and 500 cells* ***(c, d)*** *of each population, plus unsorted cells, were then plated for colony forming assay. Experiments were performed in triplicate biological replicates with triplicate technical replicate per*

*sample. Graphs and statistical analysis were performed in Graphpad. Error bars = mean +/- SEM; *P<0.05; **P<0.01 by Brown-Forsythe and Welch ANOVA test.*

To test the hypothesis that CD200 expression on BCC cells would lead to immune evasion, I next transduced the UW-BCC1 cells with a lentiviral vector containing CD200+-GFP or GFP alone. To transduce cells, a CD200 complimentary DNA (provided by the I.M.A.G.E Consortium, clone ID 5299899) was subcloned into a PINCO retroviral expression vector which co-expresses GFP from an internal CMV promoter (kindly gifted by Professor Alex Tonks). Cells were then transduced with the vectors (transduction was kindly performed by Dr. Giusy Tornillo). Following transduction, CD200 expression was validated by flow cytometry (Appendix 3.2) and xenotransplantation experiments were then performed. Experiments were performed the same as with the HeLa cells. 1 x 10^6 UW-BCC1 CD200+ and UW-BCC1 CD200- cells were injected in the left and right flanks (both n=5) of 6- to 8-week old athymic nude mice and left to establish. Tumours were first measured, which revealed no significant difference between the UW-BCC1 CD200+ and UW-BCC1 CD200- cell lines (19.46mm^3 +/- 5.76 vs 21.28mm^3 +/- 7.58, mean +/- SEM; Figure 3.6a). UW-BCC1 CD200+ tumours demonstrated ubiquitous CD200 protein expression when compared to the CD200- counterpart (Appendix 3.3).

Quantification of H&E-stained sections (Figure 3.6c) showed a small, but significant increase in the number of tumour cells (3,786.57 +/- 97.15 vs 3,258.54 +/- 143.12; $P < 0.05$; mean +/- SEM; Figure 3.6b). UW-BCC1 CD200+ tumours also demonstrated significantly lower number of immune cells when compared to CD200- tumours (18,631 +/- 1,745 vs 25,941.30 +/- 2,362.23; $P < 0.05$; mean +/- SEM; Figure 3.7a). A significantly lower ratio of necrotic to normal tumour areas was also seen in the CD200+ tumours (0.017 +/- 0.003 vs 0.038 +/- 0.006; mean +/- SEM; $P < 0.05$; Figure 3.7b). There were significantly lower numbers of NK cells in the CD200+ tumours

compared to the CD200- tumours (1,096.43 +/- 150.89 vs 2,333.04 +/- 368.69; $P < 0.05$; mean +/- SEM; Figure 3.7c). Using serial tissue sections, I was able to determine the percentage of NK cells that expressed CD200R and Cleaved Caspase 3. UW-BCC1 CD200+ tumours had significantly higher expression of cleaved caspase-3 (31.59% +/- 6.28 vs 14.08% +/- 4.11; $P < 0.05$; mean +/- SEM; Figure 3.7d) but no significant difference in CD200R expression (40.73% +/- 6.02 vs 35.82% +/- 6.50; mean +/- SEM; Figure 3.7e).

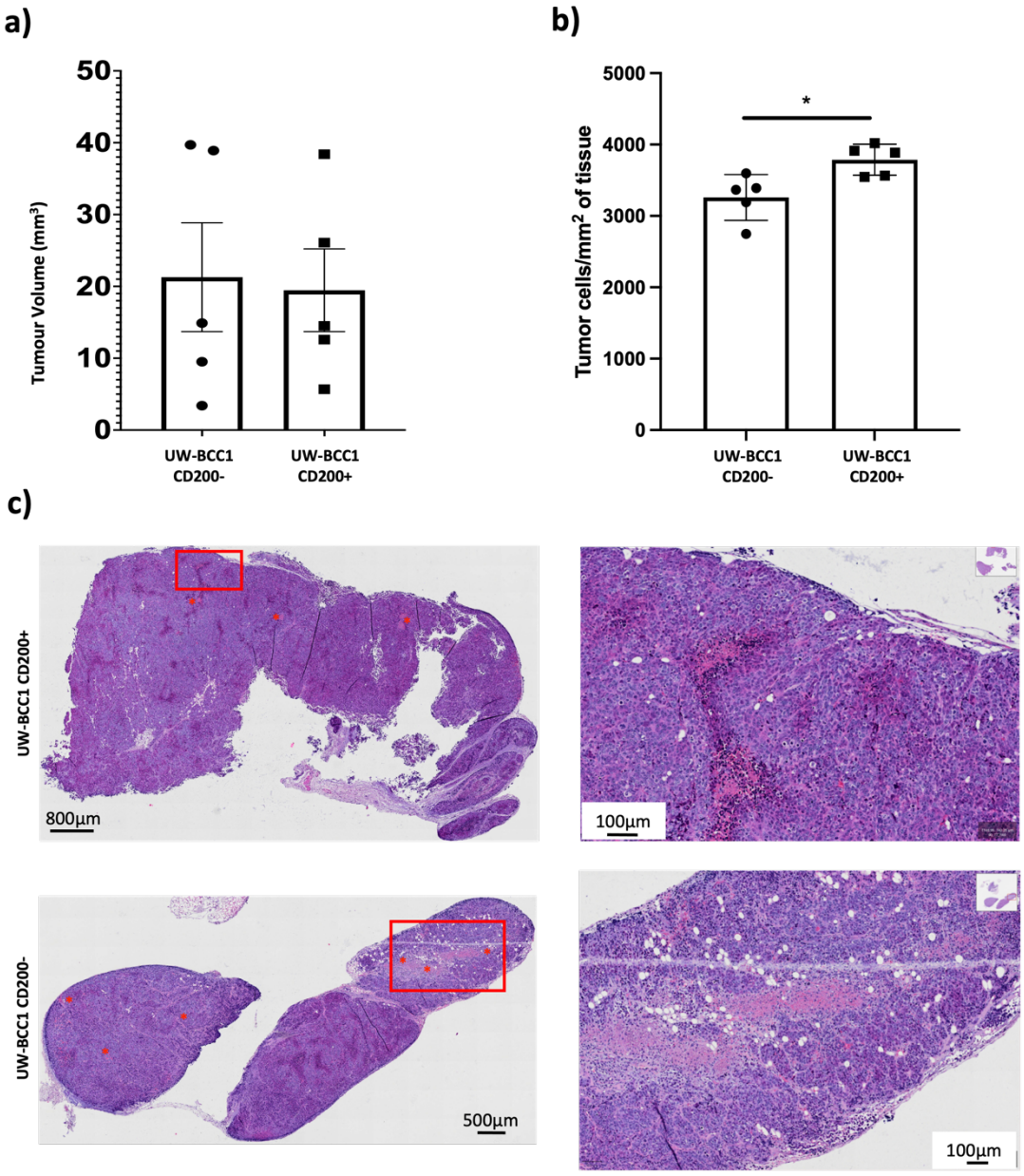


Figure 3.6 | Xenotransplantation of UW-BCC1 CD200+/- cells. *UW-BCC1 cells were transduced with CD200+-GFP or vector control-GFP and injected into the left and right flanks of athymic nude mice (both n=5) and left to establish for 5 days. Mice were then culled, and tumours harvested.* ***a)*** *Tumours were measured post-dissection and*

volumes calculated. Volumes are given in mm^3. Dissected tumours were fixed, paraffin embedded, and sectioned. ***c)*** *Haematoxylin & Eosin (H&E) staining of both tumours. Zoomed panel shows enlarged necrotic areas.* ***b)*** *Total tumour cells were counted and displayed as cells per mm^2 of tissue. *P<0.05 by Mann Whitney test. QuPath was used to analyse images. Graphs and statistical analyses were generated and performed in Graphpad. Error bars = mean +/- SEM.*

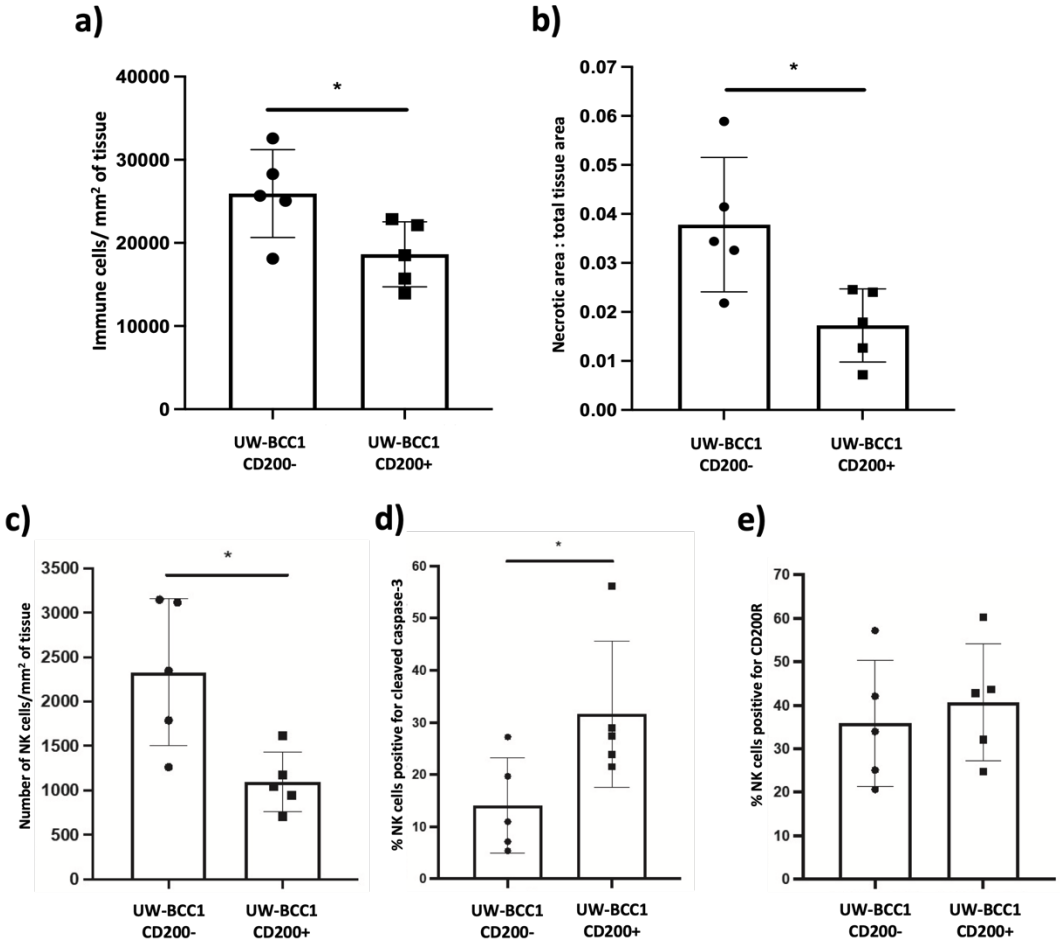


Figure 3.7 | Analysis of UW-BCC1 CD200 +/- xenograft tumours. *UW-BCC1 cells were transduced with CD200+-GFP or vector control-GFP and injected into the left and right flanks of athymic nude mice (both n=5) and left to establish for 5 days. Tumours were then harvested and sectioned for analyses.* ***a)*** *H&E section was used to count number of immune cells in the tumours. Total immune cells were counted and displayed as cells per mm^2 of tissue.* ***b)*** *Ratio of necrotic area to total tissue was calculated.* ***c)*** *Immunohistochemical (IHC) staining for NK1.1 was used to identify and quantify NK cells.* ***d and e)*** *Percentage of NK cells (NK1.1-positive) that express cleaved caspase-3*

*and CD200R, respectively. P < 0.05 (*). QuPath was used to analyse images. Graphs and statistical analyses were generated and performed in Graphpad. Error bars = mean +/- SEM; *P<0.05 by Mann Whitney test.*

3.5 Discussion

CD200 had been used to identify the presence of cancer stem cells within BCC through the enrichment of tumour-initiating cells (Colmont et al. 2013b). The utility of CD200 in BCC, beyond its ability to act as a cancer stem cell marker, had not been established. CD200 has been reported to modulate the immune response. In the context of hair follicle, a reduction in CD200 expression has been associated with the permanent destruction of the hair follicle leading to hair loss (Rosenblum et al. 2006). Herein, I sought to determine the functional significance of CD200 on BCC cancer stem cells by undertaking in vivo studies to support the in vitro data outlining the mechanisms for CD200-mediated immune evasion (Morgan et al. 2022).

First, I validated the use of a new CD200 antibody to successfully detect CD200-positive cells, which could then be used to isolate live cells by flow-sorting for functional study. Building on in vitro data generated in the lab, I was able to undertake a xenografting experiment to determine if CD200 expression in vivo led to NK cell apoptosis. Using the existing HeLa cell model of CD200 expression, I showed that cultured CD200+ and CD200- HeLa cells could propagate tumour growth in vivo after subcutaneous injection of 1 million cells. I validated CD200 expression within tumour sections. In CD200- HeLa cell tumours there were large areas of necrosis, and the number of tumour cells was greater in the CD200+ HeLa cell tumours. The was a significant reduction in the immune infiltrate surrounding CD200+ HeLa tumours, notably a reduction in NK cell numbers. I used serial-section labelling of NK cells to show NK cell apoptosis in the tumour microenvironment of CD200+ HeLa tumours.

These findings supported our hypothesis that CD200 was a novel NK cell-specific immune checkpoint.

Since BCC cells may be different from HeLa cells, I obtained a BCC cell line, UW-BCC1. These cells demonstrated equivocal CD200 labelling by antibody. I showed the flow cytometry characteristics of a CD200+ sub-population, did not express CD200 by RT-PCR and did not enhance colony-forming efficiency as one might expect from cancer stem cell enrichment. Therefore, I transduced UW-BCC1 cells with a vector containing GFP alone and another containing CD200 and GFP, which I validated by flow sorting and RT-PCR. Once more, CD200 expression did not affect tumour size. As with the HeLa cell experiments, CD200 expression on UW-BCC1 cells led to NK cell apoptosis.

Our recent work, which included the data from this chapter, showed that CD200 is functionally relevant to enable BCC immune evasion (Morgan et al. 2022). We have shown that CD200 ectodomain shedding, through matrix metalloproteinases MMP3 and MMP11, releases a biologically active soluble CD200 form into the tumour microenvironment. Thus, even though CD200 is only expressed by a small BCC population, through ectodomain shedding, it was sufficient to reduce the peritumoural NK cell numbers.

In conclusion, CD200 expression may be an important protein in mediating BCC cancer stem cell immune evasion. Even though CD200 is expressed by a small fraction of BCC cells, the tumour-initiating cell frequency is 1 in 822. However, CD200 expression in BCC may define an important subgrouping worthy of further study.

Chapter 4: Bulk RNA Sequencing of Human BCC

4.1 Introduction

CD200 is used to identify BCC cancer stem cells based on its expression on human HF bulge stem cells, from which BCCs are thought to arise. Moreover, CD200 appears to be essential for maintaining BCC tumour immune evasion. Even though CD200 is only expressed by a small tumour population, ectodomain shedding means that CD200 shields BCC from NK cell attack. Thus, CD200 defines an important functional subset within BCC which is enriched for BCC cancer stem cells.

Since we have showed that CD200 is functionally relevant for immune evasion and survival of BCC cells and simultaneously identifies a TIC population, I next performed bulk RNA sequencing on this population. The aim of this chapter was to determine novel BCC cancer stem cell pathways that promote survival. In this chapter, I performed bulk RNA sequencing on CD200+ and CD200- UW-BCC1 BCC cancer cell line populations. I also optimised a conventional BCC tumour tissue dissociation protocol and subsequently performed bulk RNA sequencing on human BCC CD200+ CD45- and CD200- CD45- populations.

4.2 Results

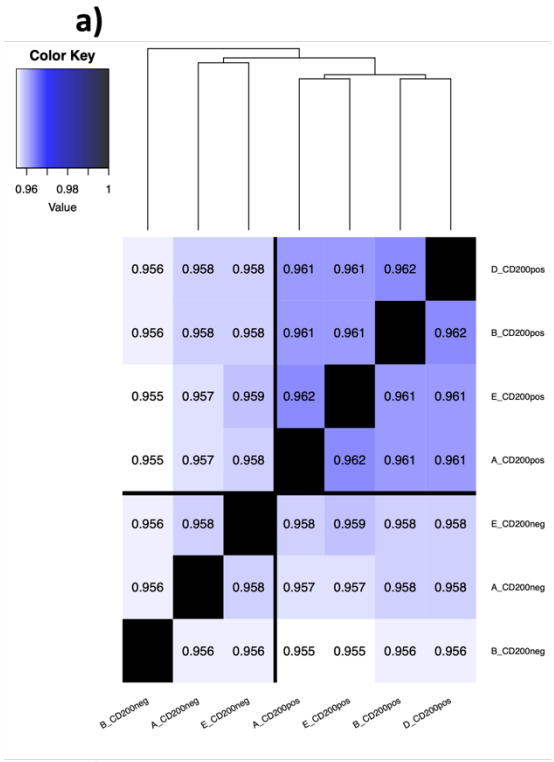
4.2.1 RNA Sequencing of Human BCC Cell Line, UW-BCC1

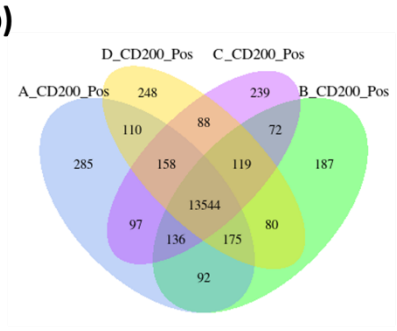
In the previous results chapter, I had identified a CD200+ population in the UW-BCC1 cell line. I next decided to undertake bulk RNA sequencing of flow sorted CD200+ and CD200- populations. I firstly isolated the CD200 population along with its negative counterpart using FACS and subsequently extracted RNA (n=4 per population). RNA quantity and quality of the samples was first determined using nanodrop™ and Bioanalyser™, respectively. All samples had an RNA concentration > 40ng/µL and a RIN quality score > 9. Samples were then sent to Novogene (Cambridge, UK) for library preparation and subsequent sequencing of the samples. Samples underwent and passed an internal quality control (QC) prior to mRNA, poly A enrichment and library preparation. One sample failed to produce a library and was subsequently removed from the sequencing run. The remaining libraries (4 x UW-BCC1 CD200+, 3 x UW-BCC1 CD200-) were then sequenced on a NovaSeq 6000 S1 flow cell (PE150) to produce 6Gb of raw sequencing data per sample. Adapter sequences were removed, and low-quality reads were removed from the raw reads using the following parameters: 1) reads containing N > 10% (where N is an undetermined base), and 2) reads where the quality value of >50% of the bases in that read is ≤ 5 (base quality is provided by the sequencing machine). High quality reads were then mapped to the human reference genome (GRCh38) using Hisat2 (v2.0.5) (Kim et al. 2019). Normalised gene expression counts (FPKM) were then generated by featureCounts (v1.5.0-p3) (Liao et al. 2014). On average, there were 52 million 150 base pair (bp) paired-end total reads per sample, 95.99% of which were uniquely mapped to the

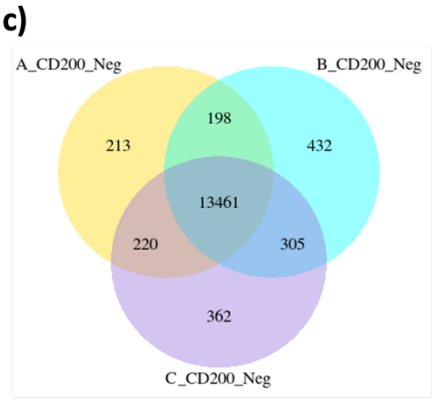
human genome and 93.42% that mapped to a unique position of the human genome. There was an average of 92.87% Q30 sequence quality and 49.71% GC content. An average of 89.41% reads mapped to exons, 8.49% mapped to introns, and 2.07% mapped to intergenic regions. There was a total of 58,735 mapped features, 25,823 (43.97%) of which contained a null count across all samples. RNAseq sample information can be found in Appendix 4.1. Data was then run through a DESeq2 pipeline (Love et al. 2014) to normalise the read counts, detect and remove any outliers, and produce CD200+ vs CD200- DEGs.

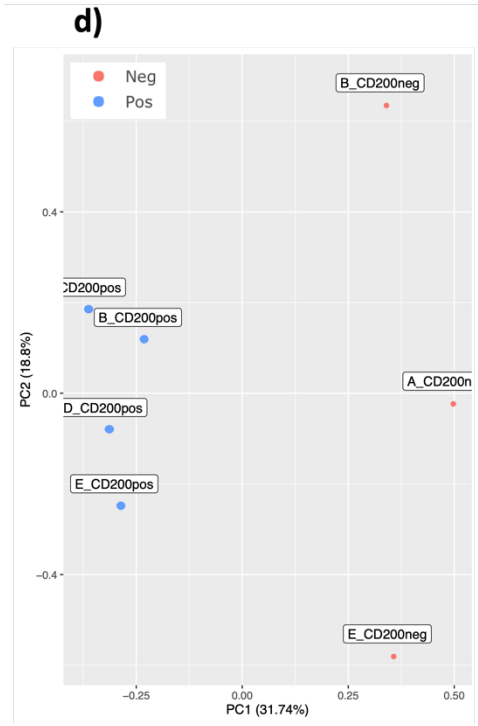
I next checked for any correlations between the sorted populations. Samples were firstly checked for normal (Gaussian) distribution using a Shapiro-Wilk's normality test (using $\alpha = 0.05$). All samples were below the α level (< 0.05) and so the null hypothesis was rejected, indicating that samples are not normally distributed. A nonparametric two-tailed Spearman's correlation test was therefore used to test for both positive and negative correlations between samples. A correlation matrix was generated and used as input for the gplots heatmap.2 function in R, which generated a heatmap of the R^2 values and a dendrogram (Figure 4.1a). As expected, samples were highly correlated with a Spearman's $R^2 > 0.95$, though intriguingly, a marginally higher correlation was seen amongst the CD200+ sorted populations compared to their negative counterparts (average $R^2 = 0.961$ and $R^2 = 0.956$, respectively) which was also demonstrated through secondary dendrogram clade separation. Venn diagrams were also generated between the CD200+ samples and CD200- samples using normalised reads, in which there was a 13,544 and 13,461 gene overlap, respectively (Figure 4.1b and c, respectively). I next wanted to observe any differences between the sample groups. To do this, a PCA was performed on the

normalised data, which generated 6 principal components. PC1 and PC2 were selected as the input for biplot generation, which represented > 50% of the dataset's variance (50.54%). Interestingly, although all samples are highly correlated (as seen in the correlation analysis), the CD200+ population was clearly separated from the CD200- population along the PC1 axis (Figure 4.1, and e, respectively) and samples were clustered together, albeit slightly spread along the PC2 axis. We also see that the CD200- samples were less clustered and were widely dispersed along the PC2 axis, suggesting that the gene expression amongst the CD200- population is more heterogeneous. A heatmap was generated on the 2,364 significant DEGs (normalised data, p.adj < 0.05) and hierarchical clustering (1 – Pearson's correlation) was performed on the rows (genes) and columns (samples). The heatmap displayed a clear separation in gene expression profiles between the CD200+ and CD200- samples and similar gene expression within the groups, which is confirmed by the hierarchical clustering (Figure 4.2).


a)
Color Key
0.96 0.98 1
Value
0.956 0.958 0.958 0.961 0.961 0.962
0.956 0.958 0.958 0.961 0.961 0.962
0.955 0.957 0.959 0.962 0.961 0.961
0.955 0.957 0.958 0.962 0.961 0.961
0.956 0.958 0.958 0.959 0.958 0.958
0.956 0.958 0.957 0.957 0.958 0.958
0.956 0.956 0.955 0.955 0.956 0.956
D_CD200pos
B_CD200pos
E_CD200pos
A_CD200pos
E_CD200neg
A_CD200neg
B_CD200neg
B_CD200neg
A_CD200neg
E_CD200neg
A_CD200pos
E_CD200pos
B_CD200pos
D_CD200pos


b)
D_CD200_Pos
C_CD200_Pos
A_CD200_Pos
B_CD200_Pos
248
239
110
88
72
285
158
119
187
13544
97
80
136
175
92


c)
A_CD200_Neg
B_CD200_Neg
198
213
432
13461
220
305
362
C_CD200_Neg


d)
Neg
Pos
B_CD200neg
B_CD200pos
A_CD200n
E_CD200pos
E_CD200neg
0.4
0.0
−0.4
PC2 (18.8%)
−0.25
0.00
0.25
0.50
PC1 (31.74%)

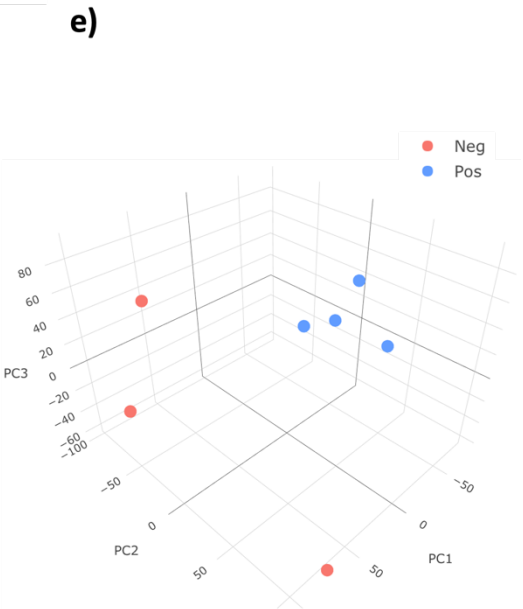

e)
Neg
Pos
PC3
80
60
40
20
0
−20
−40
−60
−100
PC2
−50
0
50
PC1
−50
0
50

Figure 4.1 | Sample correlation analyses following bulk RNA sequencing. *UW-BCC1 cell line was sorted into CD200+ and CD200- populations (n=4 and n=3, respectively) and subsequently bulk RNA sequenced.* ***a)*** *A nonparametric two-tailed Spearman's correlation test was performed using the normalised counts and heatmap was generated using the gplots package in R. The colour key represents the* R^2 *values. These values are also in each box. Values closer to 1 depict higher similarities, whereas values closer to 0 depict no similarities.* ***b, c)*** *VENN analysis was performed on the CD200+ and CD200- normalised reads, respectively, which shows the number of genes shared or unique between the different samples.* ***d, e)*** *2D and 3D principal component analysis (PCA) was performed on the normalised reads. PC1 and PC2 axes were used to plot* ***d)*** *and PC1, PC2, and PC3 axes were used to plot* ***e)****. Red points, CD200- samples. Blue points, CD200+ samples. Percentages in brackets on the axes labels in* ***d)*** *depict the percentage of variation explained by that PC. All analyses and plots were generated in R.*

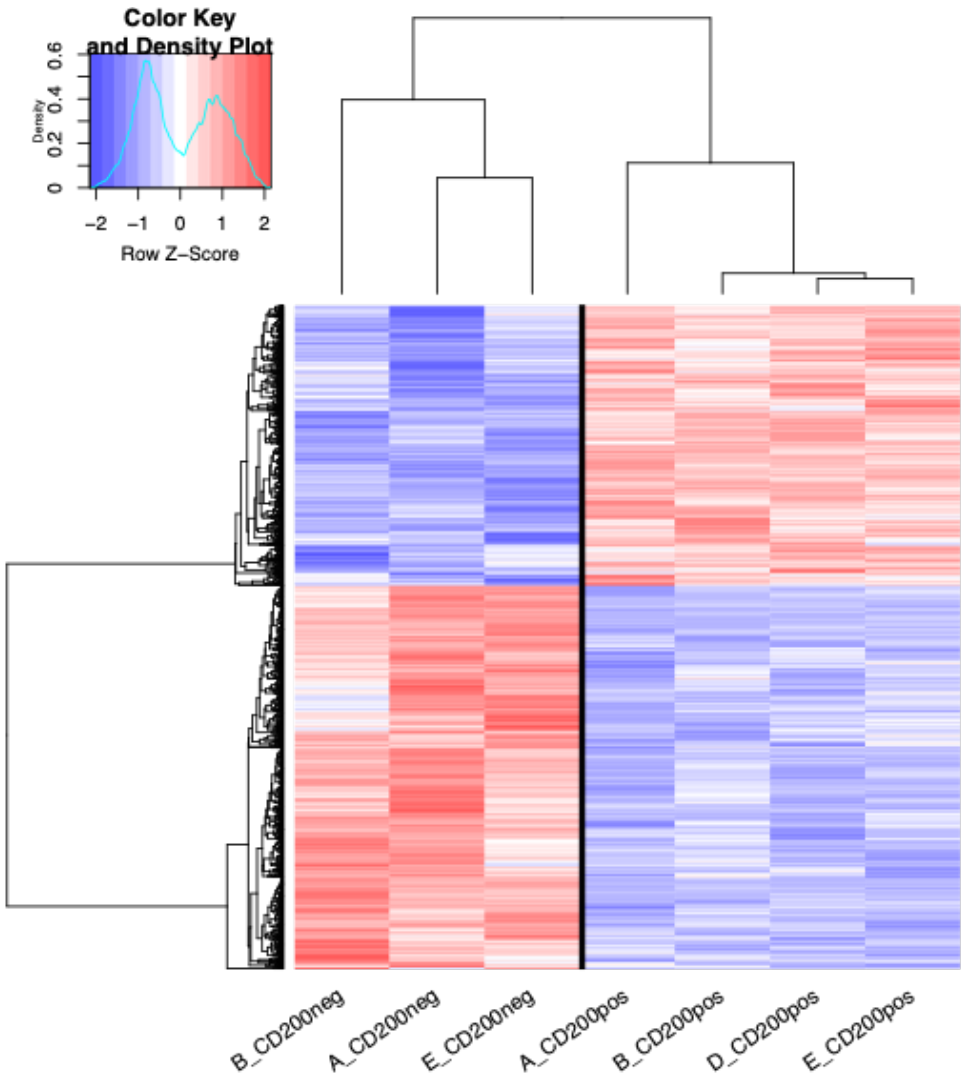


Figure 4.2 | Heatmap of UW-BCC1 CD200+ and CD200- samples. *UW-BCC1 cell line was sorted into CD200+ and CD200- populations (n=4 and n=3, respectively) and subsequently bulk RNA sequenced. Heatmap was generated using the sample normalised reads and the gplots package in R. Hierarchical clustering was also performed on the samples (columns) and genes (rows). Primary segregation was seen between the two sample groups, which have been separated by a black line. Colour*

key shows the colours associated with the row Z-score. The row Z-score is an indicator of the expression value, which has been scaled across the samples to allow for comparison.

Of the 2,364 significant DEGs (adjusted p value ≤ 0.05), 998 were up (Log2FC > 0, range 0.11 - 2.58) and 1366 were down (Log2FC < 0, range -0.1 - -4.98; Figure 4.3). CD200 was not a DEG. On further analysis of the data there were no read counts for CD200 in the negative or more importantly the CD200+ population. I validated the absence of CD200 in the sample sets by RT-PCR (Figure 3.4c). In summary, UWBCC1 population did not demonstrate a flow cytometric discrete CD200+ population, did not express CD200 transcripts, and showed no increase in colony forming efficiency. Thus, UWBCC1 was a poor surrogate for the study of tumour initiating cells and therefore we had to rely on the study of primary human tumour tissues.

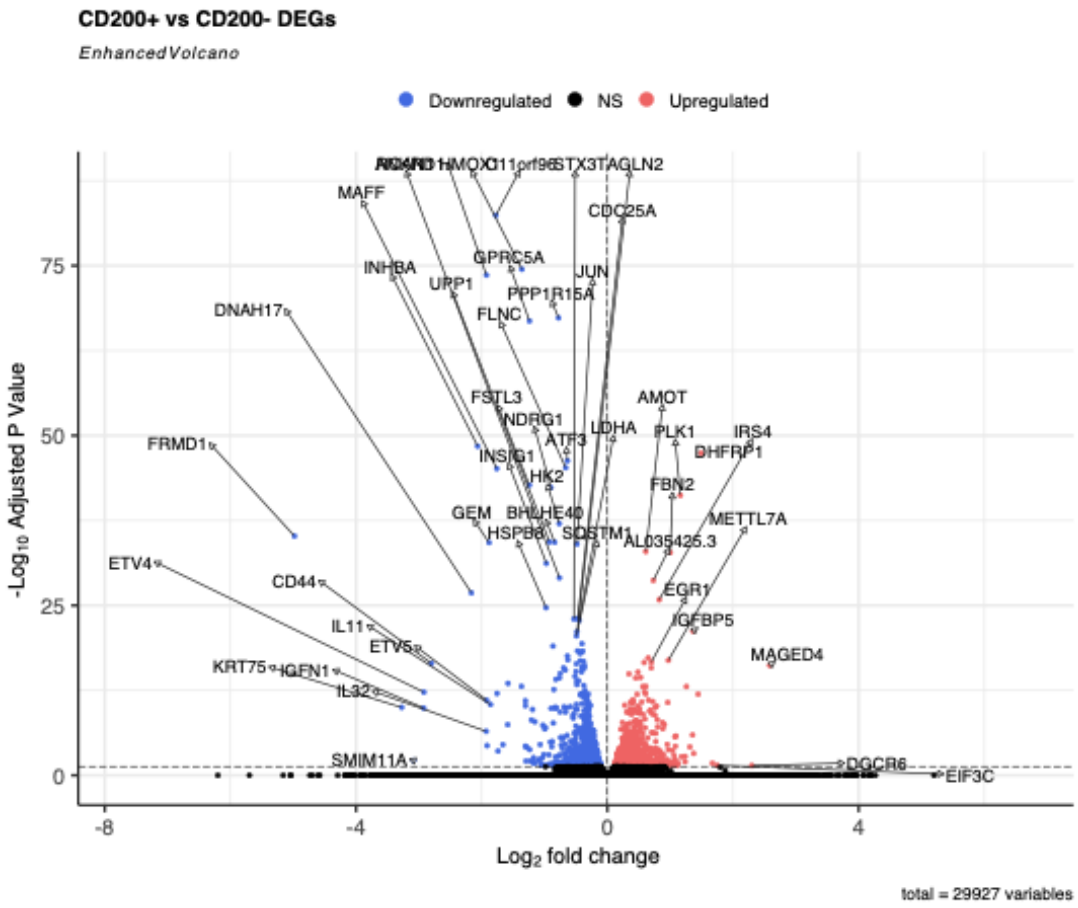


Figure 4.3 | Volcano plot of UW-BCC1 CD200+ vs CD200- differentially expressed genes. *UW-BCC1 cell line was sorted for CD200+ and CD200- populations and subsequently RNA sequenced. Differential gene expression was then performed on both populations and a volcano plot was generated using the EnhancedVolcano package in R. Dotted lines denote significance boundaries.*

4.2.2 Bulk RNA Sequencing of Primary Human BCC Cells

4.2.2.1 Optimisation of Human BCC Dissociation Protocol

While in vitro isolation of HeLa and UW-BCC1 cells requires simple trypsinisation, the isolation of human BCC cells requires both mechanical and enzymatic digestion. The original protocol for human BCC dissociation was over a decade old, so I initially compared it to the MACS tumour dissociation kit. First, I determined the impact of enzymatic degradation by comparing a 3-year-old batch of enzymes with a freshly purchased enzyme set. As expected, the new enzymes led to more efficient isolation of the human BCC cells (4.0×10^5 cells vs 1.9×10^5 cells, Figure 4.4b, a, respectively) and simultaneously preserved the CD200 epitope (2.4% vs 1.18%, Figure 4.4d, c, respectively).

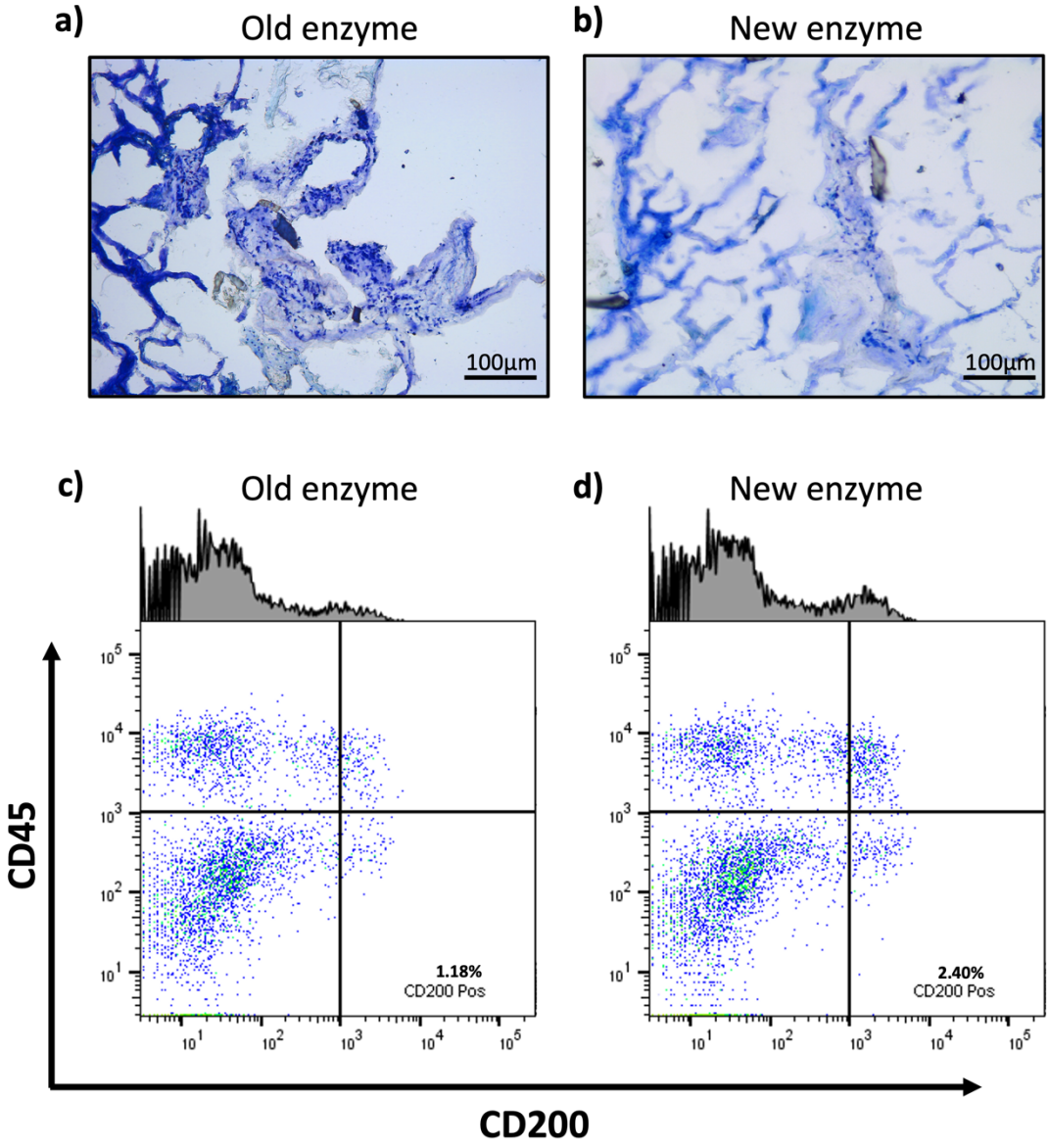


Figure 4.4 | Comparison of enzyme batches on the conventional human BCC dissociation protocol. *Two enzyme batches were compared during a dissociation protocol.* ***a, c)*** *show the data for the 3-year-old enzyme batch, hereby referred to as 'old'.* ***b, d)*** *show the data for the newly purchased enzyme batch. Using our labs*

previously published protocol, human BCC tumours were mechanically dissociated and then enzymatically digested to isolate the BCC cells. Following cell isolation, the undigested tissue was collected, cryosectioned, and stained with Toluidine Blue. ***a)*** *Undigested tissue visualised under a light microscope for the old enzyme batch.* ***b)*** *Undigested tissue visualised under a light microscope for the new enzyme batch. Isolated BCC cells were then stained for CD45, CD200, and DAPI and analysed on a flow cytometer.* ***c)*** *Flow cytometry plot for the old enzyme batch depicting percentage of live cells (DAPI-) that are CD45- (Y-axis) and CD200+ (X-axis).* ***d)*** *Flow cytometry plot for the new enzyme batch depicting live cells (DAPI-) that are CD45- (Y-axis) and CD200+ (X-axis). CD200 expression is summarised as a histogram above each plot.*

Having optimised the conventional approach, I next compared dissociation against the MACS protocol. In every experiment, the MACS protocol consistently generated greater cell numbers (2.4×10^6 +/- 1×10^6 vs 2.5×10^5 +/- 8.6×10^4, mean +/- SEM, respectively; Figure 4.5a; n=6). In two experiments where both protocols were directly compared, the MACS protocol liberated >5-fold more cells (1.02×10^6 +/- 6.75×10^5 vs 1.02×10^5 +/- 8.75×10^4, mean +/- SEM, respectively; Figure 4.5b). However, residual cells in the tumour tissue were evident, with clear epidermal cell layer structures also visible (Figure 4.5c). These results suggest that the MACS dissociation protocol appeared to be an improvement to the existing protocol when assessing the number of cells liberated from the tissue.

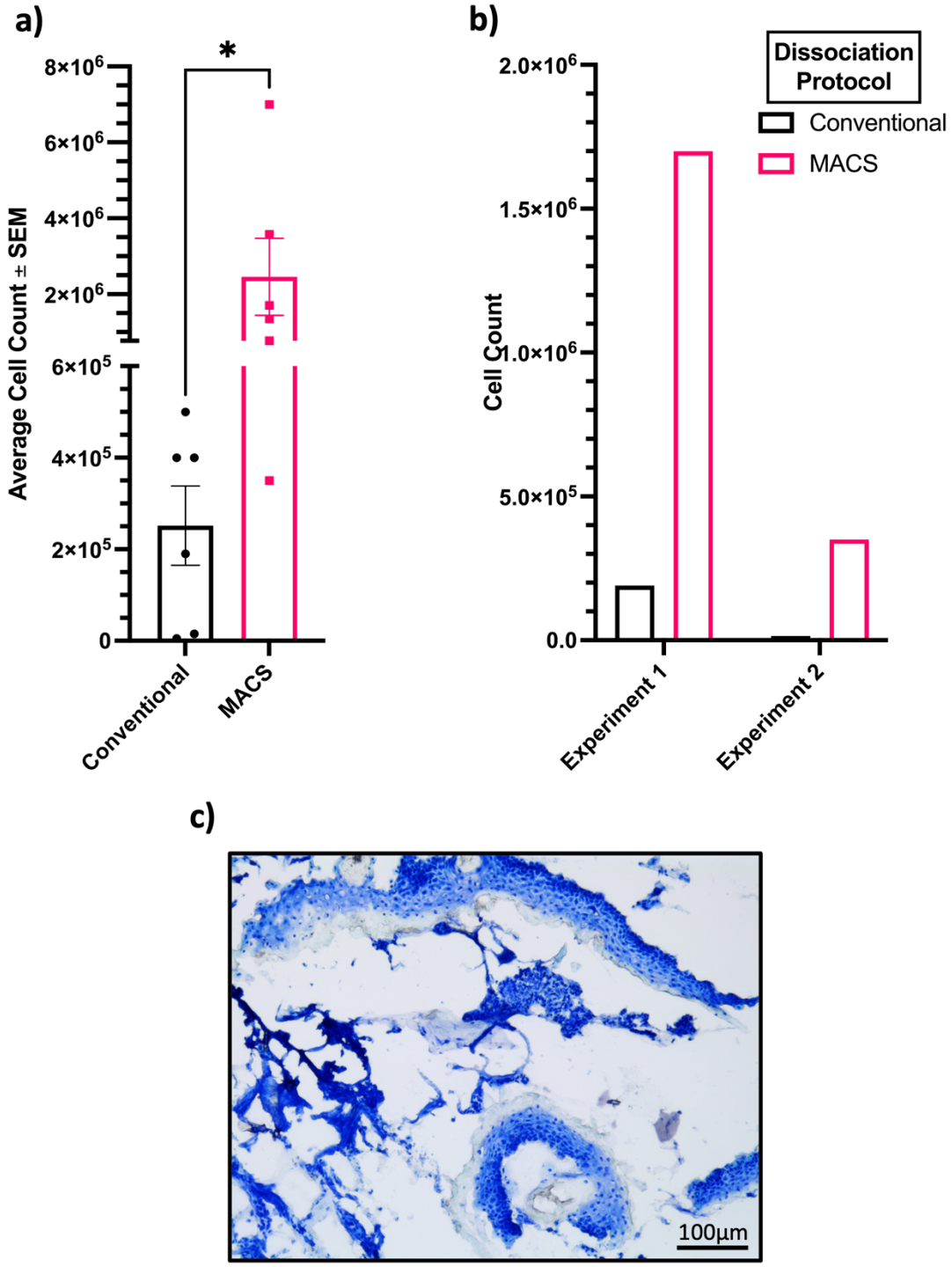

a)
*
8×10^6
6×10^6
4×10^6
2×10^6
6×10^5
4×10^5
2×10^5
0
Average Cell Count ± SEM
Conventional
MACS
b)
Dissociation Protocol
Conventional
MACS
2.0×10^6
1.5×10^6
1.0×10^6
5.0×10^5
0.0
Cell Count
Experiment 1
Experiment 2
c)
100µm

Figure 4.5| Comparison of conventional and MACS human tumour dissociation protocols in basal cell carcinoma. a) *Graph shows the average number of cells obtained from conventional and MACS tumour dissociation protocols over 6 individual experiments with varying amounts of tumour tissue. Bars represent mean +/- SEM, *P<0.05 by Mann Whitney test.* ***b)*** *Graph represents two individual experiments where equal amount of tumour tissue was used for each dissociation protocol. Experiment 1, equally divided tumour tissue; Experiment 2, 0.18g tumour tissue for each protocol. Undigested tumour tissue was taken from after the MACS protocol and was stained using Toluidine Blue.* ***c)*** *Figure shows many cells residing in undigested BCC tumour tissue following dissociation with the MACS protocol.*

Following tumour dissociation, single-cell suspensions were antibody labelled for flow cytometric analyses. Despite a greater yield of cells, there appeared to be less CD200 expression with the MACS protocol when compared to the conventional protocol (Figure 4.6a vs Figure 4.6b, CD200+ percentages 0.86% vs 2.4%, respectively). The conventional protocol (Figure 4.6b) demonstrates a distinct CD200 cancer stem cell peak, which is absent in dissociated BCC using the MACS protocol. This could also be a result of inadequate dissociation of the full epidermal layer and subsequent release of the basal cell layer. Hence, the MACS protocol was deemed not suitable for skin cancer tissue dissociation as it inadequately dissociates the tumour tissue and degrades the cell surface antigen used to select cells. My findings suggest that the original protocol for BCC dissociation remains the best for isolating all the BCC keratinocytes and maintaining CD200 expression.

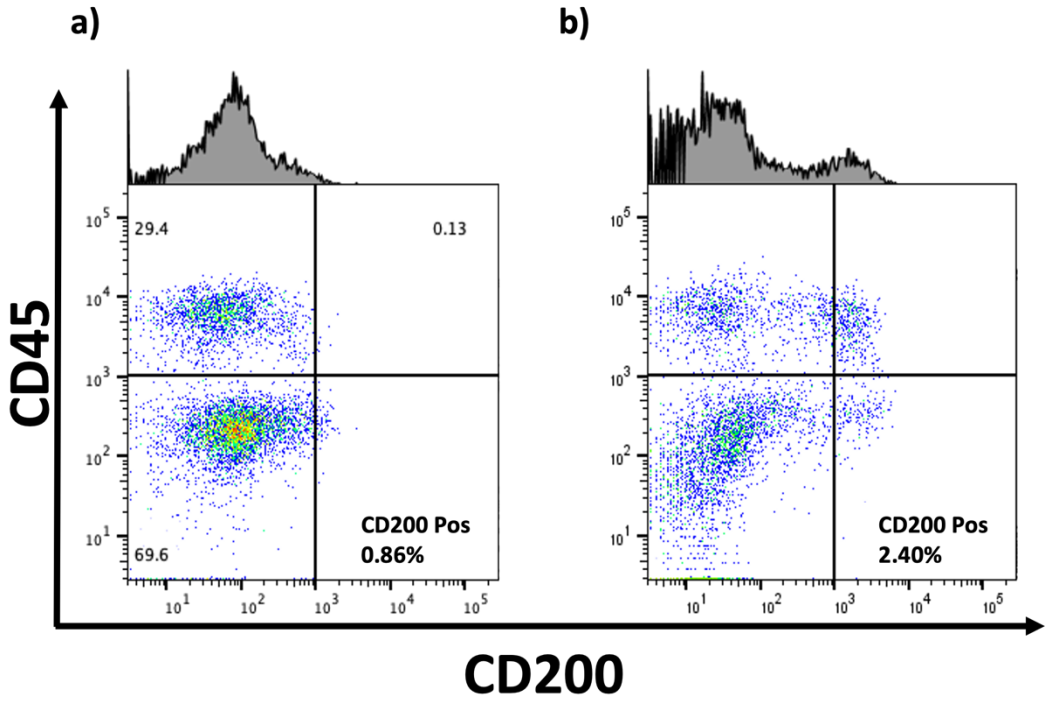


Figure 4.6 | Assessment of CD200 expression in BCC dissociated by two different protocols. *BCC tissue was dissociated and labelled with the cell surface markers CD200 and CD45. The flow cytometry plot was determined using unlabelled and single stain controls. CD45 marker is used to eliminate tumour infiltrating leukocytes. CD200+ CD45- represents the BCC cancer stem cell population* (Morgan et al. 2019)*. Histograms shows the CD200 fluorescence intensity.* ***a)*** *BCC dissociation and flow cytometry using the MACS protocol.* ***b)*** *BCC dissociated with a conventional dissociation protocol.*

Next, I undertook flow sorting for CD200+/- CD45- BCC subpopulations using the conventional protocol. As previously described, CD200+ CD45- BCC cells were identified in all tumour samples at a frequency of 1.35% - 5.2% of the total tumour (Figure 4.7; mean 2.93% +/- 0.32 SEM, n = 11; Colmont et al. 2013). Flow sorted BCC cells demonstrated strong CD200 expression by immunofluorescence on the membrane of the cells, but expression was absent in the negative-sorted cells (Figure 4.8). The BCC dissociation and CD200+ CD45- population isolation was therefore optimised. Sorted samples were then sent to Wales Gene Park for RNA sequencing.

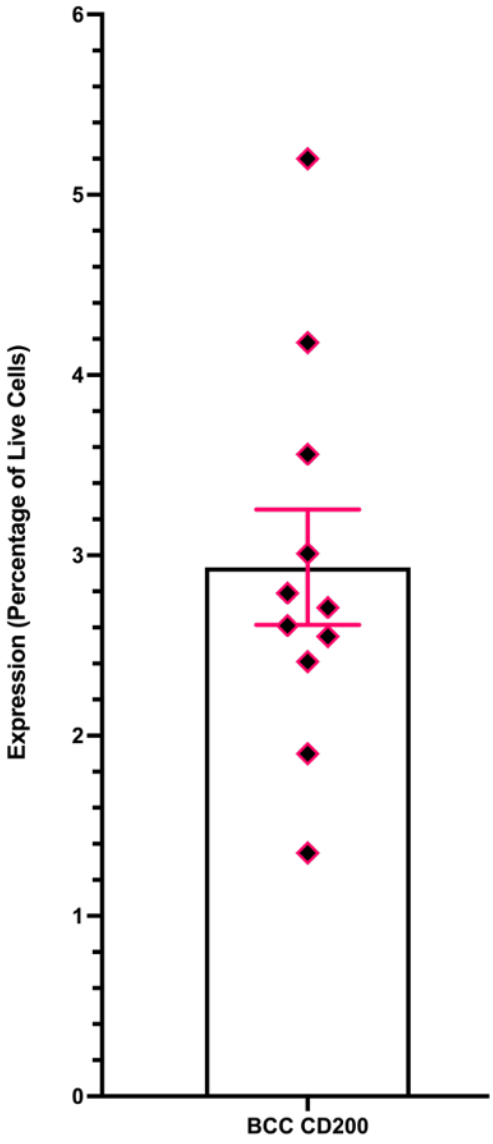


Figure 4.7 | CD200 expression in human BCCs. *11 BCC tumours were dissociated into single-cell suspensions and stained for CD45 and CD200. Live/dead stain, DAPI, was added and then CD200+/- CD45- populations were then isolated using fluorescence-activated cell sorting (FACS). Graph shows CD200 expression levels in each BCC. Each*

point represents one BCC tumour that was processed and analysed. Error bar = mean +/- SEM.

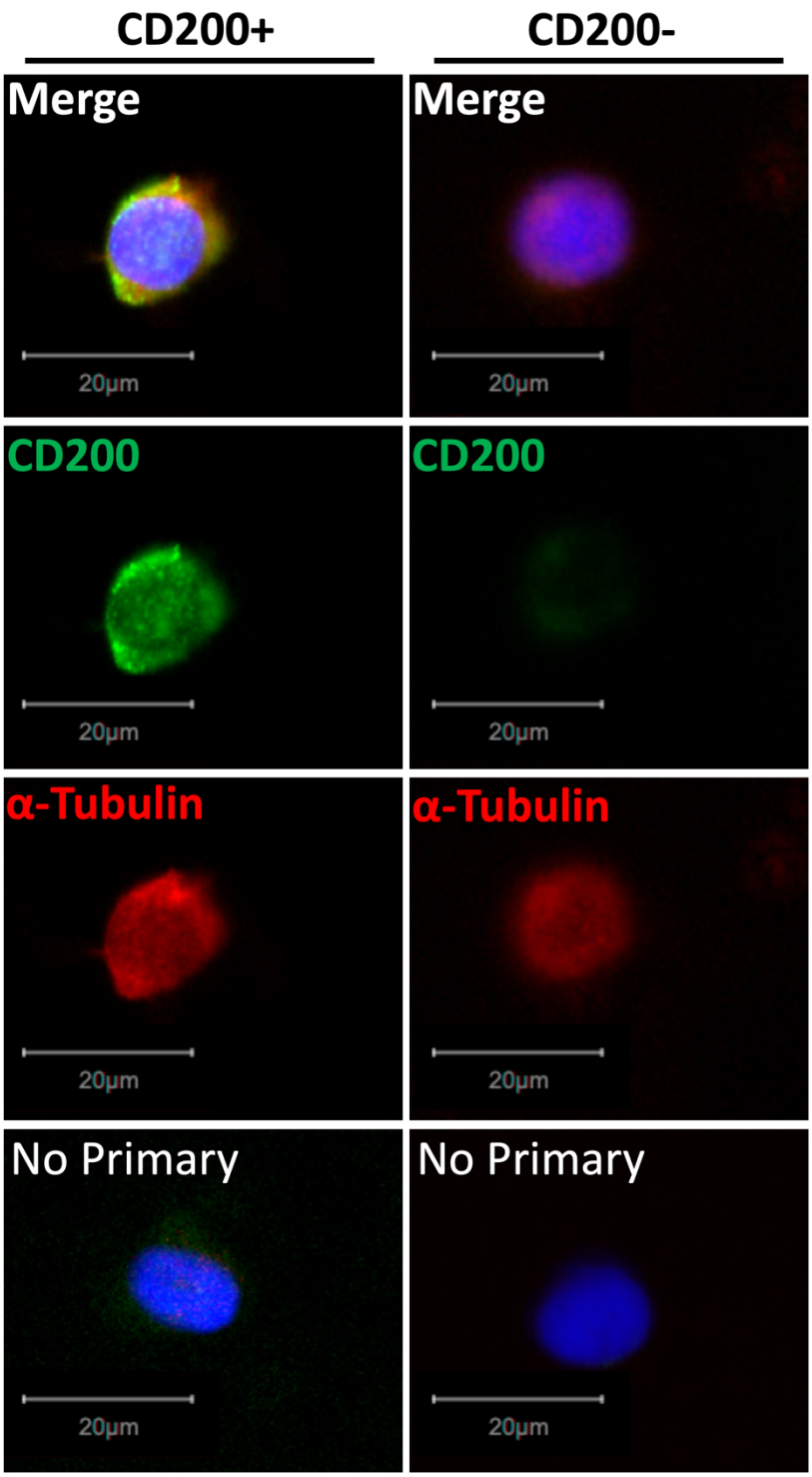

CD200+
CD200-
Merge
Merge
20μm
20μm
CD200
CD200
20μm
20μm
α-Tubulin
α-Tubulin
20μm
20μm
No Primary
No Primary
20μm
20μm

Figure 4.8 | Immunofluorescence showing expression of BCC cancer stem cell marker, CD200, on CD200+/- CD45- BCC cells. *Human BCC tumours were mechanically and enzymatically dissociated to produce single cell suspensions. Cells were then fluorescently stained with CD45, CD200, and DAPI, and CD200+/- CD45- populations were isolated using FACS. Some cells were sorted onto microscope slides and fixed in 4% formalin. Fixed cells were then stained for α-Tubulin (red), CD200 (green), and DAPI (blue).*

4.2.2.2 Bulk RNA Sequencing of Primary Human BCC Cells

A total of 11 patients with BCC were recruited. Samples were successfully dissociated and CD200+ CD45- and CD200- CD45- populations isolated. RNA was isolated and quality checked. A total of 7 CD200+/- CD45- populations (total of 14 samples) passed the RNA integrity (RIN) assessment and were therefore selected for library preparation and bulk-RNA sequencing.

Samples were then sent to the WGP, where library preparation and sequencing were performed. One sample failed to produce a sequencing library and was therefore removed from the experiment. The remaining 13 libraries were then normalised, and an additional quality control check was performed. Libraries were then sequenced on a NovaSeq 6000 S1 flow cell (PE100) to produce 10Gb of raw sequencing data per sample. Adapter sequences were removed, and low-quality reads were removed from the raw reads using Fastp with the following parameters: 1) reads containing N > 10% (where N is an undetermined base), and 2) reads where the quality value of >50% of the bases in that read is ≤ 5. High quality reads were then mapped to the human reference genome (GRCh38) using STAR (Dobin et al. 2013). Expression counts were then generated from the mapped reads using the featurecounts read quantifier against the GENCODE v27 gene model. The mapped reads data table was then explored to check for mapping features and counts. On average, there were 55.5 million, 100 base pair (bp) paired-end reads obtained from each sample with a total of 58,288 mapped features. 16,287 (27.94%) of those features contained a null count across all samples. On average, 92.43% of the fragments were uniquely mapped across the 13 samples with 62% of reads mapping to genomic features and

48.33% GC content. The data was then run through SARTools (a wrapper to DESeq2) to generate normalized reads, detect and remove outliers, and produce differentially expressed genes (DEGs; CD200+ CD45- vs CD200- CD45-). The resulting dataset consisted of 5 CD200+ CD45- and 5 CD200- CD45- samples with an average of 25,692 expressed genes per sample (sequencing data summary can be found in Appendix 4.2). Upon receipt of the data from WGP, experiment validation was performed via assessment of CD200 expression and differential expression in all samples. CD200 expression was observed in all the CD200+ CD45- samples with residual expression (10-fold lower) in the CD200- CD45- samples (CD200/ENSG00000091972; 3511 mean reads +/- 2890 SD vs. 257 mean reads +/- 208 SD, respectively) and demonstrated a significant 3.661 Log2fold- change (adjusted P value = 3.67E-16) (Figure 4.9a and b, respectively). Next, I wanted to validate that the sequencing data was from BCC basal keratinocytes. To do this, I assessed the expression of known keratinocyte markers and BCC Hedgehog signalling pathway markers. Read counts were present across all samples for keratinocyte markers KRT5, KRT17, and COL17A1 (Figure 4.10a-c, respectively). Expression of BCC Hedgehog signalling members was also present across all samples (Figure 4.10d-h). Interestingly, there appeared to be an increase in Hedgehog signalling expression amongst the CD200+ CD45- samples, indicating a further enrichment of BCC cells via CD200+ expression.

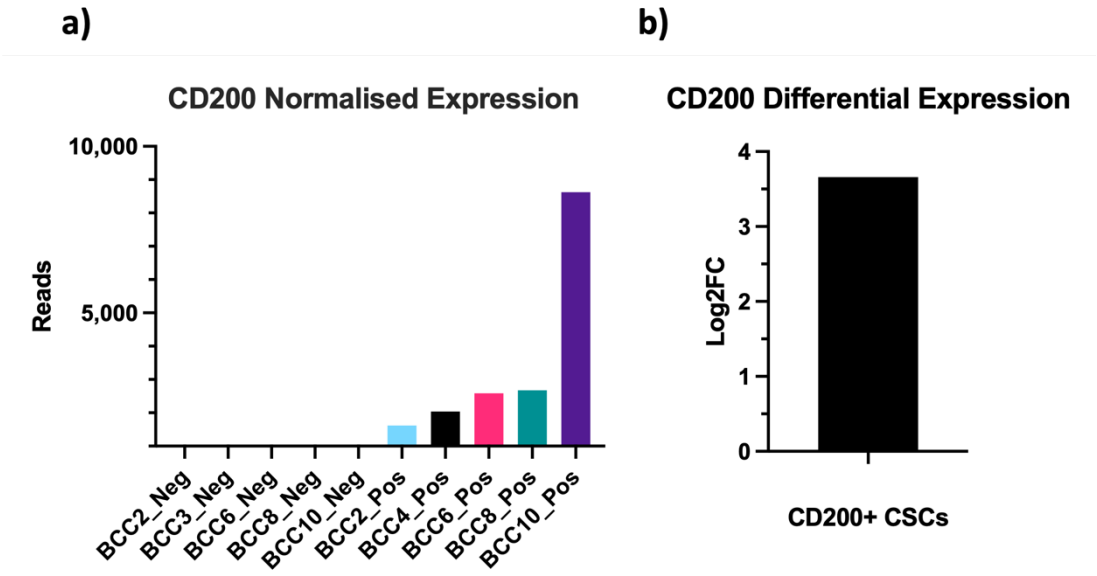


Figure 4.9 | Validation of CD200 expression in bulk RNA sequencing data from CD200+ CD45- and CD200- CD45- BCC cells. *CD200+ CD45- and CD200- CD45- BCC cells were sorted, and subsequently bulk RNA sequenced. The final dataset, following quality control, consisted of 5 CD45- CD200+ and 5 CD45- CD200- samples.* ***a)*** *CD200/ENSG00000091972 expression was assessed by checking normalised read counts across all samples.* ***b)*** *Differential Gene Expression (DGE) analysis was performed, comparing CD45- CD200+ samples to the CD45- CD200- samples. CD200 was the top differentially expressed gene (DEG). Bar graph shows the log2 Fold Change value for the gene.*

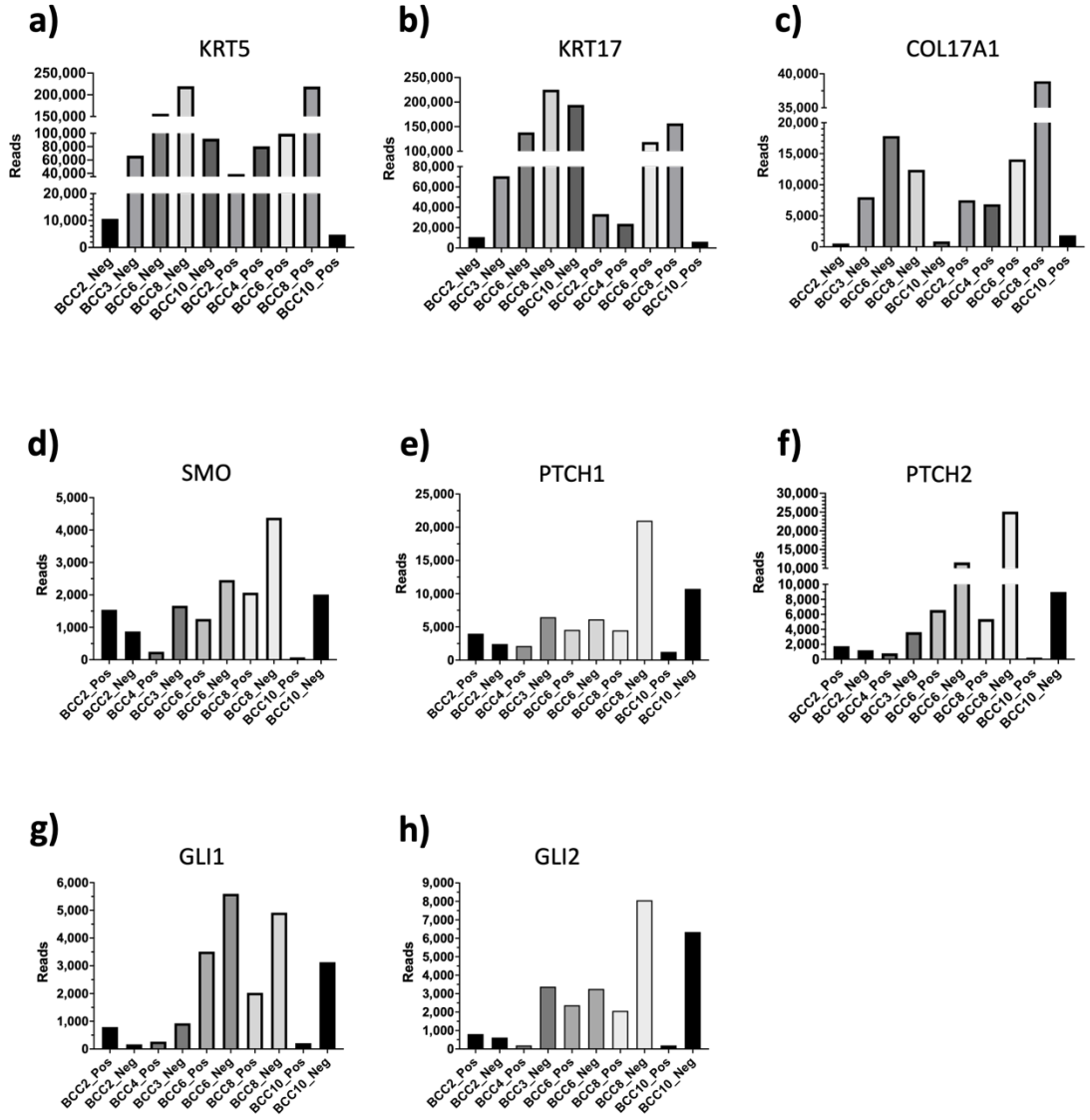


Figure 4.10 | Validation of expression of keratinocyte and BCC signalling pathway members in CD200+ CD45- and CD200- CD45- bulk RNA sequencing data. *CD200+ CD45- and CD200- CD45- BCC cells were sorted (n=5 per group), and subsequently bulk RNA sequenced. Normalised read counts were then investigated for expression of known BCC Hedgehog signalling pathway members and basal keratinocyte*

markers and plotted as bar graphs. ***a-c)*** *Normalised read counts were checked for basal keratinocyte markers KRT5, KRT17, and COL17A1, respectively.* ***d-h)*** *Normalised read counts were checked for BCC Hedgehog signalling pathway members SMO, PTCH1, PTCH2, GLI1, and GLI2, respectively.*

Having confirmed the sequencing of BCC basal keratinocytes through expression of BCC Hedgehog signalling members and basal keratinocyte markers in all samples, I next performed PCA and nonparametric two-tailed Spearman's correlation analysis on all the significant genes (normalised values used as input, adjusted P value <0.05) to check for any gene expression variation or correlations between both populations. PCA generated 5 principal components. PC1 and PC2 were selected as the input for the biplot, as combined, they explain > 60% of the dataset's variance (65.49%). The biplot clearly separated the CD200+ CD45- from the CD200- CD45- samples along the PC1 axis (Figure 4.11a and b). Interestingly, the CD200- CD45- samples form a tighter cluster along both axes, signifying less transcriptional variation in this population. CD200+ CD45- samples were distributed more along both axes compared to the CD200- CD45- samples, indicating a greater variation in gene expression in this population. For the correlation test, samples were first checked for normality using a Shapiro-Wilk's test. All samples were below the 0.5 α level and so the null hypothesis was rejected. A Spearman's non-parametric correlation test was therefore performed on the samples and a correlation metric plot was made. A greater correlation was seen amongst the CD200- CD45- samples than amongst the CD200+ CD45- samples (average R = 0.88 and R = 0.42, respectively; Figure 4.11c) and a strong inverse correlation was seen between the two sorted populations. A heatmap was generated on the 1,372 DEGs (normalised data, p.adj < 0.05) and hierarchical clustering (1-Pearson's correlation) was performed on the rows (genes) and columns (samples). The heatmap displayed a similar gene expression within populations and the hierarchical clustering displayed a clear separation between the populations (Figure 4.12).

