



Bioinformatic Analysis of Human Dendritic Cell Development

Anastasia Resteu

Thesis submitted to the Translational and Clinical Research Institute,
Newcastle University for the degree of Doctor of Philosophy

July 2020

Abstract

Dendritic cells (DCs) are professional antigen presenting cells of the mammalian immune system, and constitute a vital link between the adaptive and the innate immune systems. These cells are phenotypically and functionally highly heterogeneous, comprising at least 3 subsets in human: classical/conventional DC types 1 and 2 (cDC1, cDC2), and plasmacytoid DC (pDC). Additional heterogeneity has been described within the cDC2 compartment which can be divided into two populations, termed DC2 and DC3, more closely related to cDC1s or monocytes, respectively. DCs develop in the bone marrow (BM) under the control of lineage-specific transcription factors (TFs). However, the cellular pathways and genetic factors that govern the development of human DC subpopulations from haematopoietic stem cells are not well known, in part due to their rarity *in vivo*.

First, this work addressed the rarity of DCs via a novel *in vitro* culture system that favoured the production of large numbers of DCs from primary human CD34⁺ stem/progenitor cells. Two transcriptomic approaches were employed to verify the culture output: the NanoString assay and bulk RNA-Seq. The transcriptomic analyses confirmed that all DC subsets produced in culture exhibited appropriate transcription profiles and bore close resemblance their *ex vivo*-derived counterparts. Furthermore, these methods attested that Notch stimulation predisposed the culture output toward the production of cDC1, the rarest of the DC subsets.

The culture system, confirmed to produce *bona fide* DC subsets, facilitated the interrogation of DC haematopoiesis to establish the phenotypic identities of putative progenitor and precursor populations. These early populations, derived from human BM, along with mature DCs, were subjected to single cell transcriptomics. Pseudo-temporal ordering and lineage branching reconstruction analyses revealed two pathways of DC development, marked by differential expression of the TF IRF8 and explaining the origin of cDC2 heterogeneity. The IRF8^{high} pathway generated pDC, cDC1 and DC2, while DC3 and monocytes developed along an IRF8^{low} trajectory. Mass cytometry analysis validated the link between the two pathways in BM and DC populations found in peripheral blood.

Finally, the project focused on determining the role of IRF8 in the homeostasis of human cDC1s and pDCs, both of which develop through the IRF8^{high} pathway and retain IRF8 expression as they mature. The previously established *in vitro* culture techniques were employed to generate sufficient DCs for low-input IRF8 chromatin immunoprecipitation, followed by high-throughput DNA sequencing (ChIP-Seq). The analysis of the ChIP-Seq data revealed that IRF8 maintains both the function and surface phenotype of cDC1s, while in pDCs it controls important functional modules.

During this work, a wide variety of transcriptomic and genomic bioinformatic techniques and analyses enabled the verification of a novel human DC culture system (Kirkling and Cytlak et al., 2018), the identification of two pathways of human DC development (Cytlak and Resteu et al., 2020), and have generated new insights into the role of IRF8 in human DCs.

Acknowledgements

I am extremely grateful to Dr Venetia Bigley, who took me on as a NanoString technician and offered me the opportunity to undertake this PhD project. I appreciate her guidance and encouragement, and the inspiration to keep learning and researching. My thanks go to Prof Matthew Collin for the supervision and support during this project, as well as for offering me my current post and promoting my development as a bioinformatician. I am also grateful to Dr Roman Bauer, who offered helpful advice as part of my supervisory team.

My sincere thanks are given to all the past and current members of the Human Dendritic Cell (HuDC) Lab. I am particularly thankful to Dr Urszula Cytlak, who contributed to this project through cell culture and flow cytometry. My thanks go to Dr Preeti Singh and Dr Rachel Dickinson, who guided me through wet lab techniques. I am also grateful to Sarah Pagan and Dr Paul Milne, who were always happy to lend a helping hand, a listening ear, or a shoulder to lean on.

I appreciate the help of the Bioinformatics Support Unit at Newcastle University, who offered extremely useful learning materials. I am particularly grateful to Dr Rachel Queen, who introduced me to the world of single cell transcriptomics. The Genomics Core Facility at Newcastle University have been essential for this project and I am thankful for their help with the generation of the bulk RNA-Seq and ChIP-Seq datasets. This work would not have been possible without all the donors and clinicians who facilitated the process of obtaining material for research.

Finally, my thanks go to my family and friends, and particularly Kile Green, whose support and encouragement during the course of this project were invaluable.

Candidate's declaration

Chapter 3

The following research facilities and members of the HuDC Lab contributed to the generation and analysis of the data presented within this chapter:

- Cell culture, flow cytometry, and cell sorting were performed by Dr Urszula Cytlak (HuDC Lab).
- Library preparation and sequencing (bulk RNA-Seq) were undertaken at the Genomics Core Facility, Newcastle University.
- The RNA-Seq pipeline for read trimming, alignment, and counting was adapted from the scripts shared by Dr Rachel Queen (Bioinformatics Support Unit, Newcastle University). The analysis was performed solely by the candidate.

The candidate's contribution includes:

- Generation of NanoString data on the nCounter system
- Analysis of gene expression NanoString data, including data QC and normalisation, gene filtering, principal component analysis, heatmaps and clustering, differential expression testing, generation of volcano plots, testing for overrepresentation of Gene Ontology terms
- Processing of bulk RNA-Seq data: QC, trimming, alignment, and counting of reads
- Analysis of bulk RNA-seq data: normalisation, gene filtering, principal component analysis, heatmaps, clustering, gene set enrichment analysis, testing for overrepresentation of Gene Ontology terms

Chapter 4

The following research facilities and members of the HuDC Lab contributed to the generation and analysis of the data presented within this chapter:

- Cell culture, flow cytometry, and cell sorting were performed by Dr Urszula Cytlak (HuDC Lab).
- Library preparation and sequencing (scRNA-Seq) were undertaken at the Wellcome Centre for Human Genetics, Oxford.
- The scRNA-Seq pipeline for read trimming, alignment, and counting was adapted from the scripts shared by Dr Rachel Queen (Bioinformatics Support

Unit, Newcastle University). The analysis was performed solely by the candidate.

- Mass cytometry experiments were performed by Sarah Pagan (HuDC Lab) and the data pre-processing with the FlowJo software was undertaken by Dr Venetia Bigley (HuDC Lab).
- The majority of data, figures, methods, and findings presented in Chapter 4 have been submitted for publication as part of a joint first author paper (Cytlak and Resteu et al., 2020).

The candidate's contribution includes:

- Visualisation of flow cytometry data to generate a 3D plot of antigen expression
- Generation of NanoString data on the nCounter system
- Analysis of NanoString gene expression data: normalisation, gene filtering, principal component analysis
- Processing of single cell RNA-Seq data: QC, trimming, alignment, and counting of reads
- Analysis of single cell RNA-Seq data: gene and cell QC and filtering, normalisation, hierarchical clustering, dimensionality reduction using tSNE, lineage tracing with diffusion maps, and pseudotime analysis
- Computational analysis of mass cytometry data: visualisation of tSNE embeddings, lineage reconstruction with diffusion maps

Chapters 5 and 6

The following research facilities and members of the HuDC Lab contributed to the generation and analysis of the data presented within these chapters:

- Cell culture and cell sorting experiments were performed jointly by the candidate and Dr Urszula Cytlak (HuDC Lab).
- Sequencing (ChIP-Seq) was undertaken at the Genomic Core Facility, Newcastle University. The sequencing libraries for this experiment were prepared by the candidate.

All further data presented in Chapters 5 and 6 were produced and analysed solely by the candidate.

Publications arising from this thesis

- Cytlak U*, **Resteu A***, Pagan S, Green K, Milne P, Maisuria S, McDonald D, Hulme G, Filby A, Carpenter B, Queen R, Hambleton S, Hague R, Lango Allen H, Thaventhiran J, Doody G, Collin M, Bigley V, Differential IRF8 requirement defines two pathways of dendritic cell development in humans, *Immunity*. 2020;53(2):353-370.e8. doi:10.1016/j.immuni.2020.07.003. *Joint 1st authors
- Bellmann L, Zelle-Rieser C, Milne P, **Resteu A**, Tripp CH, Hermann-Kleiter N, Zaderer V, Wilflingseder D, Hörtnagl P, Theochari M, Schulze J, Rentzsch M, Del Frari B, Collin M, Rademacher C, Romani N, Stoitzner P, Notch-mediated generation of monocyte-derived Langerhans cells: Phenotype and function, *The Journal of Investigative Dermatology* 2020, doi: <https://doi.org/10.1016/j.jid.2020.05.098>.
- Cytlak U, **Resteu A**, Bogaert D, Kuehn HS, Altmann T, Gennery A, Jackson G, Kumanovics A, Voelkerding KV, Prader S, Dullaers M, Reichenbach J, Hill H, Haerynck F, Rosenzweig SD, Collin M, Bigley V. Ikaros family zinc finger 1 regulates dendritic cell development and function in humans. *Nature Communications* 2018, 9, 1239. doi:10.1038/s41467-018-02977-8
- Kirkling ME, Cytlak U, Lau CM, Lewis KL, **Resteu A**, Khodadadi-Jamayran A, Siebel CW, Salmon H, Merad M, Tsirigos A, Collin M, Bigley V, Reizis B. Notch signaling facilitates in vitro generation of cross-presenting classical dendritic cells. *Cell Reports* 2018, 23(12), 3658-3672.e6. doi:10.1016/j.celrep.2018.05.068

Presentations delivered by the candidate

- Oral presentation at the Institute of Cellular Medicine Research Seminar, May 2019: “Differential IRF8 requirement defines two pathways of dendritic cell development in Humans”
- Joint oral presentation at PI seminar series within the Institute of Cellular Medicine, March 2019: “Differential IRF8 requirement defines two pathways of human dendritic cell development *in vivo*”
- Oral presentation at the Institute of Cellular Medicine Research Seminar, May 2018: “Molecular control of human dendritic cell development”
- Joint oral presentation at the Dean of Research Celebratory Lecture (Faculty of Medical Sciences), June 2018: “Dendritic cells and the fall of IKAROS”
- Poster presentation at the 15th International Symposium on Dendritic Cells, June 2018: “Notch signaling facilitates *in vitro* generation of human IRF8-dependent CLEC9A⁺ classical dendritic cells”