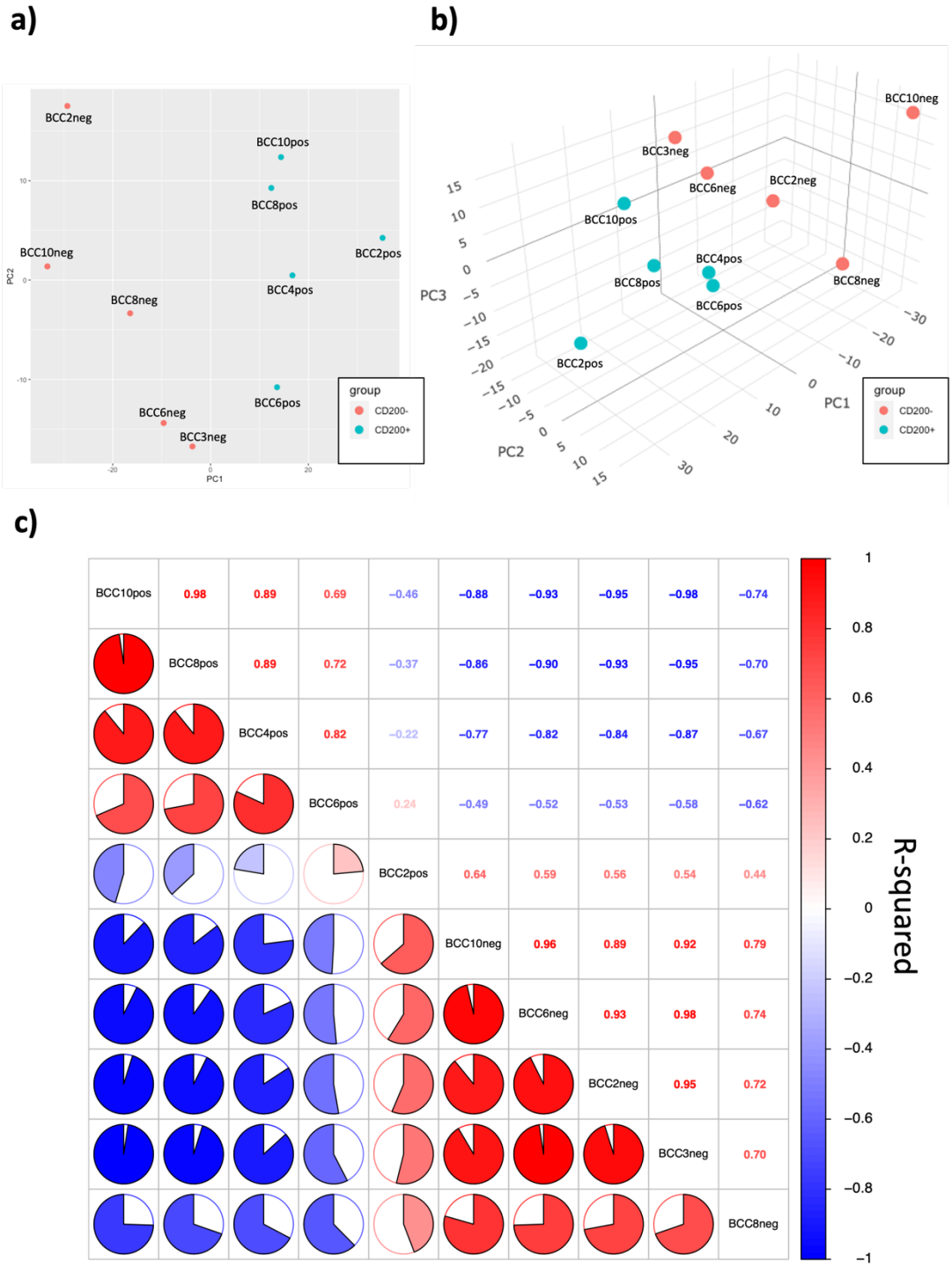

a)
b)
c)
BCC2neg
BCC10pos
BCC8pos
BCC10neg
BCC2pos
BCC4pos
BCC8neg
BCC6neg
BCC6pos
BCC3neg
PC1
PC2
group
CD200-
CD200+
PC3
BCC10pos 0.98 0.89 0.69 −0.46 −0.88 −0.93 −0.95 −0.98 −0.74
BCC8pos 0.89 0.72 −0.37 −0.86 −0.90 −0.93 −0.95 −0.70
BCC4pos 0.82 −0.22 −0.77 −0.82 −0.84 −0.87 −0.67
BCC6pos 0.24 −0.49 −0.52 −0.53 −0.58 −0.62
BCC2pos 0.64 0.59 0.56 0.54 0.44
BCC10neg 0.96 0.89 0.92 0.79
BCC6neg 0.93 0.98 0.74
BCC2neg 0.95 0.72
BCC3neg 0.70
BCC8neg
R-squared
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1

Figure 4.11 | Sample correlation analyses following bulk RNA sequencing of human BCC CD200+ CD45- and CD200- CD45- populations. *Human BCCs were sorted into CD200+ CD45- and CD200- CD45- populations (n=5 per group)) and subsequently bulk RNA sequenced. Following quality control, normalised read counts were used as input into Principal Component Analysis (PCA) and correlation analysis.* ***a, b)*** *2D and 3D PCA was performed on the normalised read using Principal Component (PC)1 and PC2, and PCs 1, 2, and 3, respectively.* ***c)*** *A nonparametric two-tailed Spearman's correlation test was performed using the normalised read counts and a heatmap was generated using the gplots package in R. The top half of triangle displays the R-squared values, and the bottom half of the triangle displays the values as a pie chart. Both halves are also coloured based on R-squared values. R-squared values closer to one are considered highly similar, whereas R-squared values close to zero are not similar. All analyses and plots were generated in R.*

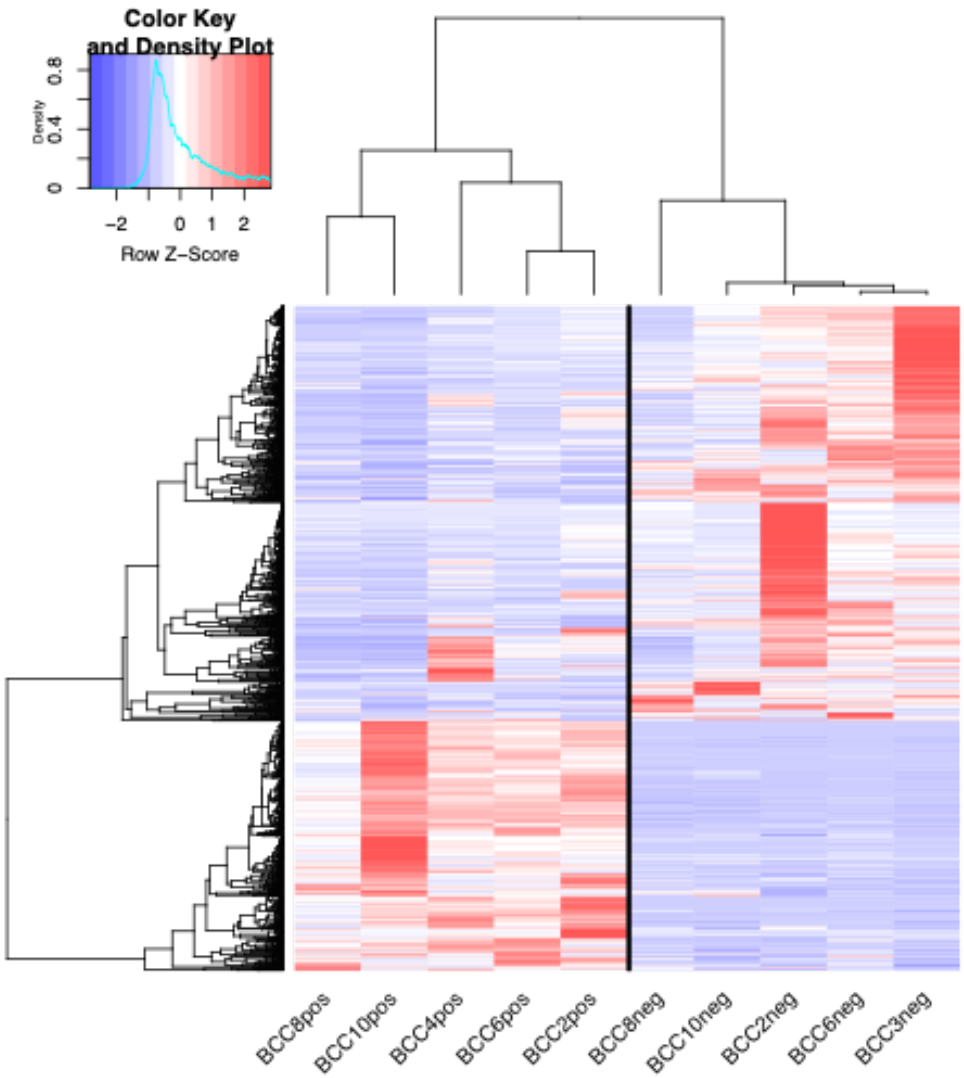


Figure 4.12 | Heatmap of human BCC CD200+ CD45- and CD200- CD45- samples. *Human BCCs were sorted into CD200+ CD45- and CD200- CD45- populations (n=5) and subsequently bulk RNA sequenced. Following quality control, normalised read counts were used for heatmap generation. Hierarchical clustering was also performed on the samples (columns) and genes (rows). Primary segregation was seen between the two sample groups, which have been separated by a black line. Colour key shows*

the colours associated with the row Z-score. The row Z-score is an indicator of the expression value, which has been scaled across the samples to allow for comparison. Heatmap was generated using the gplots package in R.

Differential expression analysis generated 1,372 DEGs, of which 517 were upregulated and 855 were downregulated (adjusted P value <0.05) (Figure 4.13).

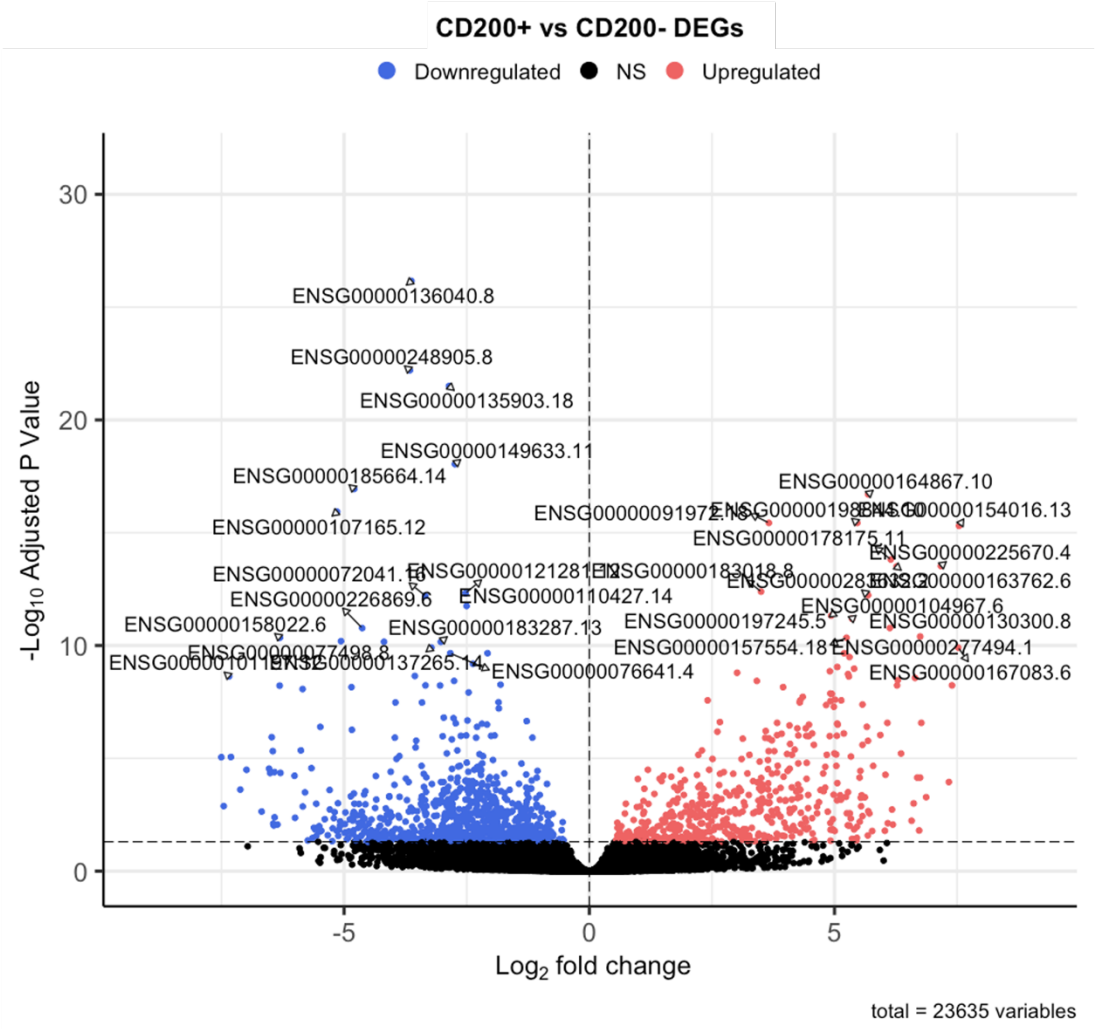


Figure 4.13 | Volcano plot of human BCC CD200+ CD45- vs CD200- CD45- differentially expressed genes. *Human BCCs were sorted for CD200+ CD45- and CD200- CD45- populations and subsequently RNA sequenced. Differential gene expression was then performed on both populations and a volcano plot was generated using the EnhancedVolcano package in R. dotted lines denote significance boundaries (Adjusted P Value < 0.05). Blue, significant downregulated genes, red, significant upregulated genes, black, non-significant genes.*

Enrichment analysis was first performed on the DEGs to identify enriched pathways. As a first pass, all 1372 DEGs were used as the input for clusterProfiler's enrichGO and enricher packages in R to perform Gene Ontology (GO) and Wikipathways enrichment analyses, respectively. Gene-Concept Network (CNET) plots were made for both analyses, which plots the linkages of the genes and the enriched concepts that they are involved in. The gene expression values were also added as an extra layer of information to the CNET plot with the addition of the foldChange argument. Enrichment dot plots were also generated, which incorporates functional enrichment scores, gene counts, gene ratios and P values into one plot. GO Enrichment analysis revealed that the CD200+ CD45- BCC cells demonstrate enrichment in cartilage development, extracellular matrix organisation, extracellular structure organisation, regulation of angiogenesis and skeletal system morphogenesis (Figure 4.14). The most significant enriched terms have been included in the dot plot, and include terms such as cellular calcium ion homeostasis, cell junction assembly, mesenchymal cell differentiation, and Rho protein signal transduction (Figure 4.15).

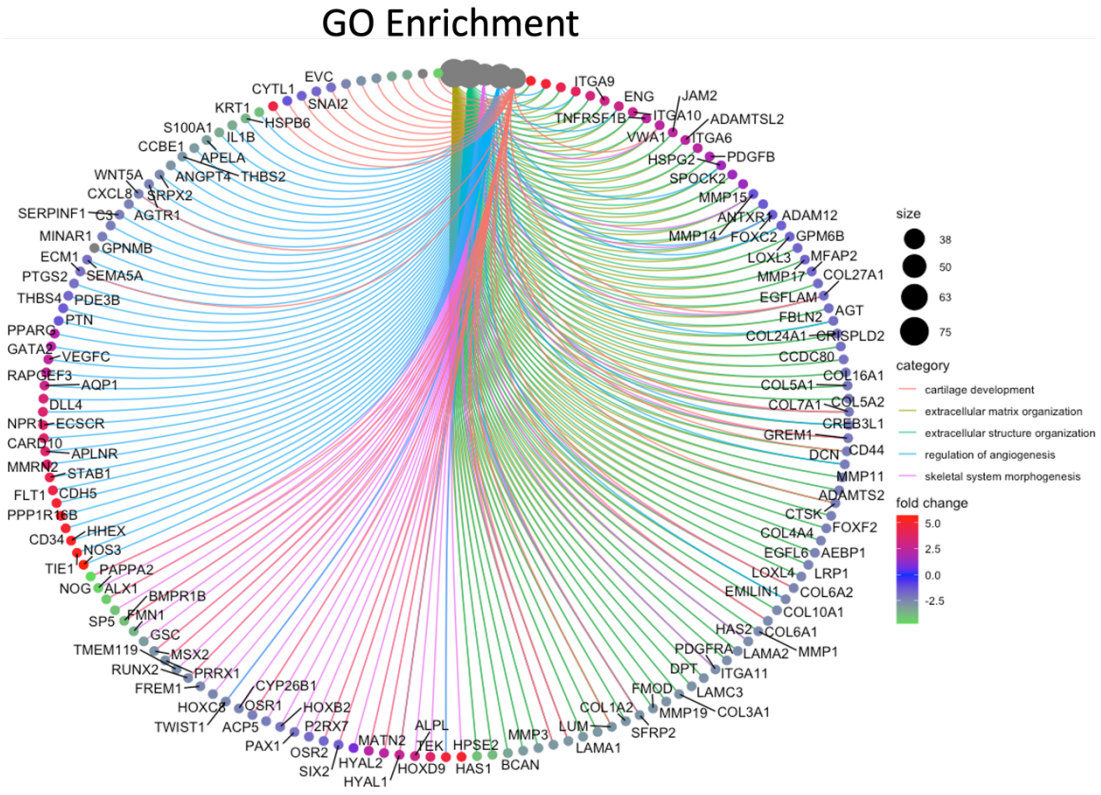


Figure 4.14 |Gene Ontology (GO) gene-concept network (CNET) plot of BCC CD200+ CD45- vs CD200- CD45- DEGs. CD200+ CD45- vs CD200- CD45- human BCC DEGs were used as input for GO analysis. CNET plot was then generated, which displays the top GO terms and genes associated with those terms. Lines and colours depict the GO category enrichment. Size of the dark grey dots (top) depict number of DEGs involved in the enriched GO term. Colour of the gene dots depict the log2 fold change value.

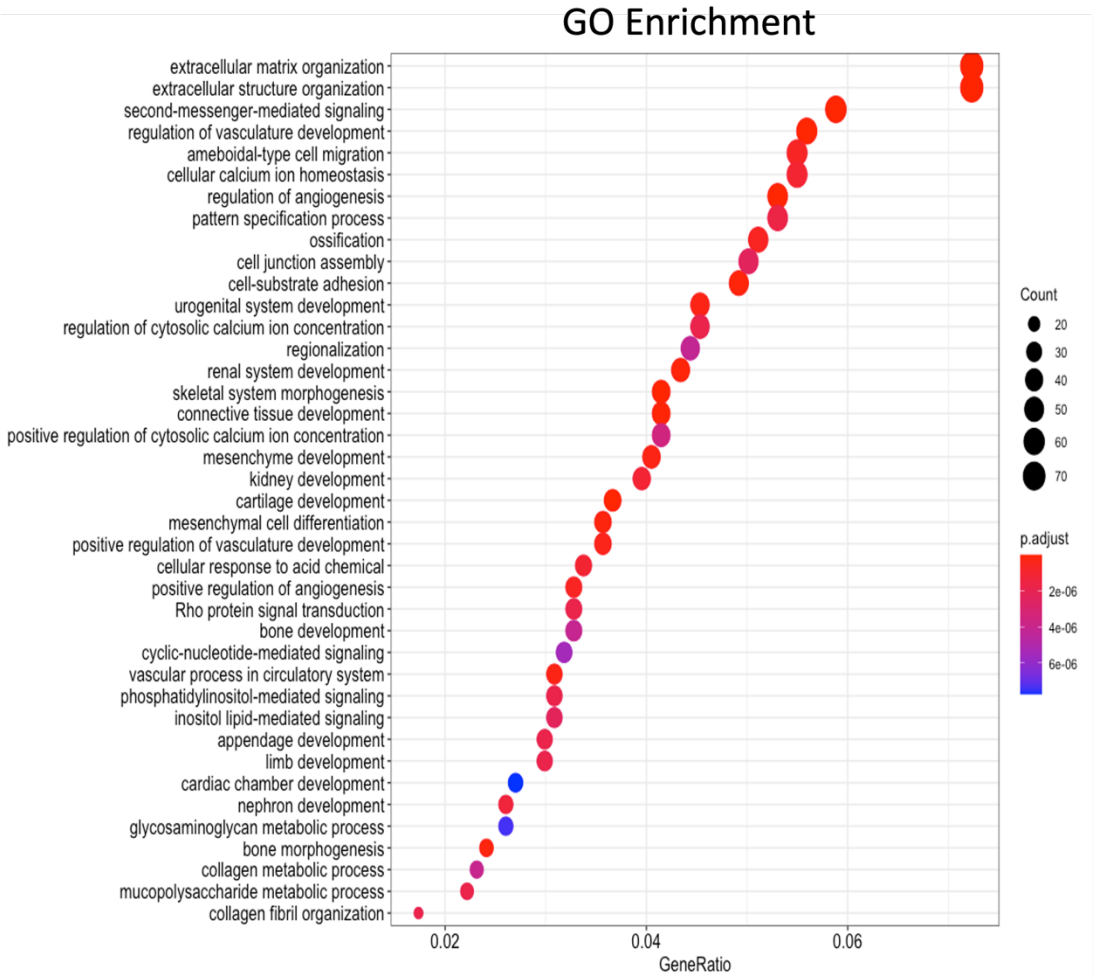


Figure 4.15 | Gene Ontology (GO) enrichment plot of BCC CD200+ CD45- vs CD200- CD45- DEGs. *CD200+ CD45- vs CD200- CD45- human BCC DEGs were used as input for GO analysis. An enrichment dot plot was then generated from the results, which depicts the top 40 significantly enriched GO terms. Size and colour of the dots depict the number of DEGs involved in the enriched GO process and the associated significance.*

Wikipathways analysis revealed enrichment in the Focal Adhesion-PI3K-Akt-mTOR-signalling pathway, PI3K-Akt-signalling pathway, Epithelial to mesenchymal transition in colorectal cancer, osteoblast differentiation and Imatinib and Chronic Myeloid Leukaemia (Figure 4.16). Amongst the most significantly enriched pathways were Matrix Metalloproteases, Hippo-Merlin signalling dysregulation and Complement activation (Figure 4.17).

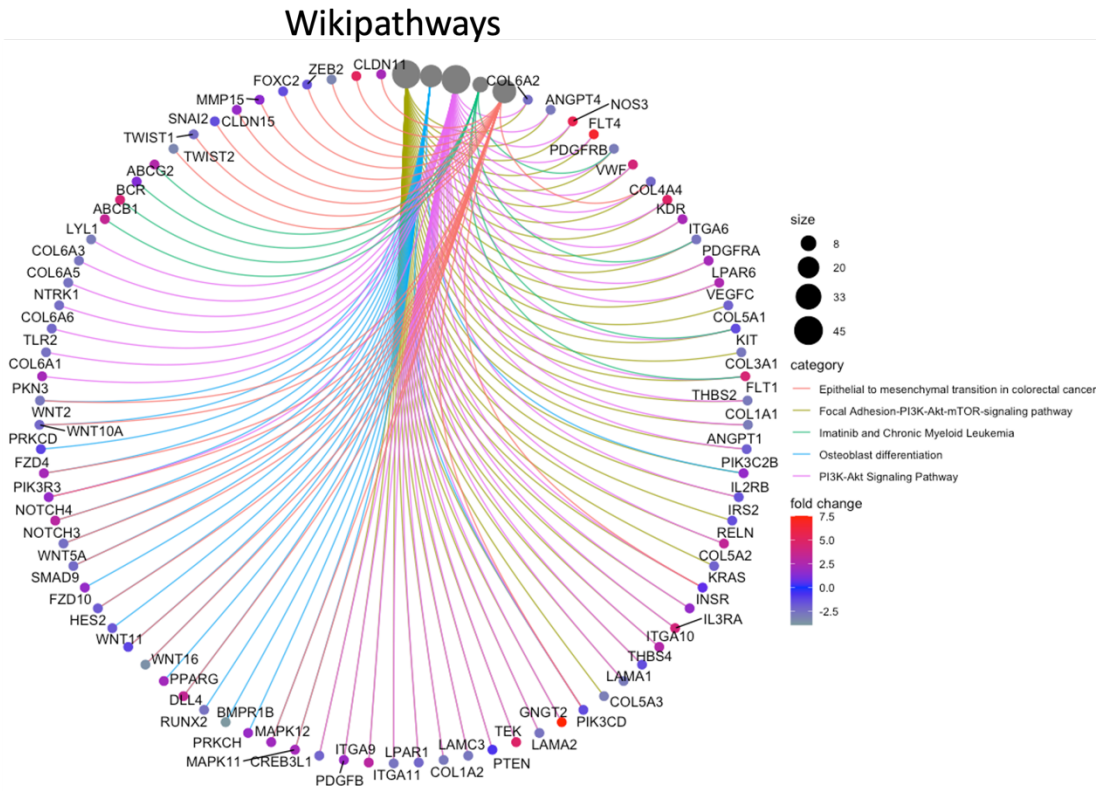


Figure 4.16 | Wiki Pathways gene-concept network (CNET) plot of BCC CD200+ CD45- vs CD200- CD45- DEGs. *CD200+ CD45- vs CD200- CD45- human BCC DEGs were used as input for Wiki Pathways analysis. CNET plot was then generated, which displays the top Wiki Pathways terms and genes associated with those terms. Lines and colours depict the Wiki Pathway category enrichment. Size of the dark grey dots (top) depict number of DEGs involved in the enriched GO term. Colour of the gene dots depict the log2 fold change value.*

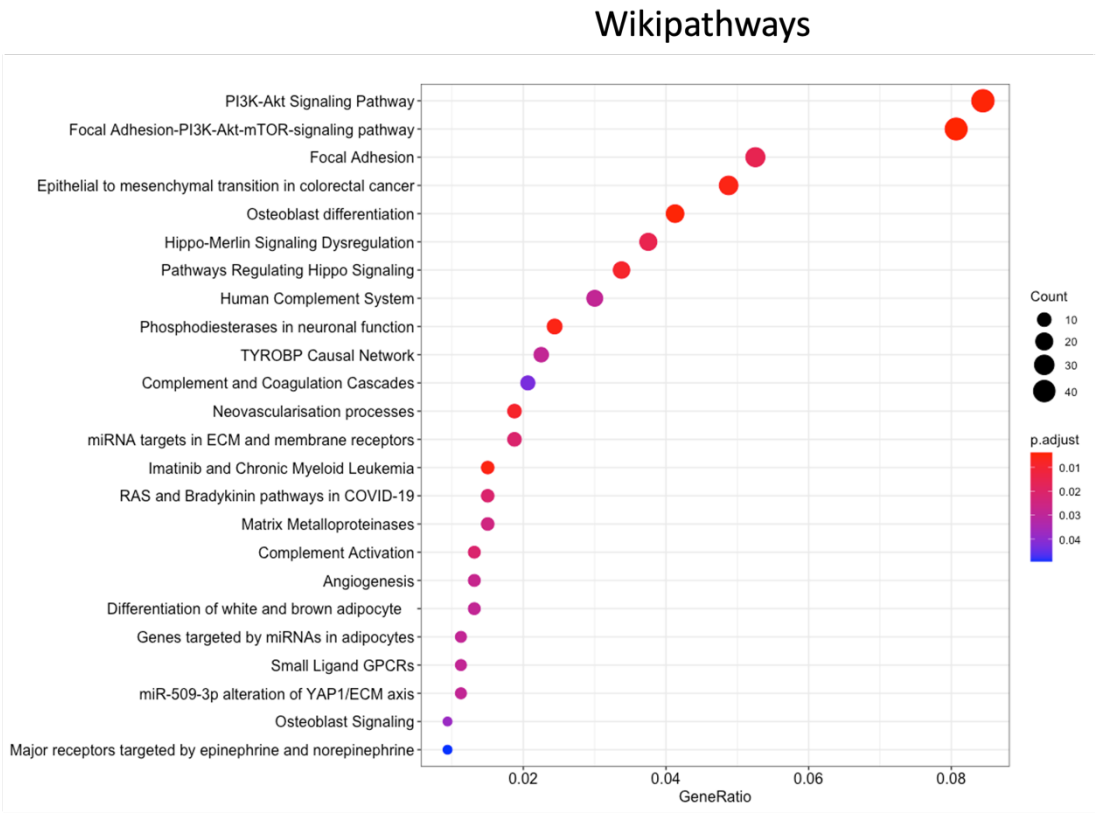
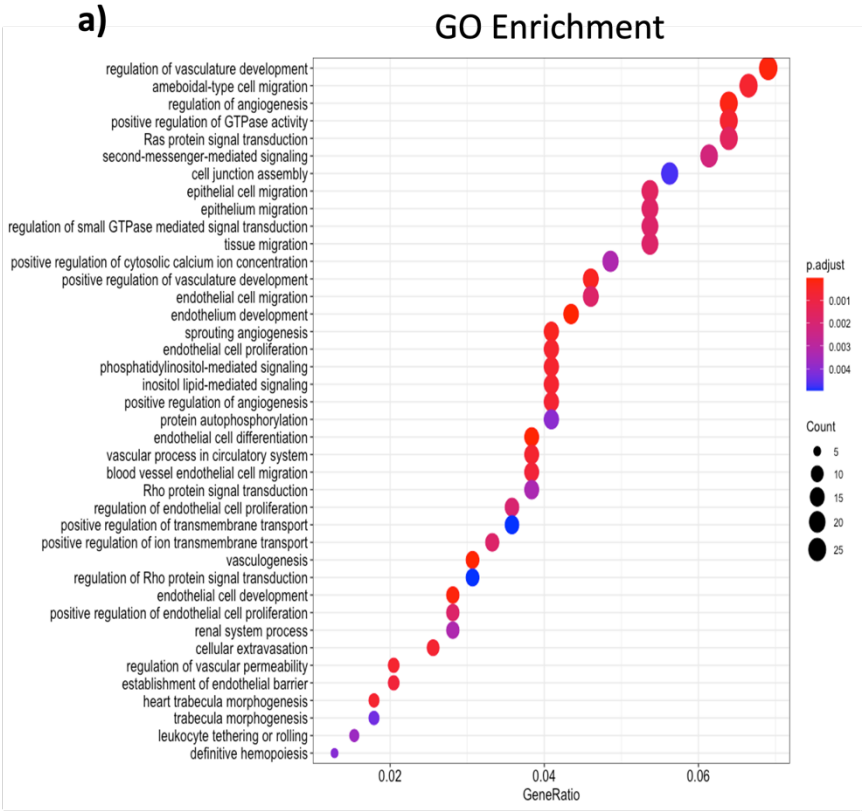


Figure 4.17 | Wiki Pathways enrichment plot of BCC CD200+ CD45- vs CD200- CD45- DEGs. *CD200+ CD45- vs CD200- CD45- human BCC DEGs were used as input for Wiki Pathways analysis. An enrichment dot plot was then generated from the results, which depicts the top 40 significantly enriched Wiki Pathways terms. Size and colour of the dots depict the number of DEGs involved in the enriched Wiki Pathways process and the associated significance.*

To investigate the dataset further, DEGs were split into upregulated (Log2FC >0) and downregulated (Log2FC<0) gene sets and GO enrichment and Wikipathways analyses were performed again. Upregulated genes showed GO enrichment in epithelial cell migration, Ras protein signal transduction, Rho protein signal transduction (Figure 4.18a) and enrichment in Wnt signalling, Regulatory circuits of the STAT3 signalling pathway, and miR-509-3p alteration of TAP1/ECM axis (Figure 4.18b).

a)


GO Enrichment
regulation of vasculature development
ameboidal-type cell migration
regulation of angiogenesis
positive regulation of GTPase activity
Ras protein signal transduction
second-messenger-mediated signaling
cell junction assembly
epithelial cell migration
epithelium migration
regulation of small GTPase mediated signal transduction
tissue migration
positive regulation of cytosolic calcium ion concentration
positive regulation of vasculature development
endothelial cell migration
endothelium development
sprouting angiogenesis
endothelial cell proliferation
phosphatidylinositol-mediated signaling
inositol lipid-mediated signaling
positive regulation of angiogenesis
protein autophosphorylation
endothelial cell differentiation
vascular process in circulatory system
blood vessel endothelial cell migration
Rho protein signal transduction
regulation of endothelial cell proliferation
positive regulation of transmembrane transport
positive regulation of ion transmembrane transport
vasculogenesis
regulation of Rho protein signal transduction
endothelial cell development
positive regulation of endothelial cell proliferation
renal system process
cellular extravasation
regulation of vascular permeability
establishment of endothelial barrier
heart trabecula morphogenesis
trabecula morphogenesis
leukocyte tethering or rolling
definitive hemopoiesis
0.02
0.04
0.06
GeneRatio
p.adjust
0.001
0.002
0.003
0.004
Count
5
10
15
20
25

b)

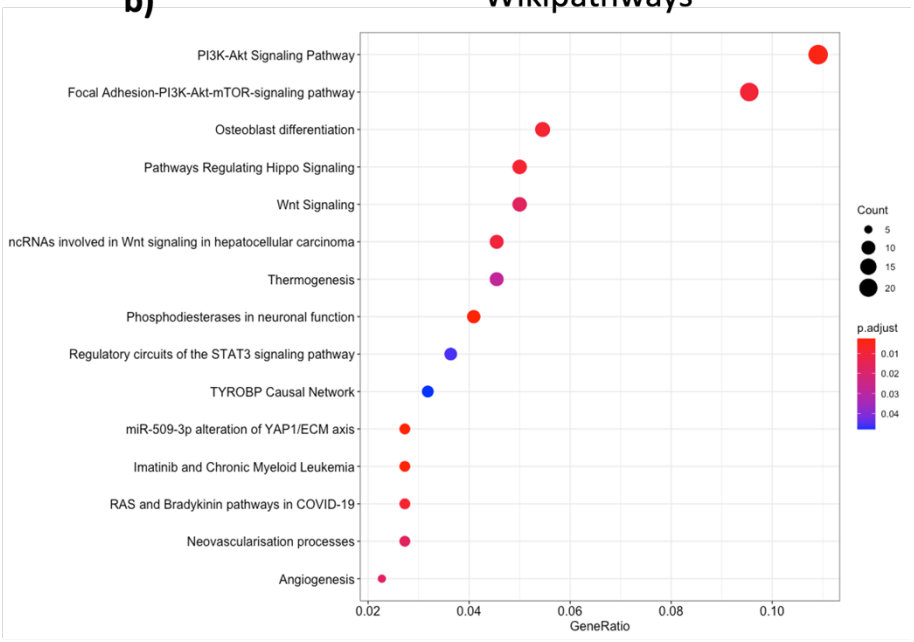

Wikipathways
PI3K-Akt Signaling Pathway
Focal Adhesion-PI3K-Akt-mTOR-signaling pathway
Osteoblast differentiation
Pathways Regulating Hippo Signaling
Wnt Signaling
ncRNAs involved in Wnt signaling in hepatocellular carcinoma
Thermogenesis
Phosphodiesterases in neuronal function
Regulatory circuits of the STAT3 signaling pathway
TYROBP Causal Network
miR-509-3p alteration of YAP1/ECM axis
Imatinib and Chronic Myeloid Leukemia
RAS and Bradykinin pathways in COVID-19
Neovascularisation processes
Angiogenesis
0.02
0.04
0.06
0.08
0.10
GeneRatio
Count
5
10
15
20
p.adjust
0.01
0.02
0.03
0.04

Figure 4.18 | GO and Wiki Pathways enrichment plots of BCC CD200+ CD45- vs CD200- CD45- upregulated DEGs. *CD200+ CD45- vs CD200- CD45- human BCC upregulated DEGs were used as input for Wiki Pathways analysis. An enrichment dot plot was then generated from the results, which depicts the top 40 significantly enriched terms from* ***a)*** *GO, and* ***b)*** *Wiki Pathways. Size and colour of the dots depict the number of DEGs involved in the enriched processes and the associated significance.*

Amongst the downregulated genes, negative enrichment was observed in extracellular matrix organisation, connective tissue development and mesenchymal cell proliferation (Figure 4.19). Interestingly, negative enrichment for regulation of Wnt signalling pathway and epithelial to mesenchymal transition was also seen. Wikipathways failed to generate enrichment results.

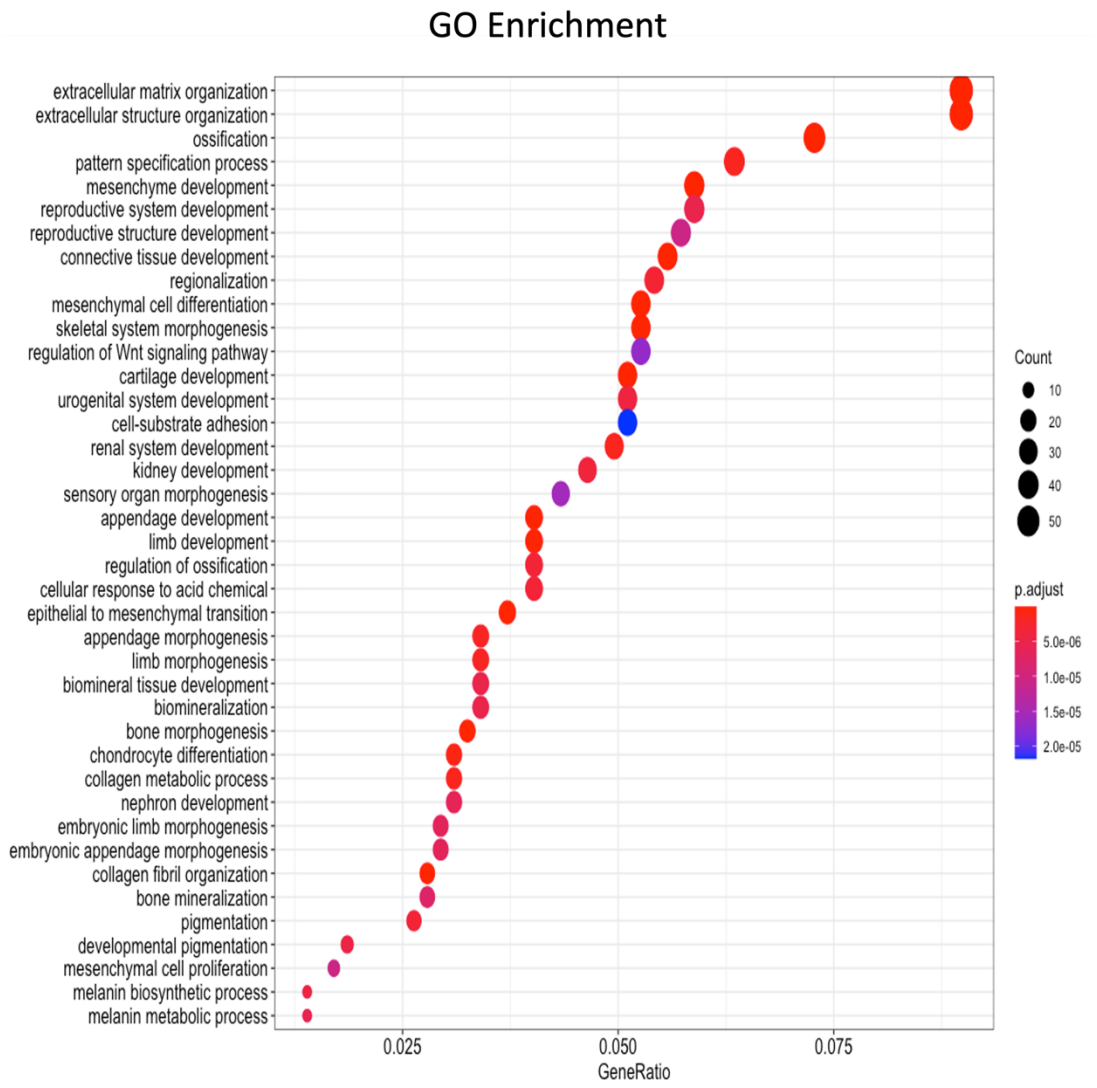


Figure 4.19 | GO enrichment plot of BCC CD200+ CD45- vs CD200- CD45- downregulated DEGs. *CD200+ CD45- vs CD200- CD45- human BCC downregulated DEGs were used as input for GO analysis. An enrichment dot plot was then generated from the results, which depicts the top 40 significantly enriched GO terms. Size and*

colour of the dots depict the number of DEGs involved in the enriched GO process and the associated significance.

Pre-ranked Gene Set Enrichment Analysis (GSEA) was also performed on the DEGs on the following curated gene set collections from the Molecular Signatures Database (MSigDB): Hallmark (H), Curated gene sets (C2), Ontology gene sets (C5), and Cell type signature gene sets (C8). These gene sets were selected as they provide the most informative enrichment results and are the most used in publications. The top 25 positive- and negative-Normalised Enrichment Scores (NES) were taken from each analysis and used as the input for Broad Institute's Morpheus to generate heatmaps. Amongst the most positively enriched Hallmark gene sets were KRAS signalling up, UV response up, reactive oxygen species pathway, P53 pathway, and WNT Beta Catenin pathway. Amongst the most negatively enriched Hallmark gene sets were TGF Beta signalling, Epithelial Mesenchymal Transition, Notch signalling, and MTORC1 signalling (Figure 4.20a). The most positive and negative enriched gene sets for C2 Curated can be found in Figure 4.20b. These top enriched gene sets seem to be disease-specific and may require further investigation in order to provide any value to the data. However, there does appear to be a positive enrichment for VEGF signalling.

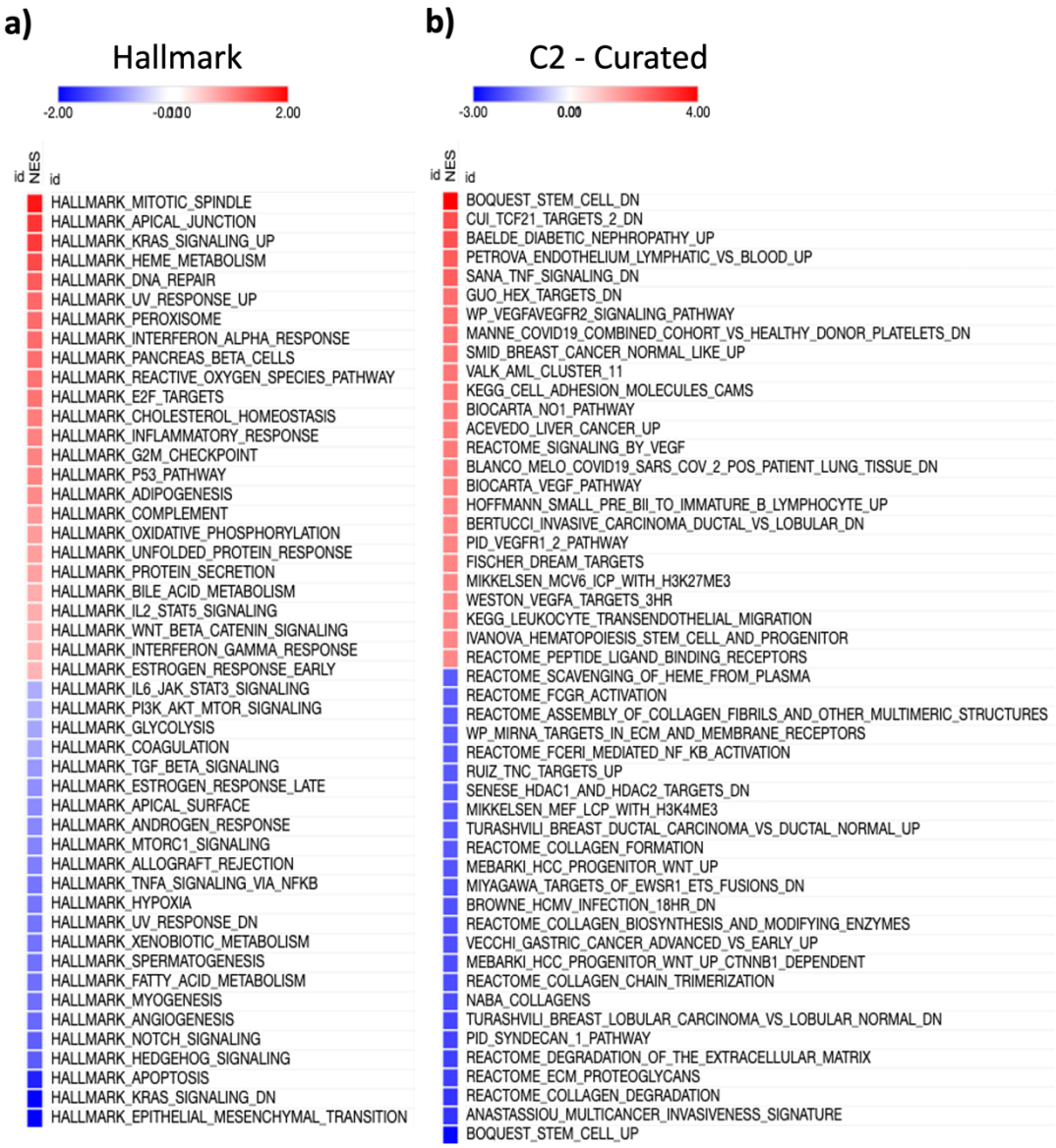


Figure 4.20 | Pre-ranked Gene Set Enrichment (GSEA) analysis of Hallmark and C2-Curated gene sets in human BCC CD200+ CD45- vs CD200- CD45- DEGs. *Significant DEGs were pre-ranked from highest-to-lowest log2 fold-change and then used as input into GSEA pre-ranked tool. The top 25 positive and negative normalised*

enrichment scores (NES) were taken from each analysis and used as input for Morpheus heatmap software (Broad Institute).

The top- and bottom-25 GSEA results for C5 – Ontology and C8 – Cell Type gene sets are displayed in Figure 4.21. Of note, amongst the most positively enriched gene sets in C5 – Ontology are Endothelial related (Endothelium Development, Endothelial Cell Development, Establishment of Endothelium Barrier) and amongst the most negatively enriched was skin development.

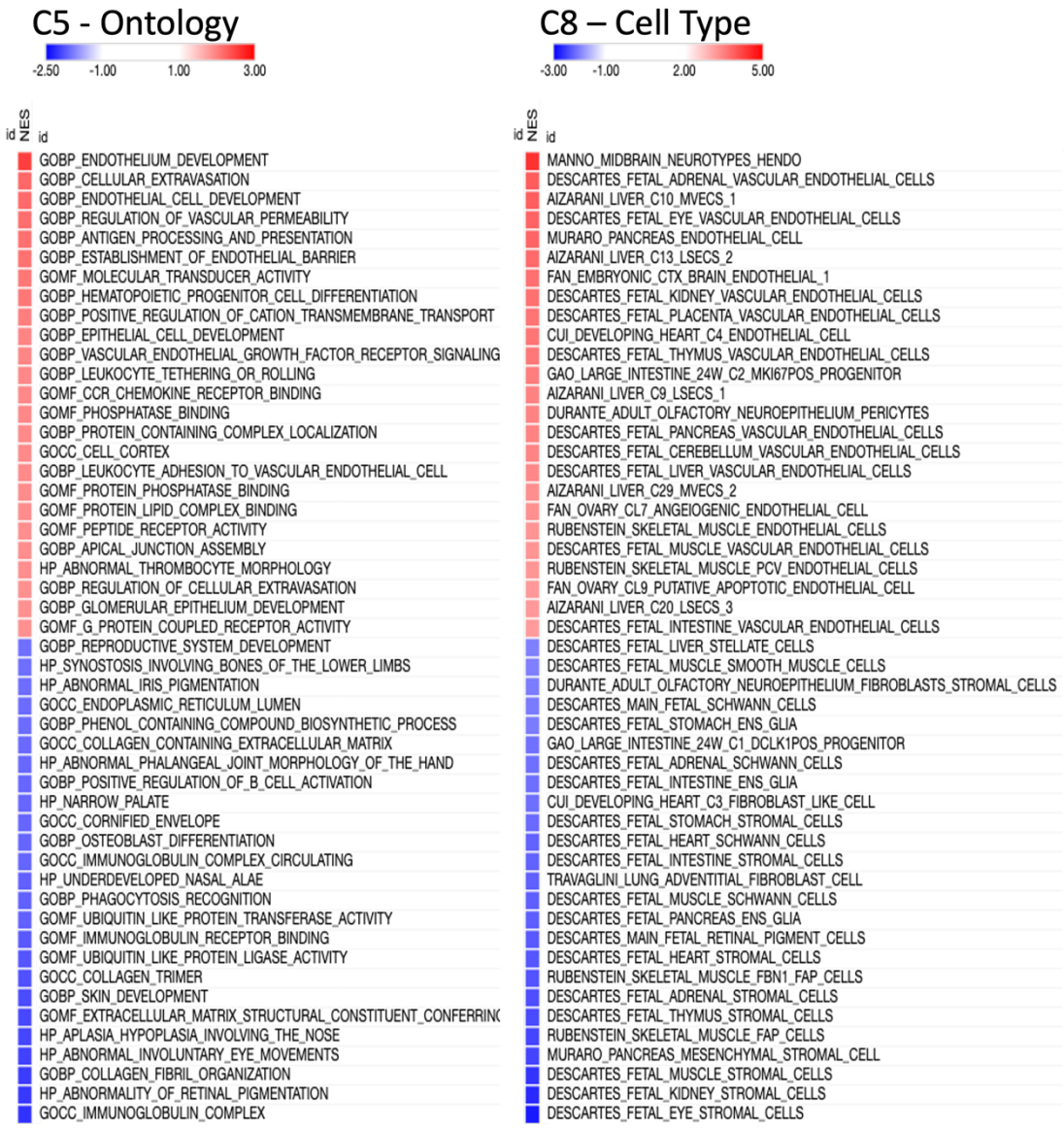
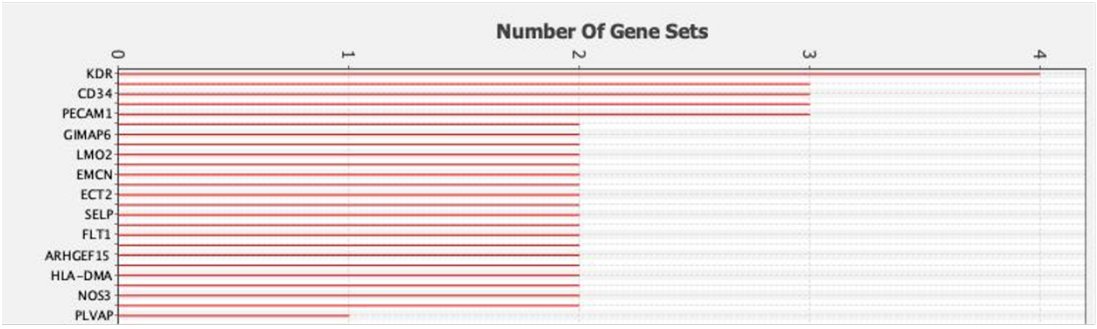


Figure 4.21 | Pre-ranked Gene Set Enrichment (GSEA) analysis of C5 – Ontology and C8 – Cell Type gene sets in human BCC CD200+ CD45- vs CD200- CD45- DEGs. *Significant DEGs were pre-ranked from highest-to-lowest log2 fold-change and then used as input into GSEA pre-ranked tool. The top 25 positive and negative normalised*

enrichment scores (NES) were taken from each analysis and used as input for Morpheus heatmap software (Broad Institute).

From the pre-ranked GSEA analyses, it was apparent that the CD200+ CD45- DEGs were enriched for the VEGF signalling pathway. To investigate which DEGs were playing a major role in the pre-ranked GSEA analyses, I next performed leading edge analysis. The leading edge is a subset of genes that appear in the ranked DEG list at or before the point in which the running enrichment score sum reaches its maximum deviation from zero, meaning that these genes can be interpreted as the core genes that account for a gene set enrichment signal. Leading edge analysis was performed on the C2 – Curated and Hallmarks top 20 positively enriched gene sets. I also performed the same analyses but performed them on a manually curated list of gene sets with the following criteria: size > 10 and enrichment score > 0.5. Both analyses confirmed that endothelial cell markers were amongst the top genes in this dataset (see Figure 4.22 for leading edge results). I also used the positively enriched DEGs as input into Rummagene. Rummagene takes a list of DEGs and searches them against 688,108 published gene sets on PubMed. Similarly, results showed matches for datasets containing cancer-associated fibroblasts and endothelial cells.

a) C2 – Curated



b) Hallmarks

Figure 4.22 | Leading-edge analysis of Pre-ranked Gene Set Enrichment Analysis of Human BCC CD200+ CD45 vs CD200- CD45- DEGs. *Significant DEGs were pre-ranked from highest-to-lowest log2 fold-change and then used as input into GSEA pre-ranked tool. Leading-edge analysis was then performed on the* ***a)*** *C2-Curated and* ***b)*** *Hallmarks results. Figures are screenshots of the top results, depicting the genes involved in the most processes.*

I then investigated the DEGs via the Ingenuity Pathway Analysis (IPA) software (Qiagen). Using all significant DEGs as input, I performed a causal analysis. The overview page was first assessed, which revealed the top canonical pathways were axonal guidance signalling, CREB signalling in neurons, hepatic fibrosis, and granulocyte adhesion and diapedesis (Figure 4.23a). Amongst the top diseases and functions was cancer and dermatological diseases and conditions (Figure 4.23bi). The top molecular and cellular functions was cellular movement, cellular development, cellular function and maintenance, cellular growth and proliferation, and cell-to-cell signalling and interaction (Figure 4.23bii). Interestingly, the top physiological system development and function was cardiovascular system development and function (Figure 4.23biii). Another interesting observation was from the top regulator effect networks, which contained regulators involved in activation of blood cells (Figure 4.23c).

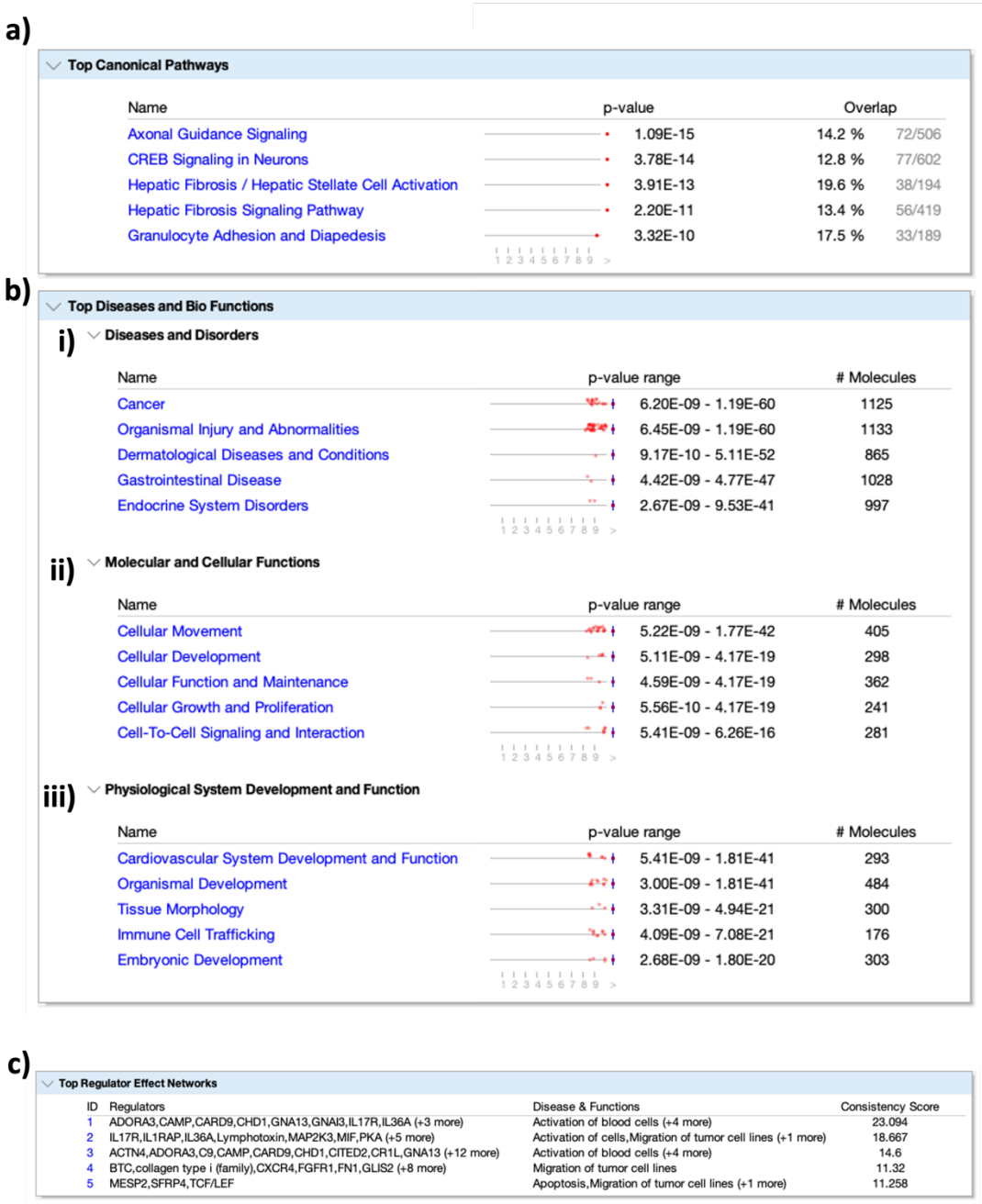
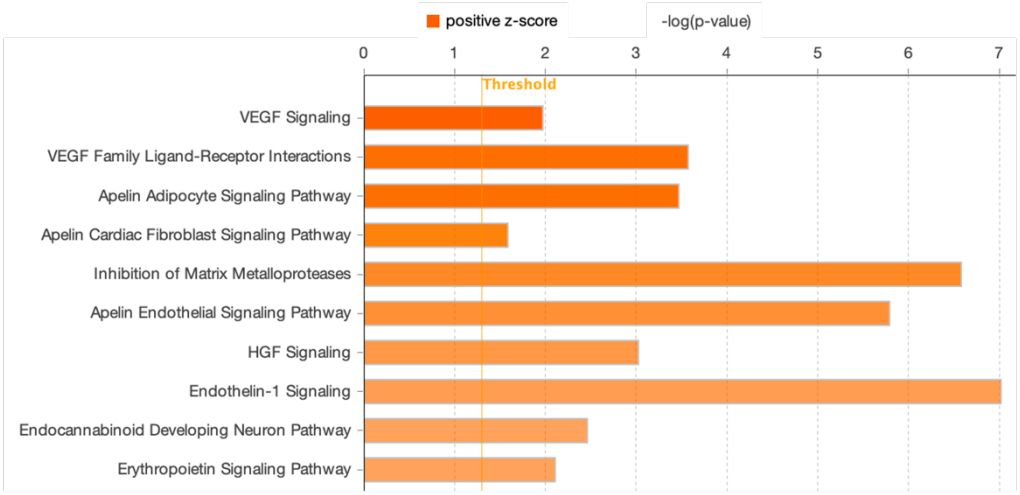


Figure 4.23 | Ingenuity Pathway Analysis (IPA) overview of human BCC CD200+ CD45- population DEGs. *Human BCC CD200+ CD45- significant DEGs were used as*

input into Ingenuity Pathway Analysis (IPA) and a causal network analysis was performed. Following analysis completion, a summary page is firstly provided. Figure shows the summary page. ***a)*** *Summary of the top canonical pathways.* ***b)*** *Summary of the top Diseases and Bio Functions.* ***c****) Summary of the top regulator effect networks. Each result is provided with an associated p-value, indicating the significance of the enrichment, and the overlap, which indicates the number of genes that are involved in other processes. The '# Molecules' depicts the number of DEGs involved in a particular process.*

When investigating and ordering the most enriched signalling pathways by activation z score, amongst the top pathways were again VEGF signalling (Figure 4.24a). There were no genes of interest when investigating the causal networks. I then performed biomarker analysis on the dataset. To do this, I ran the biomarker analysis tool in IPA, and filtered the results for molecules that were located on the plasma membrane and positively enriched. I also added the other following filters: expressed in keratinocytes, expressed in dermis, expressed in epidermis, expressed in skin. This resulted in 5 biomarkers of interest: ABCB1, ABCG1, CCRL2, KDR, and TNFRSF10A (Figure 4.24b). This analysis showed that VEGFR2 (KDR) was amongst the top differentially expressed genes, and amongst the top actionable biomarkers. I next wanted to assess the expression of the members associated with this pathway. To do this, I loaded up the VEGF pathways in IPA and used the overlay tool to overlay the BCC CD200+ CD45- DEGs (Figure 4.25). The results showed that VEGFR and VEGFR2 (KDR) are both highly enriched in the BCC CD200+ CD45- DEGs. There appears to be signalling at the cell surface via VEGF/VEGFR and VEGFC/D/VEGFR2. There was also a strong predicted activation of the ROCK signalling mechanism, and plausible predictions of PI3K/AKT and SRC/PI3K signalling. Overall, this analysis supported the model that there is VEGF signalling within these cells but may encompass more than one cell type.

a)



b)

Symbol	Location	Family	Expr Log Ratio	Expr p-value	Keratinocytes	Dermis	Epidermis	Skin
ABCB1	Plasma Membrane	transporter	4.413	4.67E-07	x	x	x	x
ABCG1	Plasma Membrane	transporter	2.097	0.000938	x	x	x	x
CCRL2	Plasma Membrane	G-protein coupled receptor	4.159	0.01	x	x	x	x
KDR	Plasma Membrane	kinase	4.902	0.00000103	x	x	x	x
TNFRSF10A	Plasma Membrane	transmembrane receptor	1.848	0.0439	x	x		x

Figure 4.24 | Ingenuity Pathway Analysis results of human BCC CD200+ CD45- population DEGs. *Human BCC CD200+ CD45- significant DEGs were input into Ingenuity Pathway Analysis (IPA) and a causal network analysis was performed. Canonical pathway analysis was investigated and ranked by activation Z-score. The*

Z-score statistic is a combination of activation (positively or negatively activated, indicated by a positive or negative score, respectively) and significance. ***a)*** *Top canonical pathways ordered by activation z-score. Bars show -log(p-value) of the pathways.* ***b)*** *Biomarker analysis was performed. This analysis takes a range of user-defined inputs and identifies biomarkers. Table shows the 5 results after querying for cell surface markers that are present on keratinocytes, dermis, epidermis, and skin. X depicts presence of marker.*

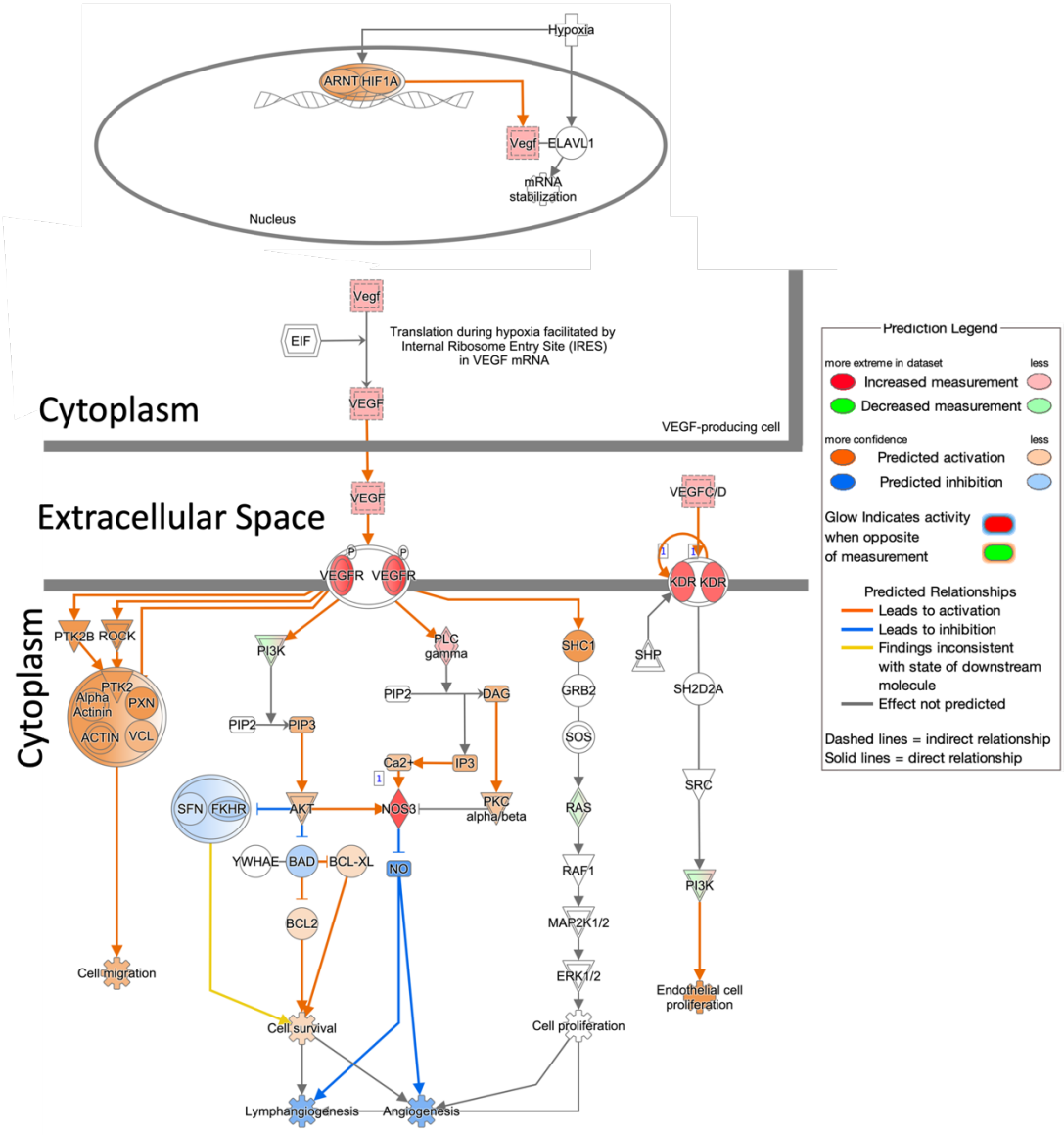


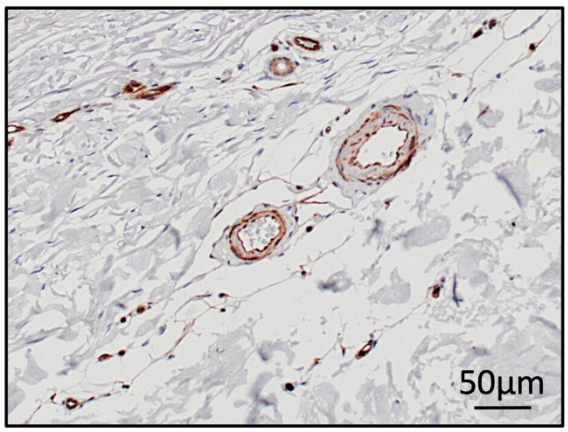
Figure 4.25 | VEFG signalling pathway with human BCC CD200+ CD45- DEGs overlaid. *Diagram shows the VEGF signalling pathway that was loaded and opened in Ingenuity Pathway Analysis (IPA) with the human BCC CD200+ CD45- DEGs overlaid. Red and green indicate overlaid DEG expression on the pathway. Orange*

and blue indicate predicted activity of the pathway in the dataset. Strength of the red/green and orange/blue indicate the size of the expression and predicted activity.