List of abbreviations

APC	Antigen Presenting Cell
AICE	AP-1–IRF Composite Element
AP-1	Activating Protein-1
ATAC-Seq	Assay for Transposase-Accessible Chromatin using Sequencing
AUROC	Area Under the Receiver Operating Characteristic Curve
BLASTP	Protein Basic Local Alignment Search Tool
BM	Bone Marrow
BMMC	Bone Marrow Mononuclear Cells
bp	Base Pair
BP	Biological Processes
Bubble GUM	GSEA Unlimited Map
bZIP	Basic Region/Leucine Zipper
CD	Cluster of Differentiation
cDC1	Classical Dendritic Cell Type 1
cDC2	Classical Dendritic Cell Type 2
CDP	Common Dendritic Cell Progenitor
ChIP	Chromatin Immunoprecipitation
ChIP-Seq	Chromatin Immunoprecipitation Sequencing
CLEC	C-Type Lectins
CLR	C-Type Lectin Receptors
CMP	Common Myeloid Progenitor
CPM	Counts per Million
CytoF	Cytometry by Time of Flight
DBD	DNA-Binding Domain
DC	Dendritic Cell
DC2	Dendritic Cell 2 (a subset of cDC2)
DC3	Dendritic Cell 3 (a subset of cDC2)
DLL1	Delta-Like Protein 1
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic Acid
DPBS	Dulbecco's Phosphate-Buffered Saline
EDTA	Ethylenediaminetetraacetic Acid
EGFP	Enhanced Green Fluorescent Protein
EICE	Ets–IRF composite element
ENCODE	Encyclopaedia of DNA Elements
EPCAM	Epithelial Cell Adhesion Molecule
ERCC	External RNA Control Consortium
FACS	Fluorescence-Activated Cell Sorting
FCS	Foetal Calf Serum
FLT3	Fms-Related Tyrosine Kinase 3
FLT3L	Fms-Related Tyrosine Kinase 3 Ligand
FSC	Forward Scatter
FSGM	Cytokine cocktail consisting of FLT3, SCF, and GM-CSF

GFP	Green Fluorescent Protein
GM-CSF	Granulocyte-Macrophage Colony-Stimulating Factor
GMP	Granulocyte-Monocyte Progenitor
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
HI-FCS	Heat-Inactivated Foetal Calf Serum
HSC	Haematopoietic Stem Cell
HuDC Lab	Human Dendritic Cell Laboratory
IAD	IRF Association Domain
ICSBP	Interferon Consensus Sequence-Binding Protein
IECS	IRF–Ets Composite Sequence
IFN	Interferon
IP	Immunoprecipitation
IRF	Interferon Regulatory Factor
ISG	Interferon-Stimulated Gene
ISRE	IFN-Stimulated Response Element
LC	Langerhans Cell
LMPP	Lymphoid-primed Multipotent Progenitor
M-CSF	Macrophage colony stimulating factor
MACS2	Model-based Analysis of ChIP-Seq version 2
MCS	Multiple Cloning Site
MDP	Macrophage-Dendritic Cell Progenitor
MEM α	Minimum Essential Medium α
MHC	Major Histocompatibility Complex
moDC	Monocyte-Derived Dendritic Cell
NLR	NOD-Like Receptor
PB	Peripheral Blood
PBMC	Peripheral Blood Mononuclear Cells
PBS	Phosphate-Buffered Saline
PC	Principal Component
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
pDC	Plasmacytoid Dendritic Cell
PRR	Pattern Recognition Receptor
QC	Quality Control
RCC	Reporter Code Count
RNA	Ribonucleic Acid
RNA-Seq	RNA Sequencing
RT	Room Temperature
SC3	Single-Cell Consensus Clustering
SCF	Stem Cell Factor
scRNA-Seq	Single Cell RNA Sequencing
SEM	Standard Error of the Mean
SLAN	6-sulfo LacNAc
SSC	Side Scatter

ssGSEA	Single Sample Gene Set Enrichment Analysis
TCR	T Cell Receptor
TF	Transcription Factor
TLR	Toll-Like Receptor
toIDC	Tolerogenic Dendritic Cell
tSNE	t-distributed Stochastic Neighbour Embedding
TSS	Transcription Start Site
w/v	Weight per Volume

Table of contents

Abstract	II
Acknowledgements	IV
Candidate's declaration	VI
Publications arising from this thesis	VIII
Presentations delivered by the candidate	IX
List of abbreviations	X
Table of contents	XIII
List of tables	XVII
List of Figures	XVII
Chapter 1. Introduction	1
1.1 <i>The roles of DCs in immunity</i>	1
1.1.1 Antigen presentation	1
1.1.2 T cell polarisation	4
1.1.3 Induction of tolerance	5
1.2 <i>Peripheral blood dendritic cell subsets</i>	5
1.2.1 Classical dendritic cells	5
1.2.1.1 Classical dendritic cell type 1	6
1.2.1.2 Classical dendritic cell type 2	6
1.2.2 Plasmacytoid dendritic cell	7
1.2.3 <i>Tissue dendritic cells</i>	7
1.2.3.1 Langerhans cells	8
1.2.3.2 Inflammatory dendritic cells	8
1.3 <i>Monocyte subsets</i>	10
1.4 <i>SLAN DC</i>	10
1.4 <i>In vitro-derived DCs</i>	10
1.5 <i>DC therapy</i>	11
1.6 <i>Murine DCs and monocytes</i>	12
1.7 <i>DC haematopoiesis</i>	13
1.7.1 Transcription regulation in the context of DC haematopoiesis	15
1.7.2 Transcription factor IRF8	17
1.7.3 Notch signaling pathway and dendritic cell development	19
1.7.4 Growth factor requirements for DC development	21
1.7.4.1 FLT3L	21
1.7.4.2 GM-CSF	21
1.7.4.3 SCF	21
1.7.4.4 M-CSF	22
1.8 <i>Technologies employed in DC research</i>	22
1.8.1 Cytometry	22
1.8.1.1 Flow cytometry	22
1.8.1.2 Mass cytometry	24
1.8.2 Transcriptomics	24
1.8.2.1 NanoString gene expression assay	24
1.8.2.2 Bulk RNA-Seq	26
1.8.2.3 Single cell RNA-Seq	27
1.8.3 Genomics	29
1.8.3.1 ChIP-Seq	29
1.9 <i>Aims and objectives</i>	29

Chapter 2. General materials and methods	31
2.1 <i>Buffers and reagents</i>	31
2.1.1 Lymphoprep Solution	31
2.1.2 Dulbecco's Phosphate-Buffered Saline	31
2.1.3 Foetal calf serum	32
2.1.4 Flow and sort buffer	32
2.1.5 Culture media	32
2.1.6 Minimum Essential Medium α	32
2.1.7 Dulbecco's Modified Eagle Medium	32
2.1.8 RF-10	33
2.1.9 RNA Lysis Buffers	33
2.1.10 Freezing solution	33
2.1.11 NanoString nCounter reagents	33
2.2 <i>Cell lines</i>	33
2.2.1 OP9 and OP9-DL1 cell lines	33
2.2.1 HEK 293T cell line	34
2.3 <i>Sample collection and processing</i>	34
2.3.1 Isolation of peripheral blood and bone marrow aspirate mononuclear cells	34
2.3.2 Isolation of cells from hip femoral bone marrow	35
2.3.3 Cryopreservation of cells	35
2.3.4 Cell counting	35
2.3.5 Co-culture of bone marrow progenitors with OP9 and OP9-DL1 cell lines	36
2.3.6 General flow cytometry and cell sorting	36
2.3.7 NanoString Assay	37
2.4 <i>Bioinformatic techniques and analysis pipelines</i>	37
2.4.1 Clustering and dimensionality reduction	37
2.4.1.1 Principal component analysis	37
2.4.1.2 t-distributed stochastic neighbour embedding	38
2.4.1.3 Diffusion maps	38
2.4.1.4 Hierarchical clustering	38
2.4.2 Analysis of transcriptomic data	38
2.4.3 Analysis of ChIP-Seq data	39
Chapter 3. Transcriptional identity of DCs generated in culture	40
3.1. <i>Introduction</i>	40
3.2. <i>Materials and methods</i>	43
3.2.1. Co-culture of bone marrow progenitors with OP9 and OP9-DL1 cells	43
3.2.2. Flow cytometry and cell sorting	43
3.2.3. NanoString gene expression assay	44
3.2.4. Bulk RNA-Seq	45
3.3. <i>Results</i>	46
3.3.1. Culture output phenotyping and cell enumeration	46
3.3.2. Confirmation of cultured cell identities via NanoString analysis	50
3.3.3. Transcriptome analysis of cDC1s from blood, tissue, and culture	56
3.4. <i>Discussion</i>	62
3.4.1. The production of human DC subsets in vitro	62
3.4.2. Confirmation of culture-derived cell identity	62
3.4.3. The cells produced in culture bear close transcriptional resemblance to tissue DCs	63
3.4.4. Notch signaling facilitates in vitro generation of human cDC1s	64
3.5. <i>Summary and further work</i>	65
Chapter 4. The two pathways of DC development defined by <i>IRF8</i> expression	68
4.1 <i>Introduction</i>	68
4.2 <i>Materials and methods</i>	70
4.2.1 Contributions	70

4.2.2 Co-culture of bone marrow progenitors with OP9 cells	71
4.2.3 Flow cytometry	71
4.2.4 Analysis of NanoString gene expression data	71
4.2.5 Generation of single cell RNA-Seq data	71
4.2.6 Processing of single cell RNA-Seq data	72
4.2.7 QC and normalisation of scRNA-Seq data	72
4.2.8 Clustering of scRNA-Seq data	73
4.2.9 Dimensionality reduction of scRNA-Seq data	73
4.2.10 Generation of mass cytometry data	74
4.2.11 Mass cytometry data analysis	75
4.3 Results	76
4.3.1 Evaluation of mature cDC2 heterogeneity in human blood	76
4.3.2 Interrogation of DC haematopoiesis in vitro	78
4.3.3 Single cell transcriptomics of human bone marrow progenitors reveals two pathways of DC development distinguished by differential IRF8 expression	82
4.3.4 IRF8 ^{high} and IRF8 ^{low} pathways connect bone marrow progenitors with mature DC subsets	88
4.3.5 Coupling the two pathways in bone marrow and peripheral blood via mass cytometry	93
4.4 Discussion	97
4.4.1 Transcription factor IRF8 defines two DC developmental pathways in human	97
4.4.2 DC2 and DC3 markers support their functional specialisation	98
4.4.3 Advances in single cell technologies help reveal the fate of DCs in human	99
4.5 Summary and further work	101
Chapter 5. Optimisation of the low cell CHIP-Seq protocol	103
5.1 Introduction	103
5.1.1 The CHIP-Seq method	104
5.1.2 Guidelines for antibody testing	107
5.2 Materials and methods	109
5.2.1 Shipment and rehydration of expression vectors	109
5.2.2 Preparation of LB medium supplemented with kanamycin	109
5.2.3 Preparation of LB agar plates supplemented with kanamycin	109
5.2.4 Bacterial culture and transformation	109
5.2.5 Plasmid DNA extraction and quantification	110
5.2.6 Restriction enzyme digestion and gel electrophoresis	110
5.2.7 Sanger sequencing of vectors	110
5.2.8 Culture and transfection of HEK293T cells	110
5.2.9 Immunoprecipitation	111
5.2.10 Western blotting	111
5.2.11 Sorting of bone marrow cells for western blotting and cell culture	112
5.2.12 Co-culture of bone marrow progenitors with OP9-DL1 cells	112
5.2.13 Chromatin immunoprecipitation	112
5.2.14 Quantification of sequencing libraries	113
5.3 Results	115
5.3.1 Candidate antibodies for characterisation	115
5.3.2 Primary characterisation	117
5.3.3 Secondary characterisation	118
5.3.3.1 Confirmation of vector identities	119
5.3.3.2 Immunoprecipitation and western blotting of epitope-tagged IRF8 protein	123
5.3.4 Optimisation of sonication settings	126
5.4 Discussion	129
5.5 Summary and further work	130
Chapter 6. The role of IRF8 in cDC1 and pDC homeostasis	131
6.1 Introduction	131

<i>6.2 Materials and methods</i>	135
6.2.1 In vitro generation of cDC1s and pDCs from human bone marrow progenitors	135
6.2.2 Chromatin immunoprecipitation, DNA extraction, and sequencing library preparation	135
6.2.3 Generation and processing of ChIP-Seq data	137
6.2.4 Differential binding analysis	137
6.2.5 Peak annotation and pathway enrichment analysis	137
6.2.6 Motif analysis	138
<i>6.3 Results</i>	138
6.3.1 Overview of the IRF8 ChIP-Seq dataset	138
6.3.2 IRF8 differentially regulates immune pathways in mature cDC1s and pDCs	142
6.3.3 cDC1s maintain their cell identity through the auto-activation of IRF8	147
<i>6.4 Discussion</i>	152
6.4.1 Role of IRF8 in cDC1s	153
6.4.2 Role of IRF8 in pDCs	155
6.4.3 Pathways likely to be downregulated by IRF8	155
6.4.4 IRF8 in immunodeficiency	155
6.4.5 Technical caveats of the study	156
<i>6.5 Summary and further work</i>	157
Chapter 7. Overview, discussion, and further work	159
7.1 <i>Summary of novel findings</i>	159
7.2 <i>Overview of the techniques used</i>	160
7.3 <i>Research impact</i>	168
7.4 <i>Limitations of the project</i>	170
7.5 <i>Future research vision</i>	173
References	175
Appendix A. Microscopy of cells produced in culture	198
Appendix B. List of antibodies used for flow cytometry, FACS and mass cytometry in Chapter 4	199
Appendix C. Detailed gating strategy for the FACS-purification of bone marrow progenitors for scRNA-Seq	203
Appendix D. Detailed gating strategy for the FACS-purification of bone marrow DC and monocyte precursors for scRNA-Seq	204

List of tables

Page 9 - Table 1.1. Characteristics of human dendritic cell subsets

Page 43 - Table 3.1. Antibodies used for flow cytometry and cell sorting

Page 144 - Table 6.1. Biological pathways enriched in cDC1s and pDCs (adjusted p-value<0.05)

Page 146 - Table 6.2. Known motifs enriched (adjusted p-value <0.005) in cDC1s within the IRF8 binding peaks

Page 147 - Table 6.3. Known motifs enriched (adjusted p-value <0.005) in pDCs within the IRF8 binding peaks

Page 150 - Table 6.4. Motif scanning results for the 7 consensus peaks from the cDC1 triplicates, highlighting the bZIP, Ets, and IRF motifs identified with GimmeMotif

List of Figures

Page 3 - Figure 1.1. Schematic of antigen presentation via MHC I and MHC II

Page 4 - Figure 1.2. Schematic of the three-signal model for T cell polarisation

Page 14 - Figure 1.3. Classical and revised models of human haematopoiesis

Page 16 - Figure 1.4. Formation of the preinitiation complex as part of transcriptional regulation

Page 17 - Figure 1.5. Model of transcription factors regulating DC haematopoiesis

Page 18 - Figure 1.6. Schematic of the IRF8 protein and multiple sequence alignment of IRF8 orthologues, illustrating conserved coding sequences across multiple species

Page 20 - Figure 1.7. Canonical Notch signaling

Page 23 - Figure 1.8. Gating strategy for flow cytometric identification of human DCs and monocytes in peripheral blood

Page 25 - Figure 1.9. Schematic of the NanoString gene expression assay hybridisation chemistry

Page 26 - Figure 1.10. The NanoString nCounter FLEX platform

Page 28 - Figure 1.11. Schematic of the plate-based approach to single cell RNA sequencing

Page 31 - Figure 2.1. Schematic of samples separated using Lymphoprep

Page 42 - Figure 3.1. Schematic of the culture conditions used for experiments in Chapter 3

Page 48 - Figure 3.2. Analysis of primary peripheral blood and in vitro-derived cells via flow cytometry

Page 49 - Figure 3.3. The number of output DCs generated per input progenitor cell in culture media supplemented with a mix of cytokines (FSGM) and in co-culture with the OP9 or OP9-DL1 feeder layers

Page 51 - Figure 3.4. Heatmaps of DC subset-specific genes encoding surface markers, transcription factors, and TLRs and chemokine receptors, as determined by NanoString nCounter analysis

Page 52 - Figure 3.5. Principal component analysis of mRNA expression from FACS-purified primary and in vitro-derived DC subsets using the NanoString Human Immunology V2 panel plus 30 custom DC-related genes

Page 54 - Figure 3.6. Differential gene expression analysis between all culture-derived versus all blood DC subsets, as determined by NanoString nCounter analysis

Page 55 - Figure 3.7. Bar chart depicting the overrepresentation of Gene Ontology terms among the culture signature genes with higher expression in blood and in culture, as determined by NanoString nCounter analysis

Page 57 - Figure 3.8. Principal component analysis of primary cDC1s from peripheral blood (PB), bone marrow (BM), and spleen, and cDC1s co-cultured with OP9 and OP9-DL1, as determined by bulk RNA-Seq

Page 58 - Figure 3.9. Heatmap and clustering of primary cDC1s from peripheral blood (PB), bone marrow (BM), and spleen, and cDC1s co-cultured with OP9 and OP9-DL1, as determined by bulk RNA-Seq

Page 59 - Figure 3.10. Enrichment scores of gene signatures across primary cDC1s from peripheral blood (PB), bone marrow (BM), and spleen, and cDC1s co-cultured with OP9 and OP9-DL1, as determined by bulk RNA-Seq

Page 61 - Figure 3.11. Bar chart depicting the overrepresentation of Gene Ontology terms among the BubbleGUM signatures for the blood cDC1s and tissue cDC1s, as determined by bulk RNA-Seq

Page 66 - Figure 3.12: Schematic of OP9-DL1 co-culture process and its cDC1 output

Page 69 - Figure 4.1. Schematic of the development of haematopoietic lineages from bone marrow HSCs, illustrating the unknown phenotypic identities of the DC progenitors

Page 77 - Figure 4.2. Phenotyping of DC populations in human peripheral blood

Page 78 - Figure 4.3. Principal component analysis of mRNA expression from FACS-purified cDC1, cDC2 (DC2 and DC3), pDC and classical monocytes from peripheral blood, as determined by the NanoString assay

Page 79 - Figure 4.4. Gating strategy used to identify in vitro-derived DCs and monocytes generated from BM CD34+ progenitors, shown with blood and BM for comparison

Page 80 - Figure 4.5. Principal component analysis of mRNA expression from FACS-purified primary cDC1, cDC2 (DC2 and DC3), pDC and classical monocytes from human peripheral blood and in vitro culture

Page 81 - Figure 4.6. Schematic of the development of haematopoietic lineages from bone marrow, showing the DC and monocyte culture output of the phenotypic spaces as bar charts

Page 84 - Figure 4.7. Hierarchical clustering of single cell transcriptomes of CD34+ progenitors isolated from BM showing marker genes that identify 10 clusters of all progenitors

Page 85 - Figure 4.8. Visualisation of single cell transcriptomes of CD34+ progenitors

Page 87 - Figure 4.9. 3D diffusion map and lineage tracing of 262 single cell transcriptomes from CD34+ progenitor subsets

Page 89 - Figure 4.10. Hierarchical clustering of single cell transcriptomes of CD14+ monocytes, DCs and precursors isolated from BM, showing signature genes that identify 15 clusters

Page 91 - Figure 4.11. tSNE of 244 single cell transcriptomes sampled from mature DCs and classical monocytes and pre-DC populations of human BM

Page 92 - Figure 4.12. tSNE of 244 single cell transcriptomes sampled from pre-DC populations of human BM

Page 92 - Figure 4.13. 3D diffusion map and lineage tracing of 244 single cell transcriptomes pre-DC populations of human BM annotated by 15 clusters from hierarchical clustering

Page 94 - Figure 4.14. 3D diffusion map generated with mass cytometry data for 14,000 granulocyte-monocyte progenitors (GMP), precursor and mature DC/monocyte cells from human bone marrow

Page 96 - Figure 4.15. tSNE analysis of mass cytometry data for a total of 75,000 progenitors, precursors and mature DCs and monocytes from human PBMC and BMMC