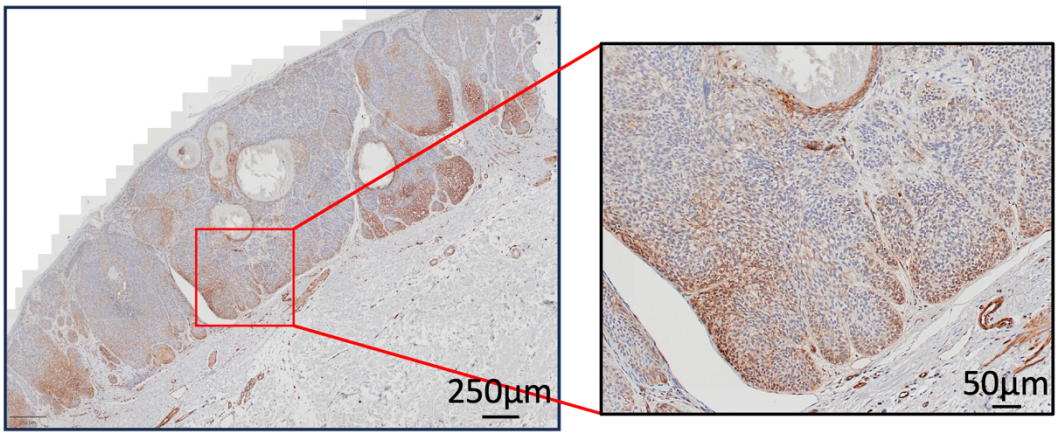
4.2.2.3 Immunohistochemical validation of VEGFR2 expression in primary human BCC tissue

As the VEGFR2 signalling pathway was the most significant finding in the bulk RNAseq analyses, I next wanted to check its expression in human BCC tissue sections. A new VEGFR2 antibody was purchased and was first validated on the tissue sections by IHC and positive labelling in the blood vessels in the dermis (Figure 4.26a). When assessing VEGFR2 expression in human BCCs, I saw in some samples there was expression within the tumour nodules (4/10 BCCs; Figure 4.26b), and in others the expression was seen in the endothelial cells surrounding the tumour nodules (6/10 BCCs; Figure 4.26c). interestingly, from the BCCs that displayed expression within the tumour nodule, the expression appeared to be mainly located around the tumour periphery. From these studies, I concluded that there was an endothelial cell contamination within the bulk RNA sequencing.

a)



b)



c)

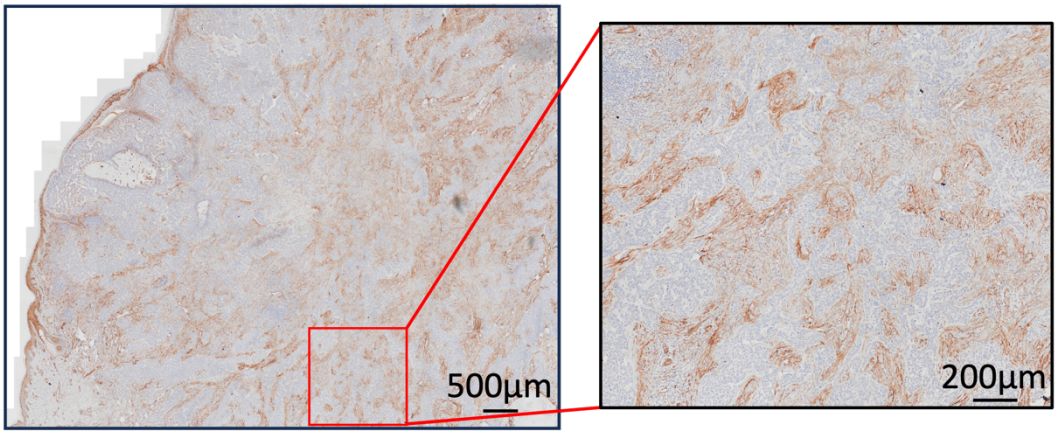


Figure 4.26 | Immunohistochemical staining for VEGFR2 in human BCCs. *Human BCC tissue was sectioned and stained for VEGFR2 and counterstained with haematoxylin. Stained sections were then imaged on a light microscope and analysed in QuPath software.* ***a)*** *The blood vessels located within the dermis of each tissue section was used as an internal positive staining control.* ***b, c)*** *VEGFR2 staining was seen within the tumour nodules and on the surrounding tumour endothelial cells, respectively. Red squares depict the zoomed areas.*

4.3 Discussion

In this chapter, I performed bulk RNA sequencing on a BCC cancer cell line, optimised the conventional human BCC dissociation protocol, and subsequently bulk RNA sequenced human BCC.

Having established that transduced CD200 expression within the UW-BCC1 cell line led to immune evasion in xenograft mice, I next decided to perform bulk RNAseq on the innate CD200-positive cells within the cell line. I first showed that the CD200 +/- populations unsurprisingly showed high correlation between them, with a slight increase in correlation between the CD200+ samples. This was also shown via PCA analyses, where CD200+ samples were tightly clustered along the PC1 axis. Generating a heatmap from the normalised reads of all samples, there was a clear separation between both groups and a clear difference in program of gene expression in both sample groups, indicating that CD200+ UW-BCC1 cells have a distinct program of transcription compared to their CD200- counterparts. This can also be demonstrated through the number of significant DEGs present. However, upon inspection of the DEGs and read counts, CD200 was not expressed in either sample group. It is clear that there was no CD200 expression within this cell line, consistent with the findings in chapter 3. As such, my analysis suggests that maybe I isolated a subpopulation that does not enrich for cancer stem cells.

I next sequenced the CD200+ CD45- and CD200- CD45- populations from human BCCs. I first needed to optimise our outdated tumour dissociation protocol. Isolating single cells in vitro requires simple trypsinisation of the cells, whereas human tissue

required both mechanical and enzymatic dissociation. At the time, there were two dissociation methods available to me, the first being the current conventional protocol that had been previously used by our lab to dissociate human skin cancers, and the second being a commercial kit developed and sold by Miltenyi Biotec. I firstly optimised the conventional dissociation protocol and demonstrated the importance of using fresh enzymes. I also showed that using this method was able to preserve the CD200 epitope. I also showed that the commercial Miltenyi kit was able to liberate more cells from the BCC tissue than the conventional method. Interestingly, despite a greater yield of cells, there were still large areas of undigested tissue present post-dissociation. When comparing this undigested tissue to the conventional method, we see that there were large areas of epidermal tissue still present. I also showed that the Miltenyi kit led to a degradation of the CD200 epitope. The company have not revealed the list of ingredients within their enzyme cocktail. There is a possibility that Trypsin is present, which is known to cleave the extracellular domain of CD200 in several regions. The Miltenyi protocol was therefore deemed not appropriate for the isolation of the CD200+ CD45- population within human BCC.

Having optimised the BCC dissociation protocol, I next sequenced the CD200+ CD45- and CD200- CD45- BCC populations. I firstly showed the accuracy of the FACS at sorting CD200+ CD45- and CD200- CD45- cells via immunofluorescence. For the RNAseq analysis, I generated a final dataset consisting of 10 samples, and showed that these were sequenced at a considerable depth with an average of 55.5 million reads per sample and 92.4% of those reads mapping to the human genome. I showed that I did in fact sequence BCC keratinocyte by assessment of keratinocyte and BCC markers. I also showed that CD200 was expressed in the CD200+ CD45- sorted

samples, and not in the CD200- CD45- sorted samples. I then checked for correlation within and between the samples through correlation and PCA analyses. Even though BCC is known to be heterogeneous, the CD200+ CD45- sorted samples showed a high degree of correlation between them, as demonstrated by tighter clustering on the PCA analyses, and higher R-squared values. There was also a clear separation between the two populations, indicating differing transcriptional programs between the two. Through heatmap analyses, I showed again that there was a clear separation between the two populations through hierarchical clustering. An interesting observation here was that there were more downregulated genes present in the CD200+ CD45- sample, as indicated by blue rows in the figure. This was also confirmed by DGE analysis between the two populations, resulting in 517 up- and 855-down regulated DEGs. Analysing the DEGs, I showed that the most significant findings were related to endothelial processes. The VEGF signalling pathway, in particular, VEGFR2 signalling, most frequently appeared amongst the results during analysis of the upregulated DEGs. Pre-ranked GSEA analysis revealed the key genes involved in the enriched gene sets, which identified VEGFR2 (KDR) as a significant player. I used a new analysis tool, Rummagene, to show that my list of significant DEGs were highly related to other published datasets. These datasets were also endothelial cell related. Using the IPA software to analyse the DEGs, I again showed that the VEGFR signalling pathway was amongst the most significant. I then showed that the VEGFR2 (KDR) gene was the amongst the results during a biomarker analysis, and indeed, when overlaying my DEGs onto the VEGF pathway in IPA, I saw an upregulation of the VEGF and VEGFR2 receptors.

The RNAseq analysis showed that there was a clear endothelial cell signature present. The IPA analysis of the VEGF pathway identified from the DEGs suggested that there

was communication between at least two different cell populations. This was confirmed when I undertook VEGFR2 labelling of BCC tissues which demonstrated positive tumour nodules and presumed endothelial cells. these findings could suggest that there are CD200-expressing endothelial cells present within BCC and that our flow sorting strategy may have included more than one cell type leading to a potential contamination of our dataset.

To conclude, I was not able to accurately identify an enriched BCC-specific signalling pathway due to the potential contamination of the dataset with endothelial cells. To address these findings in chapter 5, I undertook single cell RNA sequencing.

Chapter 5: Single Cell RNA Sequencing of Human BCC

5.1 Introduction

Bulk RNA-sequencing has long been the method of choice to investigate aberrant gene expression in disease. The method has been widely used in cancer research to classify, identify biomarkers, discover gene fusions, and optimise therapeutic interventions (Li and Wang 2021). The method is most commonly used to identify biomarkers of diseases in order to aid diagnosis, prognosis, and predictions. To date, there have been many successful applications where bulk RNAseq has been used to generate and validate disease-specific signatures (Van Allen et al. 2015; Chen et al. 2016; Shukla et al. 2017; Han et al. 2018; Zhou et al. 2019). There has however, only been a few of these signature panels that have been successfully translated into clinical practice. One of the main reasons behind this is due to the nature of the bulk RNAseq coupled with the inherent nature of the diseases being studied. Bulk RNAseq analyses the average gene expression across a large, bulk population of cells. This method operates under the assumption that all cells within a given sample have similar gene expressions and therefore the downstream results accurately represent them. However, it has been long known that cancers display a high degree of inter-tumour and intra-tumour heterogeneity (Liu et al. 2018). The problem here is that differences between individual cells and/or populations within samples are masked,

making it impossible to identify and distinguish rare cell populations. One such solution to this problem, particularly if a population of interest was being studied, is to isolate these cells through FACS, ensuring elimination of other cell populations. The problem with this approach is that the flow sorting strategy is limited by the ability to remove lineage negative cells. In our case, we only removed immune cells identified by CD45 and thus may have left behind endothelial cells that could have confounded our findings.

To combat this, single cell RNA-sequencing (scRNAseq) was developed. The method was first introduced in 2009 in which it was used in a 4-cell-stage blastomere (Tang et al. 2009). This method saw the generation of single cell libraries in individual tubes and lacked the ability to increase multiplex numbers, meaning that each sample had to be sequenced individually. The introduction of multiplexed scRNAseq was seen in 2011, where the barcoding of single cells in 96-well plates was used and then transferred to a single tube (Islam et al. 2011). In 2012, a new scRNAseq method was introduced, which demonstrated the ability to linearly amplify libraries, allowing for greater numbers of cells to be sequenced (Hashimshony et al. 2012). This method led to the first commercial application of scRNAseq in 2014, which is where the method gained traction. Since then, there have been many scRNAseq methods generated, mainly those that employ droplet-based or well-based methods (Lafzi et al. 2018). One of the more popular scRNAseq methods used today is a droplet-based technique developed by 10X Genomics. The company developed a chromium system, which uses advanced microfluidics to perform single cell partitioning and barcoding, and up to 8 samples and 20,000 cells-per-sample can be processed at once (Li and Wang 2021). The method generates Gel Bead-in-emulsion (GEMs), which contain a single

cell plus all the reagents required to amplify and specifically tag each RNA molecule that has come from that cell, which allows the identification of those amplified RNA molecules to be traced back to the cell. Post-sequencing, the company also has a software, cellranger, that performs sample demultiplexing, barcode processing, and gene mapping and counting. This allows users to directly input their sequencing data into downstream analysis pipelines.

In this chapter, I used scRNAseq to investigate the CD200+ BCC cell population. I employed the 10X Genomics scRNAseq method to sequence 13 BCC and normal skin samples. To analyse the complete dataset, I first used a smaller dataset to optimise the analysis pipeline. I then used the optimised pipeline to fully analyse the dataset. I was able to identify the cell types present within sorted and unsorted samples and accurately identify the BCC tumour keratinocytes. I was then able to identify aberrant signalling pathways that were specific to these cells and validate the activity of the TGFβ pathway in human BCC tissue sections.

5.2 BCC single-cell RNA Sequencing Analysis Protocol Optimisation

For single-cell sequencing, we recruited 11 patients, generating a total of 13 samples. The tissues were dissociated as described in chapter 2 and 4 and single cell suspensions were obtained following flow-sorting: CD45- (BCC and normal skin) and specifically the CD200+ CD45- population. Samples were then sent to Novogene (Cambridge, UK) for sequencing. A total of 4,337,304,263 raw sequencing reads were generated across 16,878 raw single cells from all thirteen samples (BCC =

3,505,339,430 raw reads across 13,836 raw cells and NS = 831,864,732 raw reads across 3,042 raw cells). On average, there were 3,144 median genes and 11,111 median UMI's per cell for the BCC samples and 3,208 median genes and 14,401 median UMI's per cell for the NS samples. On average, 82.38% of reads were confidently mapped to the human genome (BCC = 81.14% average and NS = 86.53% average) and 49.72% of reads were confidently mapped to the human transcriptome (BCC = 48.37% average and NS = 54.2% average). On average, sequencing saturation was 86.72% across all samples (BCC = 86.89% average and NS = 86.13% average). Raw sequencing statistics can be found in Table 5.1. Samples were then processed as described in Chapter 2: Methods.

Table 5.1 | Single cell RNA sequencing data information table.
Table shows the sequencing statistics for thirteen basal cell carcinoma (BCC) and normal skin (NS) samples.

Sample	Sample Information	Condition	Number of Cells	Mean reads per cell	Median UMI per cell	Median genes per cell	Total genes detected	Reads confidently mapped to genome	Reads confidently mapped to transcriptome	Total reads	Sequencing saturation
BCC_NS_1	Multiple BCC patients NS pooled together.	CD45-	1,614	140,807	10,594	2,818	24,177	85.20%	50.70%	227,262,411	76.80%
BCC_1	Multiple BCC patients and pooled together.	CD45-	2,383	97,531	12,626	3,499	26.945	91.90%	52.70%	232,903,692	68.80%
BCC_2	BCC patient 2	CD45-/CD200+	2,402	98,988	10,146	3,066	25,168	90.60%	48.80%	237,769,763	71.40%
BCC_2_NS	Taken from BCC patient 2	CD45-	888	272,393	14,018	3,172	22,728	90.90%	59.10%	242,429,873	88.10%
BCC_3	Taken from BCC patient 3	CD45-/CD200+	541	412,546	8,511	2,914	22,463	88.40%	48.30%	223,187,128	91.50%
BCC_4+	Taken from BCC patient 4	CD45-/CD200+	262	1,566,222	15,931	3,556	20,148	88.50%	52.80%	410,350,265	97.10%
BCC_4-	Taken from BCC patient 4	CD45-/CD200-	426	943,715	17,206	3,638	21,437	91.50%	63%	402,022,772	95%
BCC_5	Taken from BCC patient 5	CD45-	776	461,449	13,080	3,561	23,539	66.60%	43.20%	358,084,670	88.50%
BCC_6	Taken from BCC patient 6	CD45-	1,805	264,062	11,209	3,271	24,590	78.70%	45.50%	476,631,188	87.90%
BCC_7	Taken from BCC patient 7	CD45-	2,002	181,151	6,606	2,364	23,875	65.50%	39.10%	362,664,326	87.20%
BCC_7_NS	Taken from BCC patient 7	CD45-	540	670,690	18,590	3,635	20,222	83.50%	52.80%	362,172,448	93.50%
BCC_8+	Taken from BCC patient 8	CD45-/CD200+	1,564	271,478	8,517	3,008	23,508	72.60%	44.30%	424,591,733	91.10%
BCC_8-	Taken from BCC patient 8	CD45-/CD200-	1,675	225,155	7,274	2,560	22,949	77.10%	46%	377,133,893	90.40%

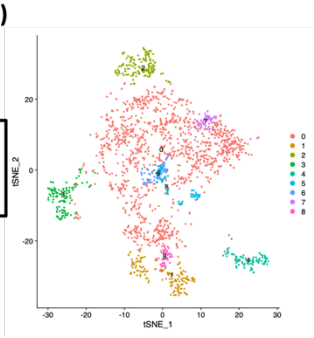
To optimise the analysis pipeline, I first used the first batch of samples that were sent to Novogene. This dataset consisted of 4 samples (3 BCCs, 1 NS). Initially, each sample was processed individually to assess any differences between clustering and how well separated cell types were. Please see Methods 2.12.2.1 for a detailed overview of the optimisation. When clustering the samples, a clustering resolution of 0.1 was chosen (see Appendix 5.1). To visualise the clustering, two common approaches were taken: t-distributed stochastic neighbour embedding (t-SNE; Maaten and Hinton 2008) and Uniform Manifold Approximation and Projection (UMAP; McInnes et al. 2018; see Appendix 5.2). Both methods were used to visualise the clustering in the four samples. Favour was initially given to UMAP, as it works better at preserving both local and global interactions, is quicker, and works well on larger datasets. The integrated dataset will also consist of a large number of cells. Downstream functions also use UMAP as the default visualisation tool. Both methods were performed using default parameters (t-SNE: perplexity = 30. UMAP: n_neighbours = 5, min_dist = 0.03) and number of dimensions = 3 to allow for three-dimensional exploration of the clustering. Figure 5.1 displays both dimension reduction plots for each sample. In samples where there were more than 1000 cells (BCC_1 and BCC_2, Figure 5.1a and c, respectively), UMAP appeared to generate more defined and well-separated clusters. For the samples with < 1000 cells, both methods produced very similar clustering. However, UMAP generated 'smoother' clusters, whereas t-SNE generated more dispersed and granular clusters. It became evident that correctly clustering samples with less than 1000 cells would be a difficult task. In the two samples where there was improved clustering, t-SNE in both cases displayed a cluster(s) within a cluster, whereas in UMAP, there was no overlapping of the clusters. However, plotting a third dimension (tSNE_3) of the data shows the separation of those

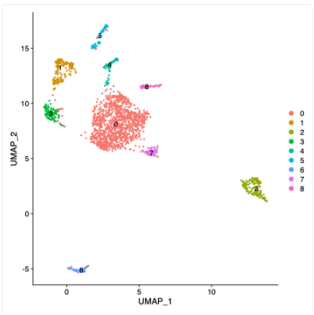
overlapping clusters for both samples (Figure 5.2). Therefore, both methods are able to distinguish all clusters, but a third dimension and subsequently more compute power is needed. UMAP was therefore chosen as the visualisation tool.

t-SNE	UMAP	Cells/Cluster

a)

BCC_1

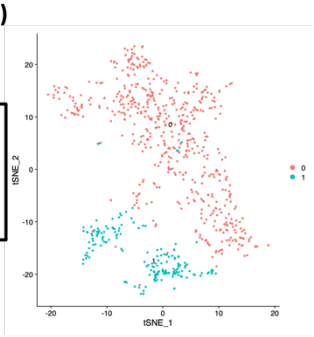


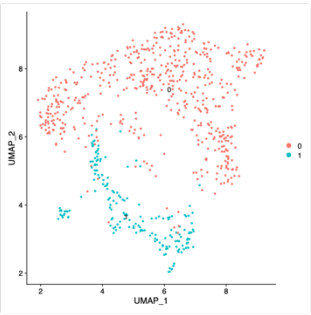


Cluster	Number of Cells
0	1293
1	164
2	151
3	124
4	90
5	87
6	83
7	68
8	59

b)

BCC_2_NS

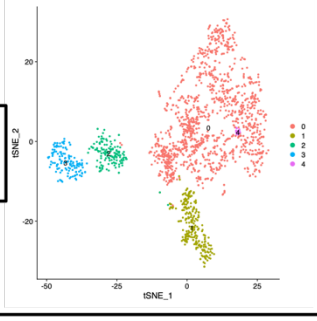


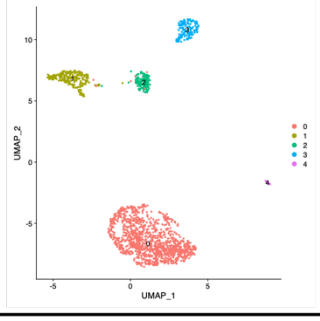


Cluster	Number of Cells
0	551
1	181

c)

BCC_2

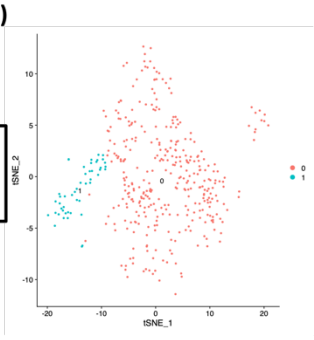


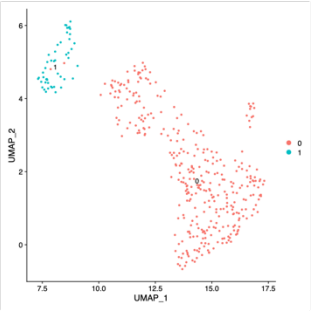


Cluster	Number of Cells
0	1211
1	256
2	148
3	142
4	16

d)

BCC_3





Cluster	Number of Cells
0	348
1	48

Figure 5.1 | t-SNE and UMAP cluster visualisations of 4 single cell samples. *Seurat's clustering algorithm was used with a resolution parameter of 0.1 and using 30 Principal Components (PCs) as input. t-Stochastic Neighbour Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP) were then used with default parameters to visualise the clustering.* ***Left****, t-SNE plots.* ***Middle****, UMAP plots.* ***Right****, number of cells per Seurat cluster.* ***a)*** *BCC_1 sample,* ***b)*** *BCC_2_NS sample,* ***c)*** *BCC_2 sample, and* ***d)*** *BCC_3 sample.*

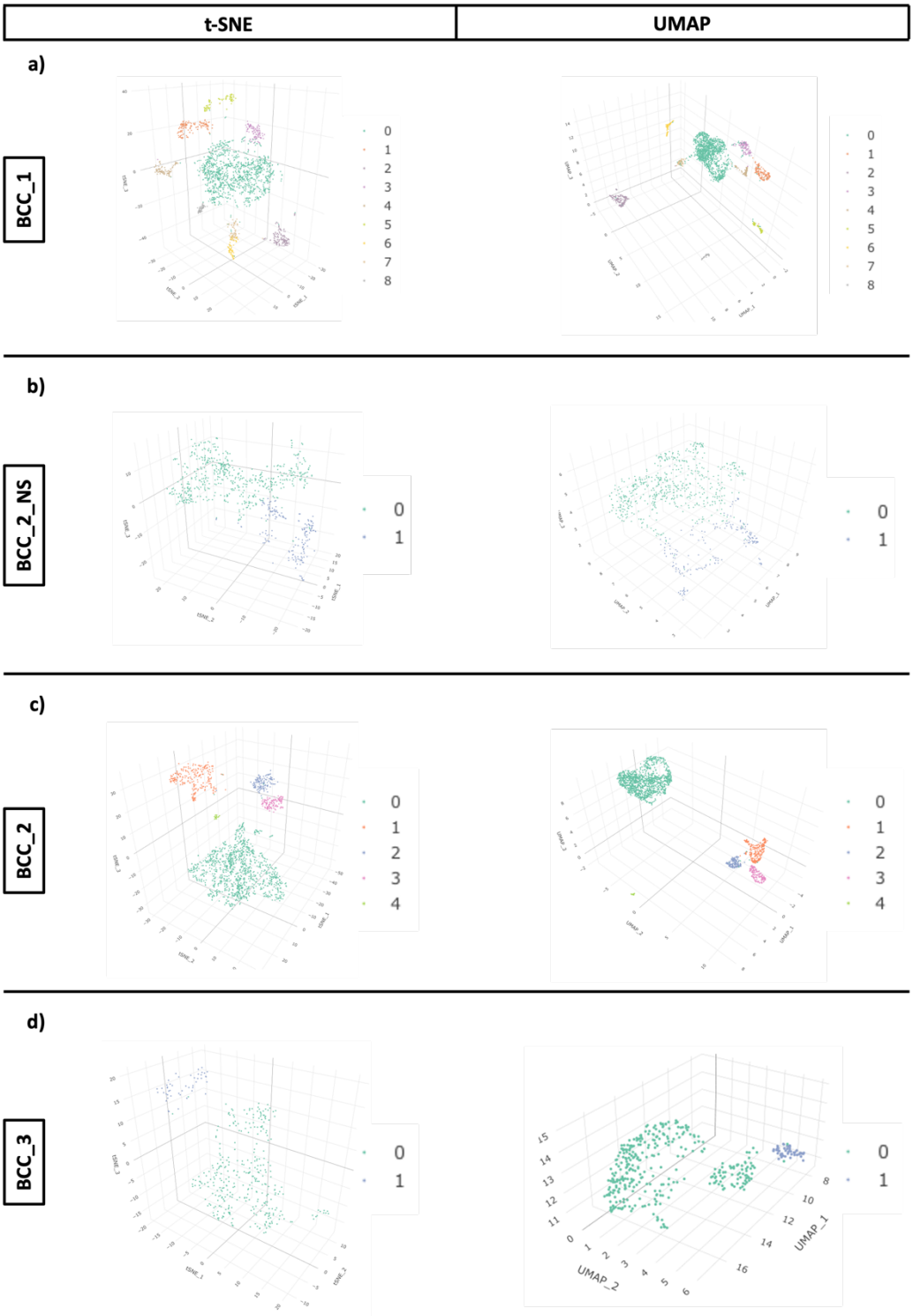

t-SNE
UMAP
a)
BCC_1
0
1
2
3
4
5
6
7
8
b)
BCC_2_NS
0
1
c)
BCC_2
0
1
2
3
4
d)
BCC_3
0
1
UMAP_1
UMAP_2

Figure 5.2 | 3-Dimensional t-SNE and UMAP cluster visualisations of 4 single cell samples. *Seurat's clustering algorithm was used with a resolution parameter of 0.1 and using 30 Principal Components (PCs) as input. t-Stochastic Neighbour Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP) were then used with default parameters plus the number of dimensions set to 3 to visualise the clustering in 3-D.* ***Left****, t-SNE plots.* ***Right****, UMAP plots.* ***a)*** *BCC_1 sample,* ***b)*** *BCC_2_NS sample,* ***c)*** *BCC_2 sample, and* ***d)*** *BCC_3 sample.*

5.2.1 Cell Type Identification on Individual Samples

To identify the clusters, a variety of cell type markers were obtained from The Human Protein Atlas (HPA) and the literature, and their expression was observed via UMAP and Violin plots (Ji et al. 2020; Solé-Boldo et al. 2020; Wang et al. 2020; Guerrero-juarez et al. 2022; Yerly et al. 2022; Quah et al. 2023; Ren et al. 2023; The Human Protein Atlas. 2023). A table of all markers used can be found in Appendix 5.3. Seurat's 'FindAllMarkers' was also performed on each sample to generate differentially expressed genes (DEGs) for each cluster. Significant DEGs (adjusted p value < 0.05) were then entered into g:Profilers g:GOSt functional profiler (Raudvere et al. 2019) and results were used in conjunction with the marker expression to determine the identity of each cluster. Finally, immune cell markers were checked to validate their removal during sample processing. 3 of 4 samples had no CD45 expression (gene = PTPRC), whilst the BCC_1 sample had 16 cells with expression ranging from 0.1-1 (Appendix 5.4).

There was a total of nine clusters for the BCC_1 sample (Figure 5.3a). Cell type marker expression revealed clusters 0, 1, 3, 4, 5, and 7 highly expressed keratinocyte markers KRT5, KRT14, and KRT17 (Figure 5.3b), whilst cluster 2 highly expressed endothelial markers CD34, PECAM1, CDH5, and KDR (Figure 5.3c), cluster 6 expressed fibroblast-like markers COL3A1, LUM, and DCN (Figure 5.3d), and cluster 8 expressed melanocyte markers MLANA and PMEL (Figure 5.3e). Enrichment analysis for each cluster confirmed the marker expression with HPA terms such as 'skin 1: cells in basal layer', 'skin 2: cells in spinous layer', 'skin 1: cells in basal layer' for the keratinocyte clusters, 'skin 1: endothelial cells' and 'skin 2: endothelial cells' for the endothelial cell cluster, 'skin 1: fibrohistiocytic cells' and 'skin 2: fibrohistiocytic cells' for the

fibroblast-like cluster, and ‘skin 1: melanocytes’, ‘skin 2: melanocytes’ and ‘melanocyte differentiation’ for the melanocyte cluster (Appendix 5.5). CD45 (PTPRC) expression was also assessed in this sample, which returned 16 expressing cells. The expression ranged from 0.1 - 1 and cells were dispersed throughout the keratinocyte clusters (Appendix 5.4). This demonstrates the limitations of FACS, where CD45+ cells are being sorted during a CD45- sort. Fortunately, I can identify these cells and remove them from the analysis. Cell types were then assigned to each cluster using the ‘RenameIdents’ Seurat function. This function renames the identities of each cell in the dataset from cluster ID (0-8) to cell type (Figure 5.3f). In line with similar studies, all four cell types are distinctly clustered from each other, with the keratinocyte clusters demonstrating a more complex cluster pattern (Figure 5.3g; Ji et al. 2020; Solé-Boldo et al. 2020; Wang et al. 2020; Guerrero-juarez et al. 2022; Yerly et al. 2022; Quah et al. 2023).

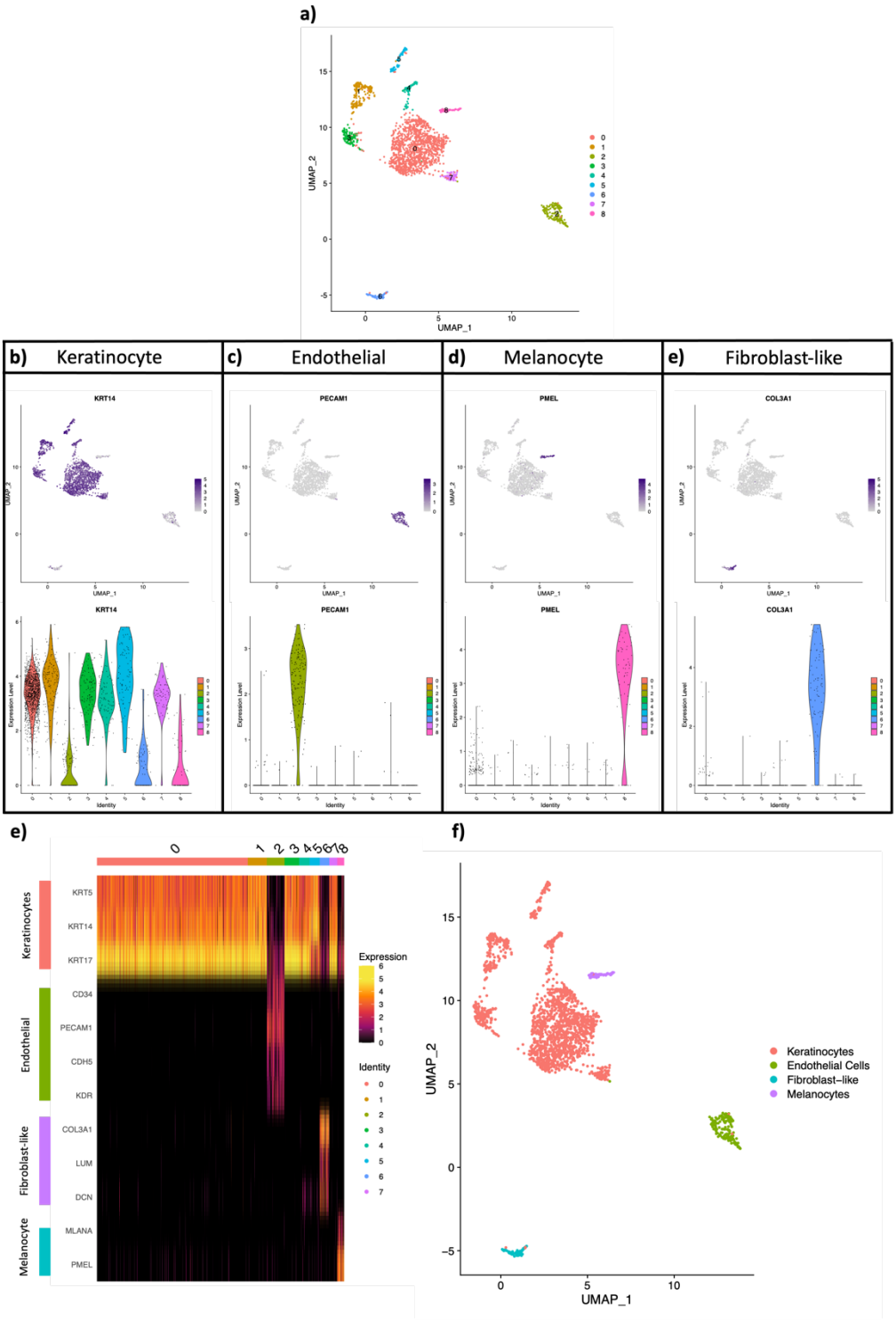

a)
UMAP_2
UMAP_1
b) Keratinocyte
c) Endothelial
d) Melanocyte
e) Fibroblast-like
KRT14
PECAM1
PMEL
COL3A1
Expression Level
Identity
e)
Keratinocytes
Endothelial
Fibroblast-like
Melanocyte
KRT5
KRT14
KRT17
CD34
PECAM1
CDH5
KDR
COL3A1
LUM
DCN
MLANA
PMEL
Expression
Identity
f)
Keratinocytes
Endothelial Cells
Fibroblast−like
Melanocytes

Figure 5.3 | Cell annotation of BCC_1 sample. *Single cell RNA sequencing was performed on the BCC_1 sample. Various markers were obtained from the literature and from the Human Protein Atlas (HPA) and used to annotate the sample.* ***a)*** *UMAP showing the number of Seurat clusters generated for the sample.* ***b-e)*** *UMAPs (top) and Violin plots (bottom) showing expression of KRT14, PECAM1, PMEL, and COL3A1, respectively, showing keratinocytes, endothelial cells, melanocytes, and fibroblast-like cells, respectively.* ***e)*** *A heatmap was generated, which shows marker expression for each cell type in each Seurat cluster.* ***f)*** *The marker analysis confidently identified specific cell types. Each cluster was then annotated for its cell type. Figure shows the annotated UMAP. Expression values represent normalised reads.*

The BCC_2 sample produced a total of five clusters (Figure 5.4a). Clusters 1, 2, and 3 expressed keratinocyte markers KRT5, KRT14, and KRT17 (Figure 5.4b), cluster 0 expressed endothelial markers CD34, PECAM1, CDH5, and KDR (Figure 4.5c), and cluster 4 expressed fibroblast-like marker FABP4 (Figure 5.4e). Whilst there were a small number of cells expressing MLANA and PMEL (melanocyte markers) in cluster 0 (Figure 5.4d), none of these cells highly expressed these markers or formed a distinct subcluster. For the fibroblast-like cluster, the usual markers either showed no expression, or variable expression across the dataset. However, FABP4 clearly identified cluster 4 as fibroblast-like. Enrichment analysis confirmed these findings with g:Profiler HPA terms similar to the BCC_1 sample (Appendix 5.6). Again, cell types were distinctly clustered apart from each other (Figure 5.4g). Interestingly, for this sample, there were more endothelial cells than keratinocytes, implying that enriching for CD200+ CD45- cells also enriched for endothelial cells. There were also fewer keratinocyte clusters for this sample.

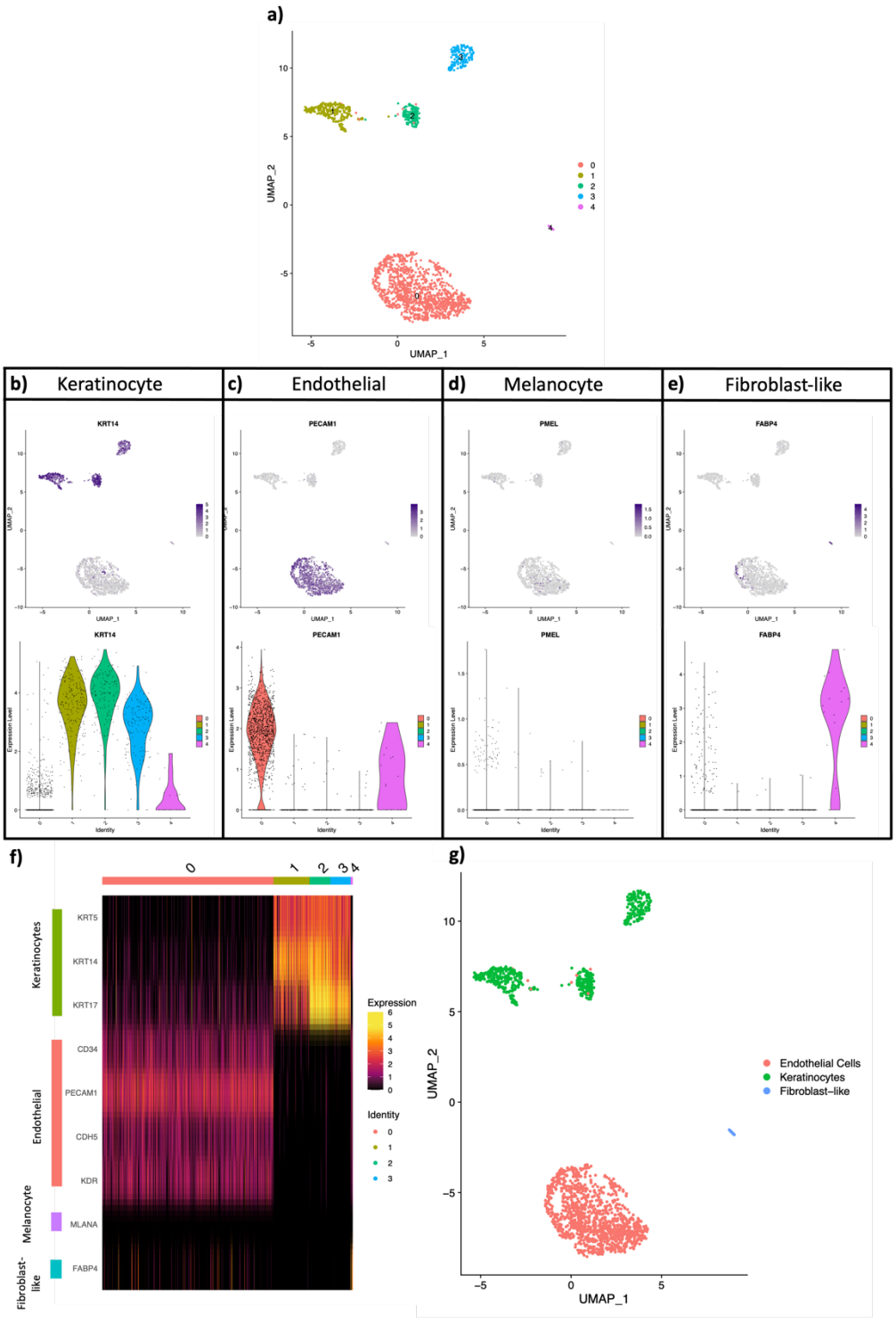

a)
UMAP_2
UMAP_1
0
1
2
3
4
b) Keratinocyte
c) Endothelial
d) Melanocyte
e) Fibroblast-like
KRT14
PECAM1
PMEL
FABP4
Expression Level
Identity
f)
Keratinocytes
Endothelial
Melanocyte
Fibroblast-like
KRT5
KRT14
KRT17
CD34
PECAM1
CDH5
KDR
MLANA
FABP4
Expression
Identity
g)
Endothelial Cells
Keratinocytes
Fibroblast–like

Figure 5.4 | Cell annotation of BCC_2 sample. *Single cell RNA sequencing was performed on the BCC_2 sample. Various markers were obtained from the literature and from the Human Protein Atlas (HPA) and used to annotate the sample.* ***a)*** *UMAP showing the number of Seurat clusters generated for the sample.* ***b-e)*** *UMAPs (top) and Violin plots (bottom) showing expression of KRT14, PECAM1, PMEL, and COL3A1, respectively, showing keratinocytes, endothelial cells, melanocytes, and fibroblast-like cells, respectively.* ***e)*** *A heatmap was generated, which shows marker expression for each cell type in each Seurat cluster.* ***f)*** *The marker analysis confidently identified specific cell types. Each cluster was then annotated for its cell type. Figure shows the annotated UMAP. Expression values represent normalised reads.*

The BCC_2_NS sample initially produced 2 clusters at resolution = 0.1 (Figure 5.5a). This is potentially due to the resolution parameter being too 'granular' coupled with the low number of cells to perform the algorithm on. After inspecting the UMAP, it was clear that the dataset can be further distilled into more clusters through the increasing of the resolution. Keratinocyte and endothelial marker expression further confirmed this as both cell types were shown to belong to cluster 1 (Figure 5.5d and e, respectfully). There was no melanocyte or fibroblast-like cells present in this sample (Figure 5.5f and g, respectfully). The Clustree results were further studied to select an appropriate resolution (Appendix 5.1.2d). Specifically, cluster 1 at resolution 0.1 could be split further. At resolution = 0.4, this cluster is split to form two new clusters, therefore this resolution was imputed and clustering and UMAP was regenerated. Unfortunately, the endothelial cluster failed to separate at this resolution, so resolution = 0.8 was then selected (Figure 5.5b). Again, this resolution failed to isolate the endothelial cell cluster. The Clustree analysis was regenerated with a higher set of resolutions (0.8 – 8). The cluster containing the endothelial cells at resolution = 0.8 (cluster 7) failed to divide until resolution = 7, which produced 35 clusters (Figure 5.5c). At this resolution, the UMAP lost a lot of its structure, and unfortunately failed to isolate the endothelial cells. To truly isolate the endothelial cells, the 'CellSelector' tool could be used to draw around the cluster and manually annotate in that fashion. However, as this is just the optimisation phase, I chose not to do this and to annotate using the resolution = 0.8. Integrating the samples may enhance the separation of these cells, as there will be more endothelial cells in the integrated dataset with similar expression matrices.

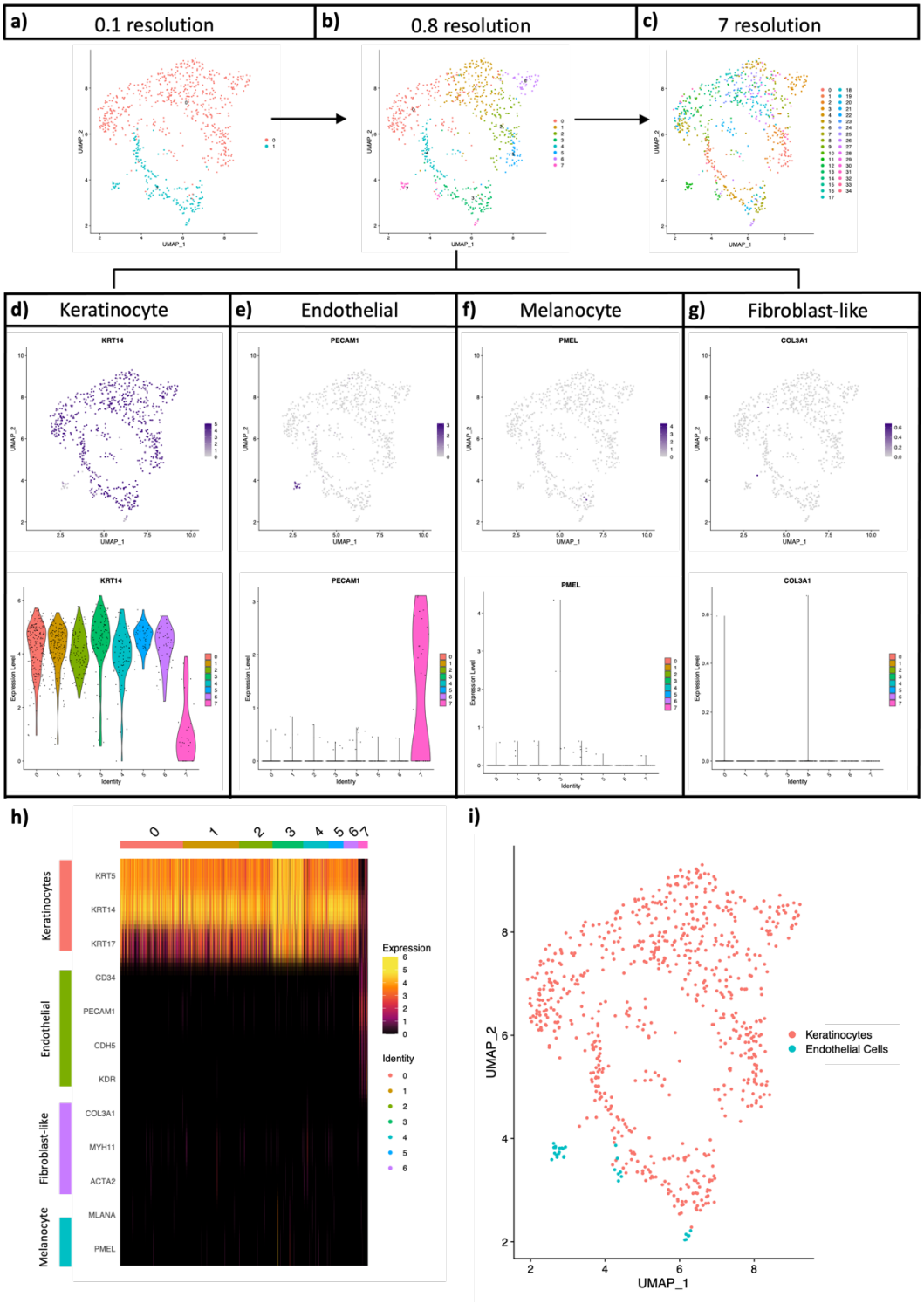

a) 0.1 resolution
b) 0.8 resolution
c) 7 resolution
d) Keratinocyte
e) Endothelial
f) Melanocyte
g) Fibroblast-like
KRT14
PECAM1
PMEL
COL3A1
UMAP_1
UMAP_2
Expression Level
Identity
h)
Keratinocytes
Endothelial
Fibroblast-like
Melanocyte
KRT5
KRT14
KRT17
CD34
PECAM1
CDH5
KDR
COL3A1
MYH11
ACTA2
MLANA
PMEL
Expression
Identity
i)
Keratinocytes
Endothelial Cells

Figure 5.5 | Cell annotation of BCC_2_NS sample. *Single cell RNA sequencing was performed on the BCC_2_NS sample. Various markers were obtained from the literature and from the Human Protein Atlas (HPA) and used to annotate the sample.* ***a-c)*** *UMAPs showing the number of Seurat clusters generated for the sample at clustering resolutions 0.1, 0.8, and 7, respectively. Clustering resolution of 0.8 was used for marker analysis.* ***d-g)*** *UMAPs (top) and Violin plots (bottom) showing expression of KRT14, PECAM1, PMEL, and COL3A1, respectively, showing keratinocytes, endothelial cells, melanocytes, and fibroblast-like cells, respectively.* ***h)*** *A heatmap was generated, which shows marker expression for each cell type in each Seurat cluster.* ***i)*** *The marker analysis confidently identified specific cell types. Each cluster was then annotated for its cell type. Figure shows the annotated UMAP.* ***Expression values represent normalised reads.***

Clustering generated two clusters at resolution = 0.1 for the BCC_3 sample (Figure 5.6a). Cell type marker expression analysis showed the presence of keratinocytes, endothelial, and fibroblast-like cells present in cluster 0 (Figure 5.6c, d, and f, respectfully), indicating that an increase in resolution was needed. There were no melanocytes present in this sample (Figure 5.6e). After inspecting the Clustree for this sample, cluster 0 (at resolution = 0.1) is subdivided into three new, smaller clusters at resolution = 0.4. This resolution was able to separate the keratinocytes and endothelial cells but failed to isolate the fibroblast-like cells. A resolution of 2.5 was sufficient to partially isolate the fibroblast-like cluster, however, increasing the resolution further than 2.5 produced the same effect as the BCC_2_NS sample, therefore resolution = 2.5 was used (Figure 5.6b, g, h). The current analysis protocol was therefore applied to the larger dataset to identify cell populations across all samples.

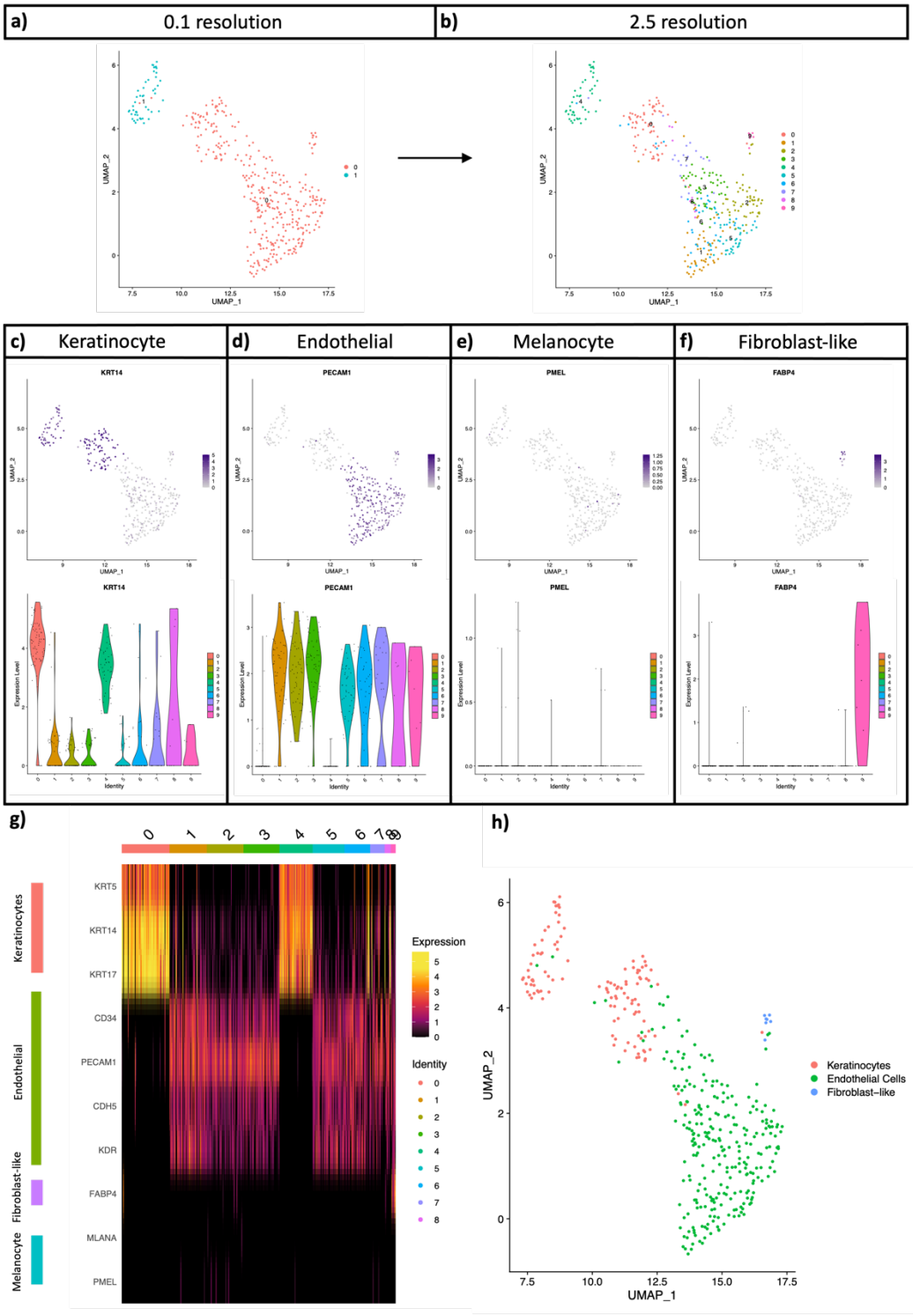

a) 0.1 resolution
b) 2.5 resolution
UMAP_1
UMAP_2
c) Keratinocyte
d) Endothelial
e) Melanocyte
f) Fibroblast-like
KRT14
PECAM1
PMEL
FABP4
Expression Level
Identity
g)
h)
Keratinocytes
Endothelial
Fibroblast-like
Melanocyte
KRT5
KRT14
KRT17
CD34
PECAM1
CDH5
KDR
FABP4
MLANA
PMEL
Expression
Identity
Keratinocytes
Endothelial Cells
Fibroblast-like

Figure 5.6 | Cell annotation of BCC_3 sample. *Single cell RNA sequencing was performed on the BCC_3 sample. Various markers were obtained from the literature and from the Human Protein Atlas (HPA) and used to annotate the sample.* ***a, b)*** *UMAPs showing the number of Seurat clusters generated for the sample at clustering resolutions 0.1 and 2.5, respectively. Clustering resolution of 2.5 was used for marker analysis.* ***c-f)*** *UMAPs (top) and Violin plots (bottom) showing expression of KRT14, PECAM1, PMEL, and COL3A1, respectively, showing keratinocytes, endothelial cells, melanocytes, and fibroblast-like cells, respectively.* ***g)*** *A heatmap was generated, which shows marker expression for each cell type in each Seurat cluster.* ***h)*** *The marker analysis confidently identified specific cell types. Each cluster was then annotated for its cell type. Figure shows the annotated UMAP****.*** *Expression values represent normalised reads.*

5.3 Cell Type Identification Within the Integrated Dataset

Highly variable genes (HVGs) were selected before being integrated using Seurat's v3 integration method (see Appendix 5.7 for integration optimisation). Following integration, the dataset was scaled, and PCA, clustering, and UMAP was performed. Low-resolution clustering was first performed to identify the putative cell populations. A resolution of 0.1 was initially used, which identified eight clusters. Cell type marker analysis placed both melanocytes and endothelial cells within cluster 1, identifying an increase in resolution was needed to correctly separate both populations. Clustree analysis was therefore performed, which revealed that resolution = 0.2 was sufficient to further split cluster 1. Clustering was therefore reperformed at resolution = 0.2, which generated nine clusters, and UMAP was re-run (Figure 5.7a). There appeared to be good integration of each sample across the clusters, however, it was apparent that the BCC_3 sample comprised the majority of cluster 2 (Figure 5.7b). Cell type marker analysis was again performed which displayed correct isolation of cell type clusters (Figure 5.7c-g). On average, 73.56% of cells in each sample accounted for keratinocytes (+/- 18.09 SD), 17.14% accounted for endothelial cells (+/- 21.66 SD), 4.13% accounted for melanocytes (+/- 4.98 SD), and 5.16% accounted for fibroblast-like cells (+/- 5.76 SD). However, looking specifically at the four CD200+ CD45- BCC samples, it was apparent that each sample had also been enriched for endothelial cells, with an average of 43.56% of cells per sample (range 15.46% - 63.73%). This is accompanied by a reduction in keratinocytes, with an average of 54.72% cells per sample (range 34.91% - 80.62%). These findings indicate a possible flaw in the bulk RNA sequencing previously performed, where a sizable population of endothelial cells may indeed mask the signals from the

keratinocyte CD200+ CD45- cancer stem cells. The superior advantage to performing single cell analyses on these samples is the pure isolation of these keratinocytes, allowing identification of true signals emanating from these cancer stem cells. In total, there were 10,024 keratinocytes, 2,180 endothelial cells, 652 melanocytes, and 730 fibroblast-like cells present in the integrated dataset. The 2D UMAP displays a clear separation between the keratinocytes and the other cell types, however, the other three cell types appear to be clustered closer together. These clusters did not fully separate during 3D UMAP analysis. To aid visualisation of the different cell isolations performed via FACS prior to sequencing, a new metadata column was added onto the integrated object, which labelled samples as BCC (CD45-), BCC CD200- (CD200- CD45-), BCC CD200+ (CD200+ CD45-), and NS (Normal Skin, CD45-). This metadata was used to plot the UMAP (Figure 5.7h) and a stacked bar graph was also generated to display the composition of each sample type within each identified cell type (Figure 5.7i).