Page 102 - Figure 4.16. Schematic summary of the main findings from Chapter 4

Page 106 - Figure 5.1. Schematic of the ChIP-Seq workflow

Page 108 - Figure 5.2. Flowchart of antibody characterisation assays, as outlined by ENCODE

Page 114 - Figure 5.3. Flow chart of the low cell ChIP-Seq process

Page 116 - Figure 5.4. Western blot of Raji and HeLa cell lines probed with the primary antibody IRF8 (D20D8) Rabbit mAb #5628

Page 118 - Figure 5.5. Western blot of primary CD34- Lineage- cells from human bone marrow, probed with IRF8 antibody ICSBP (E-9): sc-365042

Page 120 - Figure 5.6. The pIRES2-EGFP-HA-IRF8 expression construct

Page 121 - Figure 5.7. The pIRES2-EGFP expression construct

Page 122 - Figure 5.8. Gel electrophoresis of the IRF8-containing vector pIRES2-EGFP-HA-IRF8 and the empty vector pIRES2-EGFP

Page 123 - Figure 5.9. Overlay of light and fluorescence microscopy images for HEK293T cells transfected with the IRF8-containing vector and the empty vector

Page 125 - Figure 5.10. Western blot performed on lysates and IP reactions of HEK293T cells transfected with the IRF8-containing vector pIRES2-EGFP-HA-IRF8 and the empty vector pIRES2-EGFP

Page 127 - Figure 5.11. Bioanalyser electropherogram for ChIP libraries prepared from 100,000 cultured, sorted, and formaldehyde-fixed cDC1s subjected to different sonication settings and readings for an empty well and the ladder ran on the same Bioanalyser chip

Page 128 - Figure 5.12. Bioanalyser electropherogram for a ChIP library prepared from 50,000 cultured, sorted, and formaldehyde-fixed cDC1s subjected to 6 sonication cycles

Page 132 - Figure 6.1. Transcriptional regulation by IRF8 and its interacting proteins

Page 134 - Figure 6.2. Quantification of blood DC and monocyte populations of subjects carrying IRF8 mutations

Page 139 - Figure 6.3. Overview of the generated IRF8 ChIP-Seq dataset comprised of culture-derived cDC1s and pDCs with 3 biological replicates per cell type

Page 140 - Figure 6.4. Venn diagrams showing IRF8 binding site overlaps in biological replicates of cDC1 and pDC

Page 141 - Figure 6.5. Venn diagram displaying the overlap between consensus peak regions in cDC1s and pDCs

Page 141 - Figure 6.6. Visualisation of the genomic annotation for the IRF8 binding peaks in cDC1s and pDCs

Page 143 - Figure 6.7. Pathway enrichment analysis for genes regulated by IRF8 in cDC1s and pDCs

Page 149 - Figure 6.8. Overview of peaks called with MACS2 ($qval < 0.05$) in the region between -50kb and +100kb around the IRF8 TSS in individual cDC1 and pDC samples

Page 152 - Figure 6.9. Heatmap displaying the expression of cDC1 marker genes identified by Villani et al., 2017, across peripheral blood DC and monocyte subsets

Page 161 - Figure 7.1. Analysis pipeline for NanoString data

Page 163 - Figure 7.2. Analysis pipeline for bulk RNA-Seq data

Page 165 - Figure 7.3. Analysis pipeline for scRNA-Seq data

Page 167 - Figure 7.4. Analysis pipeline for CHIP-Seq data

Chapter 1. Introduction

1.1 The roles of DCs in immunity

Dendritic cells (DCs) are professional antigen-processing and presenting cells with critical roles in the regulation of immunity and inducing tolerance (Banchereau et al., 2001; Steinman et al., 2003). The role of DCs in immunity can be summarised with two terms: (1) DCs act as sentinels, able to capture, process and present antigens and to migrate to lymphoid tissues to select rare, antigen-reactive T cell clones, and (2) DCs are sensors, responding to a variety of environmental stimuli by extensive differentiation and maturation (Steinman, 2006). The type of DC and the specific response induced by different stimuli shape the immunological outcome, e.g. by driving naïve T cell polarisation to T helper subsets, such as T_H1 , T_H2 , T_H17 , and T_H22 , or inducing tolerance (reviewed by Geissmann et al., 2010). Together with monocytes and macrophages, DCs form the mononuclear phagocyte system, which encompasses leukocytes with specialised antigen processing and presentation function (Haniffa et al., 2015).

DCs reside in tissue in an immature state, continuously sampling their environment by endocytosis, macropinocytosis, and phagocytosis (British Society for Immunology, <https://www.immunology.org>). During pathogen invasion, DCs sense microbial products using a variety of pattern recognition receptors (PRRs) such as the toll-like receptors (TLRs), nucleotide-binding oligomerization domain-like receptors (NOD-like receptors), retinoic acid inducible gene 1-like receptors (RIG-I-like receptors) and C-type lectins (CLEC) (Kassianos et al., 2012). Upon interaction with pathogens, DCs undergo complex cellular processes resulting in their activation (reviewed by Tibúrcio et al., 2019). DCs then migrate from tissue to draining lymph nodes, where they can induce an adaptive immune response (Banchereau and Steinman, 1998). In order to present the antigen to lymphocytes and induce their clonal selection, the internalised pathogen-derived peptides are processed and loaded onto major histocompatibility complex (MHC) molecules (reviewed by Mantegazza et al., 2013).

1.1.1 Antigen presentation

Most of the proteins involved in antigen processing and presentation are encoded by MHC genes. Only antigens presented within MHC complexes are able to trigger an immune response, a property called MHC restriction (Zinkernagel and Doherty,

1997). Peptides loaded onto MHC I molecules are recognised by CD8⁺ T cells, while peptides loaded onto MHC II molecules are presented to CD4⁺ helper T cells. However, the pathways that regulate the presentation in the context of MHC I and MHC II are very diverse.

MHC I molecules are expressed ubiquitously on all human cell types, with the exception of erythrocytes (Lu et al., 2010). Their primary function is to display fragments from within the cell to CD8⁺ T cells as part of the cytosolic or endogenous pathway. Certain subsets of dendritic cells are also able to capture exogenous proteins and load peptides derived from them onto MHC Class I molecules (reviewed by Murphy and Weaver, 2017). The process of taking up exogenous antigens and their presentation on MHC Class I by APCs is known as cross-presentation (Bevan, 1976). Cross-presentation allows DCs to activate cytotoxic CD8⁺ T cells for immune defence against viruses that do not infect DCs and tumours that originate from non-DCs (reviewed by Kurts et al, 2010). The T cell activation in this context is referred to as cross-priming. Several pathways of cross presentation have been described. They include the endosome-to-cytosol pathway (illustrated in Figure 1.1 A) and the vacuolar pathway, involving the direct transport of antigens from the phagolysosome into a vesicular loading compartment, where peptides are allowed to be bound to mature MHC Class I molecules (reviewed by Murphy and Weaver, 2017).

MHC II proteins are expressed primarily by APCs, including DCs, macrophages, and B cells (Ting and Trowsdale, 2002). Antigens are captured via endocytic vesicles and cell-surface receptors. The source of the peptide antigen may also be pathogens that have invaded the cell to replicate in intracellular vesicles (reviewed by Murphy and Weaver, 2017). The low pH of the vesicles activates the proteases and causes degradation of the captured antigen (Silberstein et al., 2018). On their way to the cell surface, the MHC II molecules assembled in the endoplasmic reticulum pass through the vesicles and bind the peptides, transporting them to the cell surface (Figure 1.1 B). In DCs, antigen presentation by this pathway results in CD4⁺ T cell activation and polarisation. In addition, MHC II can be loaded with fragments derived from within the cells, generated via the autophagy pathway. This plays a role in the induction of tolerance to self-antigens, as well as a means of presentation of antigens derived from intracellular pathogens (reviewed by Murphy and Weaver, 2017).

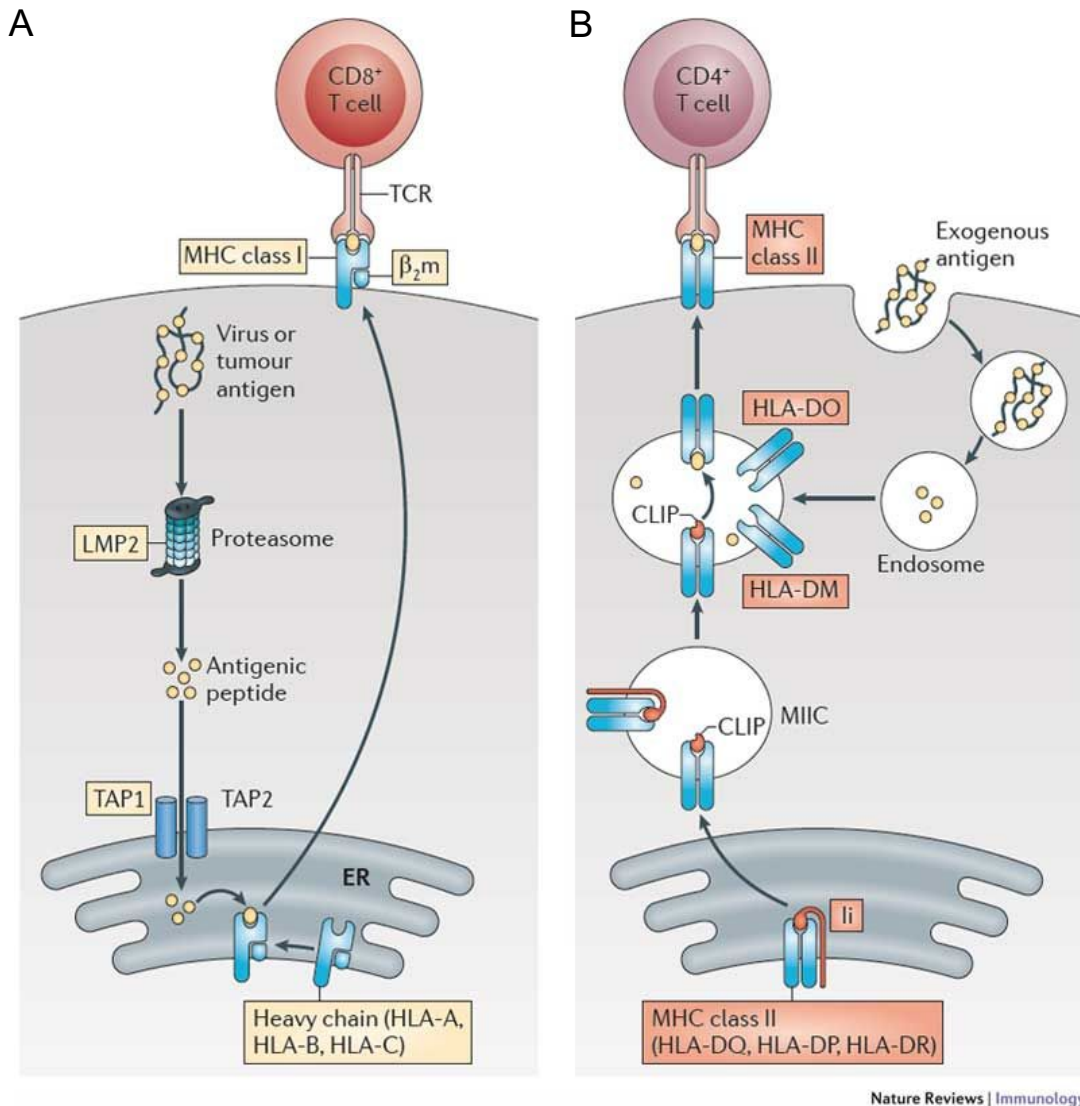


Figure 1.1. Schematic of antigen presentation via MHC I (A) and MHC II (B) (Kobayashi and van den Elsen, 2012).