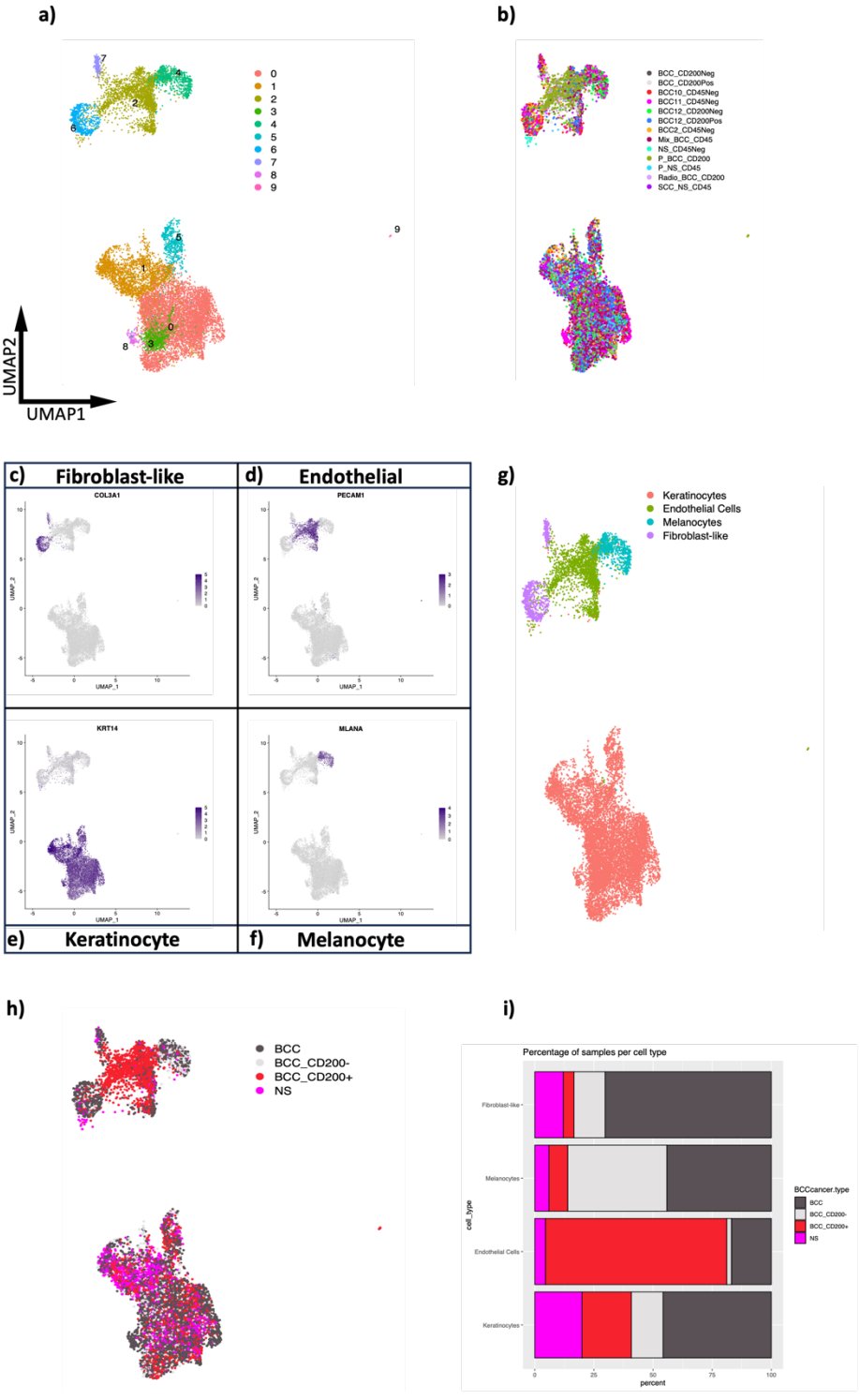

a)
b)
c) Fibroblast-like
d) Endothelial
g)
e) Keratinocyte
f) Melanocyte
h)
i)
UMAP1
UMAP2
0
1
2
3
4
5
6
7
8
9
BCC_CD200Neg
BCC_CD200Pos
BCC10_CD45Neg
BCC11_CD45Neg
BCC12_CD200Neg
BCC12_CD200Pos
BCC2_CD45Neg
Mix_BCC_CD45
NS_CD45Neg
P_BCC_CD200
P_NS_CD45
Radix_BCC_CD200
SCC_NS_CD45
COL3A1
PECAM1
KRT14
MLANA
Keratinocytes
Endothelial Cells
Melanocytes
Fibroblast-like
BCC
BCC_CD200-
BCC_CD200+
NS
Percentage of samples per cell type
cell_type
percent
BCCcancer.type

Figure 5.7 |Integration, clustering, and cell annotation of 13 single cell samples. *13 single cell samples were integrated using Seurat's v3 integration pipeline. Seurat clustering was performed using a resolution of 0.2 and UMAP was subsequently performed.* ***a)*** *UMAP showing number and distribution of Seurat generated clusters.* ***b)*** *UMAP showing sample distribution across Seurat clusters.* ***c-f)*** *UMAP's showing expression of COL3A1, PECAM1, KRT14, and MLANA which depict cell type identities of fibroblast-like, endothelial, keratinocyte, and melanocyte, respectively.* ***g)*** *The marker analysis confidently identified specific cell types. Each cluster was then annotated for its cell type. Panel shows cell type annotated UMAP. Cell metadata was used to further annotate the dataset into samples and cells that were Basal Cell Carcinoma (BCC) or Normal Skin (NS) and CD45- (BCC and NS), CD200- CD45- (BCC CD200-), and CD200+ CD45-(BCC CD200+). The new metadata was called BCC Cancer Type.* ***h)*** *UMAP showing BCC Cancer Type metadata overlay.* ***i)*** *Stacked bar graph of the BCC Cancer Type metadata. Graph shows the cell type on the Y-axis and the percentage of each sample on the X-axis. Expression values in UMAPs represent normalised reads.*

5.4 Chromosomal Analysis of BCC Cells

Next, putative tumour cells were identified in each individual sample by analysing the chromosomal landscape of inferred copy number variations using inferCNV (Tickle et al. 2019). Aberrant genomic profiles were observed purely in the keratinocyte clusters when compared to their counterpart internal reference cells in all samples (melanocytes, endothelial cells, and fibroblast-like cells). These profiles are associated with chromosomal deletions (blue) or duplications (red) (Figure 5.8, with background reduction. Appendix 5.8, without background reduction). Some tumour keratinocytes displayed substantial genomic changes across whole chromosomes (as depicted by a colour bar spanning the whole width of the chromosome), whereas others displayed very little genomic changes and had a similar CNV profile to the normal skin samples (as depicted by a coloured bar spanning only a part of the chromosome; BCC_6 chromosome 15, BCC_1 chromosome 6, BCC_8+ chromosome 15 vs BCC_5, BCC_4-, respectively). This indicates that not all BCCs have CNVs that drive tumour growth, an observation similar to those seen in a similar study (Guerrero-juarez et al. 2022). Furthermore, no BCC sample contained genomic alterations that spanned the whole of the keratinocyte population (as depicted by partial-coloured bars along the y axis/rows of a particular chromosome. For example, for the BCC_8+ sample, for chromosome 15, there is a section of cells/rows that do not have the chromosomal deletion; Appendix 5.8), suggesting the presence of normal keratinocytes within these BCC samples. To confirm this, known BCC markers (BCAM and EPCAM) were visualised in the integrated dataset via UMAP (Figure 5.9c, e, g). PTCH1 and EPCAM were strongly and solely expressed within the keratinocyte clusters, whereas BCAM was expressed strongly in both the keratinocyte and

endothelial cell clusters. Looking more specifically at the keratinocyte clusters, and subsequently subsetting them (Figure 5.9b), the BCC markers were expressed in clusters 0, 3, 5, and 8, but not in cluster 1 (Figure 5.9d, f, h). Interestingly, CD200 was strongly expressed within the endothelial clusters (Figure 5.9i), which was not expected, and in a subset of the keratinocyte clusters (Figure 5.9j).

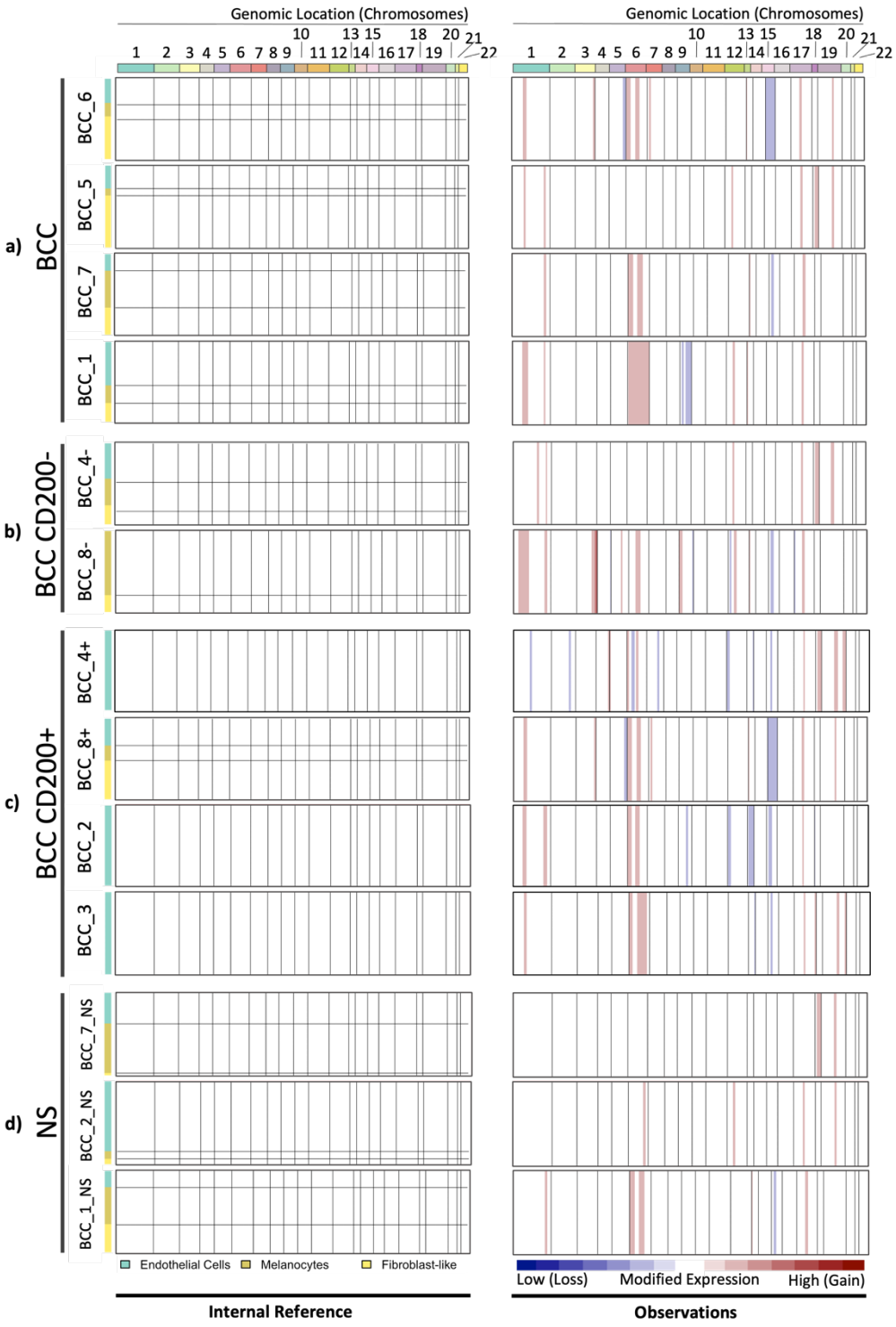

Genomic Location (Chromosomes)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
a) BCC
BCC_6
BCC_5
BCC_7
BCC_1
b) BCC CD200-
BCC_4-
BCC_8-
c) BCC CD200+
BCC_4+
BCC_8+
BCC_2
BCC_3
d) NS
BCC_7_NS
BCC_2_NS
BCC_1_NS
Endothelial Cells
Melanocytes
Fibroblast-like
Internal Reference
Low (Loss)
Modified Expression
High (Gain)
Observations

Figure 5.8 |Chromosomal landscape of inferred copy number variations and identification of putative tumour cells in single cell samples. *inferCNV analysis was performed on the keratinocyte populations of 13 single cell samples by comparing their genomic profiles to their internal reference counterparts: endothelial cells, melanocytes, and fibroblast-like cells. Chromosomal deletions are depicted as blue bars and chromosomal duplications are depicted by red bars.* ***Left,*** *chromosomal aberrations of the internal reference cells.* ***Right,*** *chromosomal aberrations of the keratinocytes (observations).* ***a)*** *genomic profiles for BCC samples (CD45-),* ***b)*** *BCC CD200- samples (CD200- CD45-),* ***c)*** *BCC CD200+ samples (CD200+ CD45-),* ***d)*** *normal skin samples (CD45-).*

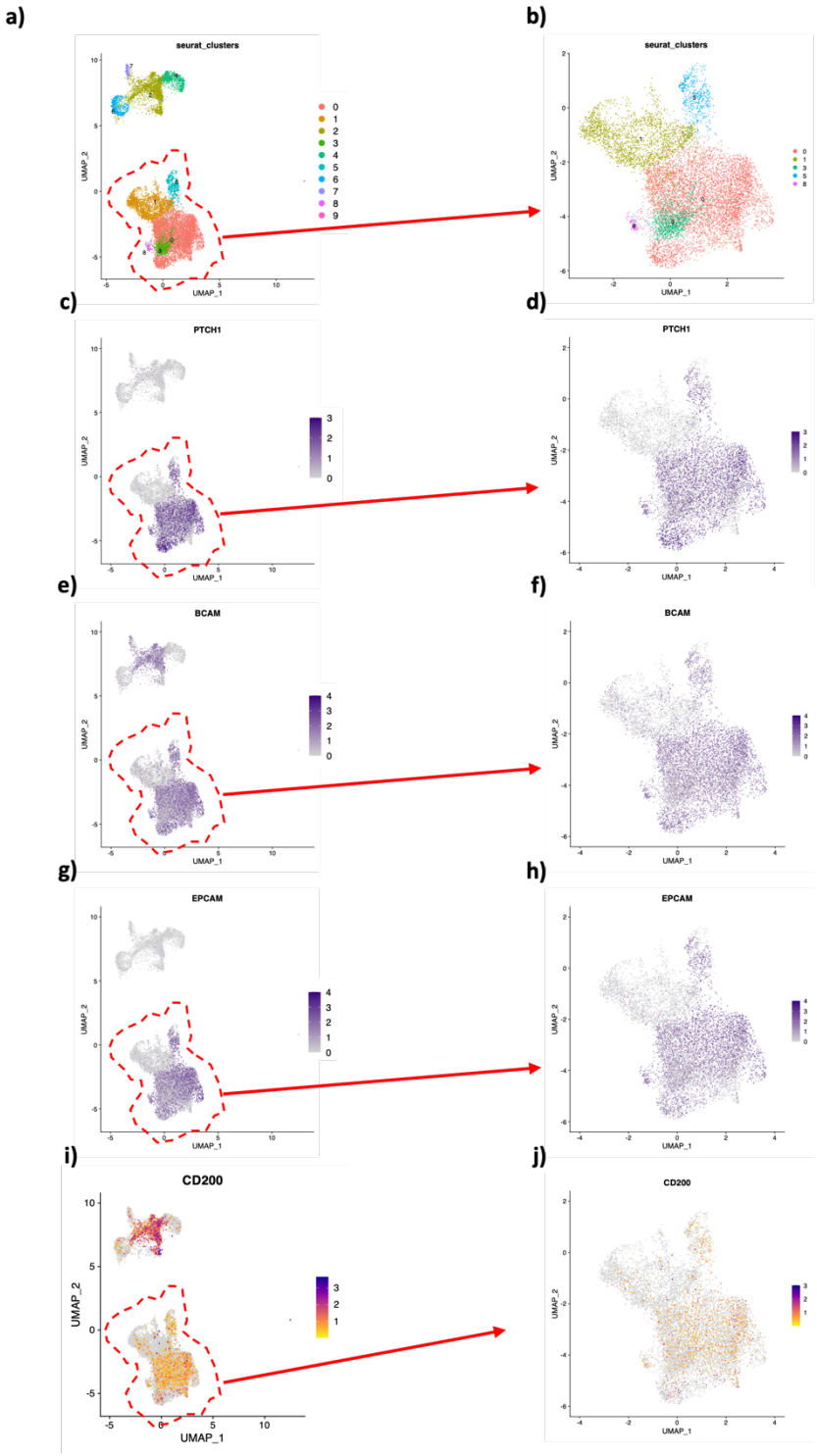

a)
b)
seurat_clusters
0
1
2
3
4
5
6
7
8
9
UMAP_1
UMAP_2
c)
d)
PTCH1
e)
f)
BCAM
g)
h)
EPCAM
i)
j)
CD200

Figure 5.9 |Identification of Basal Cell Carcinoma (BCC) cells. *Single cell RNA sequencing was performed on 13 BCC and normal skin samples. Following clustering and UMAP, known BCC markers were visualised in the integrated dataset.* ***a, c, e, g, i)*** *UMAPs showing the full dataset. The dataset was also subsetted to contain only keratinocytes.* ***b, d, f, h, j)*** *UMAPs showing the keratinocyte subsetted dataset. BCC markers were visualized in the full and keratinocyte subsetted datasets.* ***a, b)*** *UMAP of the full dataset showing 10 clusters, and the keratinocyte subsetted dataset showing five clusters, respectively.* ***c, e, g, i)*** *UMAPs of the full dataset showing expression of PTCH1, BCAM, EPCAM, and CD200, respectively.* ***d, f, h, j)*** *UMAPs of keratinocyte subsetted datasets showing expression of PTCH1, BCAM, EPCAM, and CD200, respectively. Keratinocyte clusters are circled in red with an arrow pointing to the keratinocyte subsetted dataset UMAP. Expression values in UMAPs represent normalised reads.*

5.5 Re-clustering of BCC Cells

To further investigate the BCC keratinocyte populations, the keratinocyte-only dataset was re-clustered at resolution = 1, which resulted in 17 clusters (Figure 5. 10a). Investigating sample composition of each cluster revealed clusters 14 and 16 comprised purely of keratinocytes from the BCC samples, and clusters 6 and 11 comprising mostly of keratinocytes from BCC samples with less than 5% of cells belonging to normal skin samples (Figure 5.10b). To identify and confirm the presence of BCC tumour clusters, BCC markers (Guerrero-juarez et al. 2022; Yerly et al. 2022) and Hedgehog signalling pathway and Basal Cell Carcinoma KEGG gene set pathways (from Molecular Signatures Database) were visualised. BCC marker analysis showed that BCC cells resided on the left two thirds of the UMAP (clusters 0, 1, 3, 5, 6, 7, 8, 11, 13, 14, and 16, roughly indicated by dotted line) (Figure 5.10c-h). These observations were consistent when overlaying the KEGG BCC and Hedgehog gene sets, which highlighted the same clusters (Figure 5.10i, j, respectively). Interestingly, cluster 6 showed increased expression of BCC markers and increased enrichment of the BCC and Hedgehog gene sets when compared to the other BCC clusters. CD200 also showed expression within the left two-thirds of the clusters (Figure 5.10k). A density plot of CD200 expression was generated, which revealed enrichment of CD200 at clusters 0, 1, 5,7, and 11 (Figure 5.10l).

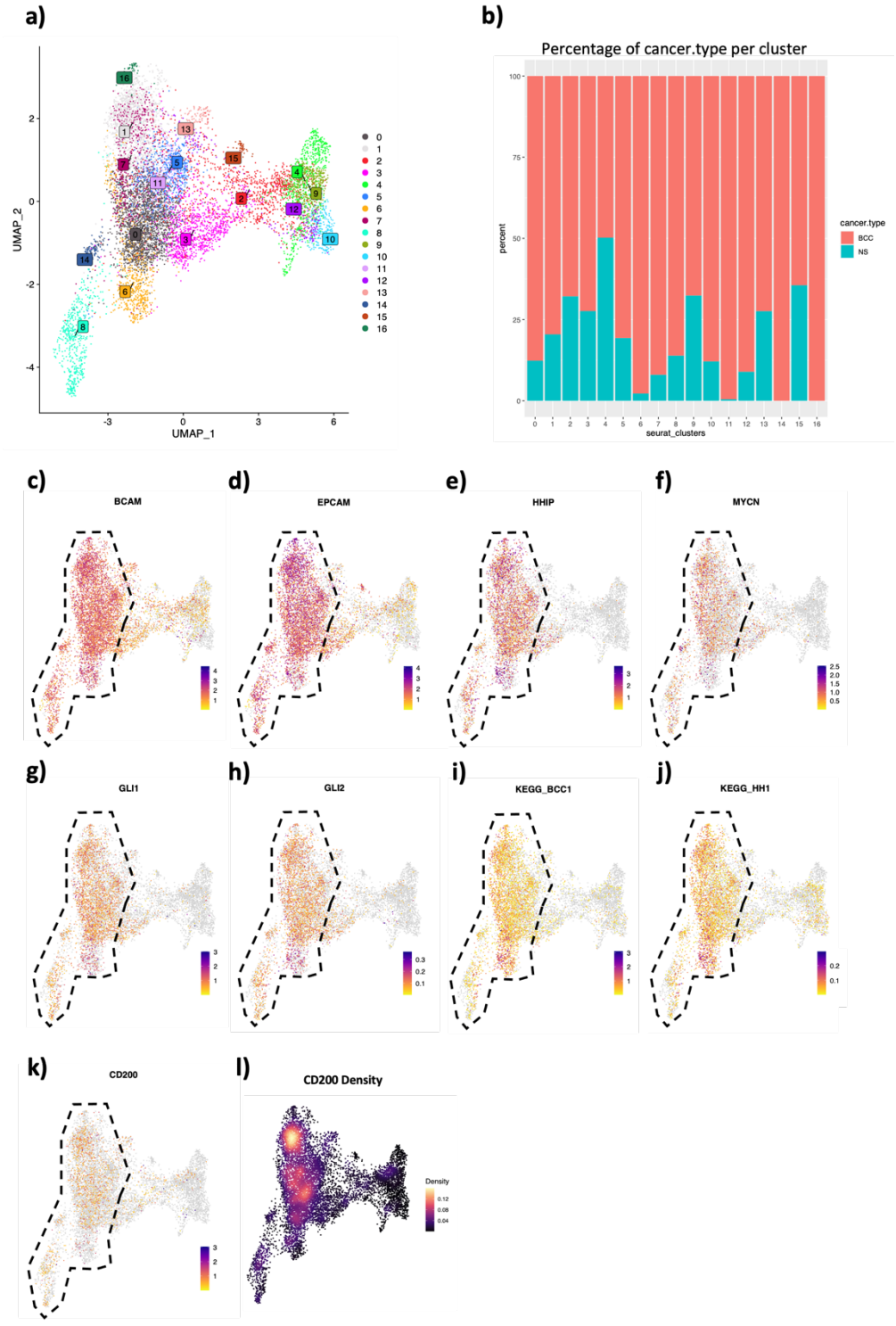

a)
b)
Percentage of cancer.type per cluster
UMAP_1
UMAP_2
percent
seurat_clusters
cancer.type
BCC
NS
c)
BCAM
d)
EPCAM
e)
HHIP
f)
MYCN
g)
GLI1
h)
GLI2
i)
KEGG_BCC1
j)
KEGG_HH1
k)
CD200
l)
CD200 Density
Density

Figure 5.10 |Identification of Basal Cell Carcinoma (BCC) cells within keratinocyte subpopulations. a) *Keratinocytes were subsetted from an integrated single cell dataset, and re-clustered using Seurat's clustering method with a resolution of 1. UMAP shows 17 clusters.* ***b)*** *stacked bar graph showing the percentage of cancer type (Y-axis) belonging to each Seurat cluster (X-axis).* ***c-h)*** *UMAP's showing expression of known BCC markers, BCAM, EPCAM, HHIP, MYCN, GLI1, and GLI2, respectively, within the keratinocyte clusters.* ***i, j)*** *UMAP's showing KEGG BCC and KEGG Hedgehog pathway expression overlaid, respectively. Seurat's 'AddModuleScore' function was used to overlay the pathway onto the dataset, which generates an expression value of the whole pathway based on the gene expression.* ***k)*** *UMAP showing CD200 expression within the keratinocyte clusters. Dotted outlines depict the left two-thirds of the UMAP where BCC marker and CD200 marker expression is greatest.* ***l)*** *UMAP showing density of CD200 expression. Expression values in* ***c-h*** *and* ***k*** *represent normalised reads.*

I next split the UMAP based on CD200 marker status (CD200- CD45- and CD200+ CD45-) and assessed the expression of the BCC markers and gene sets. GLI2, EPCAM, HHIP, BCAM, MYCN, KEGG BCC, KEGG Hedgehog signalling, and CD200 visibly showed greater expression in the CD200+ Basal/BCC clusters when compared to CD200- Basal/BCC clusters, whereas PTCH1, GLI1, and GLI2 was difficult to determine (Figure 5.11c-j and a-b, respectively). However, it was apparent that there were more cells present in the CD200+ dataset, making marker expression comparisons difficult. These differences were corrected by calculating the percentage of cells expressing the marker gene in each cluster for each sample group (Section 2.12.2.13.1). Calculating the percentages confirmed the initial findings, with 20.8% of cells expressing GLI2 (compared to 8%), 42.1% cells expressing HHIP (compared to 20.3%), 8.26 cells expressing MYCN (compared to 3.5%), 63.8% cells expressing EPCAM (compared to 31.6%), and 73.8% cells expressing BCAM (compared to 65.9%) in the CD200+ group when compared to the CD200- group. CD200+ samples also displayed greater enrichment for KEGG BCC and HH gene sets, which appeared to have a concentrated expression at the top and bottom of the Basal/BCC clusters (clusters 1, 7, 16 and cluster 6, respectively). This observation is also reflected in the BCC marker expression.

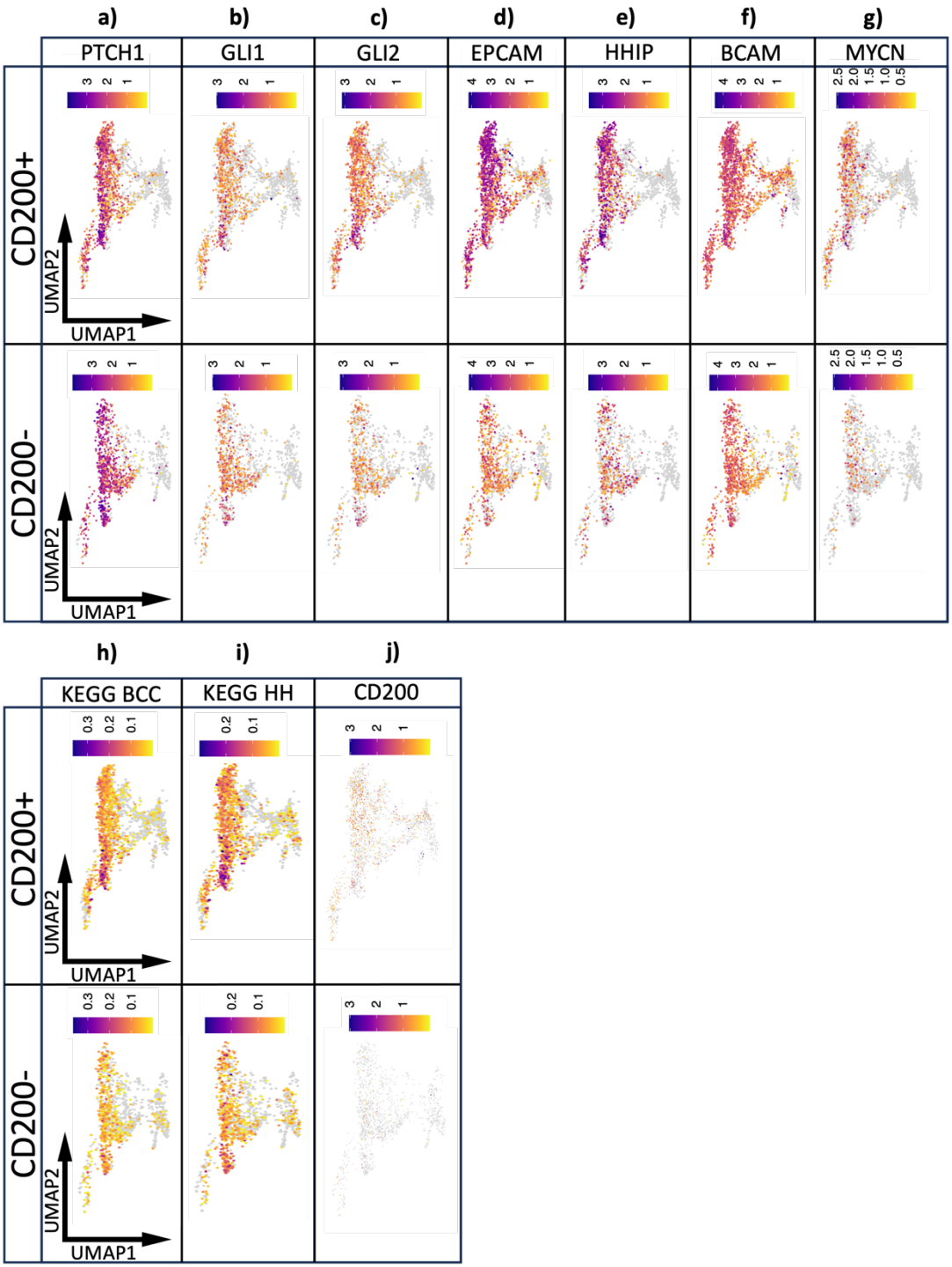

a)
b)
c)
d)
e)
f)
g)
PTCH1
GLI1
GLI2
EPCAM
HHIP
BCAM
MYCN
CD200+
CD200-
UMAP2
UMAP1
h)
i)
j)
KEGG BCC
KEGG HH
CD200

Figure 5.11 |Assessment of Basal Cell Carcinoma (BCC) markers between CD200- and CD200+ sorted populations. *Keratinocytes were subsetted from an integrated single cell dataset and re-clustered using Seurat's clustering method with a resolution of 1. The dataset was then further subset into BCC CD200- and CD200+ sorted samples and known BCC markers were assessed.* ***a-g)*** *UMAP's showing expression of PTCH1, GLI1, GLI2, EPCAM, HHIP, BCAM, and MYCN, respectively.* ***h-i)*** *UMAP's showing KEGG BCC and KEGG Hedgehog pathway overall expression. Seurat's 'AddModuleScore' function was used to overlay the pathway onto the dataset, which generates an expression value for the whole pathway based on the gene expression.* ***j)*** *UMAP showing CD200 expression. BCC CD200+ samples are the top UMAP's, CD200- samples are the bottom UMAP's. Expression values in* ***a-g*** *represent normalised reads.*

5.6 RNA Velocity of BCC Cells

Using a stochastic model, an embedding stream was generated for the normal skin samples, which displays the RNA velocities of the cells (Figure 5.12a). The RNA velocities, represented in this analysis, define development relationships and differentiation, which are vectors that represent the direction and speed of movement of cells. The embedding stream revealed that the velocity vectors are rooted at cluster one (top cluster, light blue), and terminate toward cluster six (bottom cluster, dark blue). The rooted cluster one vectors also appear to migrate toward an intermediate state at clusters twelve and thirteen (light and dark green clusters, respectively) before terminating at cluster four (right cluster, light red). To next identify the major transition states between clusters, a velocity partition-based graph abstraction (PAGA) analysis was performed (Figure 5.12b). This algorithm connects the cell clusters based on major transitions, with each edge (arrow) representing a confident probability of one cluster transitioning to the next. This analysis partially corroborated with the embedding stream, confirming the progression of cells from cluster one to cluster six. The direct progression from cluster one to cluster four was not confidently confirmed by the analysis, as indicated by faint, dashed lines. However, the indirect progression was confidently confirmed through intermediate clusters zero, six, and two. A confident progression from cluster six to cluster three and a non-confident progression from cluster three to toward cluster four was also seen. We also see that cluster eight progresses toward cluster one and seven, and toward cluster two.

In keeping with the stem cell hypothesis, it appears that clusters one and seven are a progenitor population which asymmetrically divide to give rise to two different populations: transient amplifying cells, as evidenced by velocity vectors terminating toward cluster six and PAGA analysis showing transition from cluster six towards the clusters on the right hand side, and differentiating cells, as evidenced by velocity vectors terminating toward clusters nine and ten. Cluster zero appears to be in a mid-transitional state toward the transient amplifying cells in cluster six. The clusters on the right-hand side of the UMAP appear to be in varying differentiated states, perhaps fated to form the various cell populations within the skin. Regardless of state, these clusters are either derived from the transient amplifying population (cluster six) or the progenitor population (clusters one and seven). We also see a similar progenitor behaviour from the proliferating cells in cluster eight, where they asymmetrically divide, producing cells belonging to clusters one and/or seven, and cells belonging to cluster two.

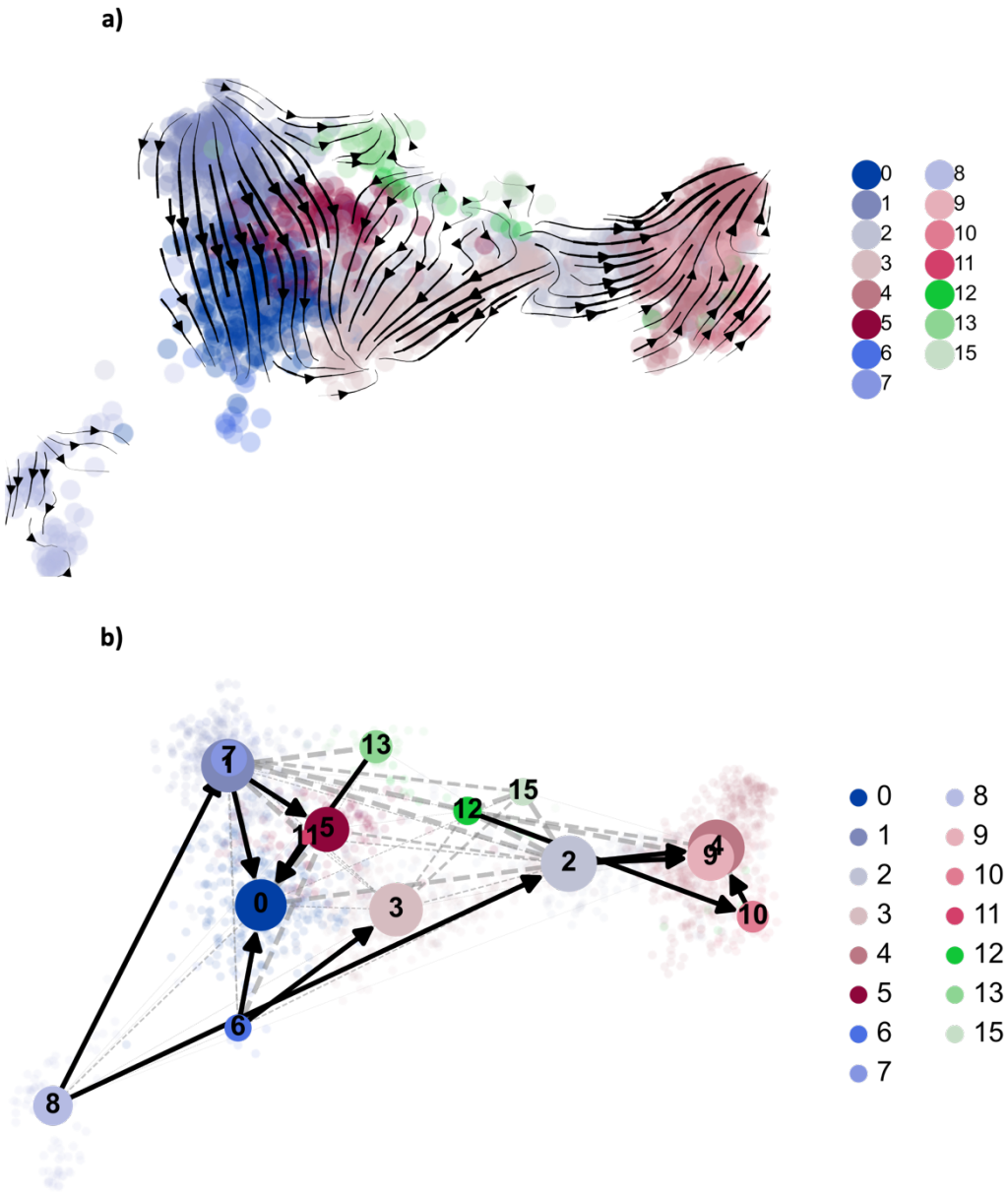


Figure 5.12 |RNA velocity of normal skin keratinocytes. *scVelo was used to project the RNA velocity of normal skin keratinocytes. Keratinocytes were first subsetted from the full integrated dataset and re-clustered using Seurat's clustering method with a*

resolution of 1, and then normal skin keratinocytes were subsetted. ***a)*** *Embedding stream displaying the cell velocities. Velocities represent the speed and direction of the cells. Arrows depict the direction of the velocities.* ***b)*** *Velocity partition-based graph abstraction (PAGA) analysis showing major transition states between clusters. Arrows represent confident transitions from one cluster to another.*

To observe BCC cellular dynamics, I again performed the same analyses on the BCC keratinocytes (Figure 5.13a). For the BCC keratinocytes, we see a similar, but more defined embedding stream, with velocity vectors rooting from the top-right of cluster one and terminating at cluster six and cluster four, again indicating cells of cluster one as a progenitor population. Interestingly, when looking at the PAGA graph, we see different cellular dynamics within the clusters on the right-hand side of the UMAP for the BCC keratinocytes (Figure 5.13b). There is no confident linkage between the apparent progenitor or transient amplifying clusters and the clusters on the right-hand side of the UMAP, however we do see confident links from the far-right clusters on the UMAP (cluster four, nine, ten, and twelve) to the various other clusters on the right-hand side (cluster two, for example). This is similar to the normal skin keratinocytes, and again is possibly where the cells of these clusters are travelling down various differentiation trajectories to recapitulate various cell types of the skin. There is a clear connection between cluster eight and the rest of the clusters for the BCC sample keratinocytes, again possibly because there is a greater number of cells to form the linkage. In this case, we see cluster eight showing asymmetrical divisional traits to form cells of two different clusters. Both analyses show clusters one and six as root and terminal vectors, which indicate progenitor and transient amplifying clusters, respectively.

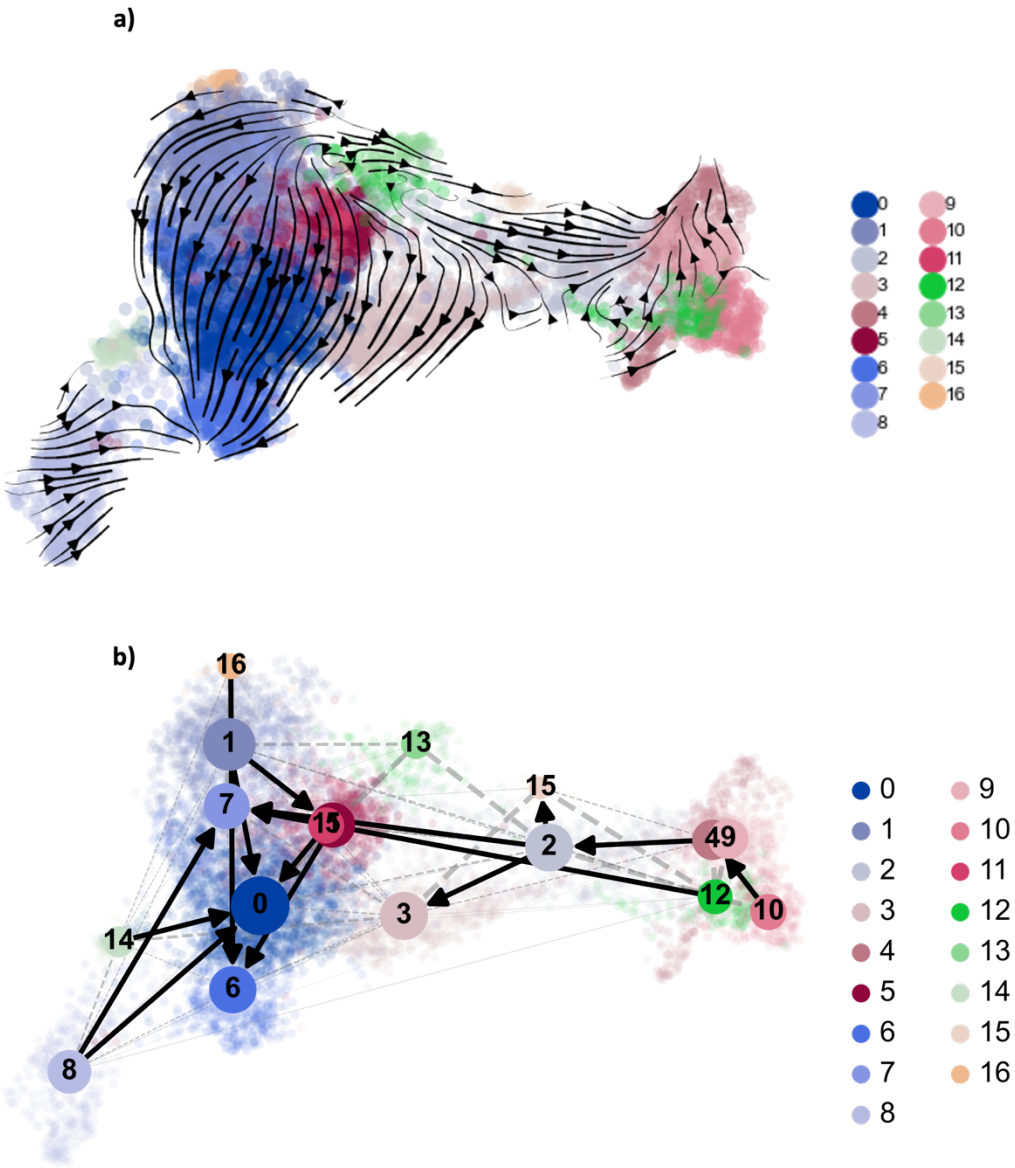


Figure 5.13 |RNA velocity of Basal Cell Carcinoma (BCC) keratinocytes. *scVelo was used to project the RNA velocity of BCC keratinocytes. Keratinocytes were first subsetted from the full integrated dataset and re-clustered using Seurat's clustering*

method with a resolution of 1, and then BCC keratinocytes were subsetted. ***a)*** *embedding stream displaying the cell velocities. Velocities represent the speed and direction of the cells. Arrows depict the direction of the velocities.* ***b)*** *Velocity partition-based graph abstraction (PAGA) analysis showing major transition states between clusters. Arrows represent confident transitions from one cluster to another.*

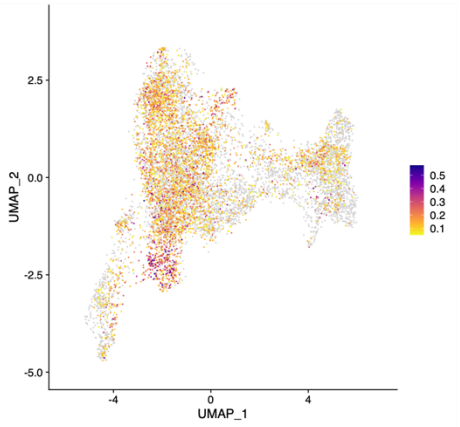
5.7 Stem-like Cluster Identification

Combining the previous analyses with the observed increase in BCC marker expression in clusters one and six, I next hypothesised an increase in stemness in these clusters. To investigate this, literature and online databases were explored for lists of stem-like markers, and then visualised via UMAP. Firstly, the molecular signatures database was searched for relevant gene sets which returned one suitable list: MALTA_CURATED_STEMNESS_MARKERS (M30411). This gene set was curated from the literature by Malta and colleagues in 2018 and includes a collection of genes that mark normal and cancer stem cells (Malta et al. 2018). Secondly, the CellMarker2.0 database identified a list of 32 genes associated with stem cells and cancer stem cells of the epidermis and hair follicle. Lastly, two lists of differentially expressed genes were obtained from Ali and colleagues in 2021, who performed microarray analysis on human epidermal keratinocytes from neonatal foreskin tissues (Ali et al. 2021). The first list was a list of DEGs uniquely upregulated in keratinocyte stem cells (KSCs), and the second list was upregulated DEGs uniquely in transient amplifying cells. These gene lists were overlaid onto the combined UMAP (Figure 5.14a-d).

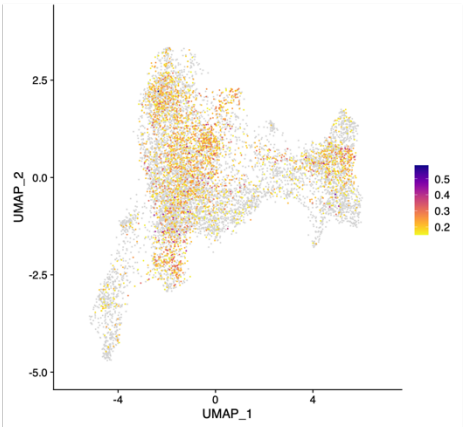
Visualising the CellMarker2.0 list of stem and cancer stem cell-related genes revealed widespread expression, with visibly heightened expression at clusters one, six, and seven (Figure 5.14a). Similarly, when visualising the KSC-related genes from Ali and colleagues, we see a similar pattern, with expression mainly focused on the central BCC clusters and a heightened expression at clusters one, six, and seven (Figure 5.14b). Interestingly, when visualising the stem and cancer stem cell-related genes

from Malta and colleagues, we see that expression is focused on the BCC and differentiating clusters but is absent or very low at clusters one and seven (Figure 5.14c). When observing the expression of the transient amplifying markers from Ali and colleagues, we see a similar expression pattern in the BCC clusters to the Malta and colleagues' markers (Figure 5.14d). If we just take the KSC and transient amplifying markers into consideration from Ali and colleagues, we see that both the top and bottom of the BCC clusters (clusters one, six, and seven) show expression of the KSC markers, which would support the theory that these clusters are progenitor and transient amplifying populations that retain stem-like properties. However, only cluster six shows expression of the transient amplifying cell markers. Coupling this with the previous velocity analyses, it appears that as cells transition from clusters one and seven, the apparent progenitor population, they steadily gain expression of the transient amplifying marker genes, supporting the hypothesis that cluster six is a population of transient amplifying cells. Adding to the complexity of these findings, perhaps the markers identified by Malta and colleagues are those that best represent a 'universal' stem and cancer stem cell, which is what the transient amplifying cells, and not the progenitor cells, show characteristics of.

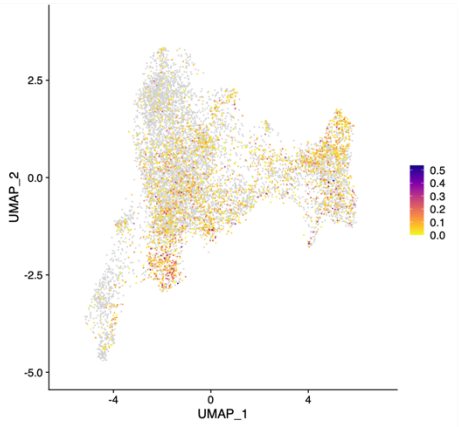
a) CellMarker 2.0 stem & cancer stem cell-related genes



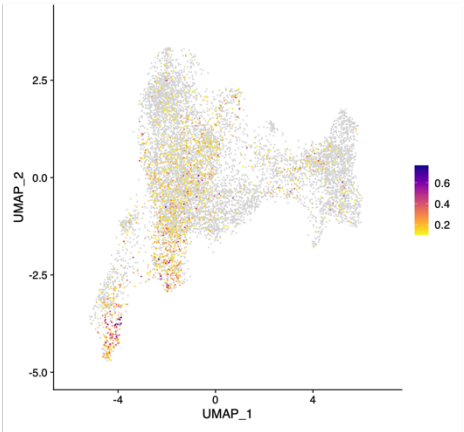
b) Keratinocyte Stem Cell-related genes



c) Malta stem & cancer stem cell-related genes



d) Transient amplifying markers



e)

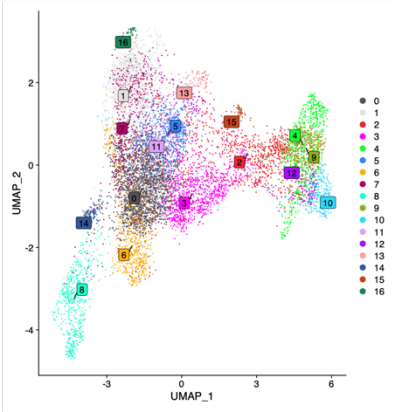


Figure 5.14 |Visualisation of stemness-related genes in Basal Cell Carcinoma (BCC) keratinocytes. *Literature and databases were used to obtain four stemness-related gene sets. These gene sets were then used to generate a gene set expression score which was then visualised via UMAP. Seurat's 'AddModuleScore' function was used to overlay the pathway onto the dataset, which generates an expression value for the whole pathway based on the gene expression.* ***a)*** *UMAP showing CellMarker2.0 database's stem and cancer stem cell related-genes expression score.* ***b)*** *UMAP showing Keratinocyte Stem Cell (KSC)-related genes expression score from Ali and colleagues (Ali et al. 2021).* ***c)*** *UMAP showing stem and cancer stem cell-related genes expression score from the molecular signatures database (M30411).* ***d)*** *UMAP showing transient amplifying markers gene expression score from Ali and colleagues (Ali et al. 2021).* ***e)*** *UMAP showing number of Seurat clusters, for reference.*

5.8 CD200+ Cancer Stem Cell Cluster Identification

I next hypothesised that there would be a CD200-expressing population residing within these progenitor and transient amplifying cell populations. As aforementioned, plotting CD200 on the UMAP revealed heterogeneous expression throughout the BCC clusters, with a focused cluster of higher-expressing cells residing in clusters one and seven (Figure 5.15a). A small number of cells with the highest CD200 expression in a small portion of cells residing in cluster six is also seen. Interestingly, when plotting a phase portrait for the CD200 gene with scVelo, we see that the gene has increased velocity in clusters one, six, and seven (Figure 5.15b). This analysis examines the genes dynamics by interpreting spliced and unspliced transcripts of that gene. A positive velocity here indicates an upregulation of the gene as there are a higher amount of unspliced transcripts which indicates transcriptional induction.

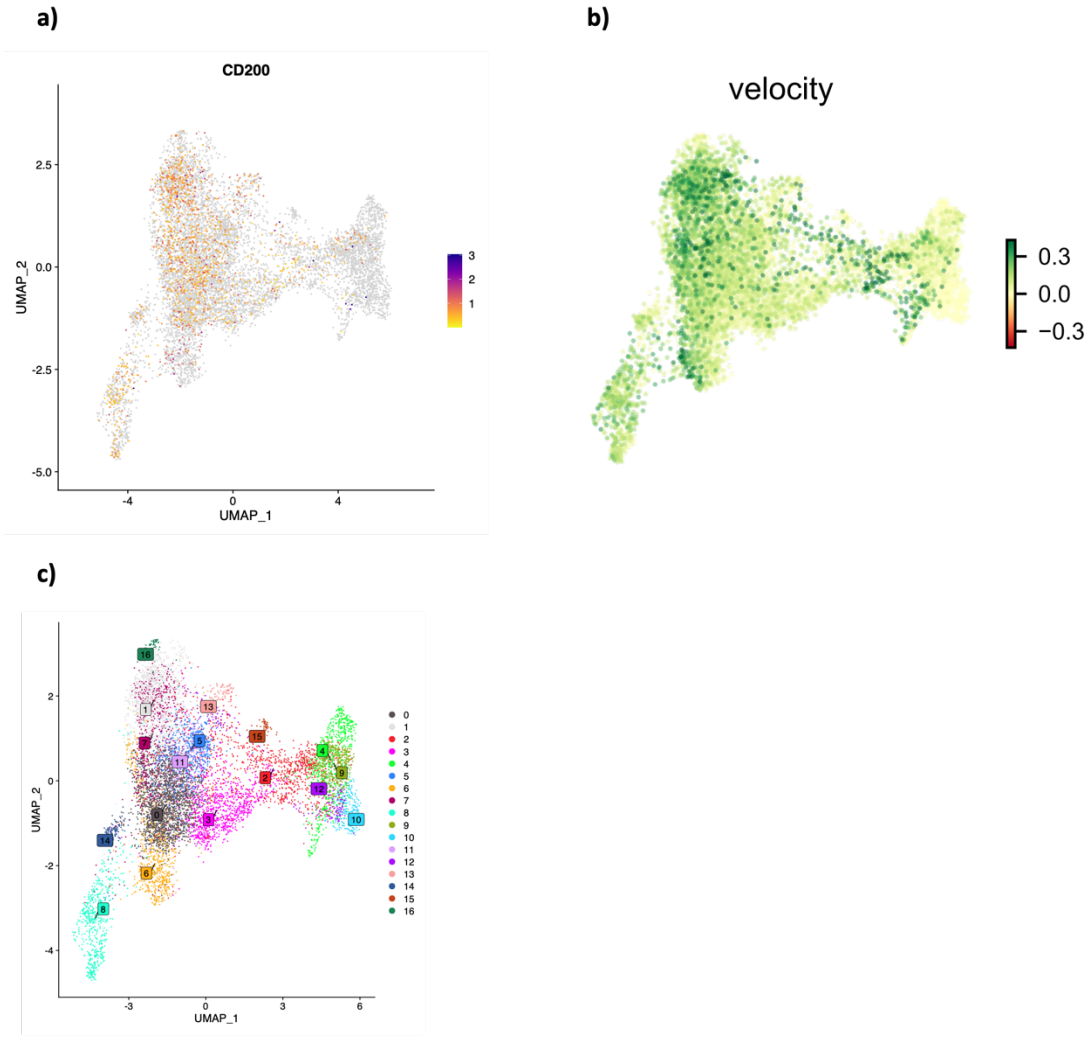


Figure 5.15 |Assessment of CD200 expression in Basal Cell Carcinoma (BCC) keratinocytes. *Keratinocytes were subsetted from an integrated dataset and re-clustering was performed using Seurat's clustering method with a resolution of 1.* ***a)*** *UMAP showing CD200 expression. Expression represents normalised reads.* ***b)*** *Phase portrait showing velocity of CD200 using scVelo. Expression scale represents velocity. Positive velocity indicates an upregulation of the gene due to higher amounts of unspliced transcripts.* ***c)*** *UMAP showing Seurat clusters, for reference.*

5.9 Characterisation of BCC Cancer Stem Cell/Stem-like Populations

I next wanted to functionally characterise the clusters and identify any aberrant signalling pathways between them. To do this, I used single cell gene set enrichment analysis (scGSEA), which provides enrichment information at the cluster level. DEGs were first generated for each cluster, which were then used as input into the 'singleseqgset' package in R (Cillo 2021). An additional benefit of this analysis is that it also uses hierarchical clustering to group the clusters based on functional similarity (see Appendix 5.9 for all scGSEA results). Whilst a common enrichment theme did not immediately present itself across the scGSEA results, the most notable finding was the common primary segregation in the hierarchical clustering of the clusters for all gene set databases. The same clusters were being primarily separated. A summarisation of the primary segregation from the hierarchical clustering for each gene set database can be found in Figure 5.16a. Using this table, I was able to identify modules of functionally related clusters. I first looked for primary hierarchical clustering segregation that was in agreement across ≥ 5 of the gene set databases. The first obvious module encompassed clusters 2, 3, 4, 9, 10, 12, and 15 (Figure 5.16a – blue) was generated and highlighted via UMAP (Figure 5.16c). Interestingly, these clusters resided at the right-hand side of the UMAP and were enriched for gene sets associated with differentiating cells. This is also in agreement with the direction of flow determined by the RNA velocity analyses. Given this functional module and the findings from the RNA velocity analyses, I hypothesised that these clusters were differentiating keratinocyte populations. To investigate this, I visualised known

keratinocyte differentiation markers via UMAP (Figure 5.16d-f). Marker analysis revealed that the far-right clusters (2, 3, 4, 9, 10, 12, and 15) were strongly enriched for these markers, indicating that these clusters are indeed differentiating keratinocytes. Given this data, and the results from the RNA velocity analyses, I then hypothesised that cluster 8 (bottom left of UMAP) was a population of proliferating keratinocytes. scGSEA analysis of this cluster revealed strong enrichment of proliferation-associated processes such as cell cycle, DNA replication, and G2M checkpoint. I then visualised expression of well-known markers of proliferation, MKI67, TOP2A, and UBE2C (Figure 5.16g-i) as well as visualised the annotation from the previously performed cell cycle scoring (Figure 5.16j), which confirmed cluster 8 as a population of proliferating keratinocytes. I next looked at clusters that were hierarchically clustered apart from the first module during the primary segregation (Figure 5.16a - grey). The most closely related clusters were 0, 5, 6, 14, and 16 (in agreement with ≥ 4 gene set databases) which populated primarily the middle-bottom portion of the central section of the UMAP (except for cluster 16 which resides at the middle-top; Figure 5.17a). When investigating the secondary hierarchical clustering, I saw that clusters 1, 7, and 11 were also closely related (Appendix 5.9 and Figure 5.17b). Intriguingly, these clusters express the highest levels of CD200 (Figure 5.17c). I have therefore shown that there are 3 functionally distinct BCC keratinocyte populations. I was therefore able to re-cluster the BCC keratinocytes based on their functional similarities (Figure 5.17d).

a)

Clusters:	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Wiki																	
Biocarta																	
PID																	
KEGG																	
Hallmark																	
Reactome																	

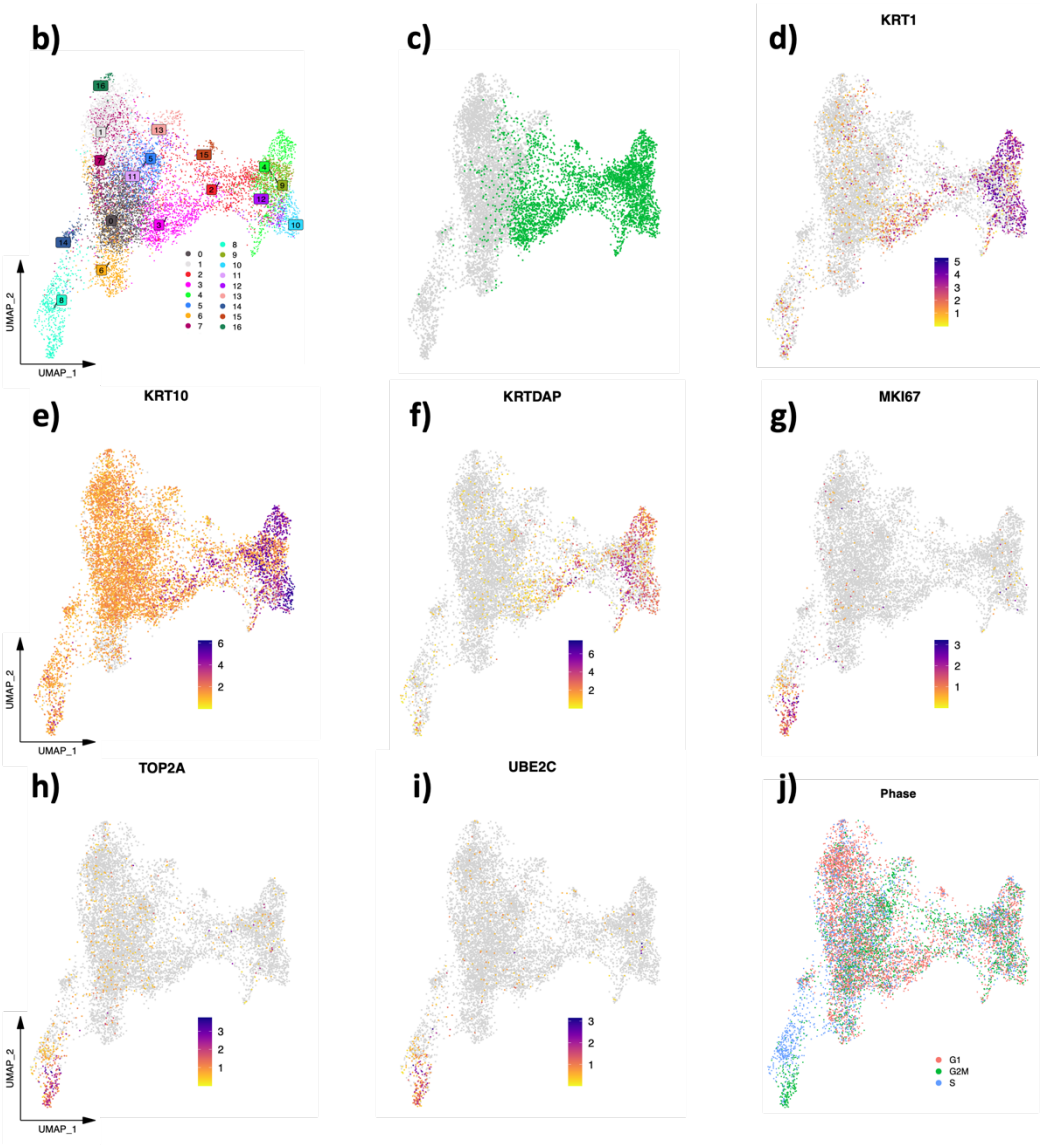


Figure 5.16 | Cluster analysis of BCC keratinocytes. *Single cell Gene Set Enrichment Analysis (scGSEA) was performed on BCC keratinocyte clusters. Wikipathways, Biocarta, PID, KEGG, Hallmark, and Reactome results were all analysed. The primary segregations from the hierarchical clustering from each result was taken and made into a table.* ***a)*** *Samples were assigned colours based on where the root node first branches into two clades (primary segregation), as either grey or blue.* ***b)*** *UMAP highlighting Seurat clusters, for reference.* ***c)*** *Using the information from* ***a****, a module of primary segregation was identified and clusters were highlighted via UMAP. Panel shows the highlighted clusters (2, 3, 4, 9, 10, 12, and 15).* ***d-i)*** *Differentiation markers were assessed in the keratinocyte clusters via UMAP. Panels show UMAPs highlighting expression of KRT1, KRT10, KRTDAP, MKI67, TOP2A, and UBE2C, respectively.* ***j)*** *UMAP highlighting the estimated cell cycle phase for each cell.*

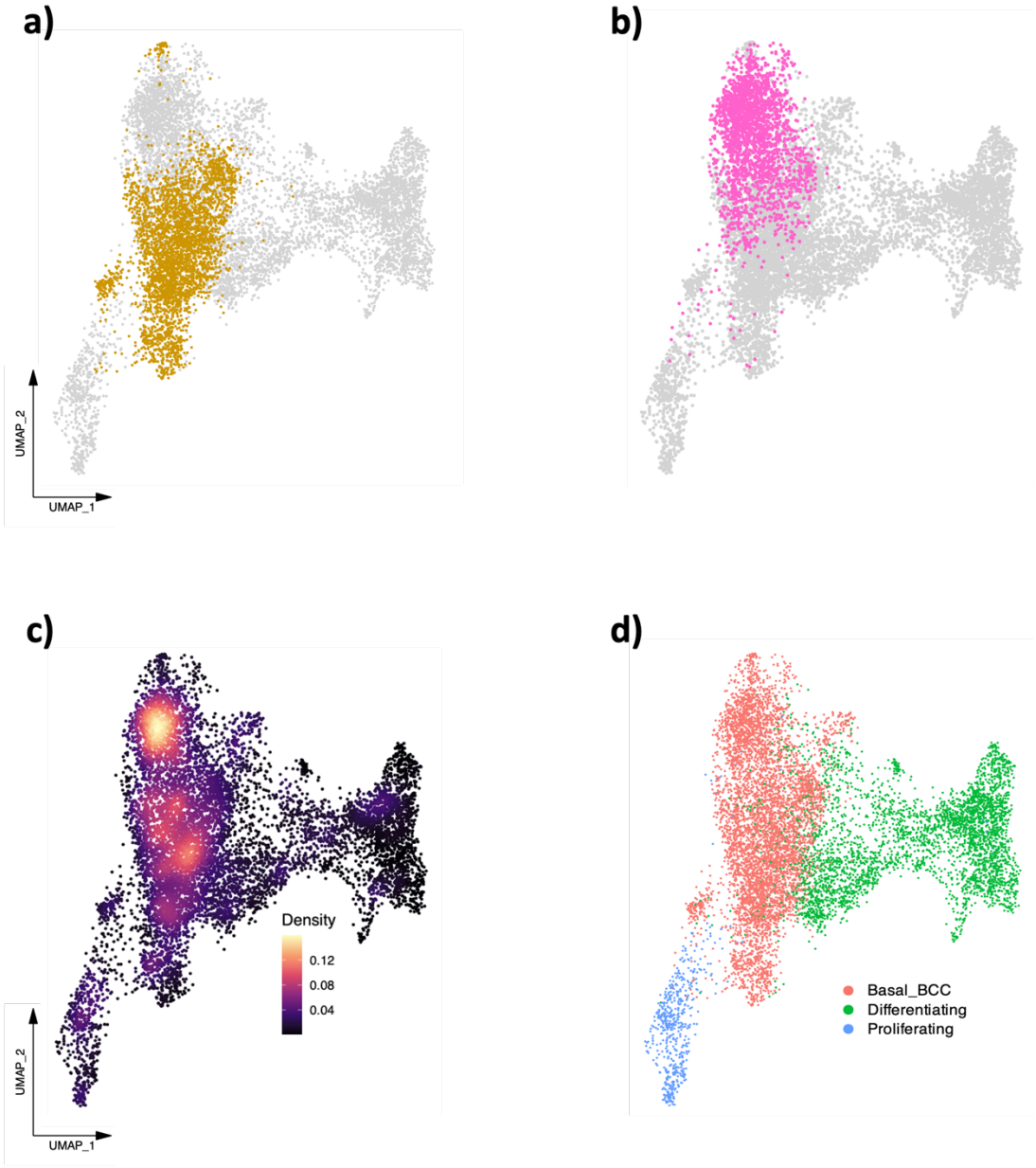


Figure 5.17 | Basal BCC keratinocytes show functional similarity. *Single cell Gene Set Enrichment Analysis (scGSEA) was performed on BCC keratinocyte clusters. Wikipathways, Biocarta, PID, KEGG, Hallmark, and Reactome results were all analysed. The primary segregations from the hierarchical clustering from each result*

was taken and made into a table. Samples were assigned colours based on where the root node first branches into two clades (primary segregation), as either grey or blue. This table is shown in the previous figure. Using this table, a module of primary segregation was identified and highlighted in the previous figure. ***a)*** *When investigating the table further, a second module of segregation was identified which was apart from the first module. Panel shows a UMAP highlighting the clusters that form this module.* ***b)*** *Secondary hierarchical clustering was then investigated, which identified a further module of closely related clusters. Panel shows a UMAP highlighting these functionally similar clusters.* ***c)*** *UMAP highlighting the CD200 expression density.* ***d)*** *Combining the information from these analyses (including previous figure), the clusters were re-annotated based on the clusters' functional similarities. Panel shows a UMAP with the new cell identities based on functional similarity.*

Having identified 3 unique functionally similar populations of BCC keratinocytes (Figure 5.17d) and showed that CD200 expression is found within the 'Basal_BCC' keratinocyte populations, I next hypothesised that BCC cancer stem cells reside within this population. I then wanted to identify any aberrant signalling pathways that distinguish this population from the proliferating and differentiating BCC populations. To address this, I firstly performed DGE analysis between the Basal_BCC clusters and both proliferating and differentiating clusters to generate a set of DEGs that were unique to the Basal_BCC clusters. As I wanted to identify aberrant signalling pathways that were investigable in the lab, I next took the positively enriched DEGs and performed g:Profiler analysis using GO: Molecular Function, KEGG, Reactome, and Wiki Pathways databases (Appendix 5.10). EnrichR analysis was also performed on the positively enriched DEGs (Appendix 5.11). g:Profiler results showed enriched TGFβ-related signalling, Notch signalling, insulin signalling, and Hedgehog signalling, amongst others. Amongst relevant results in the EnrichR analysis, the most common top hits were TGFβ signalling-related, Wnt signalling-related, Insulin-related pathways, Notch-related, and angiogenesis-related. I then performed scGSEA on the 3 populations (Appendix 5.12). Amongst the top positively enriched gene sets for the Basal_BCC population with Biocarta was Integrin pathway, TGFβ pathway, IGF1 pathway, and PTEN pathway. Amongst the top negatively enriched gene sets for biocarta were Death pathway, Inflammation pathway, Caspase pathway, and EPONNFKB pathway (Appendix 5.12.1). When looking at Hallmarks, amongst the top positively enriched gene sets were Hedgehog signalling, Notch signalling, epithelial-mesenchymal-transition, and TGFβ signalling. Amongst the top negatively enriched gene sets were allograft rejection, interferon α response, interferon gamma response, and MYC targets v1 (Appendix 5.12.2). For KEGG,

amongst the top positively enriched gene sets were Insulin signalling pathway, Basal cell carcinoma, Notch signalling pathway, TGFβ signalling pathway, and MAPK signalling. Amongst the top negatively enriched gene sets were Pyrimidine metabolism, autoimmune thyroid disease, steroid hormone biosynthesis, and oxidative phosphorylation (Appendix 5.12.3). For PID, amongst the top positively enriched gene sets were Wnt signalling pathway, hedgehog signalling pathway, AVB3 Integrin pathway, and RET TF pathway. Amongst the top negatively enriched gene sets were PKL1 pathway, E2F pathway, FOXM1 pathway, and Delta NP63 pathway (Appendix 5.12.4). For Reactome, amongst the top positively enriched gene sets were IRS mediated signalling, signalling by NTRKs, signalling by NOTCH, PI3K AKT signalling in cancer, and IL7 signalling. Amongst the top negatively enriched gene sets were gene and protein expression by JAK STAT signalling after IL12 stimulation, interferon gamma signalling, cytokine signalling in immune system, and innate immune system (Appendix 5.12.5). Lastly, for Wiki Pathways, amongst the top positively enriched gene sets were embryonic stem cell pluripotency pathways, IL4 signalling pathway, hedgehog signalling pathway, MAPK signalling pathway, Notch signalling, and TGFβ signalling pathway. Amongst the top negatively enriched gene sets were pyrimidine metabolism, complement system, IL10 anti-inflammatory signalling pathway, and prostaglandin signalling (Appendix 5.12.6). To visualise all of the above analyses, I took the most relevant enriched signalling pathways from the g:Profiler and EnrichR analyses and assessed their enrichment values in the scGSEA analyses (Figure 5.18). One pathway that routinely appeared amongst the top positively enriched pathways and displayed strong enrichment values across all the scGSEA databases during the analyses was the TGF-β pathway. Other pathways that showed strong enrichment across 5/6 databases was Notch signalling, Wnt signalling, and Insulin signalling.

Having found strongly enriched pathways to investigate in the lab, I decided to investigate the TGFβ signalling pathway. To internally validate the pathways involvement within the Basal_BCC clusters, I first checked the expression of various database TGFβ signalling gene sets, which uniformly showed increased enrichment within the Basal_BCC clusters, particularly within the lower clusters (Figure 5.19a-e). Our lab has also previously generated a TGFβ gene panel, which showed similar enrichment to the other generic gene sets (Figure 5.19f).

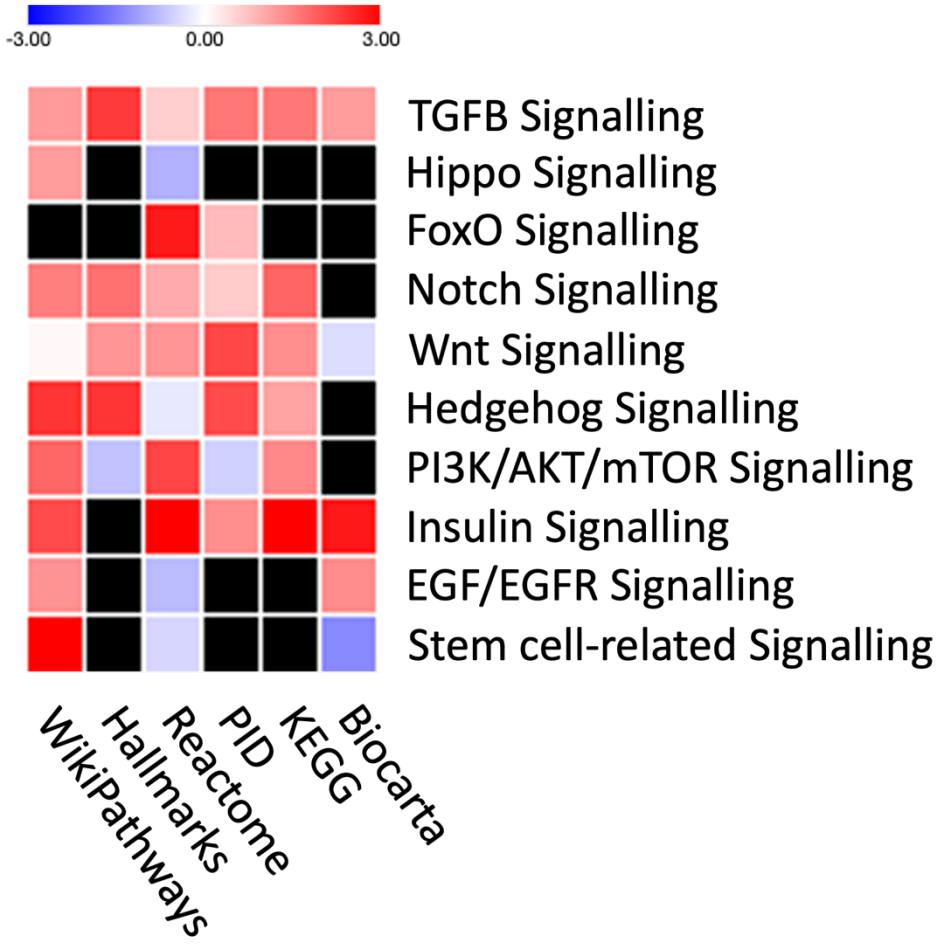


Figure 5.18 | Heatmap of commonly enriched signalling pathways in BCC basal cells. *Differential gene expression (DGE) was performed to generate BCC basal cell differentially expressed genes (DEGs). Positively enriched DEGs were then used as input into g:Profiler and EnrichR. The most commonly enriched signalling pathways were then noted. Single cell gene set enrichment analysis (scGSEA) was then performed on the BCC basal cells, and the enrichment score for the commonly upregulated signalling pathways that were found from the g:Profiler and EnrichR*

analyses were taken and used as input into Broad Institutes Morpheus heat mapping software. Black squares depict no enrichment seen.

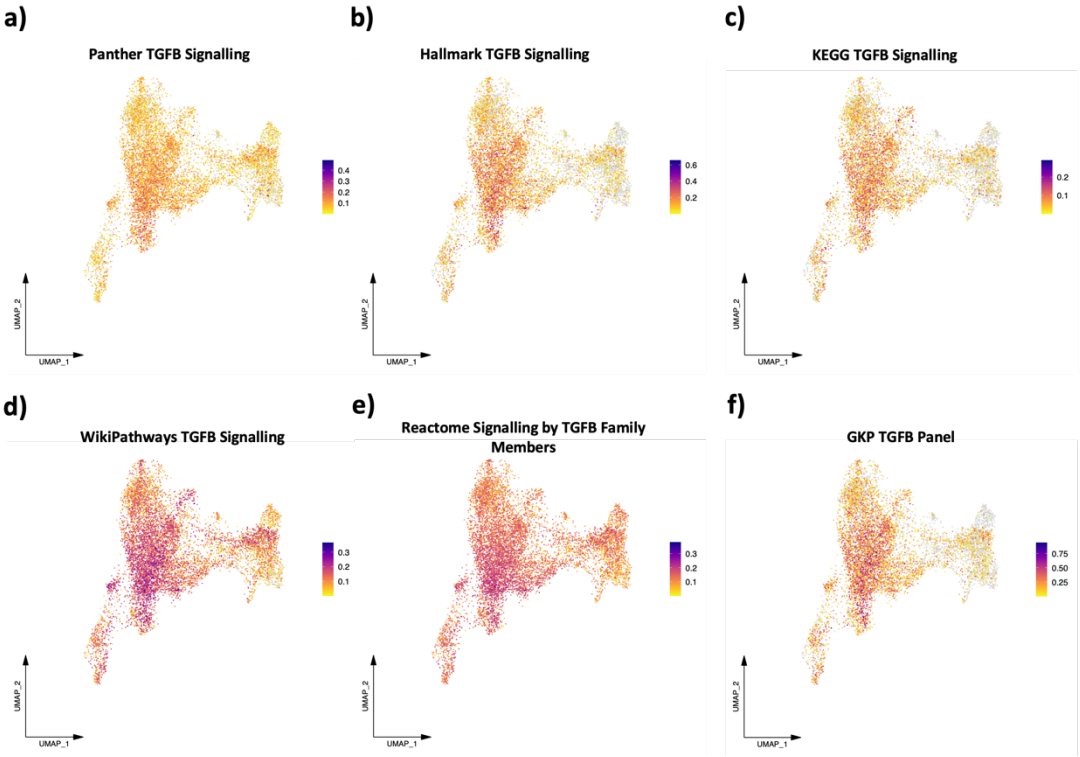


Figure 5.19 | Assessment of TGFβ Signalling in BCC keratinocytes. *TGFβ gene sets were downloaded from their respective databases and used as input to Seurat's 'AddModuleScore' function, which gives an overall enrichment score for the gene set.* ***a-f)*** *UMAPs showing module scores for Panther, Hallmark, KEGG, WikiPathways, Reactome, and GKP labs TGFβ signalling pathway gene sets, respectively.*

I next assessed the expression of SMAD2 and SMAD3, which are the major signalling components of the canonical TGFβ signalling pathway. As expected, SMAD2 and SMAD3 were enriched in the Basal_BCC cluster, with SMAD2 showing more cells with increased expression (Figure 5.20a and b, respectively). I then hypothesised that TGFβ signalling, specifically SMAD2 and SMAD3, showed co-expression with CD200. To assess this, I used density plots to plot the individual marker densities across the Basal_BCC cluster, then a co-density plot of the marker with CD200 expression. Density plots confirmed what was found within the expression UMAPs but indicated increased expression within specific areas of the Basal_BCC cluster (Figure 5.20c and d). Interestingly, both co-density plots highlighted the same populations of cells that expressed both CD200 and SMAD2/3.

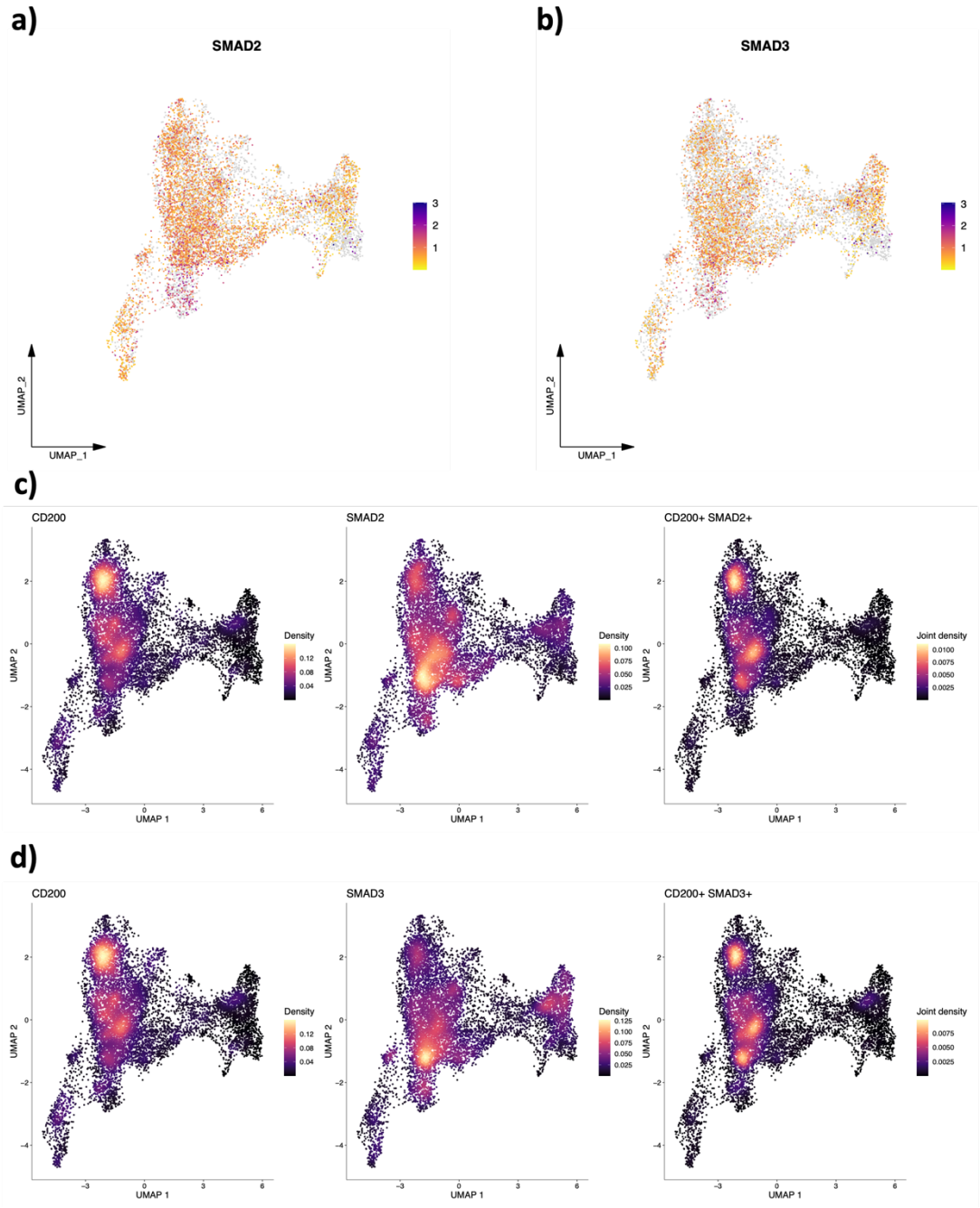


Figure 5.20 | SMAD2/3 signalling within CD200+ basal cell carcinoma basal keratinocytes. *BCC basal keratinocytes were assessed for enrichment of the TGFβ*

signalling pathway components, ***a)*** *SMAD2 and* ***b)*** *SMAD3. Co-density plots were then generated to check correlation between expression of CD200 and* ***c)*** *SMAD2 or* ***d)*** *SMAD3.* ***c, d)*** *Left, CD200 density plot. Middle, SMAD2 or SMAD3 density plot, respectively. Right, co-density plot. Expression levels in* ***a*** *and* ***b*** *represent normalised reads. Normalised reads are also used as input for the density plots. Scale bars in* ***c*** *and* ***d*** *represent the density, or co-density, of the respective genes in the cells.*

5.10 Validation of TGFβ Signalling in BCC CD200+ Basal Cells

Having found a viable enriched signalling pathway within CD200+ BCC basal cells, I next wanted to investigate where these TGFβ signalling-positive cells were located. I hypothesised that these cells would be positioned on the tumour periphery, as we see with CD200+ cells. To investigate the pathway, I used IHC on human BCC samples. Active TGFβ signalling can be seen within the bulge region of the hair follicle of the normal skin and is identified by nuclear labelling with a pSMAD3 antibody. I therefore decided to use nuclear pSMAD3 labelling as an indicator of active TGFβ signalling within the BCC tissue. pSMAD3 expression was first validated by checking for expression within the hair follicle in normal skin regions on BCC sections (Figure 5.21a). pSMAD3 expression was then checked in BCC tissues, which displayed positivity in 21.78% +/- 1.88 SEM in BCC tumour nodules (n=15 BCCs, Figure 5.21b). Interestingly, pSMAD3 expression appeared to be increased at the tumour periphery. Indeed, when counting pSMAD3-positive cells at the tumour periphery, 38.01 % +/- 2.26 SEM were positive.

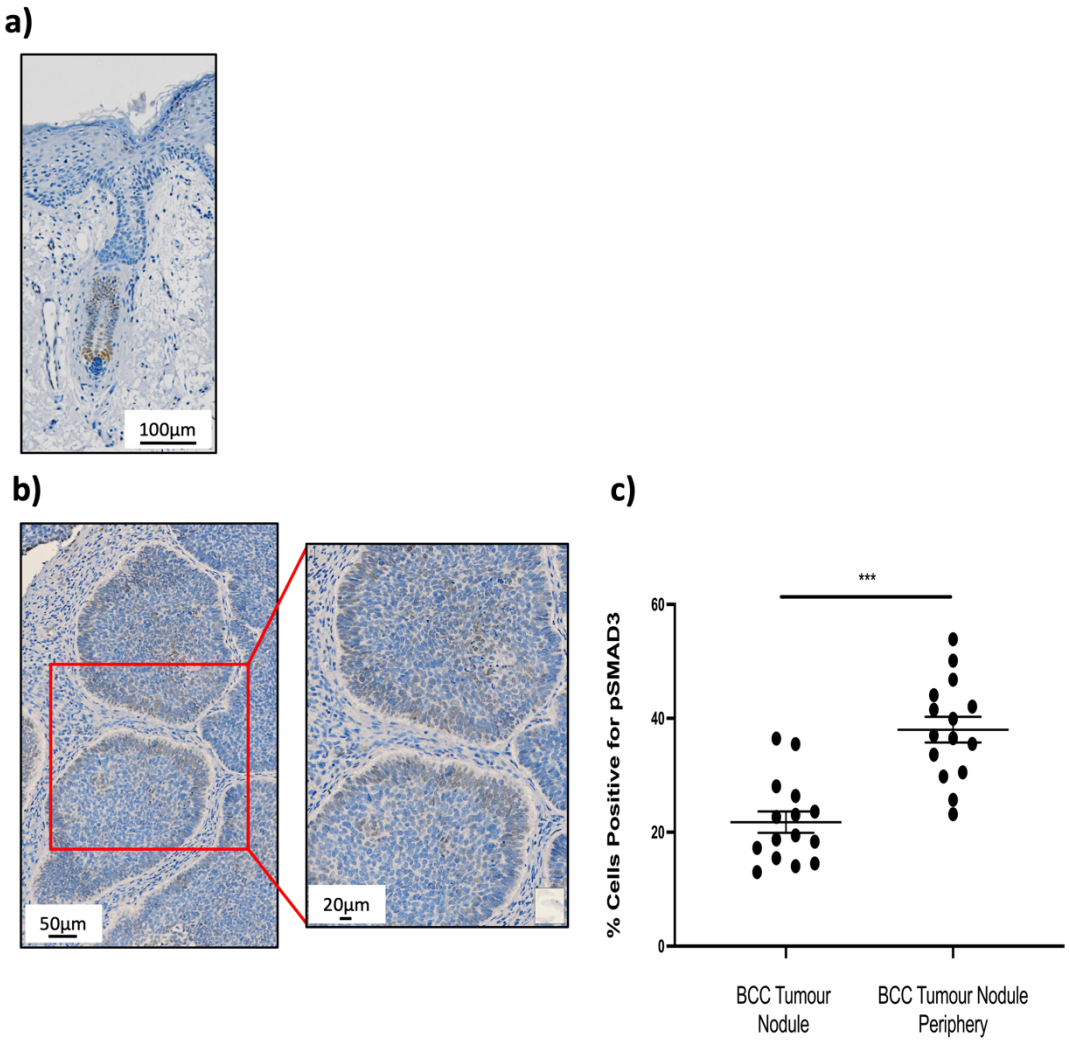


Figure 5.21 | pSMAD3 nuclear staining of human BCC tissues. *pSMAD3 staining was performed on n=14 human BCC tumours.* ***a)*** *pSMAD3 staining is positive in the bulb region of the hair follicle in normal skin, and so used as an internal positive control.* ***b)*** *pSMAD3 staining in human BCC tumour nodules. left, zoomed out image of tumour, right, zoomed in image of tumour. Each tumour nodule represents a technical*

*replicate for a given BCC tissue section. n=15 BCC tissue sections. error bars represent mean +/- SEM, ***P<0.001 by Mann Whitney test.*

5.11 Discussion

In this chapter, I optimised a single cell RNA-sequencing pipeline to analyse and characterise 13 BCC and normal skin samples. I showed that these samples were deeply sequenced, with average sequencing saturations > 80%. Using a small batch of 4 samples, I was able to optimise an analysis pipeline to accurately and reproducibly identify and cluster the various cell types present within human BCC. Using the full dataset and both supervised and unsupervised approaches, I was able to clearly identify the cell types present within human BCCs. Regardless of how many cells are present in each sample, all samples have a large keratinocyte and endothelial cell population. An interesting observation was that the CD200+ sorted samples showed an enrichment of endothelial cells rather than keratinocytes, whilst eliminating the melanocyte population. Significantly improved clustering was also observed in samples with > 1000 cells, signifying the importance of sample size for single cell analysis. I have demonstrated that all samples, regardless of how the cells were isolated via FACS (either CD45- or CD200+ CD45-), contain an endothelial cell and keratinocyte population, showing strong expression of the same markers across samples. Using chromosomal analysis, I was able to show that BCC tumour cells resided within the keratinocyte populations, as they contained significant chromosomal duplications and deletions when compared to the other cell types. An interesting finding was that not all BCC samples harboured significant genomic alterations, suggesting that BCC tumour development and progression is not purely reliant on chromosomal aberrations.

In conclusion, I have shown that CD200 is functionally important in immune cell evasion in BCC and is therefore a valuable marker for these BCC keratinocytes to possess as it provides the cells protection from immune attack. From my bulk RNAseq analyses of CD200+ BCC cells, I saw a mix of populations present. I have shown that even flow sorting cells (via FACS) results in heterogeneity in cell populations. For the first time, I have finally got to the population of interest by performing single cell RNA sequencing (and the cell type identification). Using a low-resolution approach, I was able to isolate the keratinocyte populations from the other cell types and create a keratinocyte-only dataset. By re-clustering the dataset at a higher resolution, I was able to demonstrate the number of, and heterogeneity of keratinocyte populations. I have shown that there is a large collection of populations that display known BCC gene marker expression. These populations are closely clustered and resemble the main body of the BCC keratinocyte populations. I additionally showed that there are distinct proliferating and differentiating keratinocyte populations residing within the BCC keratinocytes. Using RNA velocity analyses, I showed varying transitional states of the keratinocyte populations. I also provided evidence that points to the existence of a progenitor and a transient amplifying cell population that resides within the main BCC keratinocyte population. I then looked at CD200 expression within the BCC keratinocytes, which showed heterogeneous expression throughout with a few areas showing increased expression. These areas also showed high expression in hedgehog signalling. Using pseudo bulk DEG analyses, I showed a number of enriched pathways within these BCC keratinocyte populations. A particular pathway that showed prevalence was the TGFβ signalling pathway.

Having found a pathway of interest, I then investigated its expression within human BCC tissue sections. I used the pSMAD3 antibody to show expression of, and the activity of TGFβ signalling, which showed increased expression and activity in cells around the BCC tumour periphery. CD200+ cells are also located around the tumour periphery. TGFβ signalling therefore may be a functionally relevant signalling pathway to CD200+ cancer stem cells residing within the BCC tumour periphery.

Chapter 6: Discussion

While cancers develop as a result of DNA mutations that lead to cell transformation, the evolution of cancer growth is much more complex and dependent upon many factors. At one level, BCC is a simple cancer in that the driver mutations are clearly defined and result in the activation of a very specific growth factor pathway: the Hh growth factor pathway. Yet, SMO antagonists are confounded by both primary and secondary resistance, and rarely achieve a cure. Herein, we have hypothesised that the persistence of BCC tumour cells in the face of SMO antagonism is due to the survival of BCC cancer stem cells through alternative cell survival mechanisms.

To date, three principal mechanisms have been identified to account for the lack of response to growth factor inhibitors: 1) the activation of alternate signalling pathways that are capable of driving tumour growth, 2) target mutations that prevent drug activity, and more recently 3) the presence of intrinsically resistant cancer stem cells (Valent et al. 2012). Our work has focused on the latter mechanism, in which we have sought to identify specific survival pathways that may be amenable to therapeutic targeting.

6.1 CD200: a functionally relevant cancer stem cell-associated protein

Cell surface proteins used to isolate cancer stem cells may or may not have a functional role. This point is particularly pertinent, even though Cluster of Differentiation 200 (CD200; also known as OX-2) enriches the tumour-initiating cell population in BCC the frequency of tumour-initiating cells is still only 1 in 822 (Colmont et al. 2013a). That said, CD200 has long been considered a relevant cancer-associated protein (Moreaux et al. 2006). Of note, earlier defined cancer stem cell markers CD44 and CD133 often localise with CD200 in specific cancers: glioblastoma, breast and prostate (Kawasaki and Farrar 2008).

CD200 is a type 1 transmembrane glycoprotein comprised of two Ig-like domains and has a single transmembrane domain with a short, 19 amino acid intracellular domain (Hatherley et al. 2013). Due to its short cytoplasmic domain, CD200 provides a one-way interaction with the CD200R (Liu et al. 2016). The CD200 receptor (CD200R) has structural similarity to CD200 but has a longer cytoplasmic domain for intracellular signalling (Ngwa and Liu 2019). In contrast to CD200, CD200R expression is restricted to cells of myeloid lineages, activated T cells, B cells and NK cells (Kawasaki and Farrar 2008). The broad distribution of CD200 but immunologically restricted expression of CD200R is consistent with specificity associated with immune evasion.

CD200-/- mice had a normal lifespan and fertility but compared to WT mice, had significantly higher numbers of CD11b+ macrophage populations, highlighting its importance in immune regulation (Hoek et al. 2000). CD200 regulates immune responses in bone development and homeostasis, infection response, transplant

acceptance and reproductive biology, and pathologies including neuroinflammation, autoimmune and allergic disorders and cancer growth (Gorczynski 2012; Zhao et al. 2019). CD200 has a role in a number of cancers including multiple myeloma (MM), lymphocytic leukaemia (CLL) and acute myeloid leukaemia (AML), and solid cancers such as melanoma, breast, prostate, lung, renal and ovarian (Moreaux et al. 2008; Siva et al. 2008; Twito et al. 2013).

In studying the inflammatory infiltrate surrounding BCC and SCC, we had observed there was a near absence of NK cells (as defined by CD56+HLA-DR+) in BCC when compared to SCC and normal skin (1.53% +/- 1.00% vs 22.21% +/- 9.80% and 8.17% +/- 2.55%, both $P < 0.05$, respectively. Morgan et al. 2022). In BCC, we showed that CD200 could undergo ectodomain shedding by metalloproteinases within the tumour microenvironment. Previously it had been shown that CD200:CD200R interaction led to negative regulation of the MAPK signalling pathway, which is necessary for immune cell activation (Mihrshahi and Brown 2010). Using NK cell lines that were CD200R-positive (NK92MI; NK+) and CD200R-negative (NKL; NK-), we showed that incubation with a CD200 peptide resulted in a 4-fold reduction in p-ERK1/2 levels in NK+ but not NK-, wherein p-ERK1/2 was as an indicator of NK cell activation through MAPK signalling. Incucyte time-lapse co-culture assay with HeLa CD200+ and HeLa CD200- cells with NK+ cells demonstrated CD200 signalling-dependent NK cell killing. NK cell killing could be prevented when CD200+ cells were cocultured with NK+ cells that had been initially pre-treated with a CD200R shRNA. Western blot analysis showed that CD200 signalling led to inhibition of the MAPK pathway and, in turn, to an increase in PPARγ signalling, promoting the expression of FAS receptor and FAS ligand. Thus, continuous CD200 signalling led to NK cell

apoptosis. The in vivo experiments performed as part of this thesis validated our earlier in vitro findings.

Herein, I found that CD200+ and CD200- tumours grew to similar sizes, but that CD200- tumours had a greater anti-tumour immune response. The immune response was associated with areas of necrosis within the CD200- tumour. CD200 expression by tumour cells moderated this immune response, and importantly was accompanied by NK cell apoptosis, since NK cells are preserved in Nude mice used for xenografting. Collectively, these findings supported the functional role of CD200 in helping BCC cancer stem cells evade immune cell-mediated destruction.

6.2 CD200+ CD45- BCC subpopulation included endothelial cells and identified angiogenesis

Unfortunately, the absence of CD200 expression in the UWBCC1 cell line, precluded its use for analysis of BCC cancer stem cells. In hindsight, the flow-cytometry did not define a distinct CD200+ population. Even when I thought I had flow-sorted CD200+ cells, there was no CD200 mRNA detectable by RT-PCR or RNA sequencing.

Our earlier work to characterise BCC cancer stem cells had used a simple flow sorting strategy using CD45 as a sole linage negative marker. Using this strategy, the isolated cells had been determined to be keratinocytes by co-labelling with pan-cytokeratin (Colmont et al. 2013a). Moreover, the CD200+ CD45- BCC subpopulation demonstrated expression of Hh pathway-regulated gene expression, including keratin 17 by western blot analysis. Yet in our RNA sequencing and subsequent

immunohistochemistry experiments, we have shown that the CD200+ CD45- BCC sub-population also clearly identified endothelial cells.

Clinically, BCCs are discernible by the presence of dilated blood vessels surrounding the tumour, and many patients first acknowledge the presence of a BCC because of bleeding (Lupu et al. 2019). There is a close relationship between angiogenesis and tumour growth (Folkman et al. 1989). The term 'angiogenic switch' refers to the predominance of proangiogenic factors, resulting in the formation of new blood vessels (Folkman and Shing 1992). VEGF is the main proangiogenic factor found to be increased in BCCs when compared to normal skin and squamous cell carcinoma (Bowden et al. 2002). VEGF-associated angiogenesis in BCCs is dependent on tumour cell expression of COX-2, matrix metalloproteinase-9 (MMP-9), and Maspin (Tjiu et al. 2006).

Our bulk RNA sequencing analysis comparing CD200+ CD45- and CD200- CD45- BCC flow-sorted cells identified 1,372 DEGs, of which 517 were up-regulated. Gene ontology of up-regulated genes identified enrichment of regulation of vasculature development, angiogenesis cell and migration. Likewise, C5-Ontology and C8-cell type results showed endothelial cells and related functions. To elucidate the regulation of VEGF pathway proteins, we overlayed the DEGs onto the VEGF pathway, leading to the identification of components both within BCC cells, but also possibly endothelial cells themselves. To confirm this, we used over-expression of the VEGFR2 to label BCC tissues by immunohistochemistry and observed labelling in both tumour and endothelial cells, suggesting that there may be CD200+ CD45- endothelial cell enrichment in the BCC CD200+ CD45-sorted cells. Hence, to

determine the cell types and the interplay of cells in angiogenesis we undertook IPA analysis, which confirmed presence of multiple cell types, and therefore we undertook single-cell sequencing.

6.3 Single cell sequencing: A TGFβ signature

As suspected from the bulk RNA sequencing results, the flow-sorted CD200+ CD45- BCC subpopulation was indeed heterogenous. This sub-population, in addition to BCC keratinocytes, contained a sizable fraction of endothelial cells and occasional fibroblasts based on gene expression profiles. Compared to normal skin and unsorted BCC, the proportion of endothelial cells was much greater in the CD200+ CD45- BCC flow-sorted sub-population, unlike melanocytes and fibroblasts, suggesting that endothelial cells were positively selected. I was able to use bioinformatics analyses to demonstrate CD200 expression on endothelial cells to validate this. One of the future directions for this work is to further elucidate the interplay between BCC keratinocytes and endothelial cells to better understand the bias towards angiogenesis in BCC.

By combining the different samples to generate a single dataset, we were able to further re-cluster the BCC keratinocytes to identify 17 distinct clusters based on gene expression. Consistent with BCC keratinocytes, both the CD200+ CD45-and CD200- CD45- BCC subpopulations expressed BCC associated genes: Ptch1, Gli1, Gli2, EPCAM, Hhip, BCAM, and MycN. Concordant with these observations, the Kegg BCC and Hh gene sets were enriched in both BCC subsets. RNA velocity and expression of known stem cell gene sets were used to identify the progenitor cell clusters, which also

expressed the highest levels of CD200. The progenitor population, also called "basal", was represented by clusters 0, 1, 5, 6, 7, 11, 13, and 16. Cluster 8 represented keratinocytes undergoing proliferation, while the clusters on the right-hand side of the UMAP exhibited a differentiated cell phenotype. By undertaking gene set enrichment analysis across several gene sets, consistently identified active TGFβ signalling within the progenitor populations supported by our immunohistochemical labelling of BCC tissues.

We have shown that in BCC, a Hh-driven cancer, the cancer stem cells exhibited active TGFβ signalling. In hair follicle bulge stem cells, TGFβ signalling-associated proliferation is observed during the anagen hair follicle cycle, consistent with the potential cell of origin for BCC cancer stem cells (Youssef et al. 2010; Wang et al. 2011). Active TGFβ signalling has also been identified in cancer stem cells from other cancer types (Shipitsin et al. 2007; Peñuelas et al. 2009; Ikushima and Miyazono 2010; You et al. 2010; Bruna et al. 2012; Lo et al. 2012; Mima et al. 2012; Bhola et al. 2013; Oshimori et al. 2015). The interplay between TGFβ and Hh signalling has been previously reported in BCC, wherein GLI2 is a downstream target gene of the TGFβ signalling cascade (Dennler et al. 2007). In response to TGFβ, SMAD3 was rapidly recruited to distinct elements within the GLI2 promoter and enhanced its transcription (Dennler et al. 2009). Also following TGFβ pathway activation, GLI1 a specific marker for Hh signalling activity, was induced in a GLI2-dependent manner. Recently, Yao and colleagues found TGFβ signalling to be active in BCC demonstrating resistance to SMO antagonists, leading to activation of MRTF regulated transcription (Yao et al. 2020). Thus, there is crosstalk between the Hh and TGFβ pathways, evident

within BCC, in particular after SMO antagonism and more specifically as we have found in BCC cancer stem cells.

TGFβ is the prototypic cytokine responsible for epithelial-mesenchymal transition, which could aid BCC cancer stem cell invasion (Zavadil and Böttinger 2005). Many of the epithelial-mesenchymal transition genes regulated by the TGFβ pathway are GLI-dependent (Li et al. 2007). TGFβ-dependent GLI2 expression and GLI-dependent transcription were responsible for epithelial-mesenchymal transition to form invasive myoepithelial cells in breast ductal carcinoma in situ (Miller et al. 2000; Hu et al. 2012). Likewise, TGFβ Hh-independent GLI2 expression resulting in bladder cancer invasiveness (Mechlin et al. 2010). In ovarian cancer stem cells, TGFβ was associated with GLI1 and 2 mediated resistance to cisplatin (Steg et al. 2012). Therefore, the ability of TGFβ to maintain Hh signalling may provide a mechanism for BCC cancer stem cell survival following SMO antagonist therapy.

In conclusion we have determined that CD200 is a functionally relevant BCC cancer stem cell protein, which is responsible for a specific pattern of immune evasion observed in BCC through ectodomain shedding. In attempting to further define this population using transcriptomic approaches, we have identified VEGF angiogenesis. More importantly, using single cell sequencing, we have been able to identify highly enriched clusters of potential BCC cancer stem cells, where we have observed expression of TGFβ regulated genes. Consistent with our earlier observations to suggest that BCC cancer stem cells are not killed by SMO antagonists, we find that TGFβ signalling can support Hh pathway gene transcription.

6.4 Future Directions

In this thesis, I was able to investigate the TGFβ signalling pathway in BCC tissue sections. Expression of TGFβ signalling member, pSMAD3 was seen in the tumour periphery, the same area where the CD200+ BCC CSCs are located. Further studies should look at the following: 1) Isolate the CD200+ pSMAD3+ cells and perform stem cell assays to identify whether these cells have enhanced survival and colony forming abilities when compared to their negative and CD200+ only counterparts. 2) Perform TGFβ knockdown experiments in vitro and assess whether the canonical TGFβ signalling pathway has any functional effects on the cells, and if there are any changes to Hh signalling. These findings could ultimately lead to the testing of dual inhibitor treatment, where the Hh and TGFβ pathways are inhibited. This could potentially lead to better response rates in BCC patients.

Other enriched signalling pathways were also identified during the single cell RNA sequencing analyses. If time permitted, these pathways would have been investigated in a similar manner to the TGFβ pathway, and further analyses could have been performed as described above. Of particular interest would be the identification of cell surface markers that could enable to the further isolation of the BCC CSCs. Performing DGE analysis of the progenitor and transient amplifying clusters could potentially provide these markers, combined with IPA's biomarker analysis. These markers could then be used in conjunction with CD200 to sort cells and perform xenografting experiments.

References

Abdelwahab, R. et al. 2022. The Relationship between Vitamin D and Basal Cell Carcinoma: A Systematic Review. *Cureus* 14(9). Available at: /pmc/articles/PMC9595244/ [Accessed: 21 February 2024].

Al-Qarqaz, F. et al. 2018. Clinical and Demographic Features of Basal Cell Carcinoma in North Jordan. *Journal of skin cancer* 2018. Available at: https://pubmed.ncbi.nlm.nih.gov/30498602/ [Accessed: 5 January 2024].

Alcedo, J., Ayzenzon, M., Von Ohlen, T., Noll, M. and Hooper, J.E. 1996. The Drosophila smoothened gene encodes a seven-pass membrane protein, a putative receptor for the hedgehog signal. *Cell* 86(2), pp. 221–232. Available at: https://pubmed.ncbi.nlm.nih.gov/8706127/ [Accessed: 21 February 2024].

Ali, D., Alhattab, D., Jafar, H., Alzubide, M., Sharar, N., Bdour, S. and Awidi, A. 2021. Differential marker expression between keratinocyte stem cells and their progeny generated from a single colony. *International Journal of Molecular Sciences* 22(19), p. 10810. Available at: https://www.mdpi.com/1422-0067/22/19/10810/htm [Accessed: 21 August 2023].

Van Allen, E.M. et al. 2015. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* 350(6257), pp. 207–211. Available at: https://www.science.org/doi/10.1126/science.aad0095 [Accessed: 4 February 2024].

Alpsoy, E., Yilmaz, E., Başaran, E. and Yazar, Ş. 1996. Comparison of the effects of intralesional interferon alfa-2a, 2b and the combination of 2a and 2b in the treatment of basal cell carcinoma. *The Journal of dermatology* 23(6), pp. 394–396. Available at: https://pubmed.ncbi.nlm.nih.gov/8708151/ [Accessed: 10 February 2024].

AlSabbagh, M.M. and Baqi, M.A. 2018. Bazex-Dupré-Christol syndrome: review of clinical and molecular aspects. *International Journal of Dermatology* 57(9), pp. 1102–1106. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/ijd.14065 [Accessed: 20 February 2024].

Altman, N. and Krzywinski, M. 2018. The curse(s) of dimensionality this-month. *Nature Methods* 15(6), pp. 399–400. doi: 10.1038/S41592-018-0019-X.

Andrews, T.S. and Hemberg, M. 2018. Identifying cell populations with scRNASeq. *Molecular Aspects of Medicine* 59, pp. 114–122. doi: 10.1016/J.MAM.2017.07.002.

Apalla, Z., Nashan, D., Weller, R.B. and Castellsagué, X. 2017. Skin Cancer: Epidemiology, Disease Burden, Pathophysiology, Diagnosis, and Therapeutic Approaches. *Dermatology and Therapy* 7, pp. 5–19. doi: 10.1007/s13555-016-0165-y.

Arbiser, J.L., Bonner, M.Y. and Gilbert, L.C. 2017. Targeting the duality of cancer. *NPJ precision oncology* 1(1). Available at: https://pubmed.ncbi.nlm.nih.gov/28825045/ [Accessed: 20 February 2023].

Argelaguet, R., Cuomo, A.S.E., Stegle, O. and Marioni, J.C. 2021. Computational principles and challenges in single-cell data integration. *Nature Biotechnology 2021 39:10* 39(10), pp. 1202–1215. Available at: https://www.nature.com/articles/s41587-021-00895-7 [Accessed: 15 May 2023].

Aszterbaum, M., Epstein, J., Anthony, O., Douglas, V., Leboit, P.E., Scott, M.P. and Epstein, E.H. 1999. Ultraviolet and ionizing radiation enhance the growth of BCCs and trichoblastomas in patched heterozygous knockout mice. *Nature medicine* 5(11), pp. 1285–1291. Available at: https://pubmed.ncbi.nlm.nih.gov/10545995/ [Accessed: 21 February 2024].

Atwood, S.X. et al. 2015. Smoothened Variants Explain the Majority of Drug

Resistance in Basal Cell Carcinoma. *Cancer Cell* 27(3), pp. 342–353. Available at: http://www.cell.com/article/S1535610815000525/fulltext [Accessed: 13 January 2024].

Ayob, A.Z. and Ramasamy, T.S. 2018. Cancer stem cells as key drivers of tumour progression. *Journal of Biomedical Science 2018 25:1* 25(1), pp. 1–18. Available at: https://jbiomedsci.biomedcentral.com/articles/10.1186/s12929-018-0426-4 [Accessed: 19 February 2024].

Azizova, T. V., Bannikova, M. V., Grigoryeva, E.S. and Rybkina, V.L. 2018. Risk of malignant skin neoplasms in a cohort of workers occupationally exposed to ionizing radiation at low dose rates. *PLOS ONE* 13(10), p. e0205060. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0205060 [Accessed: 7 January 2024].

Bache, K.G., Slagsvold, T. and Stenmark, H. 2004. Defective downregulation of receptor tyrosine kinases in cancer. *The EMBO Journal* 23(14), p. 2707. Available at: /pmc/articles/PMC514952/ [Accessed: 26 January 2023].

Baghban, R. et al. 2020. Tumor microenvironment complexity and therapeutic implications at a glance. *Cell Communication and Signaling 2020 18:1* 18(1), pp. 1–19. Available at: https://biosignaling.biomedcentral.com/articles/10.1186/s12964-020-0530-4 [Accessed: 18 February 2024].

Barclay, A.N., Wright, G.J., Brooke, G. and Brown, M.H. 2002. CD200 and membrane protein interactions in the control of myeloid cells. *Trends in Immunology* 23(6), pp. 285–290. Available at: https://pubmed.ncbi.nlm.nih.gov/12072366/ [Accessed: 11 February 2024].

Barnfield, P.C., Zhang, X., Thanabalasingham, V., Yoshida, M. and Hui, C.C. 2005. Negative regulation of Gli1 and Gli2 activator function by Suppressor of fused

through multiple mechanisms. *Differentiation; research in biological diversity* 73(8), pp. 397–405. Available at: https://pubmed.ncbi.nlm.nih.gov/16316410/ [Accessed: 21 February 2024].

Becht, E., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F. and Newell, E.W. 2018. Evaluation of UMAP as an alternative to t-SNE for single-cell data. *bioRxiv* , p. 298430. Available at: https://www.biorxiv.org/content/10.1101/298430v1 [Accessed: 26 April 2023].

Belote, R.L. et al. 2021. Human melanocyte development and melanoma dedifferentiation at single-cell resolution. *Nature Cell Biology 2021 23:9* 23(9), pp. 1035–1047. Available at: https://www.nature.com/articles/s41556-021-00740-8 [Accessed: 18 April 2023].

Bendall, S.C. et al. 2014. Single-Cell Trajectory Detection Uncovers Progression and Regulatory Coordination in Human B Cell Development. *Cell* 157(3), pp. 714–725. doi: 10.1016/J.CELL.2014.04.005.

Bendinelli, B., Masala, G., Garamella, G., Palli, D. and Caini, S. 2019. Do Thiazide Diuretics Increase the Risk of Skin Cancer? A Critical Review of the Scientific Evidence and Updated Meta-Analysis. *Current cardiology reports* 21(9). Available at: https://pubmed.ncbi.nlm.nih.gov/31352643/ [Accessed: 20 February 2024].

Bergen, V., Lange, M., Peidli, S., Wolf, F.A. and Theis, F.J. 2020. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature biotechnology* 38(12), pp. 1408–1414. Available at: https://pubmed.ncbi.nlm.nih.gov/32747759/ [Accessed: 8 August 2023].

Bhola, N.E. et al. 2013. TGF-β inhibition enhances chemotherapy action against triple-negative breast cancer. *The Journal of clinical investigation* 123(3), pp. 1348–1358. Available at: https://pubmed.ncbi.nlm.nih.gov/23391723/ [Accessed: 23

February 2024].
Blondel, V.D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008(10), p. P10008. Available at: https://iopscience.iop.org/article/10.1088/1742-5468/2008/10/P10008 [Accessed: 25 April 2023].
Bowden, J., Brennan, P.A., Umar, T. and Cronin, A. 2002. Expression of vascular endothelial growth factor in basal cell carcinoma and cutaneous squamous cell carcinoma of the head and neck. *Journal of cutaneous pathology* 29(10), pp. 585–589. Available at: https://pubmed.ncbi.nlm.nih.gov/12453295/ [Accessed: 23 February 2024].
Bozic, I. et al. 2010. Accumulation of driver and passenger mutations during tumor progression. *Proceedings of the National Academy of Sciences of the United States of America* 107(43), pp. 18545–18550. Available at: https://www.pnas.org/doi/abs/10.1073/pnas.1010978107 [Accessed: 24 January 2023].
Brambullo, T. et al. 2022. Xeroderma Pigmentosum: A Genetic Condition Skin Cancer Correlated—A Systematic Review. *BioMed Research International* 2022. Available at: /pmc/articles/PMC9313971/ [Accessed: 20 February 2024].
Bray, F., Laversanne, M., Weiderpass, E. and Soerjomataram, I. 2021. The ever-increasing importance of cancer as a leading cause of premature death worldwide. *Cancer* 127(16), pp. 3029–3030. Available at: https://pubmed.ncbi.nlm.nih.gov/34086348/ [Accessed: 25 January 2023].
Breier, G. 2000. Angiogenesis in Embryonic Development—A Review. *Placenta* 21(SUPPL.1), pp. S11–S15. doi: 10.1053/PLAC.1999.0525.

Brennecke, P. et al. 2013. Accounting for technical noise in single-cell RNA-seq experiments. *Nature Methods 2013 10:11* 10(11), pp. 1093–1095. Available at: https://www.nature.com/articles/nmeth.2645 [Accessed: 6 April 2023].

Bruna, A. et al. 2012. TGFβ induces the formation of tumour-initiating cells in claudinlow breast cancer. *Nature Communications 2012 3:1* 3(1), pp. 1–12. Available at: https://www.nature.com/articles/ncomms2039 [Accessed: 23 February 2024].

Buonamici, S. et al. 2010. Interfering with resistance to smoothened antagonists by inhibition of the PI3K pathway in medulloblastoma. *Science translational medicine* 2(51). Available at: https://pubmed.ncbi.nlm.nih.gov/20881279/ [Accessed: 21 February 2024].

Butler, A., Hoffman, P., Smibert, P., Papalexi, E. and Satija, R. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology 2018 36:5* 36(5), pp. 411–420. Available at: https://www.nature.com/articles/nbt.4096 [Accessed: 16 May 2023].

Cameron, M.C. et al. 2019a. Basal cell carcinoma: Epidemiology; pathophysiology; clinical and histological subtypes; and disease associations. *Journal of the American Academy of Dermatology* 80(2), pp. 303–317. Available at: http://www.jaad.org/article/S0190962218307758/fulltext [Accessed: 8 January 2024].

Cameron, M.C. et al. 2019b. Basal cell carcinoma: Epidemiology; pathophysiology; clinical and histological subtypes; and disease associations. *Journal of the American Academy of Dermatology* 80(2), pp. 303–317. Available at: http://www.jaad.org/article/S0190962218307758/fulltext [Accessed: 4 January 2024].

Campbell, K.J. and Tait, S.W.G. 2018. Targeting BCL-2 regulated apoptosis in cancer. *Open Biology* 8(5). Available at: /pmc/articles/PMC5990650/ [Accessed: 20 February 2023].

Carneiro, B.A. and El-Deiry, W.S. 2020. Targeting apoptosis in cancer therapy. *Nature Reviews Clinical Oncology 2020 17:7* 17(7), pp. 395–417. Available at: https://www.nature.com/articles/s41571-020-0341-y [Accessed: 20 February 2023].

Castrisos, G., Lewandowski, R., Castrisos Bsc, G. and Lewandowski, ; R 2021. Narrative review of the epidemiology/biology of basal cell carcinoma: a need for public health consensus. *ANZ Journal of Surgery* 91(6), pp. 1098–1103. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/ans.16522 [Accessed: 20 February 2024].

Chahal, H.S. et al. 2016. Genome-wide association study identifies 14 novel risk alleles associated with basal cell carcinoma. *Nature communications* 7. Available at: https://pubmed.ncbi.nlm.nih.gov/27539887/ [Accessed: 20 February 2024].

Chang, A.L.S. and Oro, A.E. 2012. Initial Assessment of Tumor Regrowth After Vismodegib in Advanced Basal Cell Carcinoma. *Archives of dermatology* 148(11), p. 1324. Available at: /pmc/articles/PMC3777384/ [Accessed: 13 January 2024].

Chen, D.S. and Mellman, I. 2013. Oncology meets immunology: the cancer-immunity cycle. *Immunity* 39(1), pp. 1–10. Available at: https://pubmed.ncbi.nlm.nih.gov/23890059/ [Accessed: 18 February 2024].

Chen, P.L. et al. 2016. Analysis of immune signatures in longitudinal tumor samples yields insight into biomarkers of response and mechanisms of resistance to immune checkpoint blockade. *Cancer Discovery* 6(8), pp. 827–837. Available at: /cancerdiscovery/article/6/8/827/5660/Analysis-of-Immune-Signatures-in-

Longitudinal [Accessed: 4 February 2024].
Chen, Y., Sasai, N., Ma, G., Yue, T., Jia, J., Briscoe, J. and Jiang, J. 2011. Sonic Hedgehog Dependent Phosphorylation by CK1α and GRK2 Is Required for Ciliary Accumulation and Activation of Smoothened. *PLOS Biology* 9(6), p. e1001083. Available at: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001083 [Accessed: 21 February 2024].
Cheng, S.Y. and Michael Bishop, J. 2002. Suppressor of Fused represses Gli-mediated transcription by recruiting the SAP18-mSin3 corepressor complex. *Proceedings of the National Academy of Sciences of the United States of America* 99(8), pp. 5442–5447. Available at: https://pubmed.ncbi.nlm.nih.gov/11960000/ [Accessed: 21 February 2024].
Cheong, A. and Nagel, Z.D. 2022. Human Variation in DNA Repair, Immune Function, and Cancer Risk. *Frontiers in Immunology* 13. Available at: https://pubmed.ncbi.nlm.nih.gov/35935942/ [Accessed: 26 January 2023].
Chial, H. 2008a. Proto-oncogenes to Oncogenes to Cancer. *Nature Education* 1(1), p. 33. Available at: https://www.nature.com/scitable/topicpage/proto-oncogenes-to-oncogenes-to-cancer-883/ [Accessed: 24 January 2023].
Chial, H. 2008b. Tumor Suppressor (TS) Genes and the Two-Hit Hypothesis. *Nature Education* 1(1), p. 177. Available at: https://www.nature.com/scitable/topicpage/tumor-suppressor-ts-genes-and-the-two-887/ [Accessed: 24 January 2023].
Chiang, A.C. and Massagué, J. 2008. Molecular Basis of Metastasis. *New England Journal of Medicine* 359(26), pp. 2814–2823. Available at: https://pubmed.ncbi.nlm.nih.gov/19109576/ [Accessed: 1 February 2023].

Choe, D. and Choi, D. 2023. Cancel cancer: The immunotherapeutic potential of CD200/CD200R blockade. *Frontiers in Oncology* 13, p. 1088038. doi: 10.3389/FONC.2023.1088038/BIBTEX.

Choquet, H., Ashrafzadeh, S., Kim, Y., Asgari, M.M. and Jorgenson, E. 2020. Genetic and environmental factors underlying keratinocyte carcinoma risk. *JCI insight* 5(10). Available at: https://pubmed.ncbi.nlm.nih.gov/32434987/ [Accessed: 10 January 2024].

Chuang, T.Y., Popescu, A., Su, W.P.D. and Chute, C.G. 1990. Basal cell carcinoma. A population-based incidence study in Rochester, Minnesota. *Journal of the American Academy of Dermatology* 22(3), pp. 413–417. Available at: https://pubmed.ncbi.nlm.nih.gov/2312827/ [Accessed: 4 January 2024].

Ciążyńska, M. et al. 2021. The incidence and clinical analysis of non-melanoma skin cancer. *Scientific reports* 11(1). Available at: https://pubmed.ncbi.nlm.nih.gov/33619293/ [Accessed: 5 January 2024].

Cillo, A. 2021. singleseqgset: Single-cell RNAseq gene set enrichment analysis.

Cobb, C. and Watson, K. 2018. The Layers of Your skin.

Coles, S.J., Wang, E.C.Y., Man, S., Hills, R.K., Burnett, A.K., Tonks, A. and Darley, R.L. 2011. CD200 expression suppresses natural killer cell function and directly inhibits patient anti-tumor response in acute myeloid leukemia. *Leukemia* 25(5), p. 792. Available at: /pmc/articles/PMC3093357/ [Accessed: 21 January 2024].

Collett, D., Mumford, L., Banner, N.R., Neuberger, J. and Watson, C. 2010. Comparison of the Incidence of Malignancy in Recipients of Different Types of Organ: A UK Registry Audit. *American Journal of Transplantation* 10(8), pp. 1889–1896. Available at: http://doi.wiley.com/10.1111/j.1600-6143.2010.03181.x [Accessed: 16 August 2019].

Colmont, C.S. et al. 2013a. CD200-expressing human basal cell carcinoma cells initiate tumor growth. *Proceedings of the National Academy of Sciences of the United States of America* 110(4), pp. 1434–1439. doi: 10.1073/pnas.1211655110.
Colmont, C.S. et al. 2013b. CD200-expressing human basal cell carcinoma cells initiate tumor growth. *Proceedings of the National Academy of Sciences of the United States of America* 110(4), pp. 1434–1439. doi: 10.1073/pnas.1211655110.
Colmont, C.S., Harding, K.G., Piguet, V. and Patel, G.K. 2012. Human skin cancer stem cells: A tale of mice and men. *Experimental Dermatology* 21(8), pp. 576–580. doi: 10.1111/j.1600-0625.2012.01533.x.
Colmont, C.S., Ketah, A.B., Errington, R.J., Reed, S.H., Udey, M.C. and Patel, G.K. 2014. Human basal cell carcinoma tumor-initiating cells are resistant to Etoposide. *Journal of Investigative Dermatology* 134(3), pp. 867–870. doi: 10.1038/jid.2013.377.
Colomé-Tatché, M. and Theis, F.J. 2018. Statistical single cell multi-omics integration. *Current Opinion in Systems Biology* 7, pp. 54–59. doi: 10.1016/J.COISB.2018.01.003.
Cornel, A.M., Mimpen, I.L. and Nierkens, S. 2020. MHC Class I Downregulation in Cancer: Underlying Mechanisms and Potential Targets for Cancer Immunotherapy. *Cancers* 12(7), pp. 1–33. Available at: https://pubmed.ncbi.nlm.nih.gov/32630675/ [Accessed: 18 February 2024].
Crum-Cianflone, N. et al. 2009. Cutaneous Malignancies among HIV-Infected Persons. *Archives of internal medicine* 169(12), p. 1130. Available at: /pmc/articles/PMC2761839/ [Accessed: 10 February 2024].
Cruz-Muñoz, M.E., Valenzuela-Vázquez, L., Sánchez-Herrera, J. and Santa-Olalla Tapia, J. 2019. From the “missing self” hypothesis to adaptive NK cells: Insights of

NK cell-mediated effector functions in immune surveillance. *Journal of Leukocyte Biology* 105(5), pp. 955–971. Available at: https://onlinelibrary.wiley.com/doi/full/10.1002/JLB.MR0618-224RR [Accessed: 18 February 2024].

Curson, C. and Weedon, D. 1979. Spontaneous regression in basal cell carcinomas. *Journal of cutaneous pathology* 6(5), pp. 432–437. Available at: https://pubmed.ncbi.nlm.nih.gov/521534/ [Accessed: 23 December 2023].

Dart, A. 2017. Metastasis takes a different route. *Nature Reviews Cancer 2017 17:9* 17(9), pp. 509–509. Available at: https://www.nature.com/articles/nrc.2017.70 [Accessed: 2 February 2023].

Denef, N., Neubüser, D., Perez, L. and Cohen, S.M. 2000. Hedgehog induces opposite changes in turnover and subcellular localization of patched and smoothened. *Cell* 102(4), pp. 521–531. Available at: https://pubmed.ncbi.nlm.nih.gov/10966113/ [Accessed: 21 February 2024].

Dennler, S. et al. 2007. Induction of sonic hedgehog mediators by transforming growth factor-beta: Smad3-dependent activation of Gli2 and Gli1 expression in vitro and in vivo. *Cancer research* 67(14), pp. 6981–6986. Available at: https://pubmed.ncbi.nlm.nih.gov/17638910/ [Accessed: 23 February 2024].

Dennler, S., André, J., Verrechia, F. and Mauviel, A. 2009. Cloning of the human GLI2 Promoter: transcriptional activation by transforming growth factor-beta via SMAD3/beta-catenin cooperation. *The Journal of biological chemistry* 284(46), pp. 31523–31531. Available at: https://pubmed.ncbi.nlm.nih.gov/19797115/ [Accessed: 23 February 2024].

Denzer, L., Muranyi, W., Schroten, H. and Schwerk, C. 2023. The role of PLVAP in endothelial cells. *Cell and Tissue Research 2023* , pp. 1–20. Available at:

https://link.springer.com/article/10.1007/s00441-023-03741-1 [Accessed: 18 April 2023].
Dessinioti, C. and Stratigos, A.J. 2022. An Epidemiological Update on Indoor Tanning and the Risk of Skin Cancers. *Current Oncology* 29(11), p. 8886. Available at: /pmc/articles/PMC9689757/ [Accessed: 6 January 2024].
Dijkgraaf, G.J.P. et al. 2011. Small molecule inhibition of GDC-0449 refractory smoothened mutants and downstream mechanisms of drug resistance. *Cancer research* 71(2), pp. 435–444. Available at: https://pubmed.ncbi.nlm.nih.gov/21123452/ [Accessed: 21 February 2024].
Ding, Q. et al. 1999. Mouse suppressor of fused is a negative regulator of sonic hedgehog signaling and alters the subcellular distribution of Gli1. *Current biology : CB* 9(19), pp. 1119–1122. Available at: https://pubmed.ncbi.nlm.nih.gov/10531011/ [Accessed: 21 February 2024].
Dobin, A. et al. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1), p. 15. Available at: /pmc/articles/PMC3530905/ [Accessed: 12 February 2024].
Dunn, G.P., Old, L.J. and Schreiber, R.D. 2004. The immunobiology of cancer immunosurveillance and immunoediting. *Immunity* 21(2), pp. 137–148. Available at: https://pubmed.ncbi.nlm.nih.gov/15308095/ [Accessed: 18 February 2024].
Dyck, L. and Mills, K.H.G. 2017. Immune checkpoints and their inhibition in cancer and infectious diseases. *European journal of immunology* 47(5), pp. 765–779. Available at: https://pubmed.ncbi.nlm.nih.gov/28393361/ [Accessed: 19 February 2024].
Emami Nejad, A. et al. 2021. The role of hypoxia in the tumor microenvironment and development of cancer stem cell: a novel approach to developing treatment.

Cancer Cell International 2021 21:1 21(1), pp. 1–26. Available at: https://cancerci.biomedcentral.com/articles/10.1186/s12935-020-01719-5 [Accessed: 8 February 2023].

Epstein, E.H. 2008. Basal cell carcinomas: attack of the hedgehog. *Nature reviews. Cancer* 8(10), pp. 743–754. Available at: https://pubmed.ncbi.nlm.nih.gov/18813320/ [Accessed: 23 December 2023].

Eslami-S, Z., Cortés-Hernández, L.E. and Alix-Panabières, C. 2020. The Metastatic Cascade as the Basis for Liquid Biopsy Development. *Frontiers in Oncology* 10, p. 1055. doi: 10.3389/FONC.2020.01055/BIBTEX.

Euvrard, S. et al. 1995. Aggressive squamous cell carcinomas in organ transplant recipients. *Transplantation Proceedings* 27(2), pp. 1767–1768. Available at: https://www.ncbi.nlm.nih.gov/pubmed/7725494.

Evangelista, M. et al. 2008. Kinome siRNA screen identifies regulators of ciliogenesis and hedgehog signal transduction. *Science signaling* 1(39). Available at: https://pubmed.ncbi.nlm.nih.gov/18827223/ [Accessed: 21 February 2024].

Falcone, L.M. and Zeidler-Erdely, P.C. 2019. Skin cancer and welding. *Clinical and experimental dermatology* 44(2), p. 130. Available at: /pmc/articles/PMC6415292/ [Accessed: 7 January 2024].

Fan, H., Oro, A.E., Scott, M.P. and Khavari, P.A. 1997a. Induction of basal cell carcinoma features in transgenic human skin expressing Sonic Hedgehog. *Nature medicine* 3(7), pp. 788–794. Available at: https://pubmed.ncbi.nlm.nih.gov/9212109/ [Accessed: 23 December 2023].