A. Antigens are processed into peptides by the immunoproteasome, which is composed of multiple subunits, including LMP2. Peptides are transported into the endoplasmic reticulum (ER), where they are loaded into the groove of the MHC Class I complex, which is composed of a heavy chain and β₂-microglobulin (β₂m). MHC Class I complexes present antigens on the cell surface to CD8⁺ T cells.

B. Antigens from extracellular sources, such as bacteria, are processed by endolysosomal enzymes into peptides. These peptides bind to the groove of the MHC Class II complex by displacing the class II-associated invariant chain peptide (CLIP), which is derived from the MHC Class II-associated invariant chain (Ii). HLA-DO and HLA-DM regulate the antigen-loading process. The MHC Class II complex presents antigens to CD4⁺ T cells. MIIC, MHC Class II compartment; TAP, transporter associated with antigen processing; TCR, T cell receptor.

1.1.2 T cell polarisation

A “three-signal model” (Kapsenberg, 2003), has been proposed to describe T cell polarisation by DCs (Figure 1.2). According to the model, the first step is the recognition of the antigen displayed in the context of MHC Class II by the DC and recognised by its cognate T cell receptor (TCR). As part of the second signal, also referred to as the co-stimulatory signal, the activated DC upregulates costimulatory molecules which recognise costimulatory molecules on the T cell surface. The third and final signal is the polarising signal - through pattern recognition receptor signalling, pathogens induce the DC to produce cytokines which drive T cell polarisation (e.g. IL-12 for T_H1).

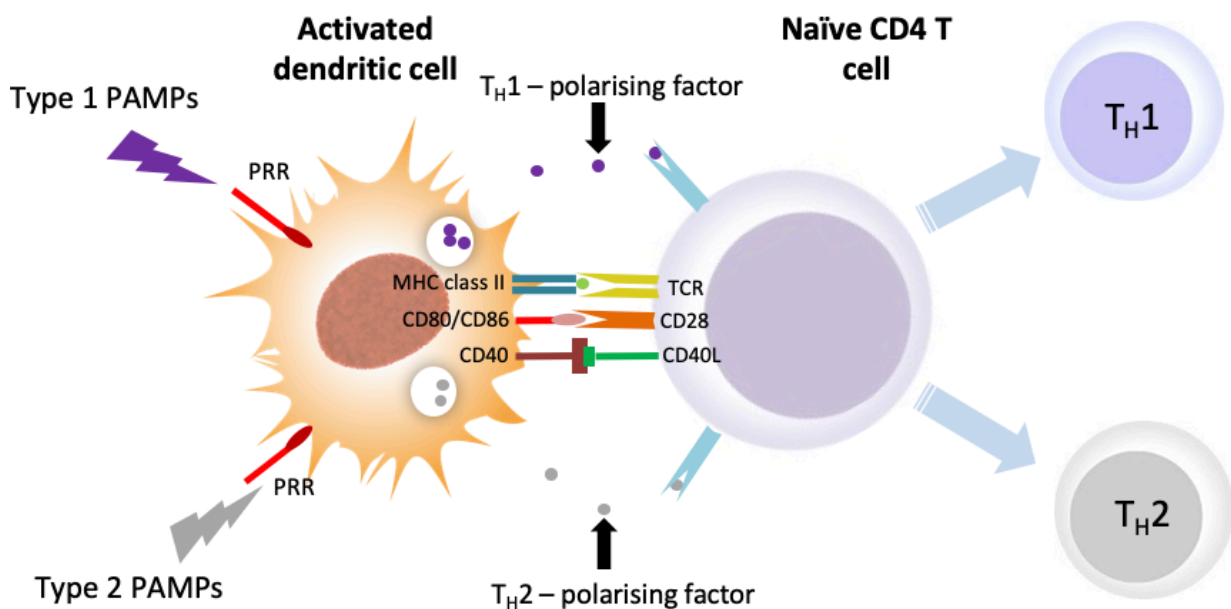


Figure 1.2. Schematic of the three-signal model for T cell polarisation.

First, the antigen in the context of MHC Class II displayed the DC and recognised by its cognate T cell receptor. Second, the activated DC upregulates costimulatory molecules (such as CD80, CD86) which recognise costimulatory molecules on the T cell surface (e.g. CD28). The third signal is the production of cytokines which drive T cell polarisation. Naïve T helper cells are able to differentiate into T_H1 cells, which promote cytotoxic $CD8^+$ T cells and cell-mediated immunity, and T_H2 cells, which promote B cells and humoral immunity (Alberts et al., 2002). PAMP, pathogen-associated molecular pattern; PRR, pattern recognition receptor, TCR, T cell receptor.

1.1.3 Induction of tolerance

An alternative outcome from a DC-T cell interaction is the induction of tolerance (Steinman et al., 2003). The mechanisms of DC-induced immune tolerance include: T cell anergy, clonal deletion, and induction of Tregs (reviewed by Hasegawa et al., 2018). This interplay occurs under steady-state conditions, when a DC has undergone homeostatic “tolerogenic” maturation, as opposed to TLR-induced “immunogenic” maturation. The transcriptomic changes occurring during both types of maturation are similar in complexity and largely overlapping, the expression of interferon-stimulated genes being among the few discriminators of immunogenic and tolerogenic murine DCs (Dalod et al., 2014; Ardouin et al., 2016). Several signaling networks, involving NF- κ B, β -catenin, and IRF4, are thought to regulate tolerogenic maturation in response to cues from the tissue microenvironment (Baratin et al., 2015; Manicassamy et al., 2010; Vander Lugt et al., 2017).

1.2 Peripheral blood dendritic cell subsets

The DC pool in peripheral organs is constantly replenished by bone marrow-derived cells, traveling through blood (Collin and Bigley, 2018). At least three human DC subsets, which differentiate under the control of specific transcription factors and perform specialised functions, have been identified in steady-state blood: 2 subsets of myeloid/classical DC (cDC1 and cDC2), and plasmacytoid DC (pDC) (Guilliams et al., 2014; Bigley et al., 2016). All human DC subsets express MHC Class II molecules (HLA-DR) and lack lineage markers characteristic for T cells, B cells, and natural killer cells (CD3, 19, 20, 56)⁻ (Collin et al., 2013). Classical dendritic cells (cDC) show a similar gene expression pattern in both blood and peripheral tissues, the blood subsets appearing less mature (Breton et al., 2015).

1.2.1 Classical dendritic cells

Classical dendritic cells were discovered in splenic tissue by Ralph Steinman and Zanvil Cohn in 1973. The extensively branched, motile, and mitochondria-rich cells were dubbed “dendritic cells” due to their distinct morphology (Katsnelson, 2006). At first, the cells seemed too rare to be relevant. However, it was later shown that DCs are extremely potent at stimulating T cell cytotoxicity and antibody responses (Nussenzweig et al., 1980; Inaba et al., 1983). All human cDCs express typical myeloid antigens CD11c, CD13, CD33 and CD11b. Both monocytes and cDCs

express CD11c, however DC can be distinguished by the lack of high CD14 expression and the absence of marker CD88 (Collin et al., 2013).

1.2.1.1 Classical dendritic cell type 1

cDC1s are present in blood in small numbers, constituting 0.1% of mononuclear cells and 10% of human cDCs (Collin et al., 2013). They express characteristic surface antigens including BDCA3/CD141/THBD, CLEC9A, XCR1, FLT3, NECL2/CADM1, and TLR3 (Table 1.1; Guilliams et al., 2014). cDC1s are specialised for the uptake and processing of material from late-apoptotic and necrotic cells due to the expression of the damaged cell-recognition molecule CLEC9A (Zhang et al., 2012). cDC1s are thought to be the most potent cross-presenting DCs *in vivo* (Embgenbroich and Burgdorf, 2018), specialised in taking up exogenous antigen and presenting it in the context of MHC Class I to CD8⁺ T cells (reviewed by Theisen and Murphy, 2017). An important feature of cDC1s is the sensing of viral nucleic acids with TLR3 (which recognises double-stranded RNA), as well as with TLR8 (which recognises foreign single-stranded RNA and short double-stranded RNA) (Gauzzi et al., 2010; Blasius and Beutler, 2010). The TLRs expressed by cDC1s also include TLR1, -2, -6 and -10 (Hemont et al., 2013).

1.2.1.2 Classical dendritic cell type 2

cDC2s comprise 1% of mononuclear cells in peripheral blood (Collin et al., 2013). The mature cells are known to express CD11c, BDCA1/CD1c, CD172/SIRPA, and ZBTB46 (Table 1.1; Guilliams et al., 2014). Additionally, cDC2s express a wide array of TLRs (TLR1-8 and -10) (Hemont et al., 2013). cDC2s determine the T helper cell response to an antigen by polarising naïve CD4⁺ T cells to drive T_H1, T_H2 or T_H17 responses (reviewed by Bigley et al., 2016). cDC2s have inferior cross presentation capacity compared to cDC1s (reviewed by Collin et al., 2013).