Fan, H., Oro, A.E., Scott, M.P. and Khavari, P.A. 1997b. Induction of basal cell carcinoma features in transgenic human skin expressing Sonic Hedgehog. *Nature medicine* 3(7), pp. 788–794. Available at:

https://pubmed.ncbi.nlm.nih.gov/9212109/ [Accessed: 21 February 2024].
Fania, L. et al. 2020. Basal Cell Carcinoma: From Pathophysiology to Novel Therapeutic Approaches. *Biomedicines* 8(11), pp. 1–38. Available at: https://pubmed.ncbi.nlm.nih.gov/33113965/ [Accessed: 20 February 2024].
Favara, D.M. et al. 2019. ADGRL4/ELTD1 Silencing in Endothelial Cells Induces ACLY and SLC25A1 and Alters the Cellular Metabolic Profile. *Metabolites* 9(12). Available at: https://pubmed.ncbi.nlm.nih.gov/31775252/ [Accessed: 18 April 2023].
Ferrara, N., Gerber, H.P. and LeCouter, J. 2003. The biology of VEGF and its receptors. *Nature medicine* 9(6), pp. 669–676. Available at: https://pubmed.ncbi.nlm.nih.gov/12778165/ [Accessed: 8 February 2023].
First, M.R. 2004. Improving long-term renal transplant outcomes with tacrolimus: speculation vs evidence. *Nephrol Dial Transplant* , pp. 17–22. Available at: https://academic.oup.com/ndt/article/19/suppl_6/vi17/1822785 [Accessed: 6 May 2024].
Fitch, M.J., Campagnolo, L., Kuhnert, F. and Stuhlmann, H. 2004. Egfl7, a novel epidermal growth factor-domain gene expressed in endothelial cells. *Developmental Dynamics* 230(2), pp. 316–324. Available at: https://onlinelibrary.wiley.com/doi/full/10.1002/dvdy.20063 [Accessed: 18 April 2023].
Fleming, I.D., Barnawell, J.R., Burlison, P.E. and Rankin, J.S. 1975. SKIN CANCER IN BLACK PATIENTS. *Cancer* 35, pp. 600–605. Available at: https://acsjournals.onlinelibrary.wiley.com/doi/10.1002/1097-0142 [Accessed: 5 January 2024].
Folkman, J. 1971. Tumor angiogenesis: therapeutic implications. *The New England journal of medicine* 285(21), pp. 1182–1186. Available at:

https://pubmed.ncbi.nlm.nih.gov/4938153/ [Accessed: 8 February 2023].
Folkman, J. and Shing, Y. 1992. Angiogenesis. *Journal of Biological Chemistry* 267(16), pp. 10931–4.
Folkman, J., Watson, K., Ingber, D. and Hanahan, D. 1989. Induction of angiogenesis during the transition from hyperplasia to neoplasia. *Nature* 339(6219), pp. 58–61. Available at: https://pubmed.ncbi.nlm.nih.gov/2469964/ [Accessed: 23 February 2024].
Forcato, M., Romano, O. and Bicciato, S. 2021. Computational methods for the integrative analysis of single-cell data. *Briefings in bioinformatics* 22(1), pp. 20–29. Available at: https://pubmed.ncbi.nlm.nih.gov/32363378/ [Accessed: 15 May 2023].
Fouad, Y.A. and Aanei, C. 2017. Revisiting the hallmarks of cancer. *American Journal of Cancer Research* 7(5), pp. 1016–1036.
Fraisl, P., Mazzone, M., Schmidt, T. and Carmeliet, P. 2009. Regulation of angiogenesis by oxygen and metabolism. *Developmental cell* 16(2), pp. 167–179. Available at: https://pubmed.ncbi.nlm.nih.gov/19217420/ [Accessed: 8 February 2023].
Fuchs, Y. and Steller, H. 2011. Programmed cell death in animal development and disease. *Cell* 147(4), pp. 742–758. Available at: https://pubmed.ncbi.nlm.nih.gov/22078876/ [Accessed: 20 February 2023].
Gailani, M.R. et al. 1996. The role of the human homologue of Drosophila patched in sporadic basal cell carcinomas. *Nature Genetics* 14(1), pp. 78–81. doi: 10.1038/NG0996-78.
Gangan, R. 2022. Basal cell carcinoma: Epidemiology. *Journal of Skin and Sexually Transmitted Diseases* 4(2), pp. 157–163. Available at: https://jsstd.org/basal-cell-

carcinoma-epidemiology/ [Accessed: 5 January 2024].
Ganier, C. et al. 2024. Multiscale spatial mapping of cell populations across anatomical sites in healthy human skin and basal cell carcinoma. *Proceedings of the National Academy of Sciences* 121(2), p. e2313326120. Available at: https://www.pnas.org/doi/abs/10.1073/pnas.2313326120 [Accessed: 21 February 2024].
Geindreau, M., Ghiringhelli, F. and Bruchard, M. 2021. Vascular Endothelial Growth Factor, a Key Modulator of the Anti-Tumor Immune Response. *International journal of molecular sciences* 22(9). Available at: https://pubmed.ncbi.nlm.nih.gov/34064508/ [Accessed: 18 February 2024].
George, E.A., Baranwal, N., Kang, J.H., Qureshi, A.A., Drucker, A.M. and Cho, E. 2021. Photosensitizing Medications and Skin Cancer: A Comprehensive Review. *Cancers* 13(10). Available at: /pmc/articles/PMC8152064/ [Accessed: 20 February 2024].
Gilbert, E.S. 2009. Ionizing Radiation and Cancer Risks: What Have We Learned From Epidemiology? *International journal of radiation biology* 85(6), p. 467. Available at: /pmc/articles/PMC2859619/ [Accessed: 6 January 2024].
Gloster, H.M. and Brodland, D.G. 1996. The epidemiology of skin cancer. *Dermatologic Surgery* 22(3), p. 2170226.
Gnagnarella, P., Pasquali, E., Serrano, D., Raimondi, S., Disalvatore, D. and Gandini, S. 2014. Vitamin D receptor polymorphism FokI and cancer risk: a comprehensive meta-analysis. *Carcinogenesis* 35(9), pp. 1913–1919. Available at: https://pubmed.ncbi.nlm.nih.gov/25053622/ [Accessed: 21 February 2024].
Go, U., Miyata, K., Fujita, M., Ohide, T. and Mitsuishi, T. 2019. Spontaneous Regression of Annular Basal Cell Carcinoma: A Case Report. *Case Reports in*

Dermatology 11(2), p. 145. Available at: /pmc/articles/PMC6587202/ [Accessed: 10 February 2024].

Goncharov, N. V., Popova, P.I., Avdonin, P.P., Kudryavtsev, I. V., Serebryakova, M.K., Korf, E.A. and Avdonin, P. V. 2020. Markers of Endothelial Cells in Normal and Pathological Conditions. *Biochemistry (Moscow) Supplement. Series A, Membrane and cell biology* 14(3), pp. 167–183. Available at: https://pubmed.ncbi.nlm.nih.gov/33072245/ [Accessed: 18 April 2023].

Gorczynski, R.M. 2012. CD200:CD200R-Mediated Regulation of Immunity. *ISRN Immunology* 2012, pp. 1–18. doi: 10.5402/2012/682168.

Gorczynski, R.M., Chen, Z., Hu, J., Kai, Y. and Lei, J. 2001. Evidence of a role for CD200 in regulation of immune rejection of leukaemic tumour cells in C57BL/6 mice. *Clinical and Experimental Immunology* 126(2), p. 220. Available at: /pmc/articles/PMC1906195/ [Accessed: 11 February 2024].

Grachtchouk, M., Mo, R., Yu, S., Zhang, X., Sasaki, H., Hui, C.C. and Dlugosz, A.A. 2000. Basal cell carcinomas in mice overexpressing Gli2 in skin. *Nature genetics* 24(3), pp. 216–217. Available at: https://pubmed.ncbi.nlm.nih.gov/10700170/ [Accessed: 23 December 2023].

Green, A. et al. 2007. The association of use of sunbeds with cutaneous malignant melanoma and other skin cancers: A systematic review. *International journal of cancer* 120(5), pp. 1116–1122. Available at: https://pubmed.ncbi.nlm.nih.gov/17131335/ [Accessed: 4 January 2024].

Griffiths, R.W., Suvarna, S.K. and Stone, J. 2005. Do basal cell carcinomas recur after complete conventional surgical excision? *British journal of plastic surgery* 58(6), pp. 795–805. Available at: https://pubmed.ncbi.nlm.nih.gov/16086990/ [Accessed: 21 February 2024].

de Groot, S.C., Ulrich, M.M.W., Gho, C.G. and Huisman, M.A. 2021. Back to the Future: From Appendage Development Toward Future Human Hair Follicle Neogenesis. *Frontiers in Cell and Developmental Biology* 9, p. 661787. doi: 10.3389/FCELL.2021.661787/BIBTEX.

Gschwind, A., Fischer, O.M. and Ullrich, A. 2004. The discovery of receptor tyrosine kinases: targets for cancer therapy. *Nature reviews. Cancer* 4(5), pp. 361–370. Available at: https://pubmed.ncbi.nlm.nih.gov/15122207/ [Accessed: 26 January 2023].

Gubin, M.M. and Vesely, M.D. 2022. Cancer Immunoediting in the Era of Immuno-oncology. *Clinical Cancer Research* 28(18), pp. 3917–3928. Available at: /clincancerres/article/28/18/3917/709011/Cancer-Immunoediting-in-the-Era-of-Immuno [Accessed: 18 February 2024].

Guerrero-juarez, C.F. et al. 2022. Single-cell analysis of human basal cell carcinoma reveals novel regulators of tumor growth and the tumor microenvironment. *Science Advances* 8(23), pp. 1–18. Available at: https://www.science.org/doi/10.1126/sciadv.abm7981.

Guerrero-Juarez, C.F. et al. 2022. Single-cell analysis of human basal cell carcinoma reveals novel regulators of tumor growth and the tumor microenvironment. *Science Advances* 8(23), p. 7981. Available at: /pmc/articles/PMC9187229/ [Accessed: 2 May 2023].

Guo, H., Carlson, J.A. and Slominski, A. 2012. Role of TRPM in melanocytes and melanoma. *Experimental dermatology* 21(9), pp. 650–654. Available at: https://pubmed.ncbi.nlm.nih.gov/22897572/ [Accessed: 18 April 2023].

Hafemeister, C. and Satija, R. 2019a. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome*

Biology 20(1), pp. 1–15. Available at: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1874-1 [Accessed: 16 May 2023].

Hafemeister, C. and Satija, R. 2019b. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *bioRxiv* , p. 576827. Available at: https://www.biorxiv.org/content/10.1101/576827v2 [Accessed: 11 April 2023].

Haghverdi, L., Büttner, M., Wolf, F.A., Buettner, F. and Theis, F.J. 2016. Diffusion pseudotime robustly reconstructs lineage branching. *Nature Methods 2016 13:10* 13(10), pp. 845–848. Available at: https://www.nature.com/articles/nmeth.3971 [Accessed: 8 August 2023].

Hahn, H. et al. 1996. Mutations of the Human Homolog of Drosophila patched in the Nevoid Basal Cell Carcinoma Syndrome. *Cell* 85(6), pp. 841–851. doi: 10.1016/S0092-8674(00)81268-4.

Han, L.O., Li, X.Y., Cao, M.M., Cao, Y. and Zhou, L.H. 2018. Development and validation of an individualized diagnostic signature in thyroid cancer. *Cancer Medicine* 7(4), pp. 1135–1140. Available at: https://onlinelibrary.wiley.com/doi/full/10.1002/cam4.1397 [Accessed: 4 February 2024].

Hanahan, D. 2022. Hallmarks of Cancer: New Dimensions. *Cancer discovery* 12(1), pp. 31–46. Available at: https://pubmed.ncbi.nlm.nih.gov/35022204/ [Accessed: 26 January 2023].

Hanahan, D. and Folkman, J. 1996. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell* 86(3), pp. 353–364. Available at: https://pubmed.ncbi.nlm.nih.gov/8756718/ [Accessed: 8 February 2023].

Hanahan, D. and Weinberg, R.A. 2000. The hallmarks of cancer. *Cell* 100(1), pp. 57–70. Available at: https://pubmed.ncbi.nlm.nih.gov/10647931/ [Accessed: 26 January 2023].

Hanahan, D. and Weinberg, R.A. 2011. Hallmarks of cancer: the next generation. *Cell* 144(5), pp. 646–674. Available at: https://pubmed.ncbi.nlm.nih.gov/21376230/ [Accessed: 26 January 2023].

Harken, E.B.O. and Fazio, J. 2022. Basal Cell Carcinoma. *Atlas of Dermatologic Diseases in Solid Organ Transplant Recipients* , pp. 331–346. Available at: https://www.ncbi.nlm.nih.gov/books/NBK482439/ [Accessed: 5 January 2024].

Hashimoto, T. and Shibasaki, F. 2015. Hypoxia-Inducible Factor as an Angiogenic Master Switch. *Frontiers in Pediatrics* 3. Available at: https://pubmed.ncbi.nlm.nih.gov/25964891/ [Accessed: 8 February 2023].

Hashimshony, T., Wagner, F., Sher, N. and Yanai, I. 2012. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell reports* 2(3), pp. 666–673. Available at: https://pubmed.ncbi.nlm.nih.gov/22939981/ [Accessed: 4 February 2024].

Hatherley, D., Lea, S.M., Johnson, S. and Barclay, A.N. 2013. Structures of CD200/CD200 receptor family and implications for topology, regulation, and evolution. *Structure (London, England : 1993)* 21(5), pp. 820–832. Available at: https://pubmed.ncbi.nlm.nih.gov/23602662/ [Accessed: 23 February 2024].

Hay, R.J. et al. 2014. The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions. *The Journal of investigative dermatology* 134(6), pp. 1527–1534. Available at: https://pubmed.ncbi.nlm.nih.gov/24166134/ [Accessed: 19 February 2024].

Hayflick, L. and Moorhead, P.S. 1961. The serial cultivation of human diploid cell strains. *Experimental Cell Research* 25(3), pp. 585–621. doi: 10.1016/0014-

4827(61)90192-6.
He, X. and Xu, C. 2020. Immune checkpoint signaling and cancer immunotherapy. *Cell Research 2020 30:8* 30(8), pp. 660–669. Available at: https://www.nature.com/articles/s41422-020-0343-4 [Accessed: 19 February 2024].
Heltoft, K.N., Slagor, R.M., Agner, T. and Bonde, J.P. 2017. Metal arc welding and the risk of skin cancer. *International archives of occupational and environmental health* 90(8), pp. 873–881. Available at: https://pubmed.ncbi.nlm.nih.gov/28766013/ [Accessed: 7 January 2024].
Van den Heuvel, M. and Ingham, P.W. 1996. smoothened encodes a receptor-like serpentine protein required for hedgehog signalling. *Nature 1996 382:6591* 382(6591), pp. 547–551. Available at: https://www.nature.com/articles/382547a0 [Accessed: 21 February 2024].
Hicklin, D.J. and Ellis, L.M. 2005. Role of the vascular endothelial growth factor pathway in tumor growth and angiogenesis. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 23(5), pp. 1011–1027. Available at: https://pubmed.ncbi.nlm.nih.gov/15585754/ [Accessed: 8 February 2023].
Hicks, S.C., Liu, R., Ni, Y., Purdom, E. and Risso, D. 2021. mbkmeans: Fast clustering for single cell data using mini-batch k-means. *PLOS Computational Biology* 17(1), p. e1008625. Available at: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008625 [Accessed: 24 April 2023].
Hockenbery, D., Nuñez, G., Milliman, C., Schreiber, R.D. and Korsmeyer, S.J. 1990. Bcl-2 is an inner mitochondrial membrane protein that blocks programmed cell death. *Nature* 348(6299), pp. 334–336. Available at:

https://pubmed.ncbi.nlm.nih.gov/2250705/ [Accessed: 20 February 2023].
Hoek, R.H. et al. 2000. Down-regulation of the macrophage lineage through interaction with OX2 (CD200). *Science (New York, N.Y.)* 290(5497), pp. 1768–1771. Available at: https://pubmed.ncbi.nlm.nih.gov/11099416/ [Accessed: 23 December 2023].
Von Hoff, D.D. et al. 2009. Inhibition of the hedgehog pathway in advanced basal-cell carcinoma. *The New England journal of medicine* 361(12), pp. 1164–1172. Available at: https://pubmed.ncbi.nlm.nih.gov/19726763/ [Accessed: 21 February 2024].
Hooper, J.E. and Scott, M.P. 1989. The Drosophila patched gene encodes a putative membrane protein required for segmental patterning. *Cell* 59(4), pp. 751–765. Available at: https://pubmed.ncbi.nlm.nih.gov/2582494/ [Accessed: 21 February 2024].
Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 24(6), pp. 417–441. doi: 10.1037/H0071325.
Hsu, Y.-C., Pasolli, A.H. and Fuchs, E. 2011. Dynamics Between Stem Cells, Niche and Progeny in the Hair Follicle. *Cell* 144(1), pp. 92–105. doi: 10.1016/j.cell.2010.11.049.
Hu, Y.L., DeLay, M., Jahangiri, A., Molinaro, A.M., Rose, S.D., Carbonell, W.S. and Aghi, M.K. 2012. Hypoxia-induced autophagy promotes tumor cell survival and adaptation to anti-angiogenic treatment in glioblastoma. *Cancer Research* 72(7), p. 1773. Available at: /pmc/articles/PMC3319869/ [Accessed: 23 February 2024].
Huang, H.W., Lee, C.H. and Yu, H.S. 2019. Arsenic-Induced Carcinogenesis and Immune Dysregulation. *International Journal of Environmental Research and Public Health* 16(15). Available at: /pmc/articles/PMC6696092/ [Accessed: 6 January

2024].
Hui, C.C. and Angers, S. 2011. Gli proteins in development and disease. *Annual review of cell and developmental biology* 27, pp. 513–537. Available at: https://pubmed.ncbi.nlm.nih.gov/21801010/ [Accessed: 21 February 2024].
Humke, E.W., Dorn, K. V., Milenkovic, L., Scott, M.P. and Rohatgi, R. 2010. The output of Hedgehog signaling is controlled by the dynamic association between Suppressor of Fused and the Gli proteins. *Genes & development* 24(7), pp. 670–682. Available at: https://pubmed.ncbi.nlm.nih.gov/20360384/ [Accessed: 21 February 2024].
Ikushima, H. and Miyazono, K. 2010. TGFbeta signalling: a complex web in cancer progression. *Nature reviews. Cancer* 10(6), pp. 415–424. Available at: https://pubmed.ncbi.nlm.nih.gov/20495575/ [Accessed: 23 February 2024].
Ilicic, T., Kim, J.K., Kolodziejczyk, A.A., Bagger, F.O., McCarthy, D.J., Marioni, J.C. and Teichmann, S.A. 2016. Classification of low quality cells from single-cell RNA-seq data. *Genome Biology* 17(1). Available at: /pmc/articles/PMC4758103/ [Accessed: 29 March 2023].
Ingham, P.W. and McMahon, A.P. 2001. Hedgehog signaling in animal development: paradigms and principles. *Genes & development* 15(23), pp. 3059–3087. Available at: https://pubmed.ncbi.nlm.nih.gov/11731473/ [Accessed: 21 February 2024].
Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.B., Lönnerberg, P. and Linnarsson, S. 2011. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome research* 21(7), pp. 1160–1167. Available at: https://pubmed.ncbi.nlm.nih.gov/21543516/ [Accessed: 4 February 2024].
Jaju, P.D., Ransohoff, K.J., Tang, J.Y. and Sarin, K.Y. 2016. Familial skin cancer

syndromes: Increased risk of nonmelanotic skin cancers and extracutaneous tumors. *Journal of the American Academy of Dermatology* 74(3), pp. 437–451. Available at: http://dx.doi.org/10.1016/j.jaad.2015.08.073.

Ji, A.L. et al. 2020. Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. *Cell* 182(2), p. 497. Available at: /pmc/articles/PMC7391009/ [Accessed: 2 May 2023].

Johnson, R.L. et al. 1996. Human Homolog of patched, a Candidate Gene for the Basal Cell Nevus Syndrome. *Science* 272(5268), pp. 1668–1671. Available at: https://www.science.org/doi/10.1126/science.272.5268.1668 [Accessed: 10 January 2024].

Ju, T., Hernandez, L., Mohsin, N., Labib, A., Frech, F. and Nouri, K. 2023. Evaluation of risk in chronic cutaneous inflammatory conditions for malignant transformation. *Journal of the European Academy of Dermatology and Venereology* 37(2), pp. 231–242. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/jdv.18663 [Accessed: 8 January 2024].

Kaplan, B. and Moy, R.L. 2000. Effect of perilesional injections of PEG-interleukin-2 on basal cell carcinoma. *Dermatologic surgery : official publication for American Society for Dermatologic Surgery [et al.]* 26(11), pp. 1037–1040. Available at: https://pubmed.ncbi.nlm.nih.gov/11096390/ [Accessed: 10 February 2024].

Kaporis, H.G. et al. 2007. Human basal cell carcinoma is associated with Foxp3+ T cells in a Th2 dominant microenvironment. *The Journal of investigative dermatology* 127(10), pp. 2391–2398. Available at: https://pubmed.ncbi.nlm.nih.gov/17508019/ [Accessed: 23 December 2023].

Karl Pearson 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of*

Science 2(11), pp. 559–572. Available at: https://www.tandfonline.com/doi/abs/10.1080/14786440109462720 [Accessed: 13 April 2023].

Kaushik, V., Kulkarni, Y., Felix, K., Azad, N., Iyer, A.K. V. and Yakisich, J.S. 2021. Alternative models of cancer stem cells: The stemness phenotype model, 10 years later. *World Journal of Stem Cells* 13(7), p. 934. Available at: /pmc/articles/PMC8316871/ [Accessed: 5 May 2024].

Kawasaki, B.T. and Farrar, W.L. 2008. Cancer stem cells, CD200 and immunoevasion. *Trends in immunology* 29(10), pp. 464–468. Available at: https://pubmed.ncbi.nlm.nih.gov/18775673/ [Accessed: 20 January 2024].

Khalesi, M., Whiteman, D.C., Tran, B., Kimlin, M.G., Olsen, C.M. and Neale, R.E. 2013. A meta-analysis of pigmentary characteristics, sun sensitivity, freckling and melanocytic nevi and risk of basal cell carcinoma of the skin. *Cancer Epidemiology* 37(5), pp. 534–543. doi: 10.1016/J.CANEP.2013.05.008.

Kibriya, M.G. et al. 2022. Interaction of Arsenic Exposure and Transcriptomic Profile in Basal Cell Carcinoma. *Cancers* 14(22), p. 5598. Available at: https://www.mdpi.com/2072-6694/14/22/5598/htm [Accessed: 6 January 2024].

Kieran, M.W., Kalluri, R. and Cho, Y.J. 2012. The VEGF Pathway in Cancer and Disease: Responses, Resistance, and the Path Forward. *Cold Spring Harbor Perspectives in Medicine* 2(12). Available at: /pmc/articles/PMC3543071/ [Accessed: 8 February 2023].

Kilgour, J.M., Jia, J.L. and Sarin, K.Y. 2021. Review of the Molecular Genetics of Basal Cell Carcinoma; Inherited Susceptibility, Somatic Mutations, and Targeted Therapeutics. *Cancers* 13(15). Available at: /pmc/articles/PMC8345475/ [Accessed: 10 January 2024].

Kim, D., Paggi, J.M., Park, C., Bennett, C. and Salzberg, S.L. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology 2019 37:8* 37(8), pp. 907–915. Available at: https://www.nature.com/articles/s41587-019-0201-4 [Accessed: 24 February 2023].

Kim, M.Y., Park, H.J., Baek, S.C., Byun, D.G. and Houh, D. 2002. Mutations of the p53 and PTCH gene in basal cell carcinomas: UV mutation signature and strand bias. *Journal of Dermatological Science* 29(1), pp. 1–9. Available at: https://pubmed.ncbi.nlm.nih.gov/12007715/ [Accessed: 21 February 2024].

Kim, S.K. and Cho, S.W. 2022. The Evasion Mechanisms of Cancer Immunity and Drug Intervention in the Tumor Microenvironment. *Frontiers in Pharmacology* 13. Available at: /pmc/articles/PMC9171538/ [Accessed: 18 February 2024].

Kiselev, V.Y., Andrews, T.S. and Hemberg, M. 2019. Challenges in unsupervised clustering of single-cell RNA-seq data. *Nature Reviews Genetics 2018 20:5* 20(5), pp. 273–282. Available at: https://www.nature.com/articles/s41576-018-0088-9 [Accessed: 24 April 2023].

Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C. and Teichmann, S.A. 2015. The Technology and Biology of Single-Cell RNA Sequencing. *Molecular Cell* 58(4), pp. 610–620. doi: 10.1016/J.MOLCEL.2015.04.005.

Kong, D., Hughes, C.J. and Ford, H.L. 2020. Cellular Plasticity in Breast Cancer Progression and Therapy. *Frontiers in Molecular Biosciences* 7, p. 72. Available at: /pmc/articles/PMC7194153/ [Accessed: 5 May 2024].

Kool, M. et al. 2008. Integrated genomics identifies five medulloblastoma subtypes with distinct genetic profiles, pathway signatures and clinicopathological features. *PloS one* 3(8). Available at: https://pubmed.ncbi.nlm.nih.gov/18769486/ [Accessed:

21 February 2024].
Kool, M. et al. 2014. Genome sequencing of SHH medulloblastoma predicts genotype-related response to smoothened inhibition. *Cancer cell* 25(3), pp. 393–405. Available at: https://pubmed.ncbi.nlm.nih.gov/24651015/ [Accessed: 21 February 2024].
Korsunsky, I. et al. 2019. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods 2019 16:12* 16(12), pp. 1289–1296. Available at: https://www.nature.com/articles/s41592-019-0619-0 [Accessed: 15 May 2023].
Kovacs, J.J. et al. 2008. Beta-arrestin-mediated localization of smoothened to the primary cilium. *Science (New York, N.Y.)* 320(5884), pp. 1777–1781. Available at: https://pubmed.ncbi.nlm.nih.gov/18497258/ [Accessed: 21 February 2024].
Kricker, A., Armstrong, B.K., English, D.R. and Heenan, P.J. 1995. Does intermittent sun exposure cause basal cell carcinoma? a case-control study in Western Australia. *International journal of cancer* 60(4), pp. 489–494. Available at: https://pubmed.ncbi.nlm.nih.gov/7829262/ [Accessed: 4 January 2024].
Labani-Motlagh, A., Ashja-Mahdavi, M. and Loskog, A. 2020. The Tumor Microenvironment: A Milieu Hindering and Obstructing Antitumor Immune Responses. *Frontiers in Immunology* 11. Available at: https://pubmed.ncbi.nlm.nih.gov/32499786/ [Accessed: 18 February 2024].
Lafzi, A., Moutinho, C., Picelli, S. and Heyn, H. 2018. Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies. *Nature Protocols 2018 13:12* 13(12), pp. 2742–2757. Available at: https://www.nature.com/articles/s41596-018-0073-y [Accessed: 4 April 2023].
Lähnemann, D. et al. 2020. Eleven grand challenges in single-cell data science. *Genome biology* 21(1). Available at: https://pubmed.ncbi.nlm.nih.gov/32033589/

[Accessed: 15 May 2023].
Lai, C.H. et al. 2021. RAMP2-AS1 Regulates Endothelial Homeostasis and Aging. *Frontiers in Cell and Developmental Biology* 9, p. 213. doi: 10.3389/FCELL.2021.635307/BIBTEX.
Lashway, S.G., Worthen, A.D.M., Abuasbeh, J.N., Harris, R.B., Farland, L. V., O'Rourke, M.K. and Dennis, L.K. 2023. A meta-analysis of sunburn and basal cell carcinoma risk. *Cancer Epidemiology* 85, p. 102379. doi: 10.1016/J.CANEP.2023.102379.
Lee, T. et al. 2015. Occupational ionising radiation and risk of basal cell carcinoma in US radiologic technologists (1983-2005). *Occupational and environmental medicine* 72(12), pp. 862–869. Available at: https://pubmed.ncbi.nlm.nih.gov/26350677/ [Accessed: 7 January 2024].
Lemmon, M.A. and Schlessinger, J. 2010. Cell signaling by receptor-tyrosine kinases. *Cell* 141(7), p. 1117. Available at: /pmc/articles/PMC2914105/ [Accessed: 26 January 2023].
Lesiak, A. et al. 2011. An enhanced risk of basal cell carcinoma is associated with particular polymorphisms in the VDR and MTHFR genes. *Experimental dermatology* 20(10), pp. 800–804. Available at: https://pubmed.ncbi.nlm.nih.gov/21732987/ [Accessed: 21 February 2024].
Lever, J., Krzywinski, M. and Altman, N. 2017. Points of Significance: Principal component analysis. *Nature Methods* 14(7), pp. 641–642. doi: 10.1038/NMETH.4346.
Li, C. et al. 2007. Identification of pancreatic cancer stem cells. *Cancer research* 67(3), pp. 1030–1037. Available at: https://pubmed.ncbi.nlm.nih.gov/17283135/ [Accessed: 23 February 2024].

Li, S. et al. 2023. A relay velocity model infers cell-dependent RNA velocity. *Nature Biotechnology 2023* , pp. 1–10. Available at: https://www.nature.com/articles/s41587-023-01728-5 [Accessed: 8 August 2023].

Li, S., Sun, Y. and Gao, D. 2013. Role of the nervous system in cancer metastasis. *Oncology Letters* 5(4), p. 1101. Available at: /pmc/articles/PMC3629128/ [Accessed: 2 February 2023].

Li, X. and Wang, C.Y. 2021. From bulk, single-cell to spatial RNA sequencing. *International Journal of Oral Science 2021 13:1* 13(1), pp. 1–6. Available at: https://www.nature.com/articles/s41368-021-00146-0 [Accessed: 4 February 2024].

Liao, Y., Smyth, G.K. and Shi, W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30(7), pp. 923–930. Available at: https://academic.oup.com/bioinformatics/article/30/7/923/232889 [Accessed: 24 February 2023].

Lichter, M.D., Karagas, M.R., Mott, L.A., Spencer, S.K., Stukel, T.A. and Greenberg, E.R. 2000. Therapeutic Ionizing Radiation and the Incidence of Basal Cell Carcinoma and Squamous Cell Carcinoma. *Archives of Dermatology* 136(8), pp. 1007–1011. Available at: https://jamanetwork.com/journals/jamadermatology/fullarticle/190516 [Accessed: 7 January 2024].

Lim, D., Roh, T., Kim, M., … Y.K.-J. of T. and 2018, undefined 2018. Non-cancer, cancer, and dermal sensitization risk assessment of heavy metals in cosmetics. *Taylor & FrancisDS Lim, TH Roh, MK Kim, YC Kwon, SM Choi, SJ Kwack, KB Kim, S Yoon, HS Kim, BM LeeJournal of Toxicology and Environmental Health, Part A,*

2018•Taylor & Francis 81(11), pp. 432–452. Available at: https://www.tandfonline.com/doi/abs/10.1080/15287394.2018.1451191 [Accessed: 6 January 2024].

Litvinov, I. V., Xie, P., Gunn, S., Sasseville, D. and Lefrançois, P. 2021. The transcriptional landscape analysis of basal cell carcinomas reveals novel signalling pathways and actionable targets. *Life Science Alliance* 4(7). Available at: /pmc/articles/PMC8200290/ [Accessed: 12 January 2024].

Liu, J.-Q. et al. 2016. A Critical Role for CD200R Signaling in Limiting the Growth and Metastasis of CD200+ Melanoma. *Journal of immunology (Baltimore, Md. : 1950)* 197(4), pp. 1489–1497. Available at: https://pubmed.ncbi.nlm.nih.gov/27385779/ [Accessed: 23 February 2024].

Liu, J., Dang, H. and Wang, X.W. 2018. The significance of intertumor and intratumor heterogeneity in liver cancer. *Experimental & Molecular Medicine 2018 50:1* 50(1), pp. e416–e416. Available at: https://www.nature.com/articles/emm2017165 [Accessed: 4 February 2024].

Liu, Y., Xie, S., Wang, Y., Luo, K., Wang, Y. and Cai, Y. 2012. Liquiritigenin Inhibits Tumor Growth and Vascularization in a Mouse Model of Hela Cells. *Molecules 2012, Vol. 17, Pages 7206-7216* 17(6), pp. 7206–7216. Available at: https://www.mdpi.com/1420-3049/17/6/7206/htm [Accessed: 9 February 2024].

Lkić, D. et al. 1991. Basal Cell Carcinoma Treated with Interferon. *International Journal of Dermatology* 30(10), pp. 734–737. doi: 10.1111/j.1365-4362.1991.tb02623.x.

Lo, P.K., Kanojia, D., Liu, X., Singh, U.P., Berger, F.G., Wang, Q. and Chen, H. 2012. CD49f and CD61 identify Her2/neu-induced mammary tumor-initiating cells that are potentially derived from luminal progenitors and maintained by the integrin-TGFβ

signaling. *Oncogene* 31(21), pp. 2614–2626. Available at: https://pubmed.ncbi.nlm.nih.gov/21996747/ [Accessed: 23 February 2024].
Lobo, D. V., Chu, P., Grekin, R.C. and Berger, T.G. 1992. Nonmelanoma Skin Cancers and Infection With the Human Immunodeficiency Virus. *Archives of Dermatology* 128(5), pp. 623–627. Available at: https://jamanetwork.com/journals/jamadermatology/fullarticle/554059 [Accessed: 23 December 2023].
Lopez, J. and Tait, S.W.G. 2015. Mitochondrial apoptosis: killing cancer using the enemy within. *British journal of cancer* 112(6), pp. 957–962. Available at: https://pubmed.ncbi.nlm.nih.gov/25742467/ [Accessed: 20 February 2023].
Love, M.I., Huber, W. and Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15(12), pp. 1–21. Available at: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8 [Accessed: 27 February 2023].
Lu, F.J. 1990. Blackfoot disease: arsenic or humic acid? *Lancet (London, England)* 336(8707), pp. 115–116. Available at: https://pubmed.ncbi.nlm.nih.gov/1975296/ [Accessed: 6 January 2024].
Lugano, R., Ramachandran, M. and Dimberg, A. 2020. Tumor angiogenesis: causes, consequences, challenges and opportunities. *Cellular and Molecular Life Sciences* 77(9), pp. 1745–1770. Available at: https://link.springer.com/article/10.1007/s00018-019-03351-7 [Accessed: 8 February 2023].
Lugassy, C. et al. 2004. Pericyte-like location of GFP-tagged melanoma cells: ex vivo and in vivo studies of extravascular migratory metastasis. *The American journal of pathology* 164(4), pp. 1191–1198. Available at:

https://pubmed.ncbi.nlm.nih.gov/15039208/ [Accessed: 2 February 2023].
Lukowiak, T.M. et al. 2020. Association of Age, Sex, Race, and Geographic Region With Variation of the Ratio of Basal Cell to Cutaneous Squamous Cell Carcinomas in the United States. *JAMA Dermatology* 156(11), pp. 1192–1198. Available at: https://jamanetwork.com/journals/jamadermatology/fullarticle/2769599 [Accessed: 5 January 2024].
Lun, A. 2018. Overcoming systematic errors caused by log-transformation of normalized single-cell RNA sequencing data. *bioRxiv* , p. 404962. Available at: https://www.biorxiv.org/content/10.1101/404962v1 [Accessed: 5 April 2023].
Lupu, M., Caruntu, C., Popa, M.I., Voiculescu, V.M., Zurac, S. and Boda, D. 2019. Vascular patterns in basal cell carcinoma: Dermoscopic, confocal and histopathological perspectives (Review). *Oncology Letters* 17(5), pp. 4112–4125. Available at: https://pubmed.ncbi.nlm.nih.gov/30944604/ [Accessed: 23 February 2024].
Lv, X. et al. 2016. The role of hypoxia-inducible factors in tumor angiogenesis and cell metabolism. *Genes & diseases* 4(1), pp. 19–24. Available at: https://pubmed.ncbi.nlm.nih.gov/30258904/ [Accessed: 8 February 2023].
Lytal, N., Ran, D. and An, L. 2020. Normalization Methods on Single-Cell RNA-seq Data: An Empirical Survey. *Frontiers in Genetics* 11, p. 41. doi: 10.3389/FGENE.2020.00041/BIBTEX.
Ma, A., McDermaid, A., Xu, J., Chang, Y. and Ma, Q. 2020a. Integrative Methods and Practical Challenges for Single-Cell Multi-omics. *Trends in biotechnology* 38(9), pp. 1007–1022. Available at: https://pubmed.ncbi.nlm.nih.gov/32818441/ [Accessed: 15 May 2023].
Ma, Y., Yu, P., Lin, S., Li, Q., Fang, Z. and Huang, Z. 2020b. The association between

nonsteroidal anti-inflammatory drugs and skin cancer: Different responses in American and European populations. *Pharmacological research* 152. Available at: https://pubmed.ncbi.nlm.nih.gov/31689521/ [Accessed: 20 February 2024].
Maaten, L. van der and Hinton, G. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9(86), pp. 2579–2605. Available at: http://jmlr.org/papers/v9/vandermaaten08a.html [Accessed: 25 April 2023].
Mackeh, R., Marr, A.K., Fadda, A. and Kino, T. 2018. C2H2-Type Zinc Finger Proteins: Evolutionarily Old and New Partners of the Nuclear Hormone Receptors. *Nuclear receptor signaling* 15. Available at: https://journals.sagepub.com/doi/full/10.1177/1550762918801071 [Accessed: 11 January 2024].
Macosko, E.Z. et al. 2015. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161(5), pp. 1202–1214. Available at: http://www.cell.com/article/S0092867415005498/fulltext [Accessed: 18 April 2023].
Mahamat-Saleh, Y., Aune, D. and Schlesinger, S. 2020. 25-Hydroxyvitamin D status, vitamin D intake, and skin cancer risk: a systematic review and dose-response meta-analysis of prospective studies. *Scientific reports* 10(1). Available at: https://pubmed.ncbi.nlm.nih.gov/32753685/ [Accessed: 21 February 2024].
Majidpoor, J. and Mortezaee, K. 2021. Angiogenesis as a hallmark of solid tumors - clinical perspectives. *Cellular oncology (Dordrecht)* 44(4), pp. 715–737. Available at: https://pubmed.ncbi.nlm.nih.gov/33835425/ [Accessed: 8 February 2023].
Malta, T.M. et al. 2018. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* 173(2), pp. 338-354.e15. Available at: https://pubmed.ncbi.nlm.nih.gov/29625051/ [Accessed: 21 August 2023].

Malumbres, M. and Barbacid, M. 2009. Cell cycle, CDKs and cancer: a changing paradigm. *Nature reviews. Cancer* 9(3), pp. 153–166. Available at: https://pubmed.ncbi.nlm.nih.gov/19238148/ [Accessed: 27 January 2023].

Mancuso, M. et al. 2004. Basal cell carcinoma and its development: insights from radiation-induced tumors in Ptch1-deficient mice. *Cancer research* 64(3), pp. 934–941. Available at: https://pubmed.ncbi.nlm.nih.gov/14871823/ [Accessed: 21 February 2024].

Mann, R.K. and Beachy, P.A. 2004. Novel lipid modifications of secreted protein signals. *Annual review of biochemistry* 73, pp. 891–923. Available at: https://pubmed.ncbi.nlm.nih.gov/15189162/ [Accessed: 21 February 2024].

La Manno, G. et al. 2018. RNA velocity of single cells. *Nature 2018 560:7719* 560(7719), pp. 494–498. Available at: https://www.nature.com/articles/s41586-018-0414-6 [Accessed: 8 August 2023].

Mantovani, A., Allavena, P., Sica, A. and Balkwill, F. 2008. Cancer-related inflammation. *Nature* 454(7203), pp. 436–444. Available at: https://pubmed.ncbi.nlm.nih.gov/18650914/ [Accessed: 8 January 2024].

Marcil, I. and Stern, R.S. 2000. Risk of Developing a Subsequent Nonmelanoma Skin Cancer in Patients With a History of Nonmelanoma Skin Cancer: A Critical Review of the Literature and Meta-analysis. *Archives of Dermatology* 136(12), pp. 1524–1530. Available at: https://jamanetwork.com/journals/jamadermatology/fullarticle/190904 [Accessed: 20 February 2024].

Marin-Acevedo, J.A., Dholaria, B., Soyano, A.E., Knutson, K.L., Chumsri, S. and Lou, Y. 2018. Next generation of immune checkpoint therapy in cancer: new developments and challenges. *Journal of hematology & oncology* 11(1). Available

at: https://pubmed.ncbi.nlm.nih.gov/29544515/ [Accessed: 19 February 2024].
Martinez, V.D., Becker-Santos, D.D., Vucic, E.A., Lam, S. and Lam, W.L. 2011. Induction of Human Squamous Cell-Type Carcinomas by Arsenic. *Journal of Skin Cancer* 2011, pp. 1–9. doi: 10.1155/2011/454157.
Martinou, J.C. and Youle, R.J. 2011. Mitochondria in apoptosis: Bcl-2 family members and mitochondrial dynamics. *Developmental cell* 21(1), pp. 92–101. Available at: https://pubmed.ncbi.nlm.nih.gov/21763611/ [Accessed: 20 February 2023].
Mattei, V. et al. 2021. The Importance of Tumor Stem Cells in Glioblastoma Resistance to Therapy. *International Journal of Molecular Sciences* 22(8). Available at: /pmc/articles/PMC8068366/ [Accessed: 5 May 2024].
Mayer, C. et al. 2018. Developmental diversification of cortical inhibitory interneurons. *Nature* 555(7697), pp. 457–462. Available at: https://pubmed.ncbi.nlm.nih.gov/29513653/ [Accessed: 11 April 2023].
Mayer, J.E. and Goldman, R.H. 2016. Arsenic and skin cancer in the USA: the current evidence regarding arsenic-contaminated drinking water. *International journal of dermatology* 55(11), pp. e585–e591. Available at: https://pubmed.ncbi.nlm.nih.gov/27420023/ [Accessed: 6 January 2024].
McInnes, L., Healy, J. and Melville, J. 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* 1802.03426(3). Available at: https://arxiv.org/abs/1802.03426v3 [Accessed: 25 April 2023].
Mechlin, C.W., Tanner, M.J., Chen, M., Buttyan, R., Levin, R.M. and Mian, B.M. 2010. Gli2 expression and human bladder transitional carcinoma cell invasiveness. *The Journal of urology* 184(1), pp. 344–351. Available at: https://pubmed.ncbi.nlm.nih.gov/20488474/ [Accessed: 23 February 2024].

Mellman, I., Chen, D.S., Powles, T. and Turley, S.J. 2023. The cancer-immunity cycle: Indication, genotype, and immunotype. *Immunity* 56(10), pp. 2188–2205. doi: 10.1016/J.IMMUNI.2023.09.011.

Meloni, A.R. et al. 2006. Smoothened signal transduction is promoted by G protein-coupled receptor kinase 2. *Molecular and cellular biology* 26(20), pp. 7550–7560. Available at: https://pubmed.ncbi.nlm.nih.gov/16908539/ [Accessed: 21 February 2024].

Melzer, C., von der Ohe, J., Lehnert, H., Ungefroren, H. and Hass, R. 2017. Cancer stem cell niche models and contribution by mesenchymal stroma/stem cells. *Molecular Cancer 2017 16:1* 16(1), pp. 1–15. Available at: https://molecular-cancer.biomedcentral.com/articles/10.1186/s12943-017-0595-x [Accessed: 19 February 2024].

Merchant, M. et al. 2004. Suppressor of fused regulates Gli activity through a dual binding mechanism. *Molecular and cellular biology* 24(19), pp. 8627–8641. Available at: https://pubmed.ncbi.nlm.nih.gov/15367681/ [Accessed: 21 February 2024].

Migden, M.R. et al. 2015. Treatment with two different doses of sonidegib in patients with locally advanced or metastatic basal cell carcinoma (BOLT): a multicentre, randomised, double-blind phase 2 trial. *The Lancet. Oncology* 16(6), pp. 716–728. Available at: https://pubmed.ncbi.nlm.nih.gov/25981810/ [Accessed: 21 February 2024].

Mihrshahi, R. and Brown, M.H. 2010. Downstream of tyrosine kinase 1 and 2 play opposing roles in CD200 receptor signaling. *Journal of immunology (Baltimore, Md. : 1950)* 185(12), pp. 7216–7222. Available at: https://pubmed.ncbi.nlm.nih.gov/21078907/ [Accessed: 23 February 2024].

Miller, F.R., Santner, S.J., Tait, L. and Dawson, P.J. 2000. MCF10DCIS.com xenograft model of human comedo ductal carcinoma in situ. *Journal of the National Cancer Institute* 92(14), pp. 1185–1186. Available at: https://pubmed.ncbi.nlm.nih.gov/10904098/ [Accessed: 23 February 2024].

Mima, K. et al. 2012. CD44s regulates the TGF-β-mediated mesenchymal phenotype and is associated with poor prognosis in patients with hepatocellular carcinoma. *Cancer research* 72(13), pp. 3414–3423. Available at: https://pubmed.ncbi.nlm.nih.gov/22552294/ [Accessed: 23 February 2024].

Mittal, A. and Colegio, O.R. 2017. Skin Cancers in Organ Transplant Recipients. *American Journal of Transplantation* 17(10), pp. 2509–2530. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28556451 [Accessed: 16 August 2019].

Mittal, D., Gubin, M.M., Schreiber, R.D. and Smyth, M.J. 2014. New insights into cancer immunoediting and its three component phases — elimination, equilibrium and escape. *Current opinion in immunology* 27(1), p. 16. Available at: /pmc/articles/PMC4388310/ [Accessed: 18 February 2024].

Mondal, S., Bhattacharya, K. and Mandal, C. 2018. Nutritional stress reprograms dedifferention in glioblastoma multiforme driven by PTEN/Wnt/Hedgehog axis: a stochastic model of cancer stem cells. *Cell Death Discovery 2018 4:1* 4(1), pp. 1–16. Available at: https://www.nature.com/articles/s41420-018-0126-6 [Accessed: 5 May 2024].

Moody, G.E. et al. 2016. Abstract 2968: A novel bispecific CD3/CDH19 antibody construct (CDH19 BiTE) directs potent killing of melanoma cells in vitro and in vivo and is enhanced by blockade of PD-L1. *Cancer Research* 76(14_Supplement), pp. 2968–2968. Available at: https://aacrjournals.org/cancerres/article/76/14_Supplement/2968/610357/Abstra

ct-2968-A-novel-bispecific-CD3-CDH19 [Accessed: 18 April 2023].
Moon, K., Guallar, E. and Navas-Acien, A. 2012. Arsenic exposure and cardiovascular disease: an updated systematic review. *Current atherosclerosis reports* 14(6), pp. 542–555. Available at: https://pubmed.ncbi.nlm.nih.gov/22968315/ [Accessed: 6 January 2024].
Moreaux, J. et al. 2006. CD200 is a new prognostic factor in multiple myeloma. *Blood* 108(13), pp. 4194–4197. Available at: https://pubmed.ncbi.nlm.nih.gov/16946299/ [Accessed: 20 January 2024].
Moreaux, J., Veyrune, J.L., Reme, T., De Vos, J. and Klein, B. 2008. CD200: a putative therapeutic target in cancer. *Biochemical and biophysical research communications* 366(1), pp. 117–122. Available at: https://pubmed.ncbi.nlm.nih.gov/18060862/ [Accessed: 23 February 2024].
Morgan, H., Olivero, C. and Patel, G.K. 2019. Identification of Human Cutaneous Basal Cell Carcinoma Cancer Stem Cells. *Methods in Molecular Biology* 1879, pp. 435–450. doi: 10.1007/7651_2018_133.
Morgan, H.J. et al. 2020a. Hair follicle differentiation-specific keratin expression in human basal cell carcinoma. *Clinical and experimental dermatology* 45(4), pp. 417–425. Available at: https://pubmed.ncbi.nlm.nih.gov/31580512/ [Accessed: 22 February 2024].
Morgan, H.J. et al. 2020b. Human basal cell carcinoma: the induction of anagen hair follicle differentiation. *Clinical and experimental dermatology* 45(3), pp. 309–317. Available at: https://pubmed.ncbi.nlm.nih.gov/31556145/ [Accessed: 22 February 2024].
Morgan, H.J. et al. 2022. CD200 ectodomain shedding into the tumor microenvironment leads to NK cell dysfunction and apoptosis. *The Journal of*

Clinical Investigation 132(21). Available at: https://doi.org/10.1172/JCI150750 [Accessed: 22 December 2023].

Morris, E.C., Neelapu, S.S., Giavridis, T. and Sadelain, M. 2022. Cytokine release syndrome and associated neurotoxicity in cancer immunotherapy. *Nature reviews. Immunology* 22(2), pp. 85–96. Available at: https://pubmed.ncbi.nlm.nih.gov/34002066/ [Accessed: 19 February 2024].

Muzic, J.G. et al. 2017. Incidence and Trends of Basal Cell Carcinoma and Cutaneous Squamous Cell Carcinoma: A Population-Based Study in Olmsted County, Minnesota, 2000 to 2010. *Mayo Clinic proceedings* 92(6), pp. 890–898. Available at: https://pubmed.ncbi.nlm.nih.gov/28522111/ [Accessed: 5 January 2024].

Naik, P.P. and Desai, M.B. 2022. Basal Cell Carcinoma: A Narrative Review on Contemporary Diagnosis and Management. *Oncology and Therapy* 10(2), pp. 317–335. Available at: https://link.springer.com/article/10.1007/s40487-022-00201-8 [Accessed: 4 January 2024].

Neagu, M., Constantin, C., Caruntu, C., Dumitru, C., Surcel, M. and Zurac, S. 2019. Inflammation: A key process in skin tumorigenesis. *Oncology Letters* 17(5), p. 4068. Available at: /pmc/articles/PMC6444305/ [Accessed: 8 January 2024].

Neagu, M., Constantin, C., Dumitrascu, G.R., Lupu, A.R., Caruntu, C., Boda, D. and Zurac, S. 2015. Inflammation markers in cutaneous melanoma - edgy biomarkers for prognosis. *Discoveries (Craiova, Romania)* 3(1), p. e38. Available at: https://pubmed.ncbi.nlm.nih.gov/32309563/ [Accessed: 8 January 2024].

Nguyen, D.X. and Massagué, J. 2007. Genetic determinants of cancer metastasis. *Nature reviews. Genetics* 8(5), pp. 341–352. Available at: https://pubmed.ncbi.nlm.nih.gov/17440531/ [Accessed: 1 February 2023].

Nguyen, L.H. and Holmes, S. 2019. Ten quick tips for effective dimensionality

reduction. *PLOS Computational Biology* 15(6), p. e1006907. Available at: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006907 [Accessed: 12 April 2023].

Ngwa, C. and Liu, F. 2019. CD200-CD200R signaling and diseases: a potential therapeutic target? *International Journal of Physiology, Pathophysiology and Pharmacology* 11(6), p. 297. Available at: /pmc/articles/PMC6971504/ [Accessed: 23 February 2024].

Nishigori, C. 2006. Cellular aspects of photocarcinogenesis. *Photochemical and Photobiological Sciences* 5(2), pp. 208–214. doi: 10.1039/b507471a.

Nüsslein-Volhard, C. and Weischaus, E. 1980. Mutations affecting segment number and polarity in Drosophila. *Natur* 287(5785), pp. 795–801. Available at: https://www.nature.com/articles/287795a0 [Accessed: 21 February 2024].

O'Donnell, M.L., Metcalf, O., Watson, L., Phelps, A. and Varker, T. 2018. A Systematic Review of Psychological and Pharmacological Treatments for Adjustment Disorder in Adults. *Journal of traumatic stress* 31(3), pp. 321–331. Available at: https://pubmed.ncbi.nlm.nih.gov/29958336/ [Accessed: 18 February 2024].

Okada, H., Tsuzuki, T., Shindoh, H., Nishigaki, A., Yasuda, K. and Kanzaki, H. 2014. Regulation of decidualization and angiogenesis in the human endometrium: Mini review. *Journal of Obstetrics and Gynaecology Research* 40(5), pp. 1180–1187. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/jog.12392 [Accessed: 8 February 2023].

Olivero, C., Morgan, H. and Patel, G.K. 2019. Identification of Human Cutaneous Squamous Cell Carcinoma Cancer Stem Cells. *Methods in Molecular Biology* 1879, pp. 415–433. doi: 10.1007/7651_2018_134.

Olsen, C.M., Wilson, L.F., Green, A.C., Bain, C.J., Fritschi, L., Neale, R.E. and Whiteman, D.C. 2015. Cancers in Australia attributable to exposure to solar ultraviolet radiation and prevented by regular sunscreen use. *Australian and New Zealand journal of public health* 39(5), pp. 471–476. Available at: https://pubmed.ncbi.nlm.nih.gov/26437734/ [Accessed: 4 January 2024].

Omodaka, T., Minagawa, A. and Okuyama, R. 2021. Ultraviolet-related skin cancers distribute differently on the face surface. *British Journal of Dermatology* 185(1), pp. 205–207. Available at: https://dx.doi.org/10.1111/bjd.19875 [Accessed: 20 February 2024].

Oro, A.E., Higgins, K.M., Hu, Z., Bonifas, J.M., Epstein, E.H. and Scott, M.P. 1997. Basal cell carcinomas in mice overexpressing sonic hedgehog. *Science (New York, N.Y.)* 276(5313), pp. 817–821. Available at: https://pubmed.ncbi.nlm.nih.gov/9115210/ [Accessed: 23 December 2023].

Oshimori, N., Oristian, D. and Fuchs, E. 2015. TGF-β Promotes Heterogeneity and Drug Resistance in Squamous Cell Carcinoma. *Cell* 160(5), pp. 963–976. Available at: http://dx.doi.org/10.1016/j.cell.2015.01.043.

Otley, C.C., Coldiron, B.M., Stasko, T. and Goldman, G.D. 2001. Decreased Skin Cancer After Cessation of Therapy With Transplant-Associated Immunosuppressants. *Archives of Dermatology* 137(4), pp. 459–463. doi: 10-1001/pubs.Arch Dermatol.-ISSN-0003-987x-137-4-dob00027.

Peñuelas, S. et al. 2009. TGF-beta increases glioma-initiating cell self-renewal through the induction of LIF in human glioblastoma. *Cancer cell* 15(4), pp. 315–327. Available at: https://pubmed.ncbi.nlm.nih.gov/19345330/ [Accessed: 23 February 2024].

Perera, E., Gnaneswaran, N., Staines, C., Win, A.K. and Sinclair, R. 2015. Incidence

and prevalence of non-melanoma skin cancer in Australia: A systematic review. *The Australasian journal of dermatology* 56(4), pp. 258–267. Available at: https://pubmed.ncbi.nlm.nih.gov/25716064/ [Accessed: 19 February 2024].

Petrova, R. and Joyner, A.L. 2014. Roles for Hedgehog signaling in adult organ homeostasis and repair. *Development (Cambridge, England)* 141(18), pp. 3445–3457. Available at: https://pubmed.ncbi.nlm.nih.gov/25183867/ [Accessed: 22 February 2024].

Pfeffer, C.M. and Singh, A.T.K. 2018. Apoptosis: A Target for Anticancer Therapy. *International Journal of Molecular Sciences* 19(2). Available at: /pmc/articles/PMC5855670/ [Accessed: 20 February 2023].

Puig, S. and Berrocal, A. 2015. Management of high-risk and advanced basal cell carcinoma. *Clinical & Translational Oncology* 17(7), p. 497. Available at: /pmc/articles/PMC4495248/ [Accessed: 20 February 2024].

Quah, H.S. et al. 2023. Single cell analysis in head and neck cancer reveals potential immune evasion mechanisms during early metastasis. *Nature Communications 2023 14:1* 14(1), pp. 1–19. Available at: https://www.nature.com/articles/s41467-023-37379-y [Accessed: 2 May 2023].

Qureshi-Baig, K., Ullmann, P., Haan, S. and Letellier, E. 2017. Tumor-Initiating Cells: a criTICal review of isolation approaches and new challenges in targeting strategies. *Molecular Cancer* 16(1). Available at: /pmc/articles/PMC5314476/ [Accessed: 19 February 2024].

Rahman, M.M., Ng, J.C. and Naidu, R. 2009. Chronic exposure of arsenic via drinking water and its adverse health impacts on humans. *Environmental geochemistry and health* 31 Suppl 1(SUPPL. 1), pp. 189–200. Available at: https://pubmed.ncbi.nlm.nih.gov/19190988/ [Accessed: 6 January 2024].

Raina, N., Rani, R. and Gupta, M. 2021. Angiogenesis: Aspects in wound healing. *Endothelial Signaling in Vascular Dysfunction and Disease: From Bench to Bedside* , pp. 77–90. doi: 10.1016/B978-0-12-816196-8.00010-2.

Razi, S., Rafiemanesh, H., Ghoncheh, M., Khani, Y. and Salehiniya, H. 2015. Changing Trends of Types of Skin Cancer in Iran. *Asian Pacific journal of cancer prevention : APJCP* 16(12), pp. 4955–4958. Available at: https://pubmed.ncbi.nlm.nih.gov/26163621/ [Accessed: 5 January 2024].

RC, C. et al. 1990. Intralesional interferon therapy for basal cell carcinoma. *Journal of the American Academy of Dermatology* 23(4 Pt 1), pp. 694–700. Available at: https://pubmed.ncbi.nlm.nih.gov/2229497/ [Accessed: 10 February 2024].

Reifenberger, J. et al. 2005. Somatic mutations in the PTCH, SMOH, SUFUH and TP53 genes in sporadic basal cell carcinomas. *The British journal of dermatology* 152(1), pp. 43–51. Available at: https://pubmed.ncbi.nlm.nih.gov/15656799/ [Accessed: 13 January 2024].

Reiter, O., Mimouni, I., Dusza, S., Halpern, A.C., Leshem, Y.A. and Marghoob, A.A. 2021. Dermoscopic Features of Basal Cell Carcinoma and its Subtypes: A systematic Review. *Journal of the American Academy of Dermatology* 85(3), p. 653. Available at: /pmc/articles/PMC9366765/ [Accessed: 20 February 2024].

Reizner, G., Chuang, T., Elpern, D., Stone, J. and Farmer, E. 1993. Basal cell carcinoma in Kauai, Hawaii: the highest documented incidence in the United States. *Journal of the American Academy of Dermatology* 29(2 Pt 1), pp. 184–189. Available at: https://pubmed.ncbi.nlm.nih.gov/8335736/ [Accessed: 4 January 2024].

Ren, J. et al. 2023. Single-cell sequencing reveals that endothelial cells, EndMT cells and mural cells contribute to the pathogenesis of cavernous malformations. *Experimental & Molecular Medicine 2023 55:3* 55(3), pp. 628–642. Available at:

https://www.nature.com/articles/s12276-023-00962-w [Accessed: 2 May 2023].
Ribatti, D. 2022. Two new applications in the study of angiogenesis the CAM assay: Acellular scaffolds and organoids. *Microvascular research* 140. Available at: https://pubmed.ncbi.nlm.nih.gov/34906560/ [Accessed: 18 February 2024].
Rich, J.N. 2016. Cancer stem cells: understanding tumor hierarchy and heterogeneity. *Medicine* 95(Suppl 1), pp. S2–S7. Available at: /pmc/articles/PMC5599210/ [Accessed: 19 February 2024].
Rodust, P.M., Stockfleth, E., Ulrich, C., Leverkus, M. and Eberle, J. 2009. UV-induced squamous cell carcinoma - A role for antiapoptotic signalling pathways. *British Journal of Dermatology* 161(SUPPL. 3), pp. 107–115. doi: 10.1111/j.1365-2133.2009.09458.x.
Rogers, H.W., Weinstock, M.A., Feldman, S.R. and Coldiron, B.M. 2015. Incidence Estimate of Nonmelanoma Skin Cancer (Keratinocyte Carcinomas) in the US Population, 2012. *JAMA Dermatology* 151(10), pp. 1081–1086. Available at: https://jamanetwork.com/journals/jamadermatology/fullarticle/2281227 [Accessed: 4 January 2024].
Rojas-Domínguez, A., Arroyo-Duarte, R., Rincón-Vieyra, F. and Alvarado-Mentado, M. 2022. Modeling cancer immunoediting in tumor microenvironment with system characterization through the ising-model Hamiltonian. *BMC Bioinformatics* 23(1), pp. 1–25. Available at: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-022-04731-w [Accessed: 18 February 2024].
Rompolas, P. and Greco, V. 2014. Stem cell dynamics in the hair follicle niche. *Seminars in Cell & Developmental Biology* 25–26, pp. 34–42. Available at: http://dx.doi.org/10.1016/j.semcdb.2013.12.005.

Rosenblum, M.D., Yancey, K.B., Olasz, E.B. and Truitt, R.L. 2006. CD200, a “no danger” signal for hair follicles. *Journal of dermatological science* 41(3), pp. 165–174. Available at: https://pubmed.ncbi.nlm.nih.gov/16386879/ [Accessed: 25 February 2024].

Rosso, S. et al. 1996. The multicentre south European study “Helios”. II: Different sun exposure patterns in the aetiology of basal cell and squamous cell carcinomas of the skin. *British journal of cancer* 73(11), pp. 1447–1454. Available at: https://pubmed.ncbi.nlm.nih.gov/8645596/ [Accessed: 4 January 2024].

Russo, G.L., Tedesco, I., Russo, M., Cioppa, A., Andreassi, M.G. and Picano, E. 2012. Cellular adaptive response to chronic radiation exposure in interventional cardiologists. *European heart journal* 33(3), pp. 408–414. Available at: https://pubmed.ncbi.nlm.nih.gov/21862464/ [Accessed: 7 January 2024].

Saintes, C. et al. 2015. Development of squamous cell carcinoma into basal cell carcinoma under treatment with Vismodegib. *Journal of the European Academy of Dermatology and Venereology : JEADV* 29(5), pp. 1006–1009. Available at: https://pubmed.ncbi.nlm.nih.gov/24980899/ [Accessed: 21 February 2024].

Sakurai, T. and Kudo, M. 2011. Signaling pathways governing tumor angiogenesis. *Oncology* 81 Suppl 1(SUPPL. 1), pp. 24–29. Available at: https://pubmed.ncbi.nlm.nih.gov/22212932/ [Accessed: 8 February 2023].

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F. and Regev, A. 2015a. Spatial reconstruction of single-cell gene expression data. *Nature biotechnology* 33(5), pp. 495–502. Available at: https://pubmed.ncbi.nlm.nih.gov/25867923/ [Accessed: 4 April 2023].