A number of studies have identified heterogeneity among blood and tissue cDC2, defined by CD1c⁺ expression. Yin et al. (2017) split cDC2s by phenotype – specifically CD5 expression, CD5^{hi} cells being more “DC-like”, and displaying higher levels of cDC2-specific genes, stronger migration, along with overrepresentation in lymph nodes, and CD5^{low} cDC2s showing greater expression of monocyte-related genes. Villani et al. (2017) analysed cDC2 from human peripheral blood using single

cell transcriptomics and revealed the presence of two subsets distinct from monocytes in this compartment. Mass cytometry analyses helped identify similar heterogeneity in blood, along with different cDC2 phenotypes in skin (Alcántara-Hernández et al., 2017). Additional cDC2 heterogeneity has recently been identified in human spleen via single cell transcriptomics (Brown et al., 2019).

1.2.2 Plasmacytoid dendritic cell

Plasmacytoid DCs were initially an enigmatic cell type observed in secondary lymphoid tissue (reviewed by Manz, 2018). Originally called “plasmacytoid T cells” or “plasmacytoid monocytes”, they were later re-classified as DCs (Grouard et al., 1997). These cells are specialised to sense viral infection and bacterial components via TLR7 or TLR9 and respond with a massive production of type I interferons (IFNs) α , β , or ω . Due to this specialised function, pDCs are also called “natural interferon-producing cells” (Colonna et al., 2004). Through IFN secretion, pDCs also support the function of other immune cells, including B cells and NK cells (Gowder, 2012). Recent studies identified heterogeneity in the phenotypic space formerly thought to be occupied exclusively by pDC. The classical pDC markers CD123, CD303, and CD304 were found on a pre-DC population, distinguished from pDC via the expression of CD33, CX3CR1, CD2, CD5, and CD327 (See et al., 2017). Pre-DCs (also termed AS DC in several studies) are phenotypically and functionally distinct from pDCs and display the ability to induce T cell proliferation and produce IL-12, while “pure” pDCs specialise in IFN α secretion (Villani et al., 2017). The contamination of the traditionally defined pDC gate with pre-DC/AS DC is therefore thought to be responsible for the T cell stimulation capabilities previously attributed to pDC (reviewed by Geissmann et al., 2010; Merad et al., 2013).

1.2.3 Tissue dendritic cells

“Migratory” cDC1s and cDC2s are present in most non-lymphoid tissue compartments (Haniffa et al., 2015). Their main function is to acquire antigen and migrate to lymph nodes in order to present it to lymphocytes (reviewed by Collin et al., 2013). During homeostasis, pDCs are present in most peripheral tissues in low numbers. However, the size of the pDC population increases in skin and mucosa during inflammation (Sisirak et al., 2011). Peripheral tissue contains other members

of the mononuclear phagocyte family, including monocyte-derived cells, tissue resident macrophages, and Langerhans cells (Haniffa et al., 2015).

In steady state, lymphoid tissue contains non-migratory “resident” cDCs and pDCs that are likely derived from blood precursors, along with “migratory” tissue-derived DC (Segura et al., 2012). The number of both the “resident” and “migratory” DCs in lymph nodes increases in inflammation (reviewed by Collin et al., 2013).

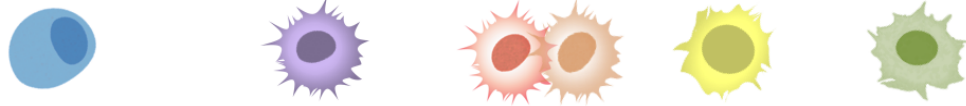
1.2.3.1 Langerhans cells

Langerhans cells (LCs) were the first DC subset to be identified. They were described in 1868 by Paul Langerhans, who thought they were cutaneous nerve cells due to their dendritic morphology (British Society for Immunology, <https://www.immunology.org>). LCs were subsequently re-classified as DC, when over a century later, classical dendritic cells were discovered, and their role in antigen presentation was demonstrated (Steinman and Cohn, 1973). LC are thought to originate prenatally from yolk sac-derived primitive myeloid progenitors and reside in epidermal tissue, where they self-renew (Hoeffel et al., 2012). Severe inflammation has been shown to recruit *de novo* bone marrow progenitors, in waves of transient classical monocytes and likely cDC2s (reviewed by Collin and Milne, 2016). LCs are marked by Langerin expression (CD207), epithelial cell adhesion molecule (EPCAM) and higher CD1a expression than cDC2s (Table 1.1; Collin et al., 2013). Due to their “strategic” location within the skin barrier, LCs play crucial functions as immune sentinels. Their skin-resident functions include the sampling of tight junctions between keratinocytes and uptake and recognition of apoptotic cells. As migratory cells, LCs are also able to travel to lymph nodes and present antigen to promote immunity or induce tolerance (reviewed by West and Bennett, 2017).

1.2.3.2 Inflammatory dendritic cells

Monocyte-derived DC (moDC) have been shown to arise from classical monocytes and expand in inflammation (Qu et al., 2014). However, the contribution of the newly recruited moDC to the initiation of immunity is an unresolved problem in humans (Collin et al, 2013). Recent murine studies showed that the priming by moDCs enhanced the memory CD8⁺ T cell differentiation during acute infection, more than priming by conventional DC, revealing an important role for moDC in immunity (Shin,

et al., 2019). More recently in humans, Dutertre and colleagues (2019) defined CD14⁺ cDC2s as an inflammatory population distinct from monocytes, dependent on FLT3L (unlike monocytes), and expressing high IRF4 (unlike moDC).



	pDC	cDC1	cDC2	moDC	LC
Surface markers	CD123/IL3RA CD303/BDCA2/CLEC4C CD304/BDCA4/NRP1 CD45RA	CD141/BDCA3/THBD CLEC9A XCR1 NECL2/CADM1	CD1C/BDCA1 CD11B CD11C CD172/SIRPA	CD1A CD1C/BDCA1 (CD14-)	CD207/CLEC4K CD1A CD1C/BDCA1 E-cadherin/CDH1 EPCAM
Major TF	IRF8 IRF4 IRF7 TCF4/E2.2 RUNX2	IRF8 BATF3 ID2	IRF4 ID2 ZEB2 KLF4 NOTCH2	KLF4 MAFB	ID2 RUNX3
TLR	TLR7 TLR9	TLR1,3 TLR6,8 TLR10	TLR2,4 TLR5,6 TLR8,9	TLR4 TLR9	TLR4 TLR9
Cytokines	IFN- α TNF IL-6 IDO	IFN- λ TNF- α IL-12 CXCL9 CXCL10	IL-8 IL-1 IL-12 IL-23 TNF- α IL-10	IL-12	IL-15
Location	Blood Lymphoid tissue	Blood Lymphoid tissue Non-lymph tissue	Blood Lymphoid tissue Non-lymph tissue	Tissue in inflammation	Epidermis Mucosa

Table 1.1. Characteristics of human dendritic cell subsets.

Dendritic cells can be divided into multiple subpopulations, based on surface markers, major transcription factors (TF) involved in their development and homeostasis, Toll-like receptors (TLR), secreted cytokines, and location (Bigley, et al., 2016; Collin & Bigley, 2018; Schlitzer et al., 2018).

1.3 Monocyte subsets

Like DCs, monocytes develop in the bone marrow, then circulate through the bloodstream, where they account for approximately 5% of leukocytes, before being recruited into tissues (Blumenreich, 1990; Patel et al., 2017). Monocytes are highly plastic and heterogeneous, able to differentiate into macrophages or moDC under inflammatory conditions (Chomarat et al., 2000). Human monocytes are divided in 3 subsets: classical (CD14⁺⁺CD16⁻), intermediate (CD14⁺⁺CD16⁺), and non-classical (CD14^{dim}CD16⁺⁺) (Ziegler-Heitbrock et al., 2010). These subsets have differential abilities for cytokine production, migration, and promoting T cell proliferation (Tolouei Semnani et al., 2014). CD14⁺ monocytes specialise in phagocytosis, innate sensing and migration, CD14⁺CD16⁺ monocytes excel at antigen presentation, cytokine secretion, apoptosis regulation, and differentiation, while CD14^{dim}CD16⁺ monocytes are involved in phagocytosis and adhesion (reviewed by Kapellos et al., 2019).

1.4 SLAN DC

The CD16⁺ monocyte subset, marked by the expression of 6-sulfo LacNAc (SLAN), has been the subject of debate, as it was also classified by some authors as a type of dendritic cell (Hofer et al., 2019; reviewed by Collin and Bigley et al., 2018).

However, transcriptional profiling of monocytes and myeloid DCs from human peripheral blood proposed their role in inflammatory processes and confirmed the relationship of SLAN⁺ cells with the monocytic compartment rather than with DCs (van Leeuwen-Kerkhoff et al., 2017). Furthermore, recent studies used a human *in vivo* cell tracing experiment to show that the non-classical monocytes originate from classical CD14⁺ monocytes (Patel et al., 2017). Contamination of the HLA-DR⁺Lin⁻ flow cytometry gate with non-classical CD14^{dim} CD16⁺⁺ monocytes, has also led to the discovery of a new subset of dendritic cells named DC4 (Villani et al., 2017). Further studies identified that these cells differ from all DC subsets and classical monocytes and behave like CD16⁺⁺ SLAN⁺ monocytes (Calzetti et al., 2017, Dutertre et al 2019).

1.4 *In vitro*-derived DCs

The first *in vitro*-generated DCs were produced from monocytes (Gieseler et al, 1998). As monocytes are found in great numbers in peripheral blood (100,000 to 500,000 per mL), this approach can generate large numbers of monocyte-derived

DCs. However, natural DCs are heterogeneous, and comprise multiple subsets with different functions, while the exact role of moDCs *in vivo* is not well understood (Collin et al, 2013). The DC generation methods shifted to their manufacture from bone marrow progenitors, in the presence of cytokines, such as GM-CSF and FLT3L (Naik et al., 2005; Helft et al., 2015). Further studies revealed that co-culture of bone marrow progenitors with a feeder layer, such as the MS5 or OP9 stromal cell lines, allowed the production of all major DC subsets (Lee et al. 2015; Kirkling and Cylak, et al., 2018). Chapter 3 of this thesis elaborates further on the *in vitro* generation of DCs, aiming to determine the transcriptional identity of DCs generated in culture.