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F. and Regev, A. 2015b. Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology 2015 33:5*

33(5), pp. 495–502. Available at: https://www.nature.com/articles/nbt.3192 [Accessed: 6 April 2023].
Schneider, I., Cepela, J., Shetty, M., Wang, J., Nelson, A.C., Winterhoff, B. and Starr, T.K. 2021. Use of “default” parameter settings when analyzing single cell RNA sequencing data using Seurat: a biologist’s perspective. *Journal of Translational Genetics and Genomics* 5(1), pp. 37–49. Available at: https://jtggjournal.com/article/view/3809 [Accessed: 6 April 2023].
Schneider, M.R., Schmidt-Ullrich, R. and Paus, R. 2009. The hair follicle as a dynamic miniorgan. *Current biology : CB* 19(3). Available at: https://pubmed.ncbi.nlm.nih.gov/19211055/ [Accessed: 19 February 2024].
Schulman, J.M. and Fisher, D.E. 2009. Indoor UV tanning and skin cancer: health risks and opportunities. *Current opinion in oncology* 21(2), p. 144. Available at: /pmc/articles/PMC2913608/ [Accessed: 4 January 2024].
Scott, J.G. et al. 2019. Recasting the Cancer Stem Cell Hypothesis: Unification Using a Continuum Model of Microenvironmental Forces. *Current Stem Cell Reports* 5(1), pp. 22–30. Available at: https://link.springer.com/article/10.1007/s40778-019-0153-0 [Accessed: 19 February 2024].
Sekulic, A. et al. 2012. Efficacy and Safety of Vismodegib in Advanced Basal-Cell Carcinoma. *New England Journal of Medicine* 366(23), pp. 2171–2179. doi: 10.1056/NEJMOA1113713.
Setty, M., Kiseliovas, V., Levine, J., Gayoso, A., Mazutis, L. and Pe’er, D. 2019. Characterization of cell fate probabilities in single-cell data with Palantir. *Nature Biotechnology 2019 37:4* 37(4), pp. 451–460. Available at: https://www.nature.com/articles/s41587-019-0068-4 [Accessed: 8 August 2023].
Sharma, A., Boise, L.H. and Shanmugam, M. 2019. Cancer Metabolism and the

Evasion of Apoptotic Cell Death. *Cancers* 11(8). Available at: /pmc/articles/PMC6721599/ [Accessed: 8 February 2023].
Sharpe, H.J. et al. 2015. Genomic Analysis of Smoothened Inhibitor Resistance in Basal Cell Carcinoma. *Cancer Cell* 27(3), pp. 327–341. Available at: http://www.cell.com/article/S1535610815000513/fulltext [Accessed: 4 January 2024].
Shay, J.W. and Wright, W.E. 2011. Role of telomeres and telomerase in cancer. *Seminars in cancer biology* 21(6), pp. 349–353. Available at: https://pubmed.ncbi.nlm.nih.gov/22015685/ [Accessed: 6 February 2023].
Sheng, T. et al. 2004. Activation of the hedgehog pathway in advanced prostate cancer. *Molecular Cancer* 3(1), pp. 1–13. Available at: https://molecular-cancer.biomedcentral.com/articles/10.1186/1476-4598-3-29 [Accessed: 21 February 2024].
Shipitsin, M. et al. 2007. Molecular definition of breast tumor heterogeneity. *Cancer cell* 11(3), pp. 259–273. Available at: https://pubmed.ncbi.nlm.nih.gov/17349583/ [Accessed: 23 February 2024].
Shukla, S. et al. 2017. Development of a RNA-Seq Based Prognostic Signature in Lung Adenocarcinoma. *JNCI: Journal of the National Cancer Institute* 109(1). Available at: https://dx.doi.org/10.1093/jnci/djw200 [Accessed: 4 February 2024].
Sinx, K.A.E. et al. 2018. Vismodegib-resistant basal cell carcinomas in basal cell nevus syndrome: Clinical approach and genetic analysis. *JAAD Case Reports* 4(5), p. 408. Available at: /pmc/articles/PMC6031482/ [Accessed: 23 December 2023].
Siva, A., Xin, H., Qin, F., Oltean, D., Bowdish, K.S. and Kretz-Rommel, A. 2008. Immune modulation by melanoma and ovarian tumor cells through expression of the immunosuppressive molecule CD200. *Cancer immunology, immunotherapy : CII*

57(7), pp. 987–996. Available at: https://pubmed.ncbi.nlm.nih.gov/18060403/ [Accessed: 23 February 2024].
Skvara, H. et al. 2011. Topical treatment of Basal cell carcinomas in nevoid Basal cell carcinoma syndrome with a smoothened inhibitor. *The Journal of investigative dermatology* 131(8), pp. 1735–1744. Available at: https://pubmed.ncbi.nlm.nih.gov/21430703/ [Accessed: 21 February 2024].
Sleeman, J.P., Nazarenko, I. and Thiele, W. 2011. Do all roads lead to Rome? Routes to metastasis development. *International journal of cancer* 128(11), pp. 2511–2526. Available at: https://pubmed.ncbi.nlm.nih.gov/21365648/ [Accessed: 2 February 2023].
Solé-Boldo, L. et al. 2020. Single-cell transcriptomes of the human skin reveal age-related loss of fibroblast priming. *Communications Biology 2020 3:1* 3(1), pp. 1–12. Available at: https://www.nature.com/articles/s42003-020-0922-4 [Accessed: 2 May 2023].
Sollberger, G., Strittmatter, G.E., Grossi, S., Garstkiewicz, M., Auf Dem Keller, U., French, L.E. and Beer, H.D. 2015. Caspase-1 activity is required for UVB-induced apoptosis of human keratinocytes. *The Journal of investigative dermatology* 135(5), pp. 1395–1404. Available at: https://pubmed.ncbi.nlm.nih.gov/25562666/ [Accessed: 8 January 2024].
Srivastava, A., Bencomo, T., Das, I. and Lee, C.S. 2023. Unravelling the landscape of skin cancer through single-cell transcriptomics. *Translational Oncology* 27, p. 101557. Available at: /pmc/articles/PMC9576539/ [Accessed: 22 February 2024].
Stary, G., Bangert, C., Tauber, M., Strohal, R., Kopp, T. and Stingl, G. 2007. Tumoricidal activity of TLR7/8-activated inflammatory dendritic cells. *The Journal of experimental medicine* 204(6), pp. 1441–1451. Available at:

https://pubmed.ncbi.nlm.nih.gov/17535975/ [Accessed: 23 December 2023].
Steele, C.D., Pillay, N. and Alexandrov, L.B. 2022. An overview of mutational and copy number signatures in human cancer. *The Journal of pathology* 257(4), pp. 454–465. Available at: https://pubmed.ncbi.nlm.nih.gov/35420163/ [Accessed: 26 January 2023].
Van Steensel, M.A.M., Jaspers, N.G.J. and Steijlen, P.M. 2001. A case of Rombo syndrome. *The British journal of dermatology* 144(6), pp. 1215–1218. Available at: https://pubmed.ncbi.nlm.nih.gov/11422044/ [Accessed: 20 February 2024].
Steg, A.D. et al. 2012. Stem cell pathways contribute to clinical chemoresistance in ovarian cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 18(3), pp. 869–881. Available at: https://pubmed.ncbi.nlm.nih.gov/22142828/ [Accessed: 23 February 2024].
Stone, D.M. et al. 1996. The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. *Nature* 384(6605), pp. 129–134. Available at: https://pubmed.ncbi.nlm.nih.gov/8906787/ [Accessed: 21 February 2024].
Stratton, M.R., Campbell, P.J. and Futreal, P.A. 2009. The cancer genome. *Nature 2009 458:7239* 458(7239), pp. 719–724. Available at: https://www.nature.com/articles/nature07943 [Accessed: 25 January 2023].
Stuart, T. et al. 2019. Comprehensive Integration of Single-Cell Data. *Cell* 177(7), pp. 1888-1902.e21. Available at: http://www.cell.com/article/S0092867419305598/fulltext [Accessed: 15 May 2023].
Stuart, T. and Satija, R. 2019. Integrative single-cell analysis. *Nature reviews. Genetics* 20(5), pp. 257–272. Available at: https://pubmed.ncbi.nlm.nih.gov/30696980/ [Accessed: 15 May 2023].

Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A. and Bray, F. 2021. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians* 71(3), pp. 209–249. Available at: https://onlinelibrary.wiley.com/doi/full/10.3322/caac.21660 [Accessed: 26 January 2023].

Suraneni, M. and Badeaux, M. 2013. Tumour-Initiating Cells, Cancer Metastasis and Therapeutic Implications. In: Jandial, R. ed. *Madame Curie Bioscience Database*. Austin, TX: Landes Bioscience

Talvalkar, G. V. 1970. Squamous cell carcinoma of the skin. Its incidence and etio-pathogenesis in 625 cases. *Indian Journal of Cancer* 7(1), pp. 24–33. Available at: https://europepmc.org/article/med/5472324 [Accessed: 5 January 2024].

Tang, F. et al. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nature methods* 6(5), pp. 377–382. Available at: https://pubmed.ncbi.nlm.nih.gov/19349980/ [Accessed: 4 February 2024].

Tang, H., Fu, S., Zhai, S., Song, Y., Asgari, M.M. and Han, J. 2018. Use of antihypertensive drugs and risk of keratinocyte carcinoma: A meta-analysis of observational studies. *Pharmacoepidemiology and drug safety* 27(3), pp. 279–288. Available at: https://pubmed.ncbi.nlm.nih.gov/29318704/ [Accessed: 20 February 2024].

Tang, J.Y. et al. 2012. Inhibiting the Hedgehog Pathway in Patients with the Basal-Cell Nevus Syndrome. *New England Journal of Medicine* 366(23), pp. 2180–2188. Available at: https://www.nejm.org/doi/full/10.1056/nejmoa1113538 [Accessed: 21 February 2024].

Tang, L. and Wang, K. 2016. Chronic Inflammation in Skin Malignancies. *Journal of*

Molecular Signaling 11(1), pp. 1–13. Available at: /pmc/articles/PMC4911907/ [Accessed: 8 January 2024].
Taylor, B.C. and Balko, J.M. 2022. Mechanisms of MHC-I Downregulation and Role in Immunotherapy Response. *Frontiers in Immunology* 13. Available at: https://pubmed.ncbi.nlm.nih.gov/35296095/ [Accessed: 18 February 2024].
Teng, Y., Yu, Y., Li, S., Huang, Y., Xu, D., Tao, X. and Fan, Y. 2021. Ultraviolet Radiation and Basal Cell Carcinoma: An Environmental Perspective. *Frontiers in Public Health* 9, p. 666528. Available at: /pmc/articles/PMC8339433/ [Accessed: 4 January 2024].
The Human Protein Atlas. 2023. Available at: https://www.proteinatlas.org/.
Tickle, T., Tirosh, I., Georgescu, C., Brown, M. and Haas, B. 2019. inferCNV of the Trinity CTAT Project. Available at: https://github.com/broadinstitute/inferCNV.
Tjiu, J.W. et al. 2006. Cyclooxygenase-2 overexpression in human basal cell carcinoma cell line increases antiapoptosis, angiogenesis, and tumorigenesis. *The Journal of investigative dermatology* 126(5), pp. 1143–1151. Available at: https://pubmed.ncbi.nlm.nih.gov/16528365/ [Accessed: 23 February 2024].
Tsuyuzaki, K., Sato, H., Sato, K. and Nikaido, I. 2020. Benchmarking principal component analysis for large-scale single-cell RNA-sequencing. *Genome Biology* 21(1), pp. 1–17. Available at: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1900-3 [Accessed: 12 April 2023].
Tukachinsky, H., Lopez, L. V. and Salic, A. 2010. A mechanism for vertebrate Hedgehog signaling: recruitment to cilia and dissociation of SuFu-Gli protein complexes. *The Journal of cell biology* 191(2), pp. 415–428. Available at: https://pubmed.ncbi.nlm.nih.gov/20956384/ [Accessed: 21 February 2024].

Twito, T., Chen, Z., Khatri, I., Wong, K., Spaner, D. and Gorczynski, R. 2013. Ectodomain shedding of CD200 from the B-CLL cell surface is regulated by ADAM28 expression. *Leukemia research* 37(7), pp. 816–821. Available at: https://pubmed.ncbi.nlm.nih.gov/23643150/ [Accessed: 23 February 2024].

Valent, P. et al. 2012. Cancer stem cell definitions and terminology: the devil is in the details. *Nature Reviews Cancer 2012 12:11* 12(11), pp. 767–775. Available at: https://www.nature.com/articles/nrc3368 [Accessed: 22 February 2024].

Venables, Z.C. et al. 2019. Epidemiology of basal and cutaneous squamous cell carcinoma in the U.K. 2013–15: a cohort study. *British Journal of Dermatology* 181(3), pp. 474–482. doi: 10.1111/bjd.17873.

Verkouteren, J.A.C., Ramdas, K.H.R., Wakkee, M. and Nijsten, T. 2017. Epidemiology of basal cell carcinoma: scholarly review. *The British journal of dermatology* 177(2), pp. 359–372. Available at: https://pubmed.ncbi.nlm.nih.gov/28220485/ [Accessed: 5 January 2024].

Vesely, M.D., Kershaw, M.H., Schreiber, R.D. and Smyth, M.J. 2011. Natural innate and adaptive immunity to cancer. *Annual review of immunology* 29, pp. 235–271. Available at: https://pubmed.ncbi.nlm.nih.gov/21219185/ [Accessed: 18 February 2024].

Vitasa, B.C. et al. 1990. Association of Nonmelanoma Skin Cancer and Actinic Keratosis With Cumulative Solar Ultraviolet Exposure in Maryland Watermen. *Cancer* . Available at: https://acsjournals.onlinelibrary.wiley.com/doi/10.1002/1097-0142 [Accessed: 20 February 2024].

Vivier, E. et al. 2011. Innate or Adaptive Immunity? The Example of Natural Killer Cells. *Science* 331(6013), pp. 44–49. Available at:

https://www.science.org/doi/10.1126/science.1198687 [Accessed: 23 December 2023].

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A. and Kinzler, K.W. 2013. Cancer genome landscapes. *Science (New York, N.Y.)* 339(6127), pp. 1546–1558. Available at: https://pubmed.ncbi.nlm.nih.gov/23539594/ [Accessed: 26 January 2023].

Vornicescu, C., Ungureanu, L., Șenilă, S., Vesa, Ștefan, Cosgarea, R., Baican, C. and Mihu, M. 2020. Assessment of sun-related behavior and serum vitamin D in basal cell carcinoma: Preliminary results. *Experimental and therapeutic medicine* 20(6), pp. 1–1. Available at: https://pubmed.ncbi.nlm.nih.gov/33101477/ [Accessed: 21 February 2024].

Walcher, L. et al. 2020. Cancer Stem Cells—Origins and Biomarkers: Perspectives for Targeted Personalized Therapies. *Frontiers in Immunology* 11, p. 539291. Available at: www.frontiersin.org [Accessed: 19 February 2024].

Walker, P. and Hill, D. 2006. Surgical treatment of basal cell carcinomas using standard postoperative histological assessment. *The Australasian journal of dermatology* 47(1), pp. 1–12. Available at: https://pubmed.ncbi.nlm.nih.gov/16405477/ [Accessed: 21 February 2024].

Wan, J. et al. 2021. Distinct transcriptomic landscapes of cutaneous basal cell carcinomas and squamous cell carcinomas. *Genes & Diseases* 8(2), pp. 181–192. doi: 10.1016/J.GENDIS.2019.10.004.

Wang, G.Y., Wang, J., Mancianti, M.L. and Epstein, E.H. 2011. Basal cell carcinomas arise from hair follicle stem cells in Ptch1+/- mice. *Cancer Cell* 19(1), pp. 114–124. Available at: http://www.cell.com/article/S1535610810004721/fulltext [Accessed: 4 January 2024].

Wang, J. -X, Inskip, P.D., Boic, J.D., Li, B. -X, Zhang, J. -Y and Fraumeni, J.F. 1990. Cancer incidence among medical diagnostic X-ray workers in China, 1950 to 1985. *International journal of cancer* 45(5), pp. 889–895. Available at: https://pubmed.ncbi.nlm.nih.gov/2335392/ [Accessed: 7 January 2024].

Wang, J.X., Zhang, L.A., Li, B.X., Zhao, Y.C., Wang, Z.Q., Zhang, J.Y. and Aoyama, T. 2002. Cancer incidence and risk estimation among medical x-ray workers in China, 1950-1995. *Health physics* 82(4), pp. 455–466. Available at: https://pubmed.ncbi.nlm.nih.gov/11906134/ [Accessed: 7 January 2024].

Wang, S. et al. 2020. Single cell transcriptomics of human epidermis identifies basal stem cell transition states. *Nature Communications 2020 11:1* 11(1), pp. 1–14. Available at: https://www.nature.com/articles/s41467-020-18075-7 [Accessed: 18 April 2023].

Wang, W. et al. 2014. Dynamics between Cancer Cell Subpopulations Reveals a Model Coordinating with Both Hierarchical and Stochastic Concepts. *PLOS ONE* 9(1), p. e84654. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0084654 [Accessed: 5 May 2024].

Wang, Y., Shen, Y., Wang, S., Shen, Q. and Zhou, X. 2018. The Role of STAT3 in Leading the Crosstalk between Human Cancers and the Immune System. *Cancer letters* 415, p. 117. Available at: /pmc/articles/PMC5748258/ [Accessed: 18 February 2024].

Weedon, M.N. et al. 2008. Genome-wide association analysis identifies 20 loci that influence adult height. *Nature genetics* 40(5), pp. 575–583. Available at: https://pubmed.ncbi.nlm.nih.gov/18391952/ [Accessed: 21 February 2024].

Weeraratna, A.T. et al. 2004. Generation and analysis of melanoma SAGE libraries:

SAGE advice on the melanoma transcriptome. *Oncogene 2004 23:12* 23(12), pp. 2264–2274. Available at: https://www.nature.com/articles/1207337 [Accessed: 18 April 2023].

Welch, D. 2006. Defining a cancer metastasis in: Research. In: AAfC ed. *AACR Education Book*. Philadelphia, PA, USA.: AACR, pp. 111–115.

Welch, D.R. and Hurst, D.R. 2019. Defining the Hallmarks of Metastasis. *Cancer research* 79(12), pp. 3011–3027. Available at: https://pubmed.ncbi.nlm.nih.gov/31053634/ [Accessed: 1 February 2023].

Willenbrink, T.J., Ruiz, E.S., Cornejo, C.M., Schmults, C.D., Arron, S.T. and Jambusaria-Pahlajani, A. 2020. Field cancerization: Definition, epidemiology, risk factors, and outcomes. *Journal of the American Academy of Dermatology* 83(3), pp. 709–717. Available at: https://pubmed.ncbi.nlm.nih.gov/32387665/ [Accessed: 5 May 2024].

Wong, D.A., Bishop, G.A., Lowes, M.A., Cooke, B., Barnetson, R.S. and Halliday, G.M. 2000. Cytokine profiles in spontaneously regressing basal cell carcinomas. *The British journal of dermatology* 143(1), pp. 91–98. Available at: https://pubmed.ncbi.nlm.nih.gov/10886141/ [Accessed: 23 December 2023].

Wright, G.J., Jones, M., Puklavec, M.J., Brown, M.H. and Barclay, A.N. 2001. The unusual distribution of the neuronal/lymphoid cell surface CD200 (OX2) glycoprotein is conserved in humans. *Immunology* 102(2), pp. 173–179. Available at: https://pubmed.ncbi.nlm.nih.gov/11260322/ [Accessed: 11 February 2024].

Xie, J. et al. 1998. Activating Smoothened mutations in sporadic basal-cell carcinoma. *Nature* 391(6662), pp. 90–92. Available at: https://pubmed.ncbi.nlm.nih.gov/9422511/ [Accessed: 23 December 2023].

Xie, P., Lefrançois, P., Sasseville, D., Parmentier, L. and Litvinov, I. V. 2022. Analysis

of multiple basal cell carcinomas (BCCs) arising in one individual highlights genetic tumor heterogeneity and identifies novel driver mutations. *Journal of Cell Communication and Signaling* 16(4), pp. 633–635. Available at: https://link.springer.com/article/10.1007/s12079-022-00679-x [Accessed: 9 January 2024].

Xu, W. et al. 2015. Bax-PGAM5L-Drp1 complex is required for intrinsic apoptosis execution. *Oncotarget* 6(30), pp. 30017–30034. Available at: https://pubmed.ncbi.nlm.nih.gov/26356820/ [Accessed: 20 February 2023].

Yang, Y., Sun, M., Wang, L. and Jiao, B. 2013. HIFs, angiogenesis, and cancer. *Journal of cellular biochemistry* 114(5), pp. 967–974. Available at: https://pubmed.ncbi.nlm.nih.gov/23225225/ [Accessed: 8 February 2023].

Yao, C.D. et al. 2020. AP-1 and TGFß cooperativity drives non-canonical Hedgehog signaling in resistant basal cell carcinoma. *Nature Communications 2020 11:1* 11(1), pp. 1–17. Available at: https://www.nature.com/articles/s41467-020-18762-5 [Accessed: 14 January 2024].

Ye, X. and Weinberg, R.A. 2015. Epithelial-Mesenchymal Plasticity: A central regulator of cancer progression. *Trends in cell biology* 25(11), p. 675. Available at: /pmc/articles/PMC4628843/ [Accessed: 2 February 2023].

Yerly, L. et al. 2022. Integrated multi-omics reveals cellular and molecular interactions governing the invasive niche of basal cell carcinoma. *Nature Communications 2022 13:1* 13(1), pp. 1–16. Available at: https://www.nature.com/articles/s41467-022-32670-w [Accessed: 2 May 2023].

Yost, K.E. et al. 2019. Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nature medicine* 25(8), p. 1251. Available at: /pmc/articles/PMC6689255/ [Accessed: 22 February 2024].

You, H., Ding, W. and Rountree, C.B. 2010. Epigenetic regulation of cancer stem cell marker CD133 by transforming growth factor-beta. *Hepatology (Baltimore, Md.)* 51(5), pp. 1635–1644. Available at: https://pubmed.ncbi.nlm.nih.gov/20196115/ [Accessed: 23 February 2024].

Yousef, H., Alhajj, M. and Sharma, S. 2022. Anatomy, Skin (Integument), Epidermis. *StatPearls* . Available at: https://www.ncbi.nlm.nih.gov/books/NBK470464/ [Accessed: 31 December 2023].

Youssef, K.K. et al. 2010. Identification of the cell lineage at the origin of basal cell carcinoma. *Nature Cell Biology 2010 12:3* 12(3), pp. 299–305. Available at: https://www.nature.com/articles/ncb2031 [Accessed: 4 January 2024].

Yu, Y., Flint, A., Dvorin, E. and Bischoff, J. 2002. AC133-2: A Novel Isoform of Human AC133 Stem Cell Antigen. *Journal of Biological Chemistry* 277(23), pp. 20711–20716. doi: 10.1074/jbc.M202349200.

Zaman, S., Wang, R. and Gandhi, V. 2014. Targeting the apoptosis pathway in hematologic malignancies. *Leukemia & lymphoma* 55(9), pp. 1980–1992. Available at: https://pubmed.ncbi.nlm.nih.gov/24295132/ [Accessed: 20 February 2023].

Zappia, L. and Oshlack, A. 2018. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *GigaScience* 7(7), pp. 1–9. Available at: https://academic.oup.com/gigascience/article/7/7/giy083/5052205 [Accessed: 25 April 2023].

Zavadil, J. and Böttinger, E.P. 2005. TGF-β and epithelial-to-mesenchymal transitions. *Oncogene 2005 24:37* 24(37), pp. 5764–5774. Available at: https://www.nature.com/articles/1208927 [Accessed: 23 February 2024].

Zhang, H. et al. 2022. The Role of Keratin17 in Human Tumours. *Frontiers in Cell and Developmental Biology* 10, p. 338. doi: 10.3389/FCELL.2022.818416/BIBTEX.

Zhao, C. et al. 2009. Hedgehog signalling is essential for maintenance of cancer stem cells in myeloid leukaemia. *Nature* 458(7239), pp. 776–779. Available at: https://pubmed.ncbi.nlm.nih.gov/19169242/ [Accessed: 21 February 2024].

Zhao, X. et al. 2015. RAS/MAPK Activation Drives Resistance to Smo Inhibition, Metastasis, and Tumor Evolution in Shh Pathway-Dependent Tumors. *Cancer research* 75(17), pp. 3623–3635. Available at: https://pubmed.ncbi.nlm.nih.gov/26130651/ [Accessed: 21 February 2024].

Zhao, X., Li, J. and Sun, H. 2019. CD200-CD200R interaction: An important regulator after stroke. *Frontiers in Neuroscience* 13(JUL), p. 468717. Available at: www.frontiersin.org [Accessed: 23 February 2024].

Zhao, Y., Tong, C. and Jiang, J. 2007. Hedgehog regulates smoothened activity by inducing a conformational switch. *Nature 2007 450:7167* 450(7167), pp. 252–258. Available at: https://www.nature.com/articles/nature06225 [Accessed: 21 February 2024].

Zhivagui, M. et al. 2023. DNA damage and somatic mutations in mammalian cells after irradiation with a nail polish dryer. *Nature Communications 2023 14:1* 14(1), pp. 1–14. Available at: https://www.nature.com/articles/s41467-023-35876-8 [Accessed: 4 January 2024].

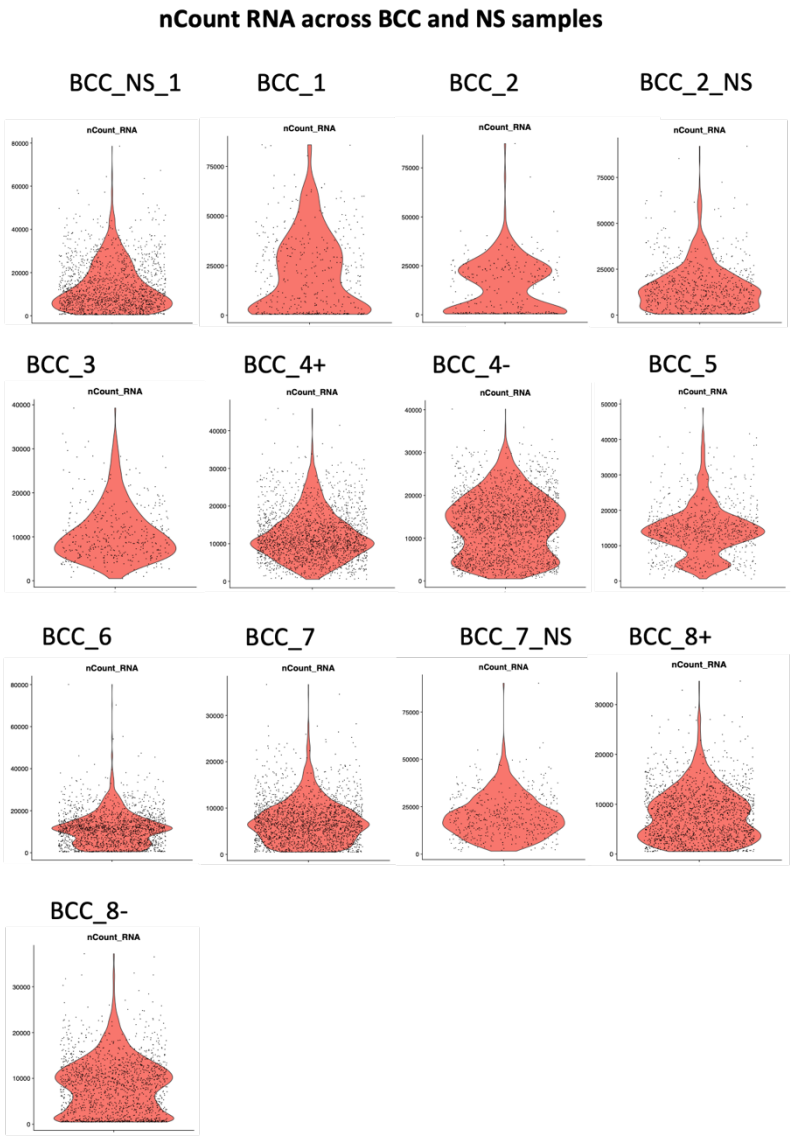
Zhou, J.G. et al. 2019. Development and Validation of an RNA-Seq-Based Prognostic Signature in Neuroblastoma. *Frontiers in Oncology* 9, p. 465917. Available at: https://ocg.cancer.gov/programs/ [Accessed: 4 February 2024].

Zhu, G.A., Sundram, U. and Chang, A.L.S. 2014. Two Different Scenarios of Squamous Cell Carcinoma Within Advanced Basal Cell Carcinomas: Cases Illustrating the Importance of Serial Biopsy During Vismodegib Usage. *JAMA Dermatology* 150(9), pp. 970–973. Available at:

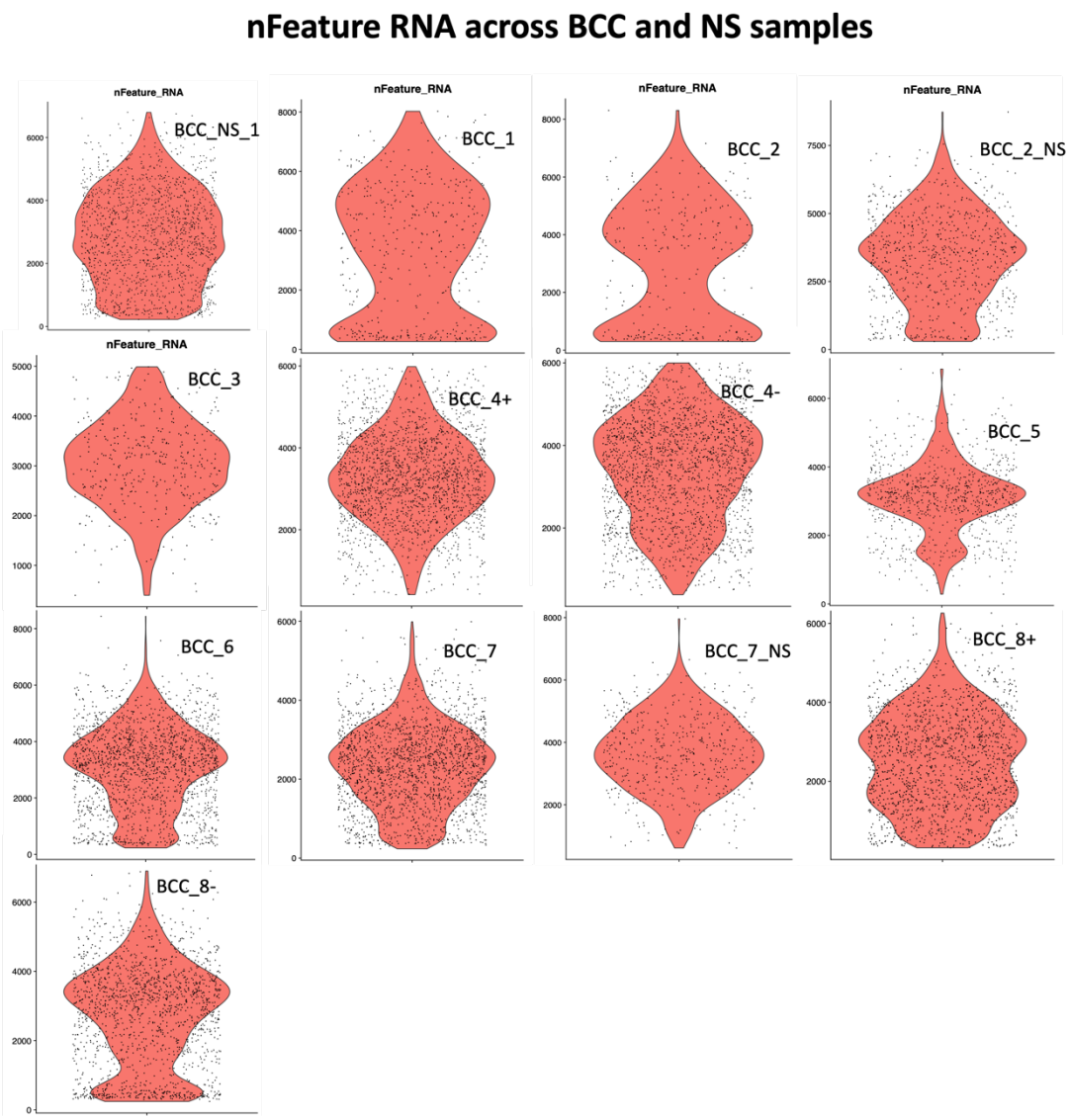
https://jamanetwork.com/journals/jamadermatology/fullarticle/1860843 [Accessed: 21 February 2024].

Appendix

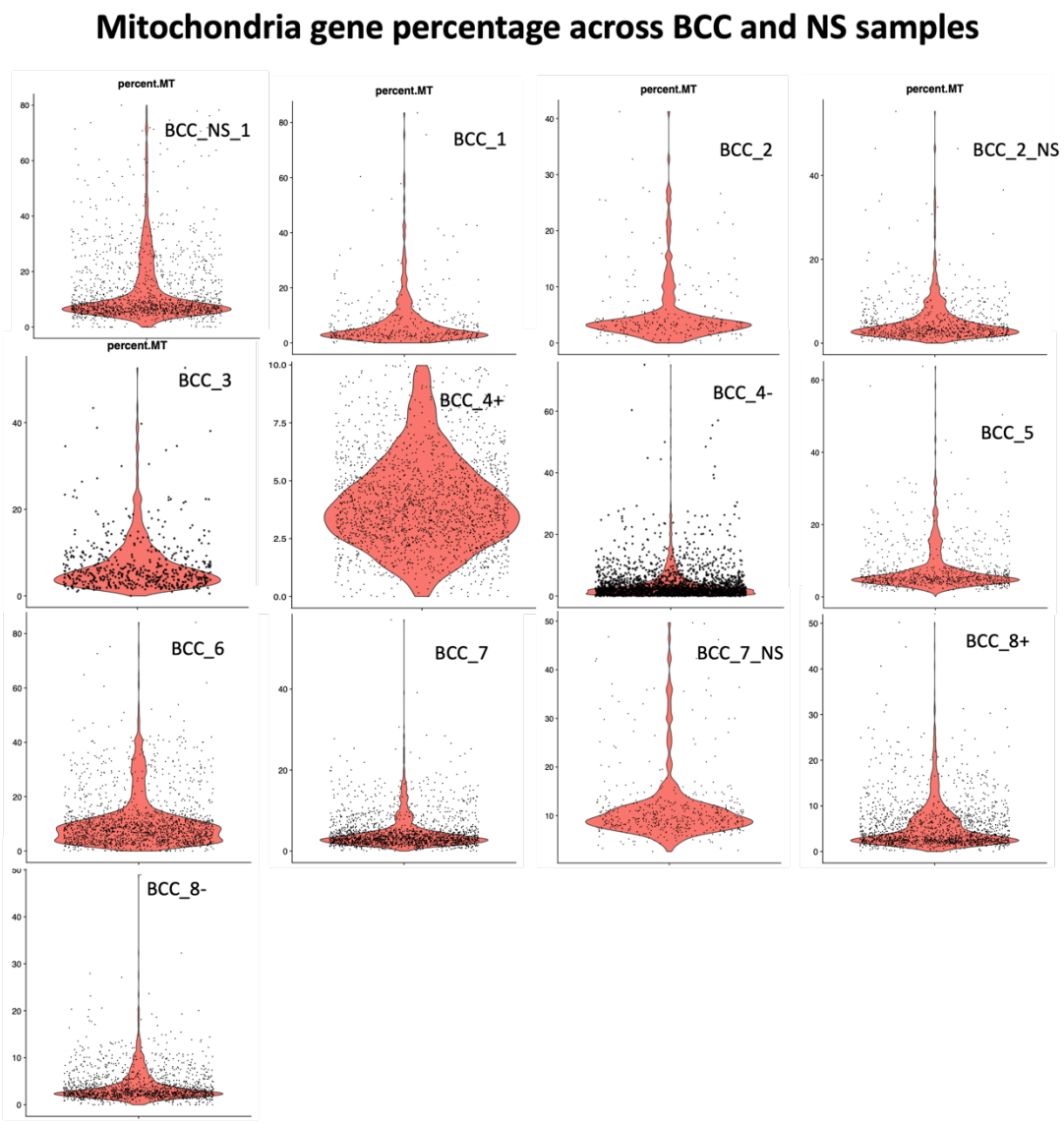
Appendix 2.2 | nCount RNA violin plots used for QC.



Appendix 2.3 | nFeature RNA violin plots used for QC.



Appendix 2.4 | Mitochondria gene percentage violin plots used for QC.

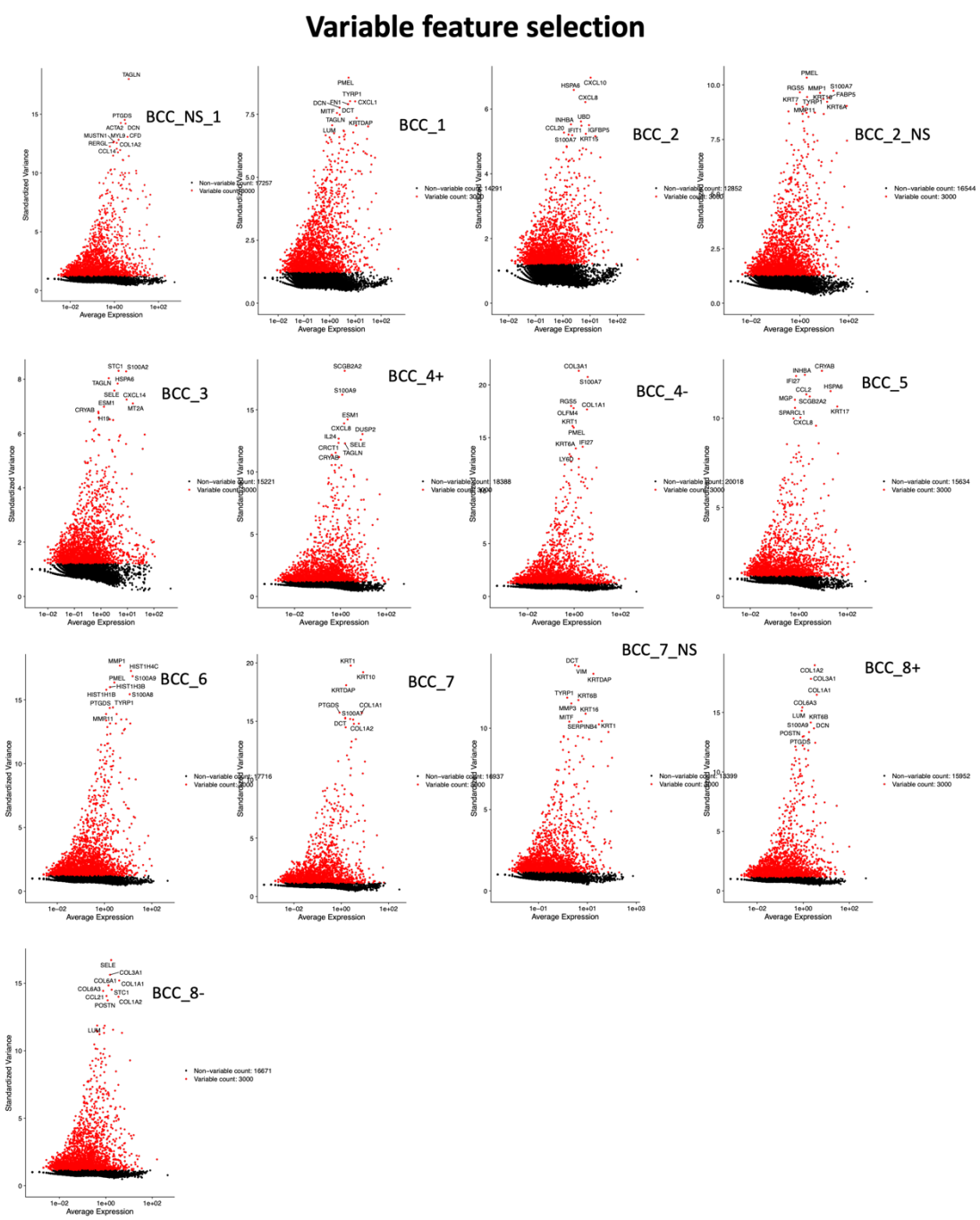


Appendix 2.5 | QC filter parameters

QC filter parameters

Sample	nFeature_RNA	percent.MT
BCC_NS_1	>200 <6000	<10%
BCC_1	>200 <6000	<10%
BCC_2	>200 <4500	<10%
BCC_2_NS	>200 <6000	<10%
BCC_3	>200 <5000	<10%
BCC_4+	>200 <7000	<10%
BCC_4-	>200 <7000	<10%
BCC_5	>200 <7500	<10%
BCC_6	>200 <6000	<10%
BCC_7	>200 <5000	<10%
BCC_7_NS	>200 <6000	<15%
BCC_8+	>200 <6000	<10%
BCC_8-	>200 <5000	<10%

Appendix 2.6 | Variable feature selection plots.



Appendix 2.7 | PCA Theory.

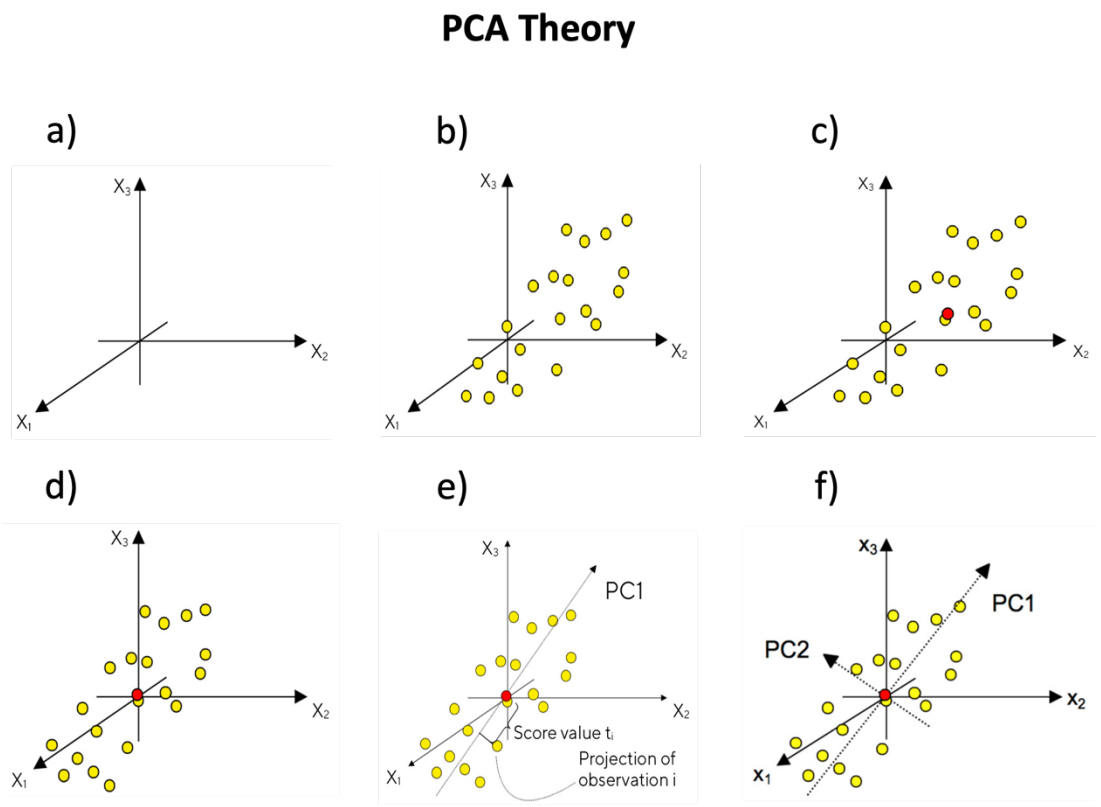
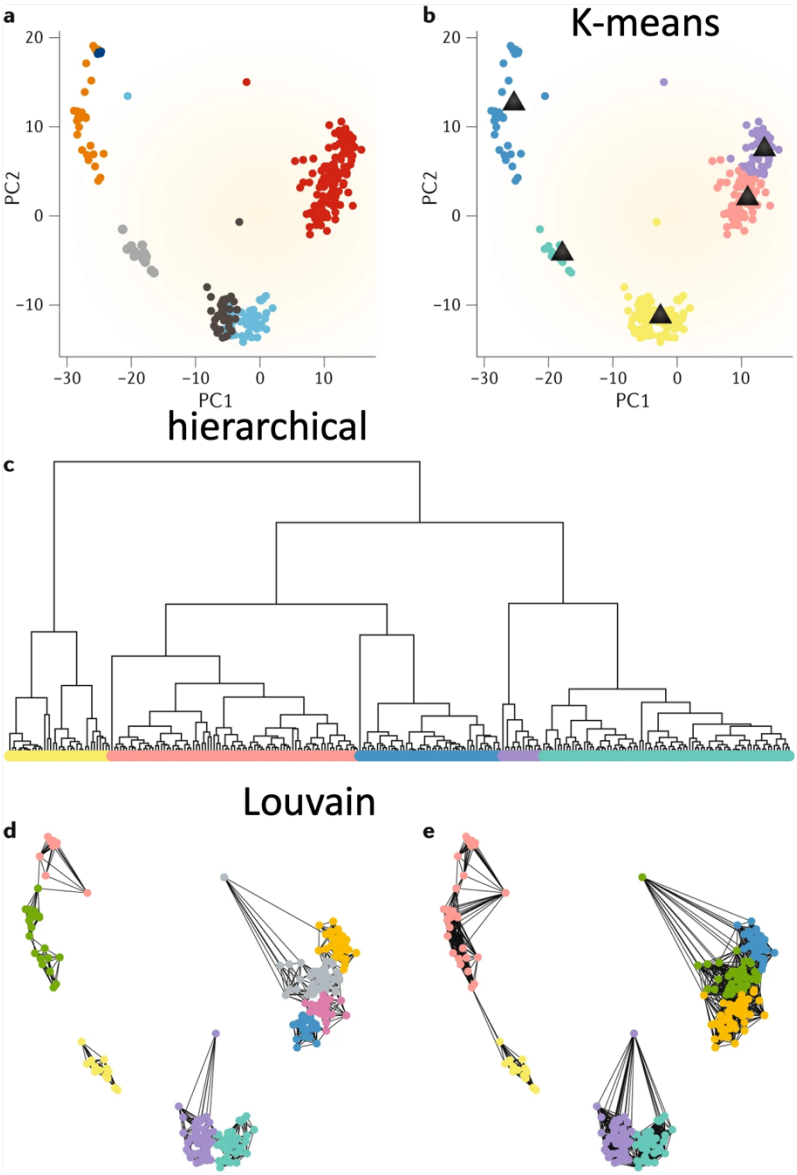


Figure adapted from https://www.sartorius.com/en/knowledge/science-snippets/what-is-principal-component-analysis-pca-and-how-it-is-used-507186

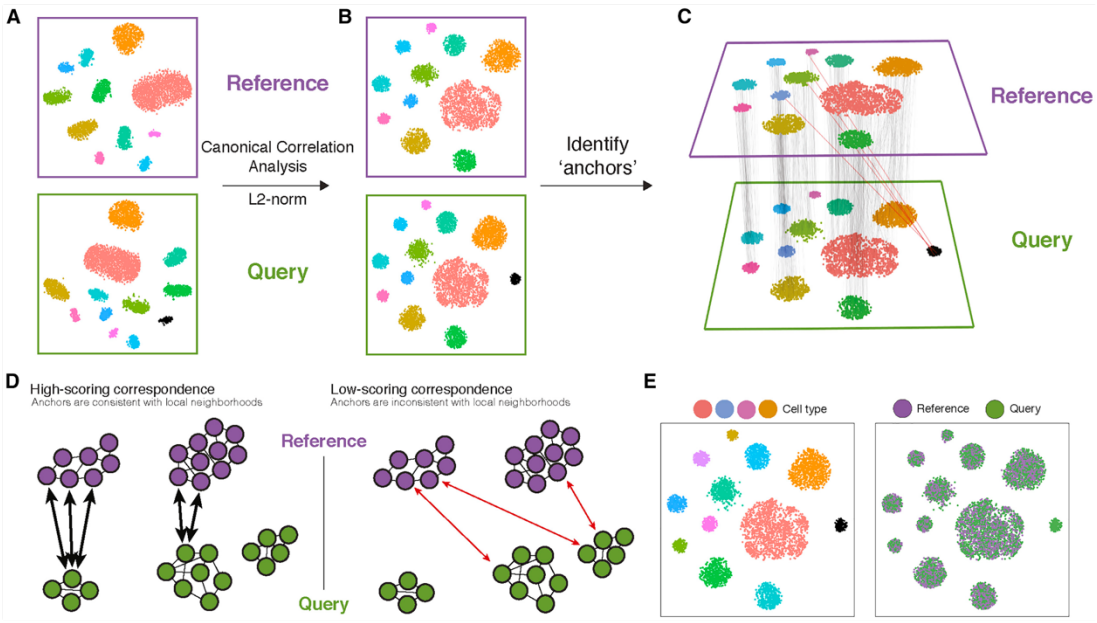
Appendix 2.8 | Clustering methods overview.



Overview of clustering methods – taken from Kiselev, V.Y., Andrews, T.S. and Hemberg, M. 2019. Challenges in unsupervised clustering of single-cell RNA-seq data. *Nature Reviews Genetics 2018 20:5* 20(5), pp. 273–282. Available at: https://www.nature.com/articles/s41576-018-0088-9 [Accessed: 24 April 2023].

***"Representation of different clustering approaches for single-cell RNA sequencing (scRNA-seq) using the Deng data set of early mouse embryo development. a** | True clusters, as defined by the authors, are based on the developmental stage. **b** | k-means separates cells into k = 5 groups. Because k-means assumes equal-sized clusters, the larger group of blastocysts is split from the other cell groups before the 8-cell and 16-cell stages are separated from each other. **c** | Complete-linkage hierarchical clustering creates a hierarchy of cells that can be cut at different levels (the result for k = 5 is indicated by the coloured bars at the bottom). Cutting farther down the tree would reveal finer substructures within the clusters. **d,e** | Louvain community detection[29] is applied to a shared-nearest-neighbour graph connecting the cells and finds tightly connected communities in the graph (number of nearest neighbours used to construct the graph is five for part **d** and ten for part **e**). Increasing the number of neighbours when constructing the cell–cell graph indirectly decreases the resolution of graph-based clustering. Each clustering algorithm was implemented in R (igraph for parts **d** and **e**) and applied to the first two principal components (PCs) of the data."*

Appendix 2.9 | Seurat integration method overview



Taken from Stuart, T. et al. 2019. Comprehensive Integration of Single-Cell Data. *Cell* 177(7), pp. 1888-1902.e21. Available at: http://www.cell.com/article/S0092867419305598/fulltext [Accessed: 15 May 2023].

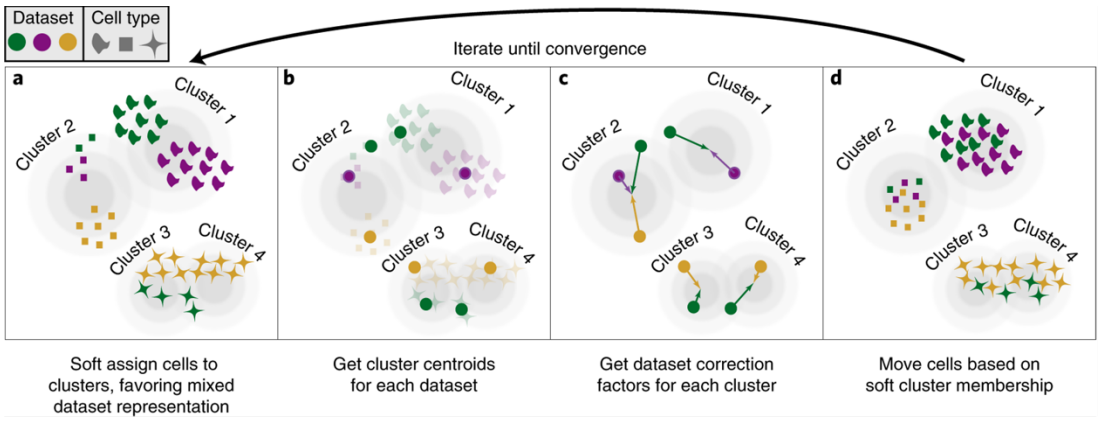
"Figure 1Schematic Overview of Reference "Assembly" Integration in Seurat v3

(A) Representation of two datasets, reference and query, each of which originates from a separate single-cell experiment. The two datasets share cells from similar biological states, but the query dataset contains a unique population (in black).

(B) We perform canonical correlation analysis, followed by L2 normalization of the canonical correlation vectors, to project the datasets into a subspace defined by shared correlation structure across datasets. (C) In the shared space, we identify pairs of MNNs across reference and query cells. These should represent cells in a shared biological state across datasets (gray lines) and serve as anchors to guide dataset integration. In principle, cells in unique populations should not participate in anchors, but in practice, we observe "incorrect" anchors at low frequency (red lines). (D) For

each anchor pair, we assign a score based on the consistency of anchors across the neighbourhood structure of each dataset. (E) We utilize anchors and their scores to compute "correction" vectors for each query cell, transforming its expression so it can be jointly analysed as part of an integrated reference."

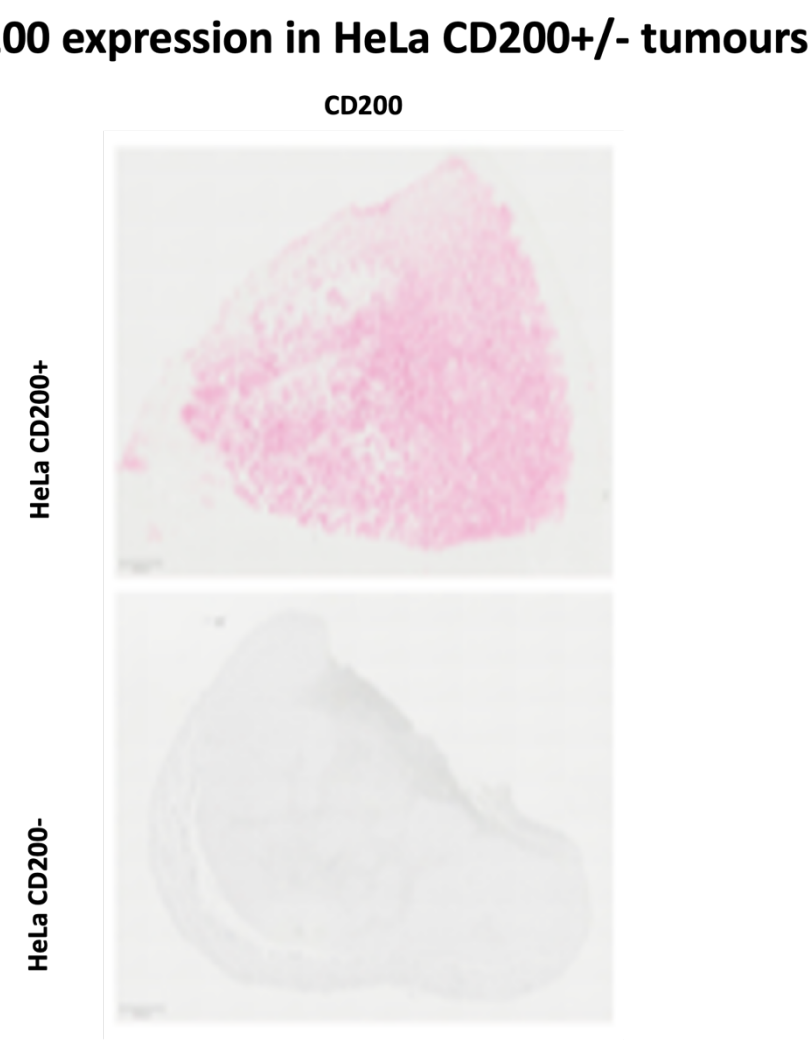
Appendix 2.10 | Harmony integration overview



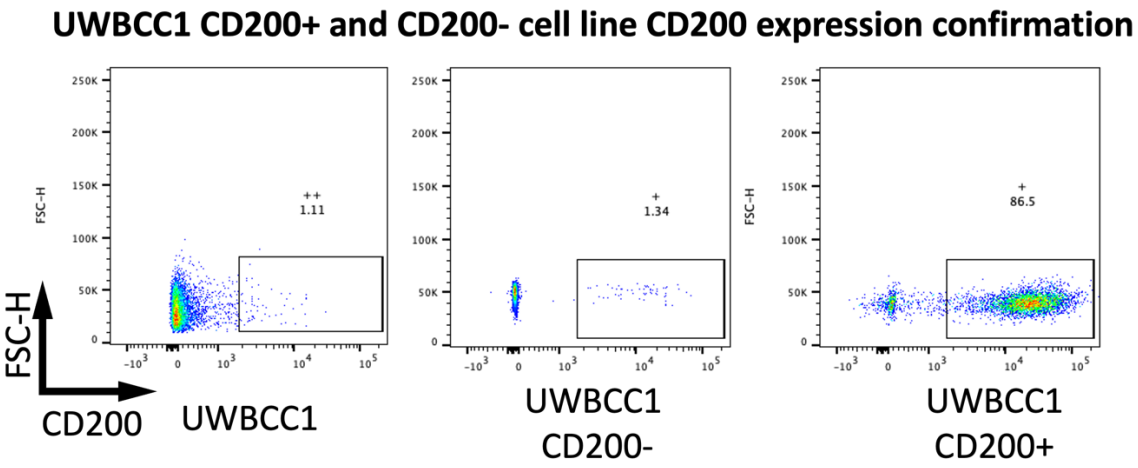
Taken from Korsunsky, I. et al. 2019. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods 2019 16:12* 16(12), pp. 1289–1296. Available at: https://www.nature.com/articles/s41592-019-0619-0 [Accessed: 15 May 2023].

"PCA embeds cells into a space with reduced dimensionality. *Harmony accepts the cell coordinates in this reduced space and runs an iterative algorithm to adjust for dataset specific effects.* ***a****, Harmony uses fuzzy clustering to assign each cell to multiple clusters, while a penalty term ensures that the diversity of datasets within each cluster is maximized.* ***b****, Harmony calculates a global centroid for each cluster, as well as dataset-specific centroids for each cluster.* ***c****, Within each cluster, Harmony calculates a correction factor for each dataset based on the centroids.* ***d****, Finally, Harmony corrects each cell with a cell-specific factor: a linear combination of dataset correction factors weighted by the cell's soft cluster assignments made in step* ***a****. Harmony repeats steps* ***a*** *to* ***d*** *until convergence. The dependence between cluster assignment and dataset diminishes with each round. Datasets are represented with colours, cell types with different shapes."*

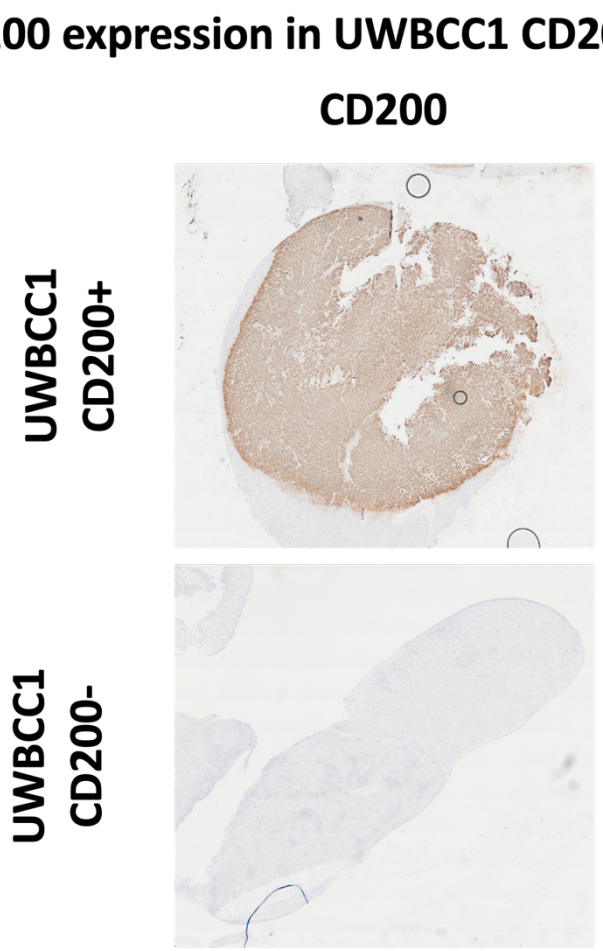
Appendix 3.1 | CD200 expression by IHC in HeLa CD200+/- tumours



Appendix 3.2 | CD200 expression by flow cytometry in UW-BCC1 and UW-BCC1 CD200+/- Cell lines



Appendix 3.3 | CD200 expression by IHC in UW-BCC1 CD200+/- tumours



Appendix 4.1 | UW-BCC1 RNAseq sample information table

Sample	Sample Information	Condition	Clean Reads	Total Reads Mapped to Genome	Total Reads Uniquely Mapped to Genome	Q30	GC %	Reads Mapped to Exons	Reads Mapped to Introns	Reads Mapped to Intergenic Regions
A-UW-BCC1-CD200+	UW-BCC1 Sample 1	CD200+	47,198,342	95.91%	93.39%	92.36	49.68	88.54%	9.29%	2.16%
A-UW-BCC1-CD200-	UW-BCC1 Sample 1	CD200-	64,872,416	96.0%	93.37%	93.13	49.9	91.01%	7.03%	1.95%
B-UW-BCC1-CD200+	UW-BCC1 Sample 2	CD200+	55,926,810	95.85%	93.28%	92.7	49.81	89.32%	8.50%	2.08%
B-UW-BCC1-CD200-	UW-BCC1 Sample 2	CD200-	43,903,096	96.15%	93.6%	93.08	49.57	89.20%	8.67%	2.11%
C-UW-BCC1-CD200+	UW-BCC1 Sample 3	CD200+	50,007,820	96.06%	93.51%	92.99	49.53	88.94%	8.97%	2.08%
D-UW-BCC1-CD200+	UW-BCC1 Sample 4	CD200+	52,429,274	95.99%	93.4%	92.85	49.83	89.04%	8.87%	2.08%
D-UW-BCC1-CD200-	UW-BCC1 Sample 4	CD200-	50,722,986	95.95%	93.39%	93.01	49.64	89.84%	8.11%	2.04%
Average			**52,151,535**	**95.99%**	**93.42%**	**92.87%**	**49.71%**	**89.41%**	**8.49%**	**2.07%**

Appendix 4.2 | Human BCC RNAseq sample information table

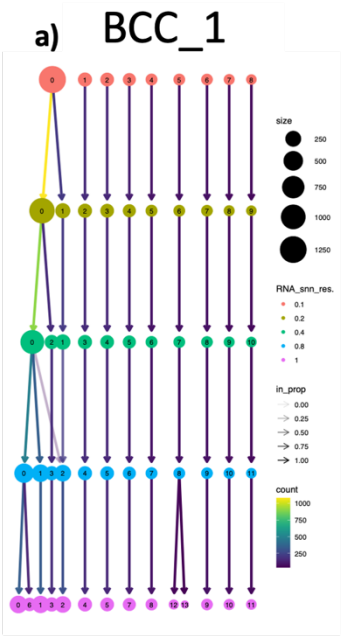
Sample	Sample Information	Condition	Total Reads Mapped to Genome	Total Reads Uniquely Mapped to Genome (%)	GC (%)
BCC2+	BCC sample 2	CD200pos	67,360,164	94.75	47.6
BCC2-	BCC sample 2	CD200neg	56,085,185	93.88	47.2
BCC3-	BCC sample 3	CD200neg	42,764,595	93.8	48.3
BCC4+	BCC sample 4	CD200pos	52,503,233	92.58	49.4
BCC6+	BCC sample 6	CD200pos	75,918,774	88.35	47.9
BCC6-	BCC sample 6	CD200neg	63,755,440	86.25	49.2
BCC8-	BCC sample 8	CD200neg	53,810,954	94.98	48.2
BCC8+	BCC sample 8	CD200pos	44,208,345	89.08	49.2
BCC10-	BCC sample 10	CD200neg	48,739,875	94.32	48.4
BCC10+	BCC sample 10	CD200pos	49,916,431	96.28	47.9
Average			55,506,299	92.43	48.33

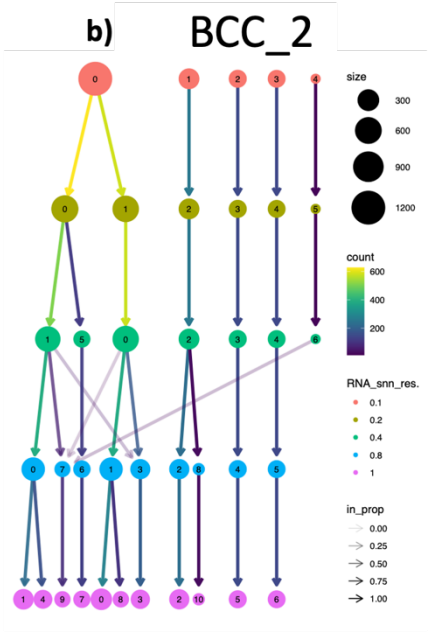
Appendix 5.1 | Low resolution clustering to identify cell types

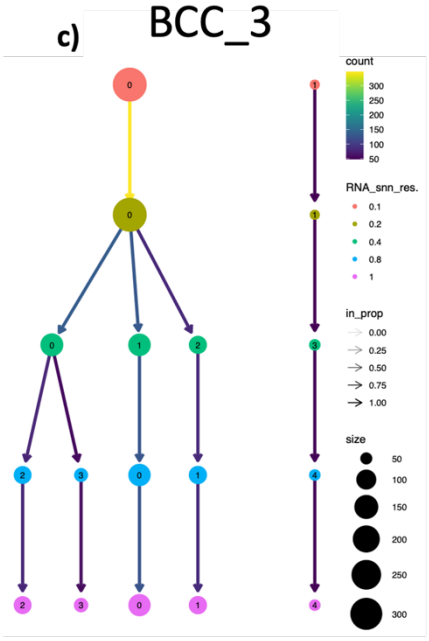
Appendix 5.1.1 | Writing

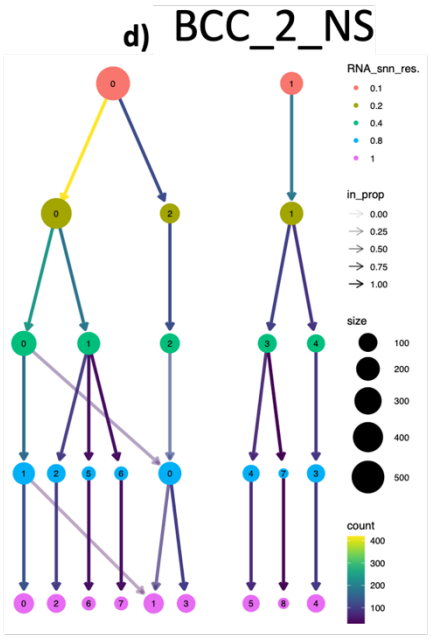
I first hypothesised that a low clustering resolution would be able to identify the major cell types in the samples. To evaluate this hypothesis, Seurat's clustering method was chosen to initially cluster the four samples individually using 30 PCs. Seurat uses the PCA results to generate the graph-like structure, where each PC embodies a 'meta feature' which aggregates information across a correlated feature set. To select the optimal resolution to identify the major cell types for each sample, the 'clustree' package was used (Zappia and Oshlack 2018). Clustree takes a range of provided resolution parameters (0.1, 0.2, 0.4, 0.8, 1) as input and generates a clustering tree which visualises how cell membership of clusters changes between different resolution parameters. Each row in the tree represents a different resolution parameter, with each node representing a unique cluster and the size of the node representing the number of cells present within that cluster. Going from top to bottom of the tree (from low to high resolution), the arrows represent the redistribution of the cells into a new node(s). The clustering tree therefore allows for the identification of 'stable' resolution parameter(s) and distinct or unstable clusters. Clustree analysis for each sample can be seen in Appendix 5.1.2. Initial observations demonstrate that across all samples, there is a large cluster of cells at resolution = 0.1. Cells from this larger cluster redistribute into smaller clusters as the resolution increases. Interestingly, cells from the other clusters at resolution = 0.1 do not redistribute and form smaller subclusters with cells that were redistributed from the large cluster, with exception for BCC_2 sample, where cells from cluster 4 at resolution = 0.1 redistribute at resolution = 0.8 and form a subcluster with cells that

were belonging to the large cluster 0 at resolution = 0.1 (Appendix 5.1.2b). This suggests that cells belonging to the large cluster (cluster 0 for all samples at resolution = 0.1) belong to the same cell type, and increasing the resolution enables the sub clustering of these cells into more specific subtypes of those cells. I hypothesised that cluster 0 at resolution = 0.1 for all four samples here represent the keratinocytes, and higher resolutions are able to subcluster the various different types of keratinocytes present within the epidermis, such as basal, spinous, and granular keratinocytes. Four cell types were expected to be present within all samples: keratinocytes, endothelial cells, melanocytes, and fibroblasts/fibroblast-like cells. Given the relatively low cell numbers present in the clustering of each sample, resolution = 0.1 was chosen as the resolution parameter, as increasing the resolution resulted in even smaller clusters, potentially making it difficult to identify the clusters downstream.


a)
BCC_1
size
250
500
750
1000
1250
RNA_snn_res.
0.1
0.2
0.4
0.8
1
in_prop
0.00
0.25
0.50
0.75
1.00
count
1000
750
500
250


b)
BCC_2
size
300
600
900
1200
count
600
400
200
RNA_snn_res.
0.1
0.2
0.4
0.8
1
in_prop
0.00
0.25
0.50
0.75
1.00


c)
BCC_3
count
300
250
200
150
100
50
RNA_snn_res.
0.1
0.2
0.4
0.8
1
in_prop
0.00
0.25
0.50
0.75
1.00
size
50
100
150
200
250
300


d)
BCC_2_NS
RNA_snn_res.
0.1
0.2
0.4
0.8
1
in_prop
0.00
0.25
0.50
0.75
1.00
size
100
200
300
400
500
count
400
300
200
100

Appendix 5.1.2 | Clustree analysis of 4 single cell samples reveal number of clusters available at different resolution parameters. *A range of resolution parameters (0.1-1) was used along with 30 Principal Components (PCs) as input into Seurat's clustering algorithm. The Clustree package was then used to generate a clustree of the clustering, which shows the differences in cell membership of clusters between each resolution parameter. The clustrees read from top-to-bottom and are ordered from lowest-to-highest resolution parameter. Clustrees generated in this figure are from* ***a)*** *BCC_1 sample,* ***b)*** *BCC_2 sample,* ***c)*** *BCC_3 sample, and* ***d)*** *BCC_2_NS sample.*

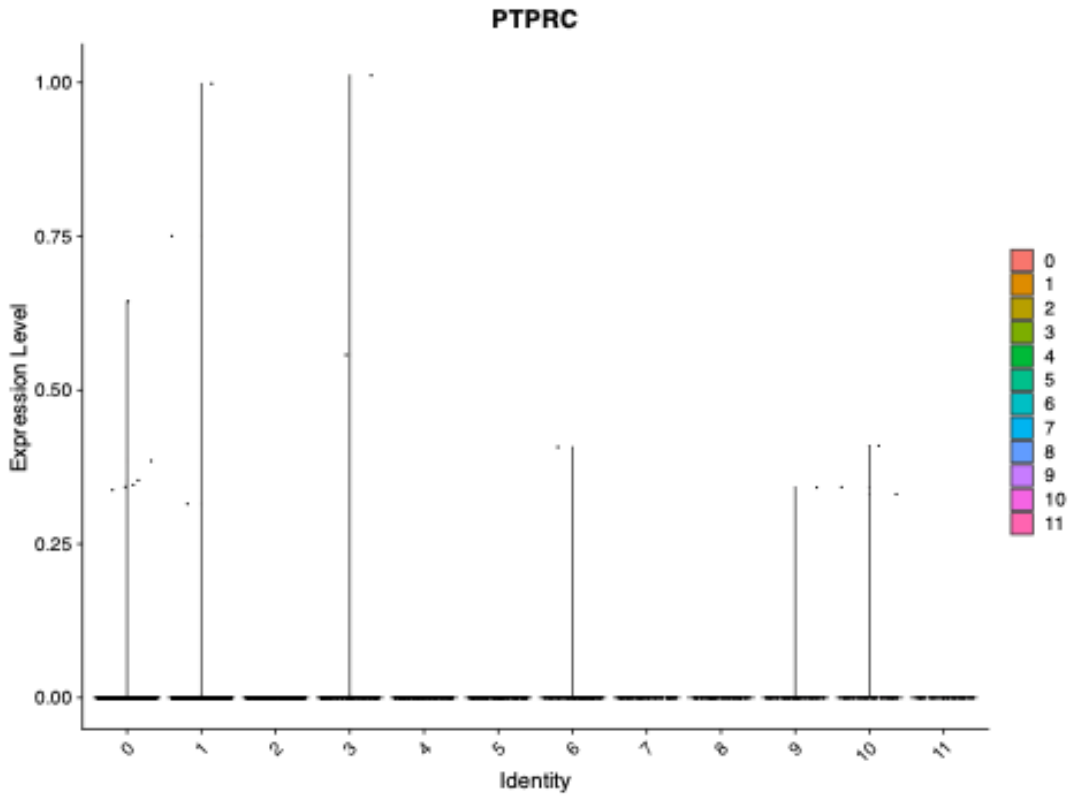
Appendix 5.2 | Visualisation

Both methods are non-linear dimensionality reduction techniques which overcome one of the major pitfalls to other dimensionality reduction techniques, such as PCA. A major problem with PCA, whilst effective at reducing the dimensionality, identifying sources of heterogeneity of the data, and preserving global structure of the data, is overcrowding. The linear nature of PCA means that where two cells may be well separated in a highly dimensional space, reducing them down to one PC may result in them overlapping or becoming very close together. T-SNE and UMAP overcome these limitations through non-linear methods and are effective at preserving local structure. The algorithms allow the separation of data that cannot be separated by a straight line (as in linear methods). Importantly, cells within the generated clusters from Seurat's graph-based clustering tend to co-localize on both dimensional reduction plots due to the use of the PCs as the input data.