1.5 DC therapy

Currently, the field of immunotherapy is dominated by immune checkpoint inhibitors that target molecules with inhibiting functions (such as CTLA-4, PD-L1, and PD-1) in order to allow T cells to kill cancer cells (van Willigen et al., 2018). As the most potent antigen presenting cells, able to shape adaptive immunity, dendritic cells also offer a promising prospect in cancer immunotherapy. Numerous clinical trials for DC vaccines have been conducted within the last decades, and the first anticancer vaccine (sipuleucel-T, used to treat prostate cancer) was approved by the United States Food and Drug Administration in 2010 (Hammerstrom et al., 2011). A common approach in most DC vaccine protocols is the harvest of DCs from the peripheral blood of a patient. As DCs are extremely rare, blood monocytes are often harvested instead, and used to generate DCs *in vitro* via the addition of GM-CSF and IL-4. The *bona fide* DCs or the monocyte-derived DCs are then matured and loaded with tumour antigens or synthetic peptides in the presence of an adjuvant, and administered to the patient, in order to induce the tumour-specific effector T cells (Sadeghzadeh et al., 2020). A number of advantages of using natural circulating DCs have been noted, including the preservation of their function due to shorter culture time compared to monocyte-derived DCs (van Willigen et al., 2018) and the ability to migrate. However, monocytes are still used in most protocols due to their availability, and the use of natural DCs (cDC1, cDC2, and pDC) has been underexplored.

In recent years, the ability of immature DCs to induce tolerance has been investigated to treat autoimmune conditions. Phase I trials have been conducted at Newcastle University, with the aim to establish the safety and efficiency of tolerogenic dendritic cells (toIDC) in treating rheumatoid and inflammatory arthritis

affecting the knee joint (Bell et al., 2017). This method used monocyte-derived tolDC, produced from the peripheral blood of the patients using a novel protocol containing immunosuppressive and immunomodulatory drugs, and suggested that injection of these DCs into the affected joint offered a safe and promising treatment (Harry et al., 2010; Bell et al., 2017).

1.6 Murine DCs and monocytes

Dendritic cells are a vital component of the mammalian immune system. pDCs and the two cDC subsets have been identified in numerous species, including humans, mice, macaques, and pigs (Guilliams, et al., 2016; Edwards et al., 2017). Efforts have been made to unify the classification of DCs across species and identify common markers of the DC subsets. The identification of murine homologues of human mononuclear phagocytes is of particular importance, as most of the knowledge about dendritic cells was gained from studies in mice.

cDC1s, or CD141⁺ DCs in human and CD8/CD103⁺ DCs in mouse, are marked by the expression of CLEC9A, XCR1, CADM1, TLR3, BAFT3 and IRF8 (Haniffa et al., 2015; Crozat et al., 2010; reviewed by Edwards et al., 2017). In both species, cDC1s are potent cross-presenters (Gutiérrez-Martínez et al., 2015; Embgenbroich and Burgdorf, 2018). Similarly, cDC2s are conserved across the two species, and are marked by CD1c in human and CD11b in mouse (Haniffa et al., 2015). These cells express CD11c, FLT3, CD11b, CX3CR1, and SIRPA/CD172, and present exogenous antigens to CD4⁺ T cells in both species (Vu Manh et al, 2015). In humans and mice, pDCs are major interferon producers, and can be distinguished by expression of CD4, CD123, and CD45RA (Vu Manh et al, 2015; Haniffa et al., 2015). pDC markers in mouse also include Ly-6C, Siglec-H, and BST2 (Haniffa et al., 2015).

Two major monocyte subsets have been described in both species: classical (Ly6C^{hi} CD43⁻ monocytes in mice and CD14⁺⁺ CD16⁻ monocytes in humans) and non-classical (Ly6C^{lo} CD43⁺ monocytes in mice and CD14^{dim} CD16⁺⁺ monocytes in humans) (reviewed by Wolf et al., 2019). In addition, the equivalents of moDC, LC, and macrophages have also been identified in mouse (Vu Manh et al, 2015; Haniffa et al., 2015).

1.7 DC haematopoiesis

Haematopoiesis is the process by which haematopoietic stem cells give rise to immune cellular components. In the developing human, the primitive wave of haematopoiesis begins in the yolk sack, while the first multipotent adult-type haematopoietic stem cells (HSCs) emerge in the aorta/gonad/mesonephros region inside the embryo (Tavian et al., 2010). The HSCs then migrate to the liver, spleen and then into the bone marrow, where they are located in adults (Jagannathan-Bogdan et al., 2013). All natural DCs originate from bone marrow progenitors, independently from monocytes, through progenitors and precursors with restricted DC potential (Geissman et al., 2010; Pühr et al., 2015). In the classical haematopoietic model, involving sequential bifurcation, DCs have an apparent “dual” origin (Figure 1.3 A), arising from both myeloid and lymphoid progenitors (Doulatov et al., 2010). Human DCs descending from common myeloid progenitors (CMPs) and their counterparts arising from common lymphoid progenitors (CLPs), have been observed to be exhibiting the same phenotype, function and gene expression in a mouse xenotransplantation model (Ishikawa, 2007). The traditional model is based on the analysis of cell populations, also referred to as gates, predefined by flow cytometry. The apparent dual origin may therefore be explained by the inclusion of DC-primed progenitors in both myeloid-defined and lymphoid-defined analysis gates. Presently, the traditional hierarchical model of haematopoiesis has largely been replaced by the early lineage priming model for haematopoietic lineages (Notta et al., 2016; Velten et al., 2017). According to the early lineage priming model, cell fate specification occurs in the early progenitors and development progresses in a continuous manner along increasingly stable unilineage-restricted trajectories (Figure 1.3 B).

Recent studies have identified precursors with DC potential in human. Breton et al. (2015) discovered a peripheral blood population of cells lacking mature DC markers but expressing CD117 (also known as KIT or stem cell growth factor receptor), CD116 (CSF2RA) and CD135 (FLT3). Via a single cell transcriptomic study of peripheral blood cells, Villani et al. (2017) identified a DC population, expressing AXL and SIGLEC6, later reclassified as pre-DCs (See et al., 2017). Villani and colleagues also described a small population of CD34⁺CD100⁺CD116⁻ negative cells which could give rise to cDC1s and cDC2s in culture. It has not been established how these cells relate to other progenitors or precursors, including the AXL⁺ SIGLEC6⁺ cells.

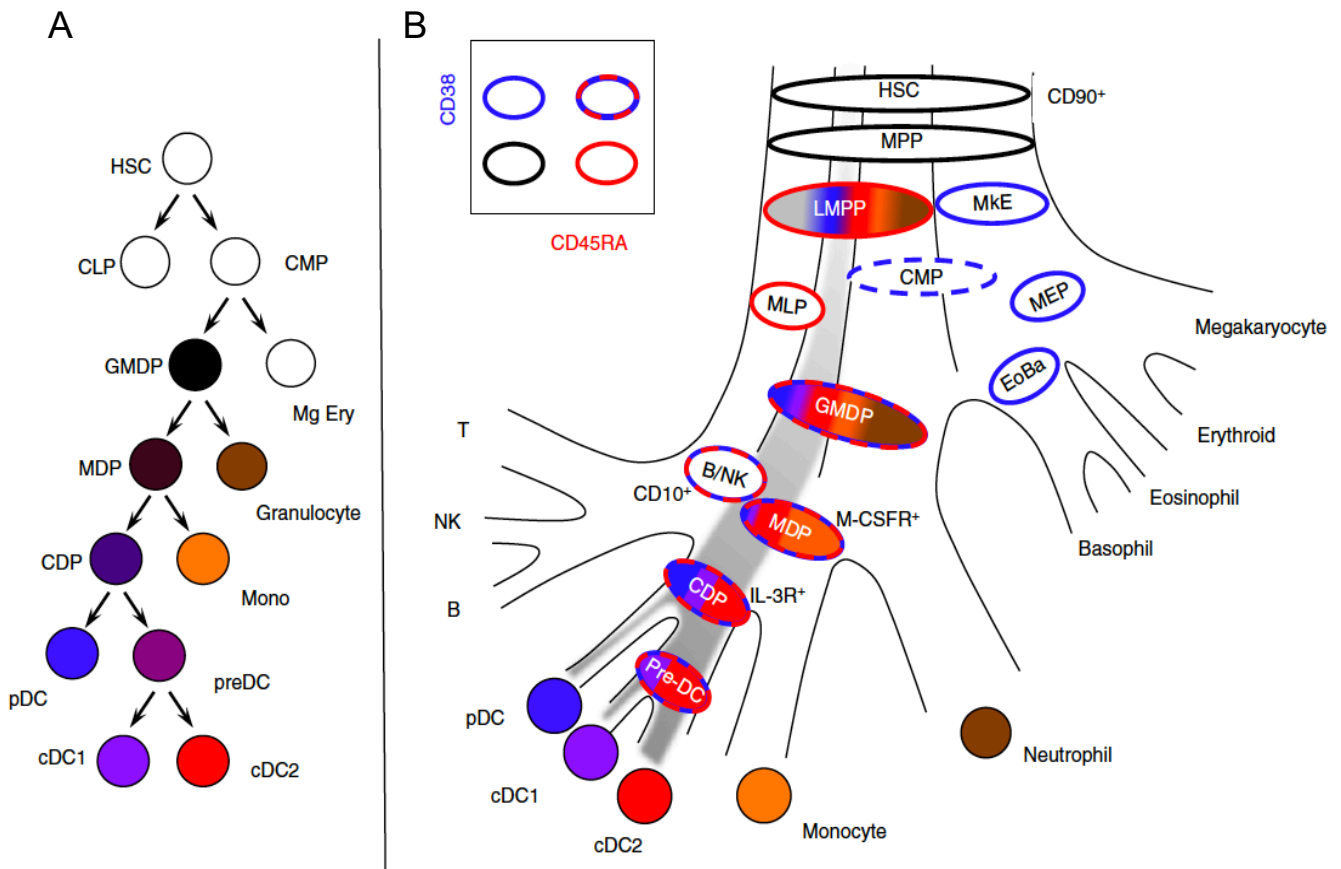


Figure 1.3. Classical and revised models of human haematopoiesis (Collin and Bigley, 2018).