Appendix 5.3 | Markers used to identify cell types

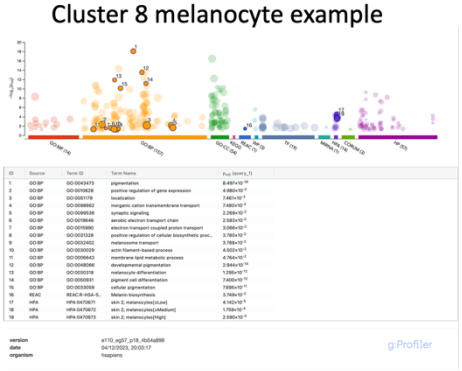
Keratinocytes	Langerhans	Endothelial Cells	Fibroblasts	SMCs	Melanocytes
ELOVL3	CD1A	CD34	COL3A1	ACTA2	PMEL
CA2	CD207	PECAM1	FBN1	ACTG2	DCT
KRT75	SPN	SELE	LUM	MYH11	GPR143
KRT34	SPI1	SLC2A1	FST	CNN1	MLANA
FASN	FCGBP	VWF	FABP4	TAGLN	TYRP1
THRSP		ABCB1	DCN		APOE
ELOVL5		ICAM2	COL1A1		FDFT1
FADS2		CDH5			
ACSBG1		KDR			
APOC1		VEGFA			
KRT6B		VEGFB			
KRT1		LYVE1			
KRT6A					
KRT16					
KRTDAP					
CASP14					
CAPNS2					
SPINK5					
COL17A1					
TP73					
KRT14					
KRT1					
KRT10					
KRTDAP					
CXCL14					
DST					

Appendix 5.4 | BCC_1 PTPRC expression via (CD45) violin plot

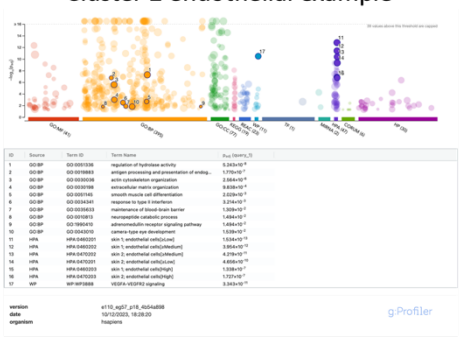


Appendix 5.5 | g:Profiler results examples for BCC_1

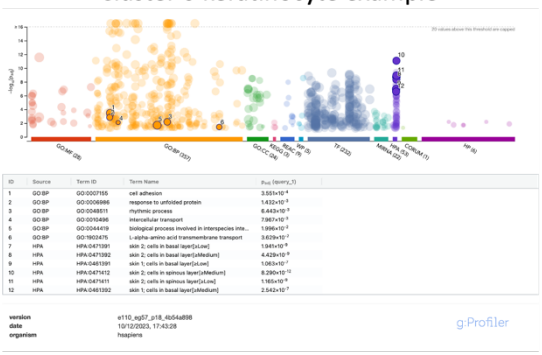
Cluster 8 melanocyte example



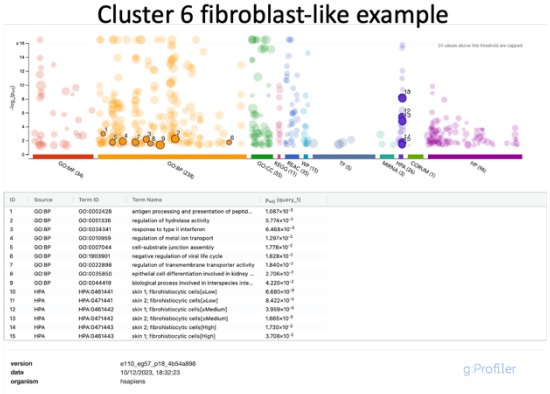
Cluster 2 endothelial example



Cluster 0 keratinocyte example



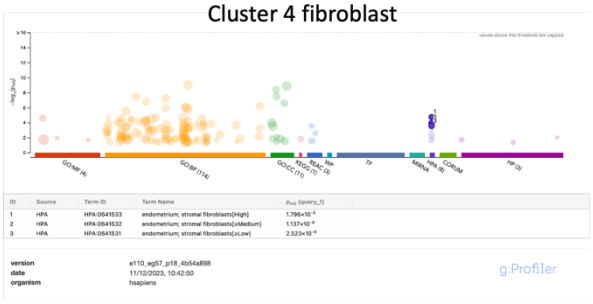
Cluster 6 fibroblast-like example



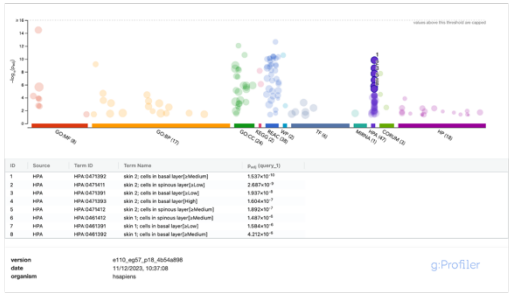
Appendix 5.6 | g:Profiler results examples for BCC_3

Cluster 0 endothelial

Cluster 4 fibroblast



Cluster 1 keratinocyte



Appendix 5.7 | Dataset integration

Appendix 5.7.1 | Writing

The primary goal of Integration is to allow for the identification of shared and/or unique subpopulations across samples (Yu et al. 2002; Colomé-Tatché and Theis 2018; Stuart et al. 2019; Lähnemann et al. 2020; Ma et al. 2020a; Forcato et al. 2021). This method takes all inputted samples and aligns cells with similar gene expression, resulting in one 'integrated' dataset with clusters representing a common cell type/subtype between all samples. Following the mostly successful identification of cell types in the individual samples, I next hypothesised that integration of the samples into one dataset would generate more accurate clustering and better separate the cell types. To test this hypothesis, I employed and tested three horizontal integration methods; Seurat v3 (Stuart and Satija 2019) , scTransform (Hafemeister and Satija 2019b), and Harmony (Korsunsky et al. 2019).

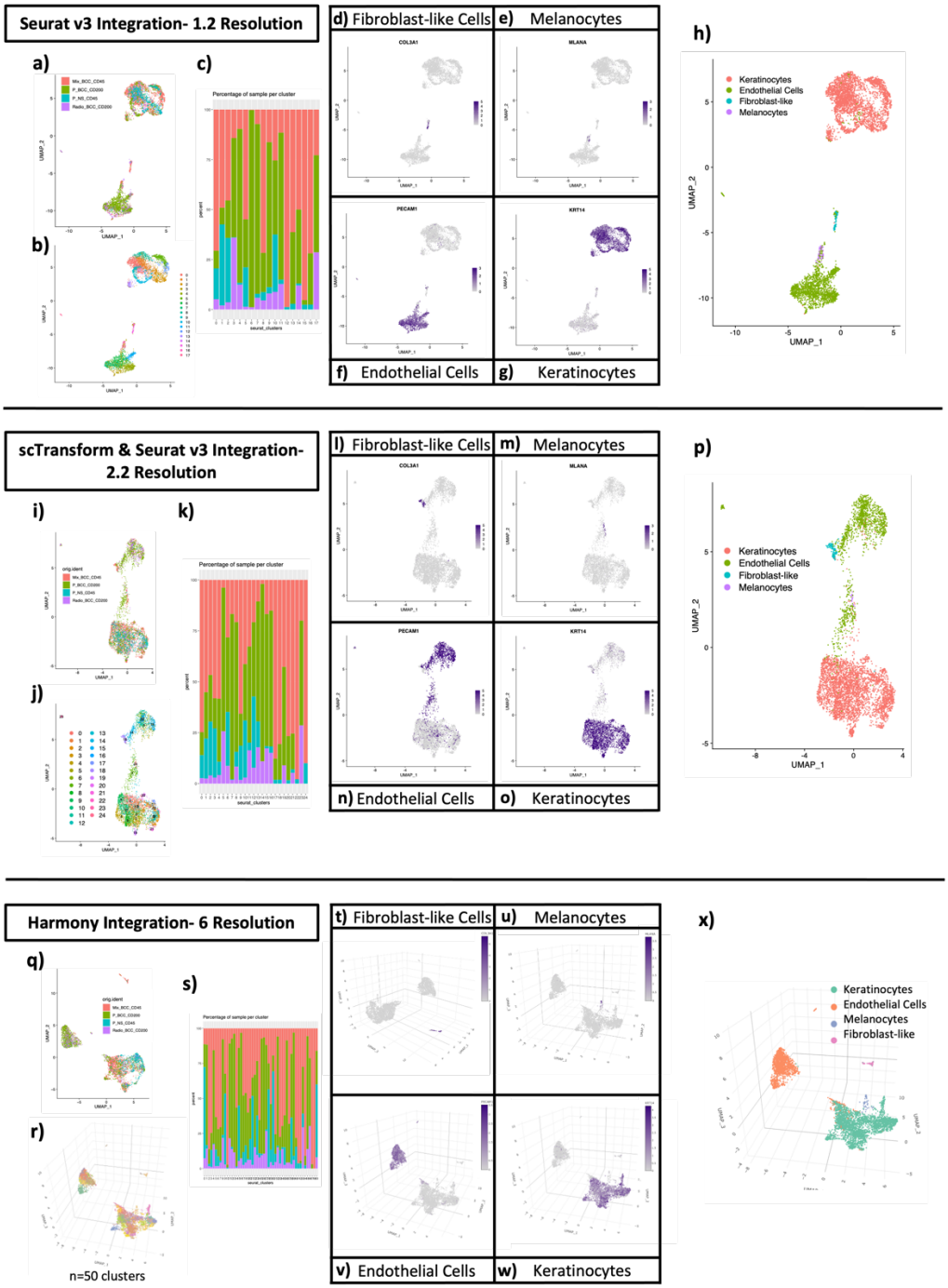
Initially, low resolution clustering (resolution = 0.1) was performed to identify the putative cell populations and assess how well the integration worked. Determining the success of the integration across all three methods proved difficult. The main factor contributing to this was the number and type of cells in each sample. Samples with more cells (BCC_1 and BCC_2) appeared to dominate in the clustering and made it difficult to see the presence of the other samples. Additionally, I was not expecting a full integration of each sample into each cluster, as some samples had been treated differently to the others. Sorting for CD200+ CD45- cells would undoubtedly cause a skew in the cell types present in those samples and vice versa. Low resolution clustering was unable to fully separate out the main cell populations in all three

methods. However, Seurat's v3 method was the only method that completely separated the endothelial cell clusters from the keratinocyte clusters. To fully isolate each cell population, a resolution of 1.2 was needed for the Seurat v3 method (Appendix 5.7.2a-h), 2.2 was needed for the scTransform method (Appendix 5.7.2i-p), and 6 was needed for the Harmony (Appendix 5.7.2q-x). Clustering at these resolutions generated 18, 25, and 50 clusters, respectively (Appendix 5.7.2b, j, and r, respectively). 3D UMAPs were required to successfully isolate the populations for the Harmony method (Appendix 5.7.2t-w). All three methods generated similar melanocyte clusters that required increased resolution to successfully isolate (Appendix 5.7.2e, m, and u). In the Seurat v3 and scTransform methods, this melanocyte population resided within/immediately adjacent to the endothelial cell clusters, whereas this population was amongst the keratinocyte clusters in the Harmony method (Appendix 5.7.2h, p, and x, respectively). The fibroblast-like population was completely isolated from all other populations in the Seurat v3 and Harmony methods but was immediately adjacent to the endothelial cell population in the scTransform method (Appendix 5.7.2h, x, and p, respectively). The endothelial cell populations failed to completely separate from the keratinocyte populations in the scTransform and Harmony methods, whereas they were completely separated in the Seurat v3 method (Appendix 5.7.2p, x, and h, respectively). In terms of complexity of the clustering, both Seurat v3 and Harmony produced more complex-appearing clusters for the keratinocyte populations, an attribute that will be beneficial during the investigations of those populations in the main dataset.

Seurat's v3 method was selected as the integration method of choice as it successfully isolated each putative cell population with a lower resolution and at this resolution, produced more complex clustering of the keratinocytes. This method also

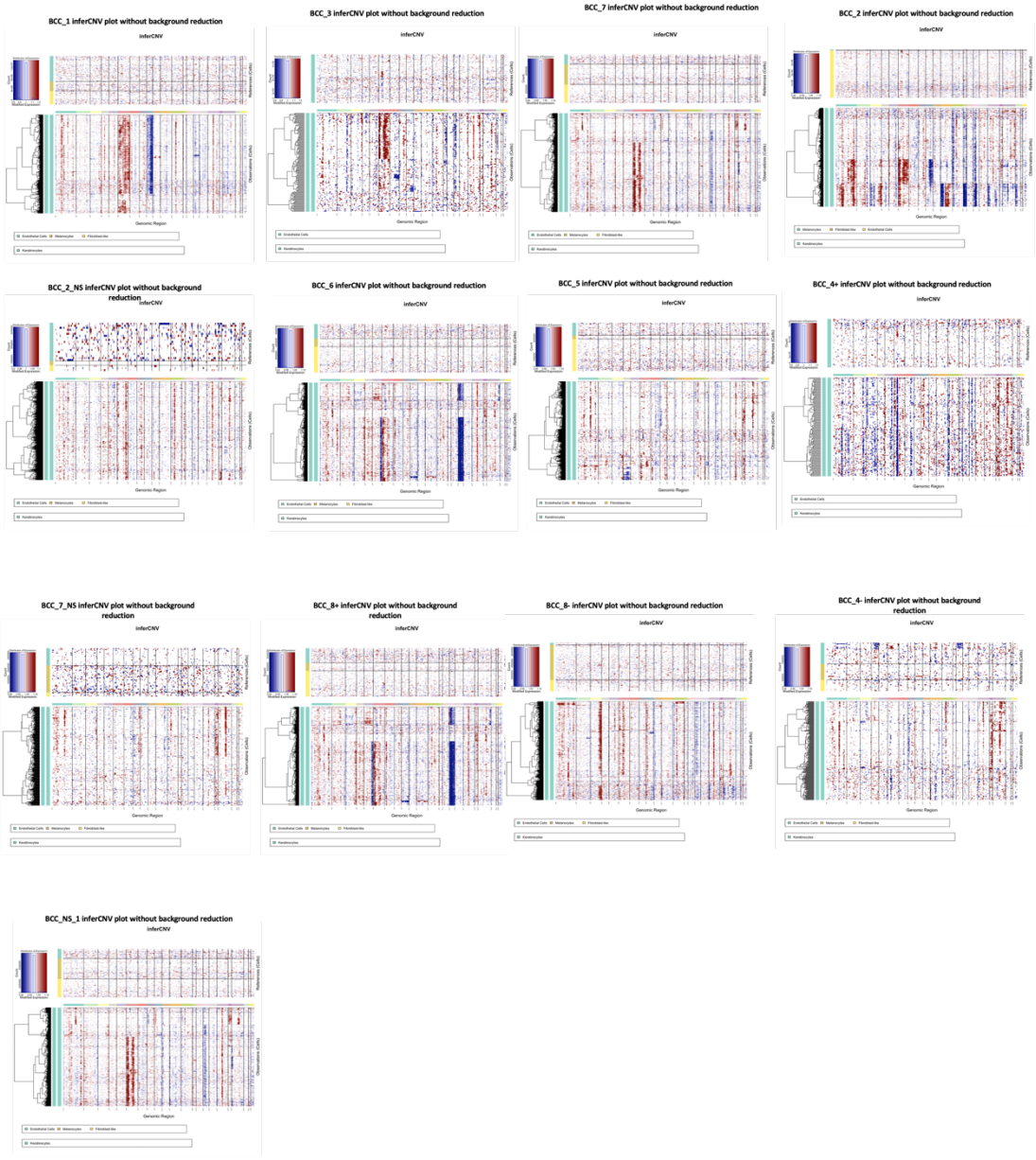
runs significantly faster than the other two methods, so will be beneficial when working with the larger dataset. In terms of complexity of the clustering, both Seurat v3 and Harmony produced more complex-appearing clusters for the keratinocyte populations, an attribute that will be beneficial during the investigations of those populations in the main dataset.

Seurat's v3 method was selected as the integration method of choice as it successfully isolated each putative cell population with a lower resolution and at this resolution, produced more complex clustering of the keratinocytes. This method also runs significantly faster than the other two methods, so will be beneficial when working with the larger dataset.


Seurat v3 Integration- 1.2 Resolution
a)
b)
c)
d) Fibroblast-like Cells
e) Melanocytes
f) Endothelial Cells
g) Keratinocytes
h)
Keratinocytes
Endothelial Cells
Fibroblast-like
Melanocytes
scTransform & Seurat v3 Integration- 2.2 Resolution
i)
j)
k)
l) Fibroblast-like Cells
m) Melanocytes
n) Endothelial Cells
o) Keratinocytes
p)
Keratinocytes
Endothelial Cells
Fibroblast-like
Melanocytes
Harmony Integration- 6 Resolution
q)
r)
n=50 clusters
s)
t) Fibroblast-like Cells
u) Melanocytes
v) Endothelial Cells
w) Keratinocytes
x)
Keratinocytes
Endothelial Cells
Melanocytes
Fibroblast-like

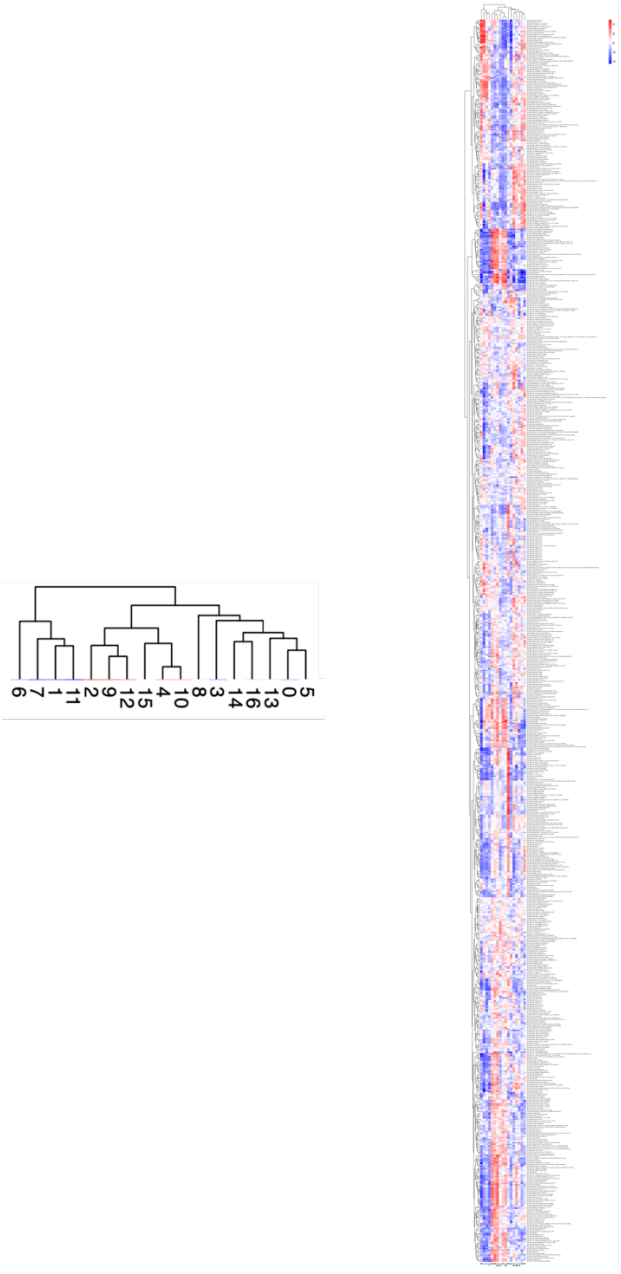
Appendix 5.7.2 | Comparison of three single cell integration methods. *Three dataset integration/batch correction methods were used and compared to integrate 4 single cell samples.* ***a-h****) Seurat's v3 integration method was used with a clustering resolution of 1.2.* ***a-b****) UMAPs showing sample distributions and Seurat cluster distributions, respectively.* ***c****) bar chart showing the percentage of samples per Seurat cluster,* ***d-g****) UMAPs showing expression of COL3A1, MLANA, PECAM1, and KRT14 that represent fibroblast-like, melanocyte, endothelial cell, and keratinocyte populations, respectively,* ***h****) annotated umap.* ***i-p****) scTransform normalisation method with Seurat's v3 integration method with a clustering resolution of 2.2.* ***i-j****) UMAPs showing sample distributions and Seurat cluster distributions, respectively.* ***k****) bar chart showing the percentage of samples per Seurat cluster,* ***l-o****) UMAPs showing expression of COL3A1, MLANA, PECAM1, and KRT14 that represent fibroblast-like, melanocyte, endothelial cell, and keratinocyte populations, respectively,* ***p****) annotated umap.* ***q-x****) Harmony integration method was used with a clustering resolution of 6.* ***q-r****) UMAPs showing sample distributions and Seurat cluster distributions (n=50 clusters), respectively.* ***s****) bar chart showing the percentage of samples per Seurat cluster,* ***t-w****) UMAPs showing expression of COL3A1, MLANA, PECAM1, and KRT14 that represent fibroblast-like, melanocyte, endothelial cell, and keratinocyte populations, respectively,* ***x****) annotated umap.*

Appendix 5.8 | InferCNV analysis without background reduction

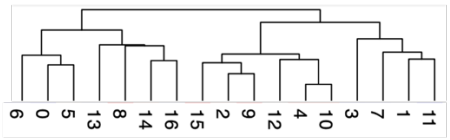


Appendix 5.9 | single cell GSEA results

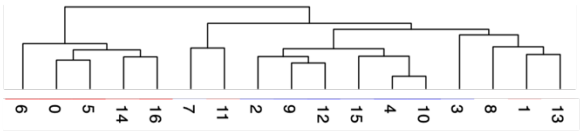
Appendix 5.9.1 | Reactome

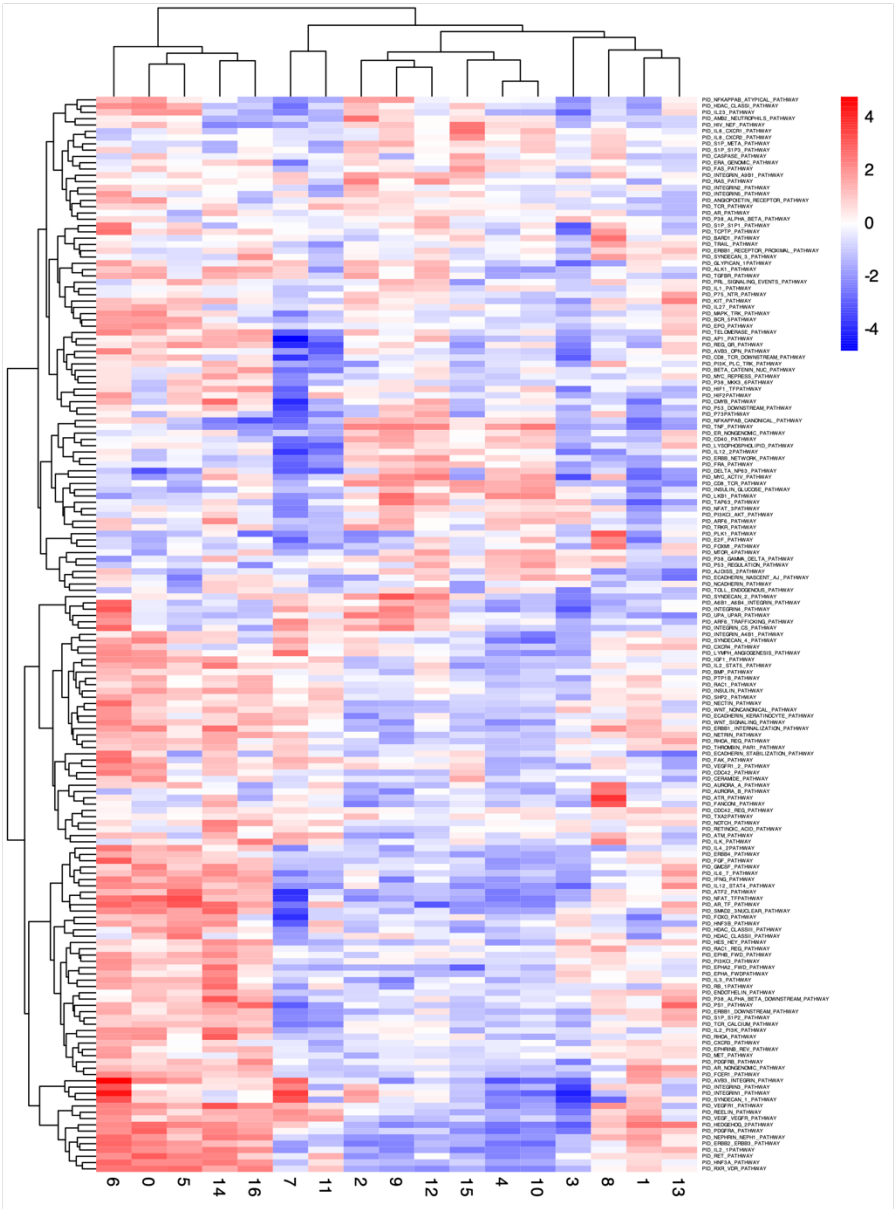


Appendix 5.9.2 | WikiPathways

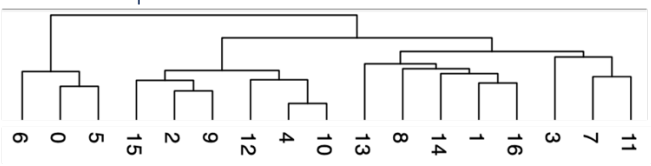


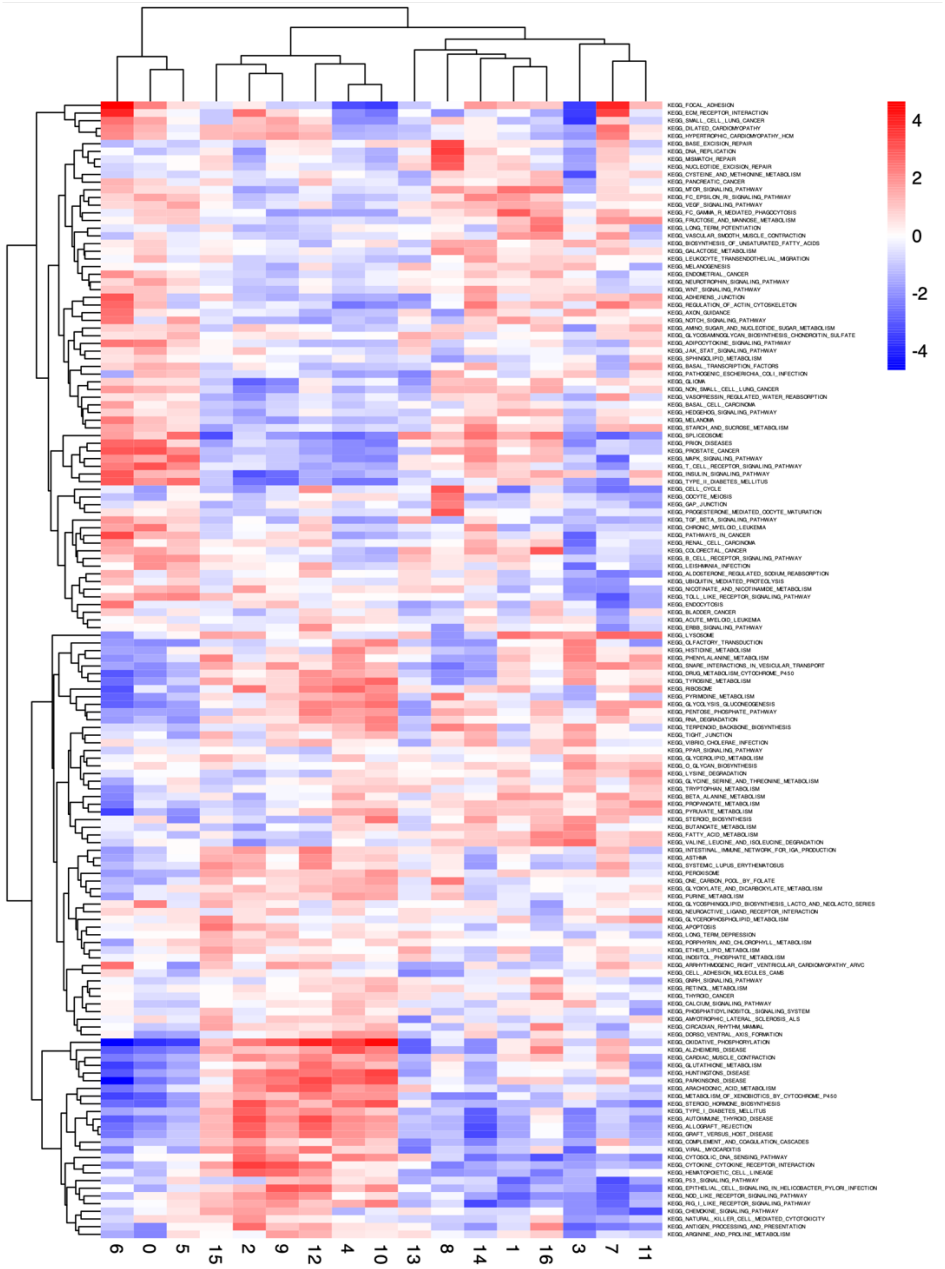
Appendix 5.9.3 | PID



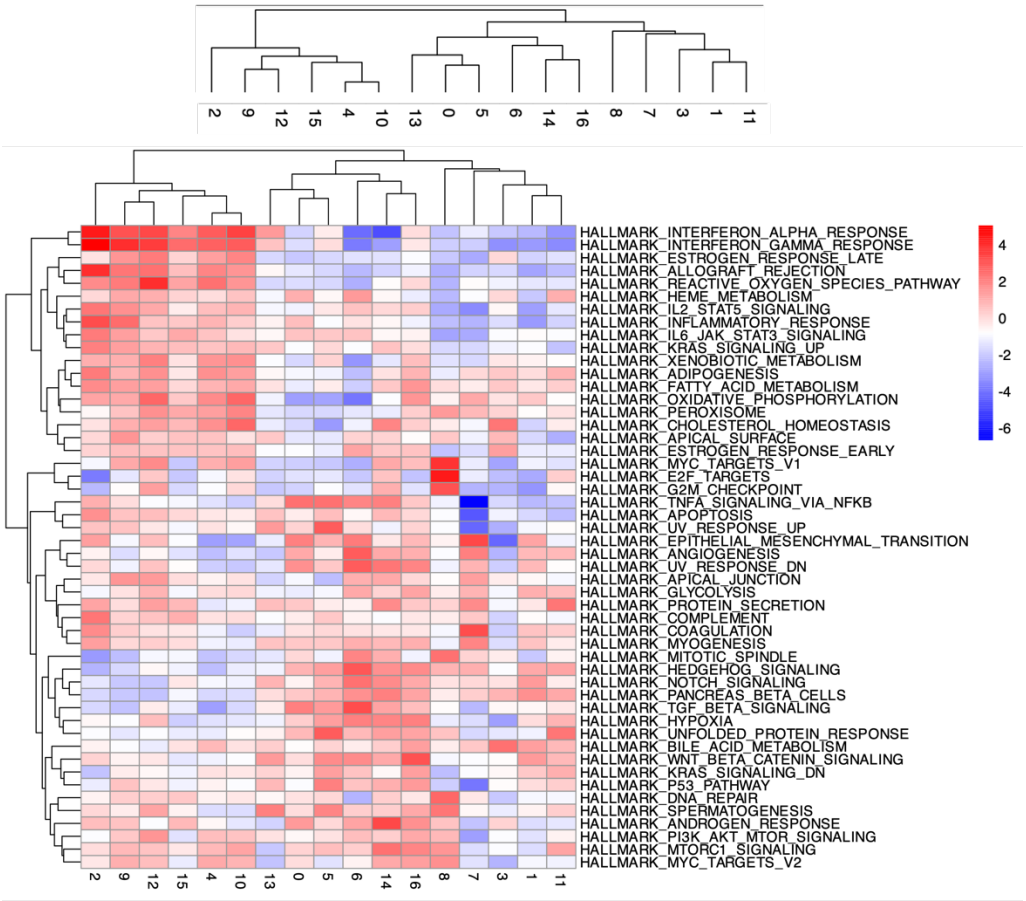


Appendix 5.9.4 | KEGG

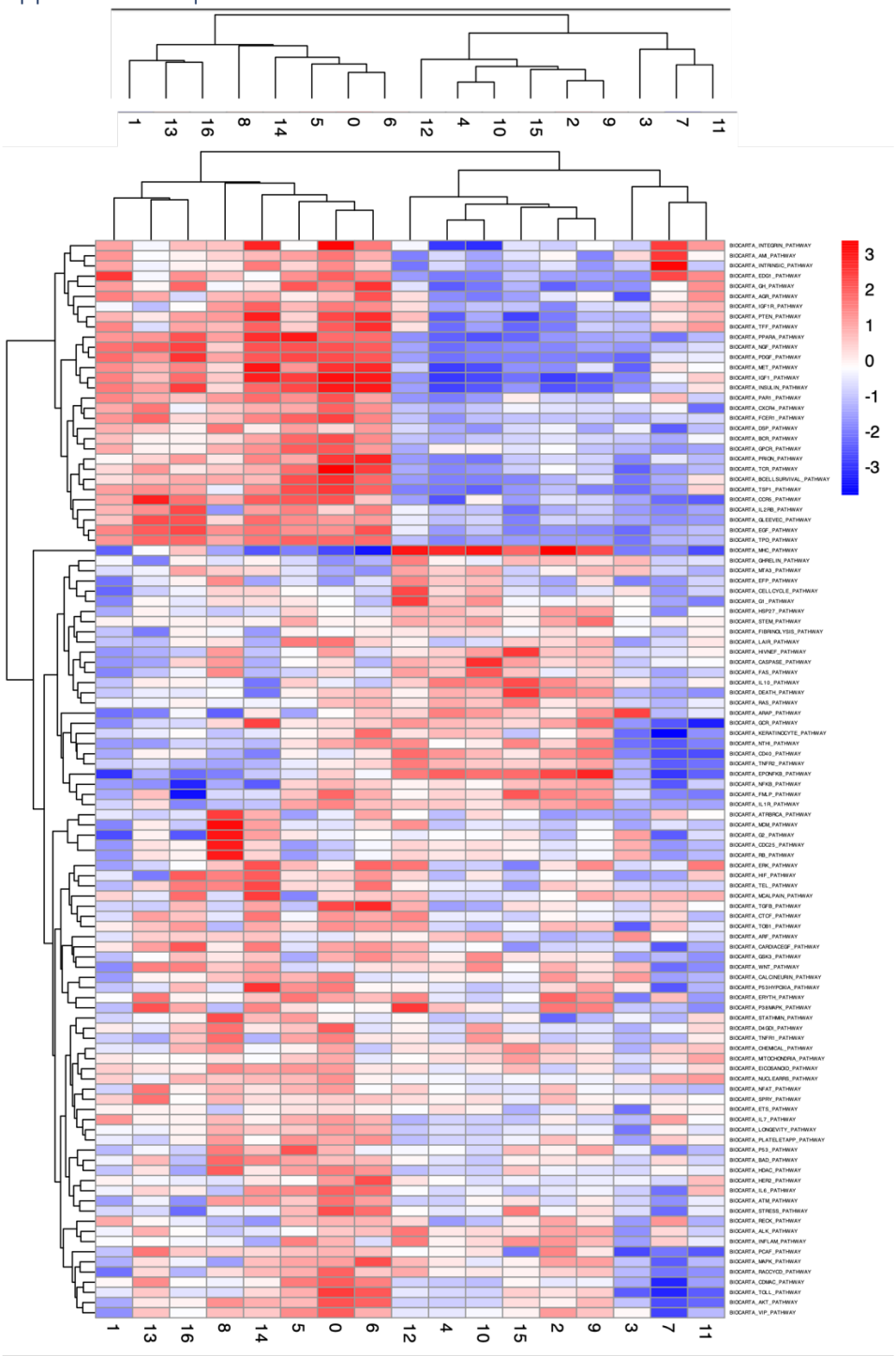




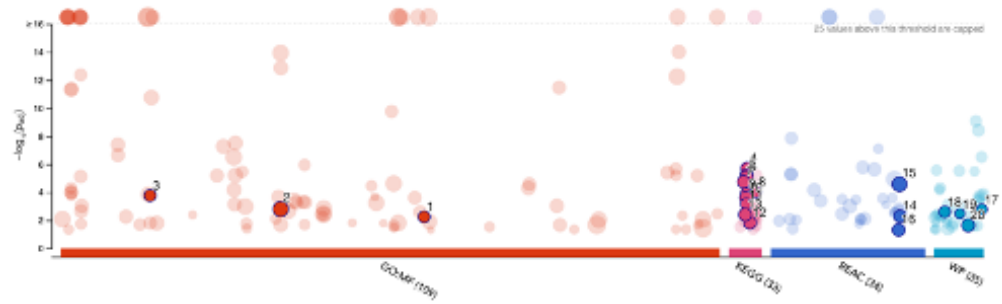
Appendix 5.9.5 | Hallmarks



Appendix 5.9.6 | Biocarta



Appendix 5.10 | g:Profiler results of Basal_BCC cluster DEGs

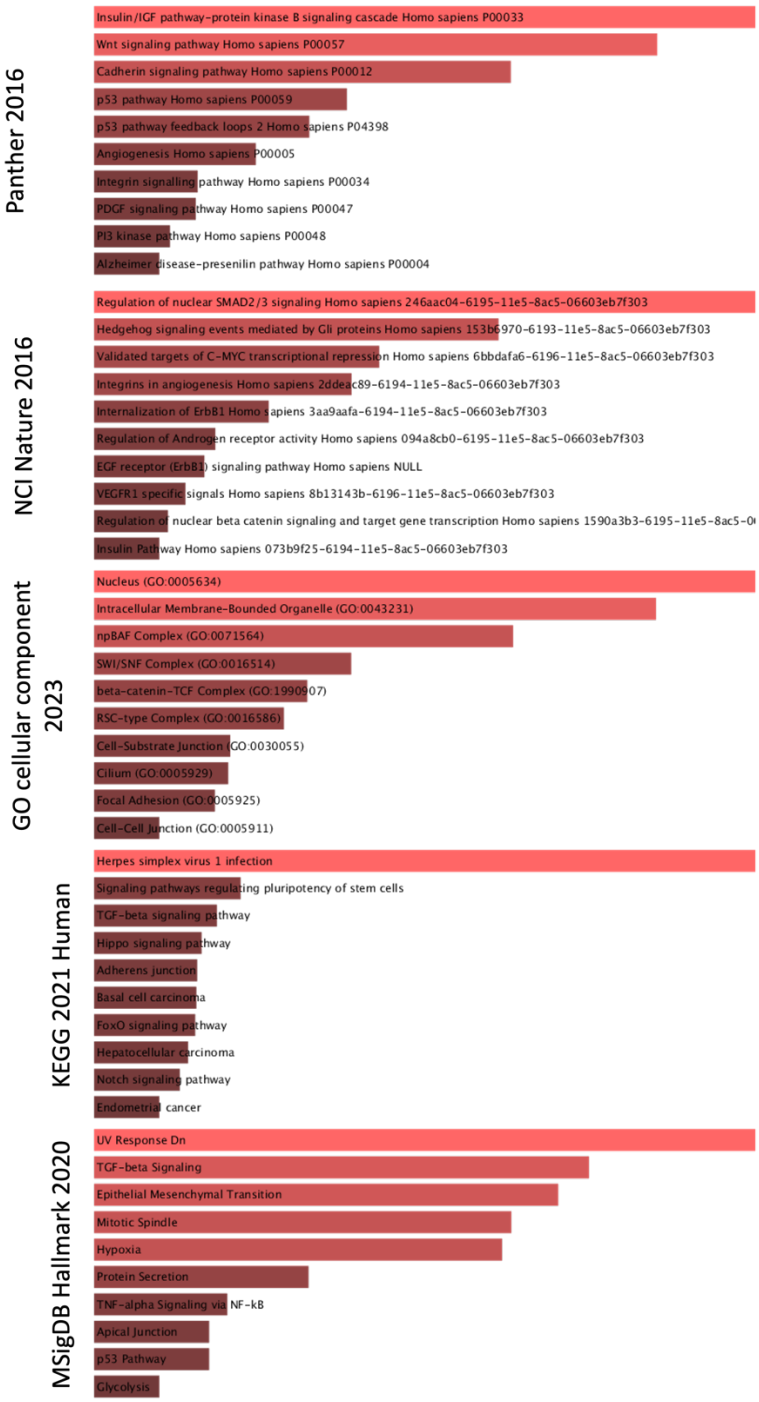


ID	Source	Term ID	Term Name	p_{adj} [query_1]
1	GO:MF	GO:0046332	SMAD binding	5.602×10^{-3}
2	GO:MF	GO:0019900	kinase binding	1.554×10^{-3}
3	GO:MF	GO:0008013	beta-catenin binding	1.634×10^{-3}
4	KEGG	KEGG:04550	Signaling pathways regulating pluripotency of ...	2.027×10^{-5}
5	KEGG	KEGG:04350	TGF-beta signaling pathway	5.521×10^{-4}
6	KEGG	KEGG:04390	Hippo signaling pathway	7.972×10^{-4}
7	KEGG	KEGG:04068	FoxO signaling pathway	1.931×10^{-3}
8	KEGG	KEGG:05217	Basal cell carcinoma	3.136×10^{-3}
9	KEGG	KEGG:04330	Notch signaling pathway	1.238×10^{-2}
10	KEGG	KEGG:04310	Wnt signaling pathway	2.438×10^{-2}
11	KEGG	KEGG:04340	Hedgehog signaling pathway	6.751×10^{-2}
12	KEGG	KEGG:04910	Insulin signaling pathway	1.457×10^{-1}
13	KEGG	KEGG:04150	mTOR signaling pathway	3.956×10^{-1}
14	REAC	REAC:R-HSA-9...	Signaling by TGFB family members	5.053×10^{-3}
15	REAC	REAC:R-HSA-9...	Signaling by Receptor Tyrosine Kinases	2.649×10^{-2}
16	REAC	REAC:R-HSA-1...	Signaling by NTRKs	4.812×10^{-2}
17	WP	WP:WP4249	Hedgehog signaling pathway	1.651×10^{-2}
18	WP	WP:WP61	Notch signaling pathway	2.436×10^{-2}
19	WP	WP:WP4537	Hippo-YAP signaling	3.341×10^{-2}
20	WP	WP:WP437	EGFR/EGFR signaling pathway	2.274×10^{-1}

version	e110_eg57_p18_4b54a898
date	01/02/2024, 23:19:49
organism	hsapiens

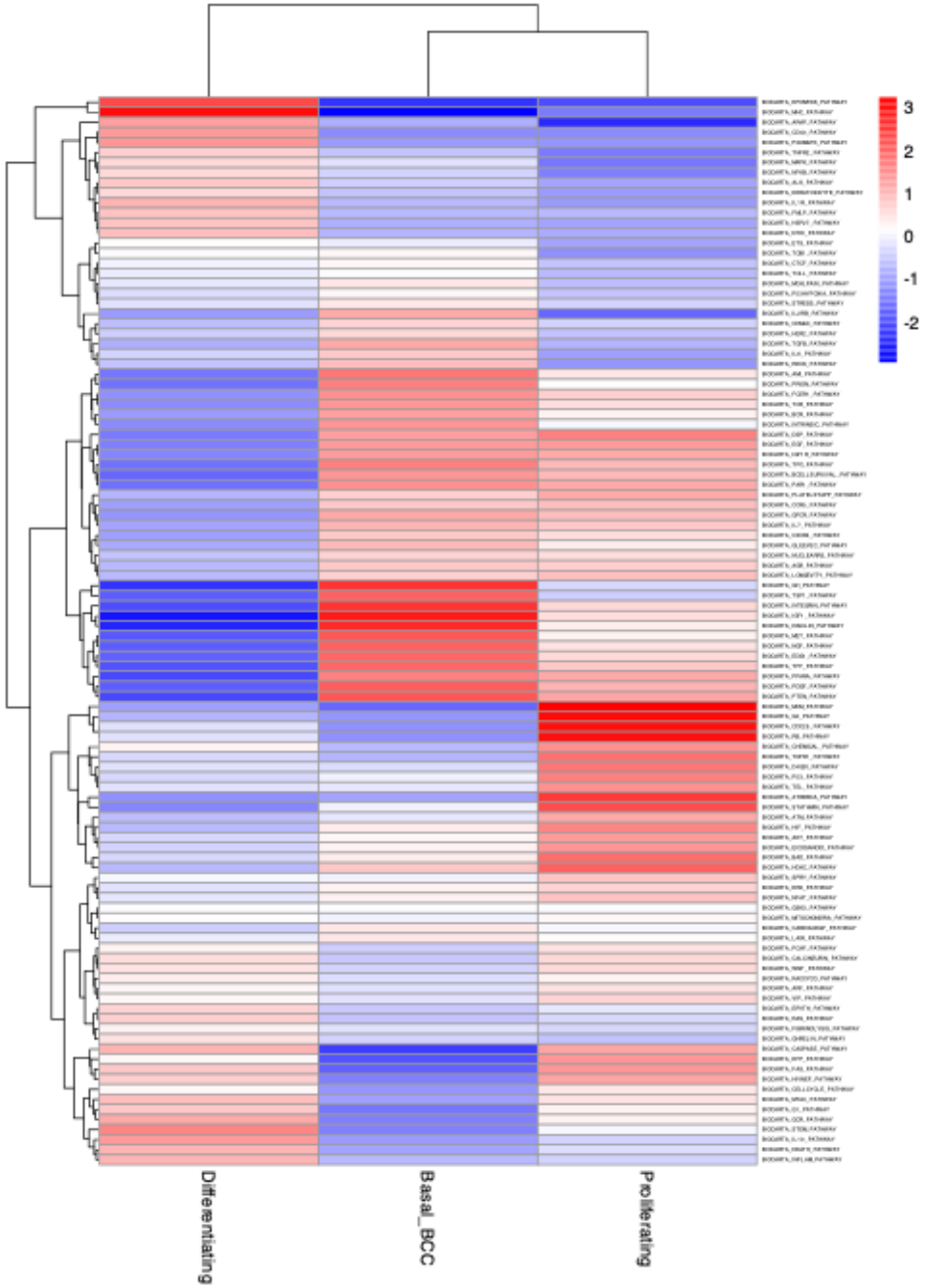
g:Profiler

Appendix 5.11 | EnrichR results of Basal_BCC cluster DEGs

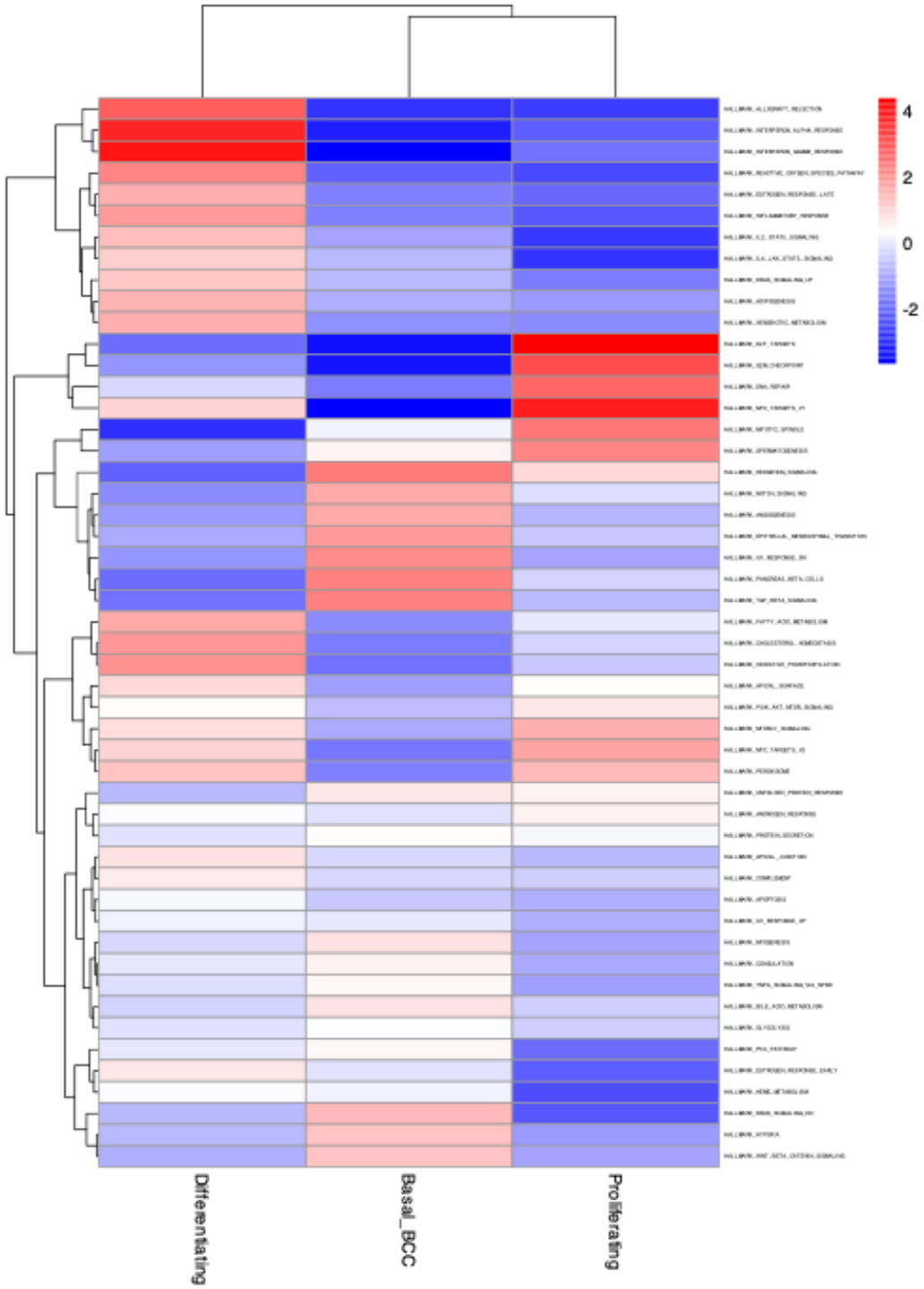


Appendix 5.12 | single cell GSEA results for Basal_BCC cluster DEGs

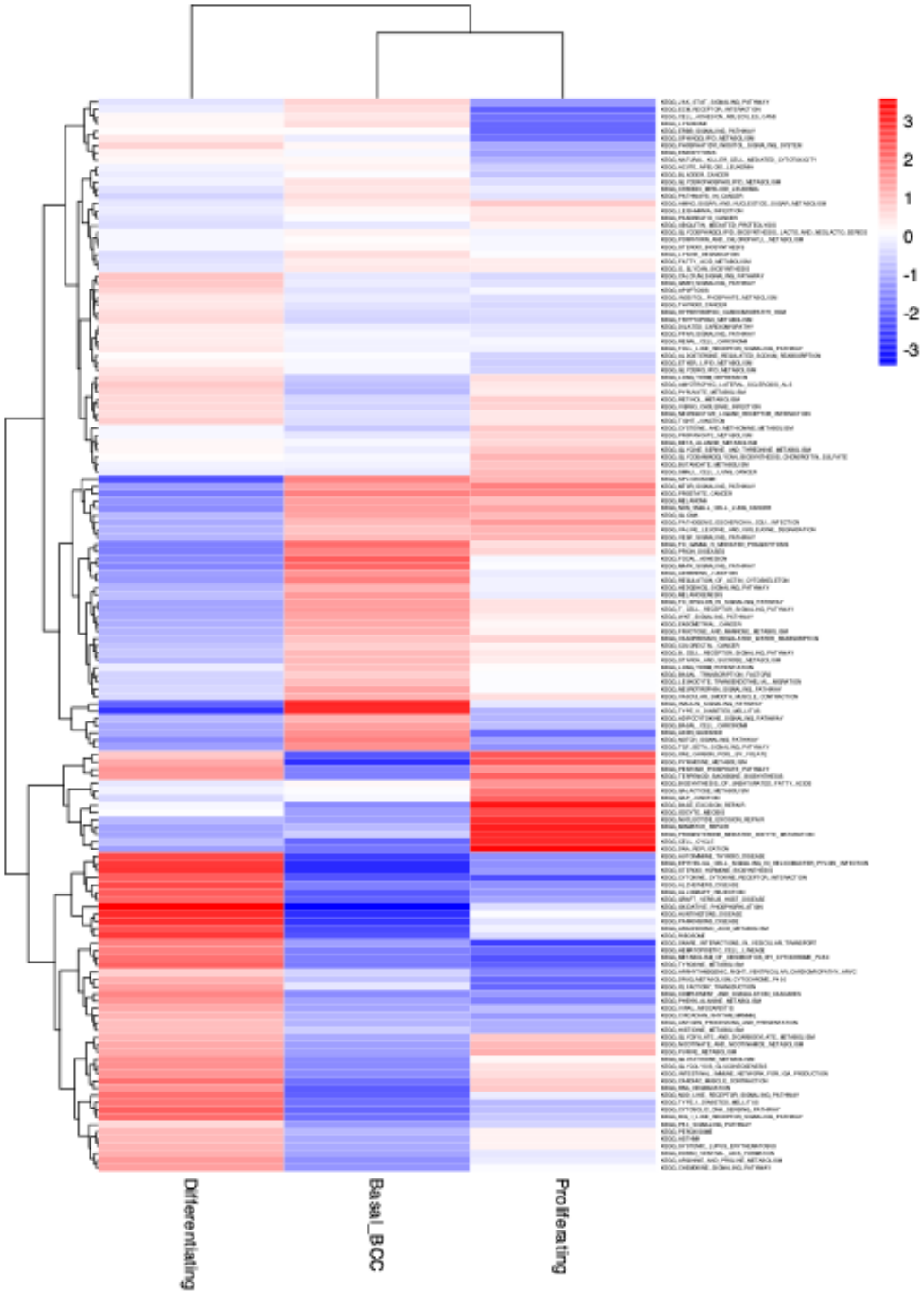
Appendix 5.12.1 | Biocarta



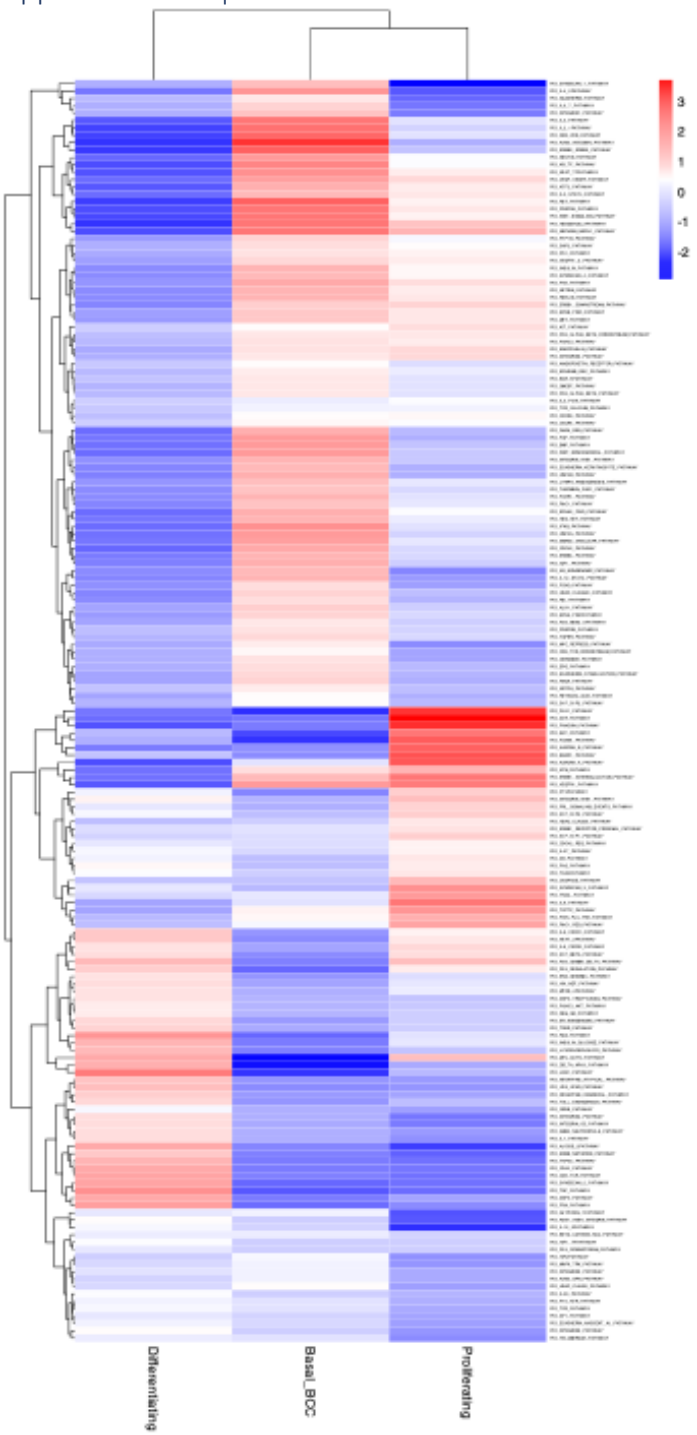
Appendix 5.12.2 | Hallmark



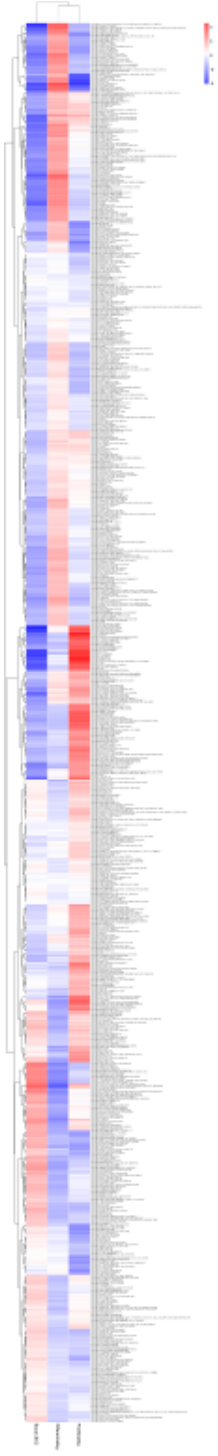
Appendix 5.12.3 | KEGG



Appendix 5.12.4 | PID



Appendix 5.12.5 | Reactome



Appendix 5.12.6 | WikiPathways