A. In classical models of haematopoiesis, cell potential partitions by successive bifurcations descending from the apex where common lymphoid and common myeloid progenitors (CLP; CMP) arise from the haematopoietic stem cell (HSC). Each progenitor population has homogeneous differentiation potential such that every cell has an equal probability of two mutually exclusive fates. Hence, dendritic cells (DC) were proposed to arise in the sequence: CMPs, granulocyte–macrophage DC progenitor (GMDP), macrophage DC progenitor (MDP), common DC progenitor (CDP) with a final pre-DC stage leading to conventional cDC1 and cDC2. Each population is given a uniform colour to indicate homogeneous potential.

B. Recent studies support several adjustments to the classical model. First, lineage is primed in early progenitors so that most populations contain only cells with a single potential. Second, lymphoid and myeloid potential run together originating as the lymphoid primed multi-potent progenitor (LMPP) that separates from megakaryocyte and erythroid potential (MkE) at the apex. Hence the gates defined by CD38 (blue borders) and CD45RA (red borders) contain phenotypically related cells but with restricted potentials, indicated by bands of colour each corresponding to a discrete lineage.

1.7.1 Transcription regulation in the context of DC haematopoiesis

In humans, gene expression programmes that establish specific cell states and maintain cellular homeostasis are controlled by transcription factors, cofactors, and chromatin regulators (Lee and Young, 2013). RNA polymerase II, which synthesizes precursors of mRNAs in eukaryotes, collaborates closely with proteins termed transcription factors in order to initiate transcription (Hampsey, 1998).

Transcription factors can be divided into two categories: (1) basal, or general, transcription factors, which are expressed ubiquitously and participate in the formation of the transcription preinitiation complex, and (2) gene-specific transcription factors (sometimes referred to as simply transcription factors) that activate or repress basal transcription (Villard, 2004). All gene-specific transcription factors contain at least one DNA-binding domain (DBD), while some of the general TFs do not bind DNA directly. DBDs bind specific DNA sequences, also called motifs. Gene-specific TFs often have a preference for certain motifs over other sequences, driven by features such as high or low guanine-cytosine content and DNA shape (Dror et al., 2016). Over 1,600 TFs are known in humans, classified into “families” based on their DBDs (Lambert et al., 2018).

Gene-specific transcription factors can act as activators or repressors of gene expression. Gene transcription starts with the assembly of the transcription preinitiation complex near the transcription start site (TSS) (Figure 1.4). Enhancer regions recruit gene-specific TFs in order to enhance transcription of a regulated gene, which is typically located downstream. Regions of DNA with high levels of transcription factor binding, comprised of multiple enhancers, are entitled super-enhancers. Super-enhancers significantly increase the expression of the genes they are regulating; however, these genes are very sensitive to transcription perturbation (Villicaña et al., 2014). The TFs bind to the enhancers or super-enhancers and act as activators, causing the DNA to bend and come closer to the gene promoter. TFs then bind coactivators, which bind to RNA polymerase II, which in turn binds to general TFs at the TSS, forming the preinitiation complex (Lee and Young, 2013). Conformational changes in the complex lead to the positioning of single-stranded DNA to the active site of RNA polymerase II (Gupta et al., 2016).

The repression of gene expression can be achieved via a variety of mechanisms. One example is the physical blocking of the binding sites of the basal TFs or RNA polymerase II by the gene-specific TFs (Donev, 2017). Often, a set of activators and repressors are involved in the regulation of the same gene, a mechanism termed combinatorial regulation (Reece, et al., 2011).

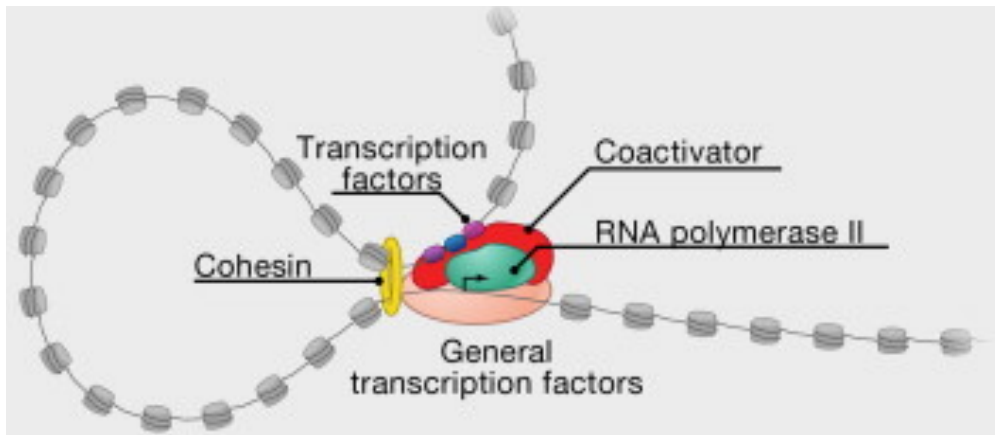


Figure 1.4. Formation of the preinitiation complex as part of transcriptional regulation (adapted from Lee and Young, 2013).

Transcription factors bind to specific DNA elements (enhancers) and to coactivators, which bind to RNA polymerase II, which in turn binds to general transcription factors at the transcription start site (TSS, shown as a bent arrow). The DNA loop formed between the enhancer and the start site is stabilised by cofactors, such as the Mediator complex and cohesin.

As described in section 1.7, DCs develop through increasing commitment, where lineage determination is mediated by transcription factors (Figure 1.5). The discovery of patients with primary immunodeficiency caused by mutations within the genes encoding for transcription factors GATA2, IRF8, and IKZF1, offered valuable insights into transcription factor requirements for DC development in human (Dickinson et al., 2014; Hambleton et al., 2011; Cytlak et al., 2018). Transcription factors essential for the development of DC subsets include IRF8, BATF3, and ID2 for cDC1s, IRF4 for cDC2s, and IRF8, E2-2/TCF4, and IKZF1 for pDCs (reviewed by Colonna et al., 2004 and Colin and Bigley et al., 2018; Cytlak, et al., 2018).

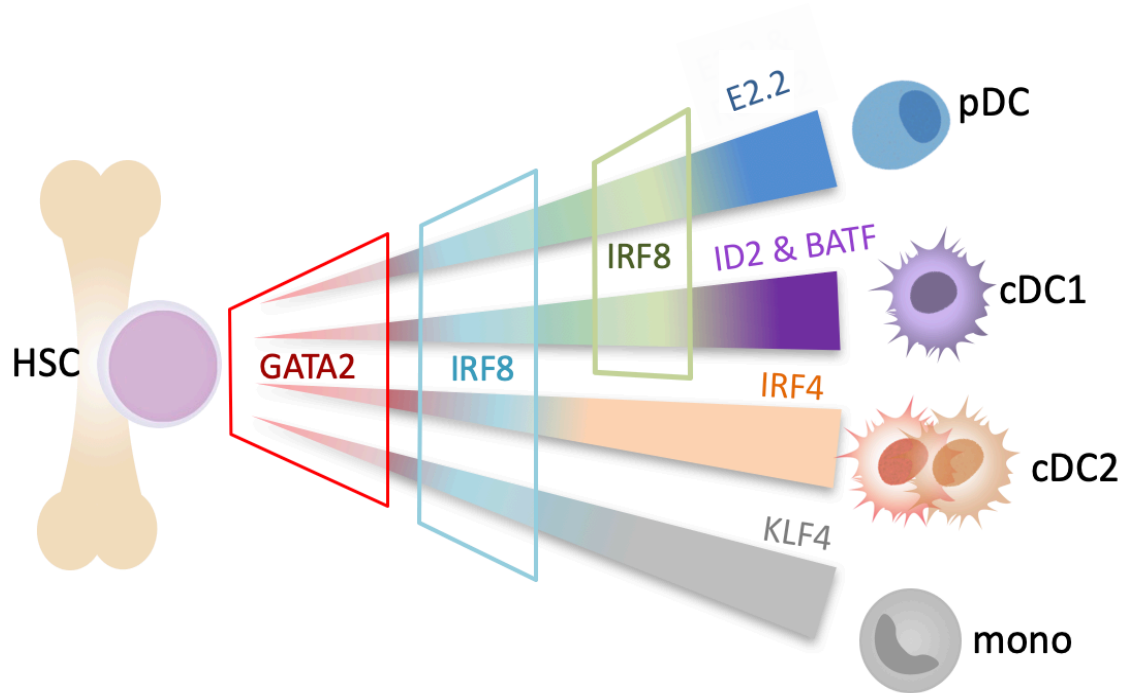


Figure 1.5. Model of transcription factors regulating DC haematopoiesis.

Human dendritic cell subsets develop in the bone marrow from HSC (haematopoietic stem cells), under the control of a set of transcription factors. Transcription factor GATA2 is required for multiple haematopoietic lineages, and mutations in this gene are the cause of dendritic cell, monocyte, B, and natural killer lymphoid deficiency (Dickinson et al., 2014). IRF8 is required at early stages of DC development, and is expressed at later stages in pDCs and cDC1s. Lineage-specific factors include E2.2/TCF4 for pDC, BATF3/BATF and ID2 for cDC1s, IRF4 for cDC2s, and KLF4 for monocytes (reviewed by Collin and Bigley, 2018 and Haniffa et al, 2015; Feinberg et al., 2007).

1.7.2 Transcription factor IRF8

IRF8 is part of the interferon regulatory factor (IRF) family which consists of nine members in mammals (Tamura et al., 2015). The IRF family regulate expression of type I Interferons and interferon-stimulated genes (ISGs) by binding to the IFN-stimulated response element (ISRE). Several members of the IRF family also play critical roles in the cellular differentiation of hematopoietic cells and in inducing innate pattern recognition receptors and antigen-specific immune responses (Tamura et al., 2008). Members of the IRF family contain a conserved N-terminal region with a DNA-binding domain that binds to the core IRF binding motif, GAAA (Figure 1.6; Fujii et al., 1999). The C-terminus region of IRFs is less well conserved and is thought to

mediate the protein–protein interaction of a specific IRF with other family members and TFs (Yanai et al., 2012). The regulatory C-terminus region is also referred to as the IRF association domain (IAD).

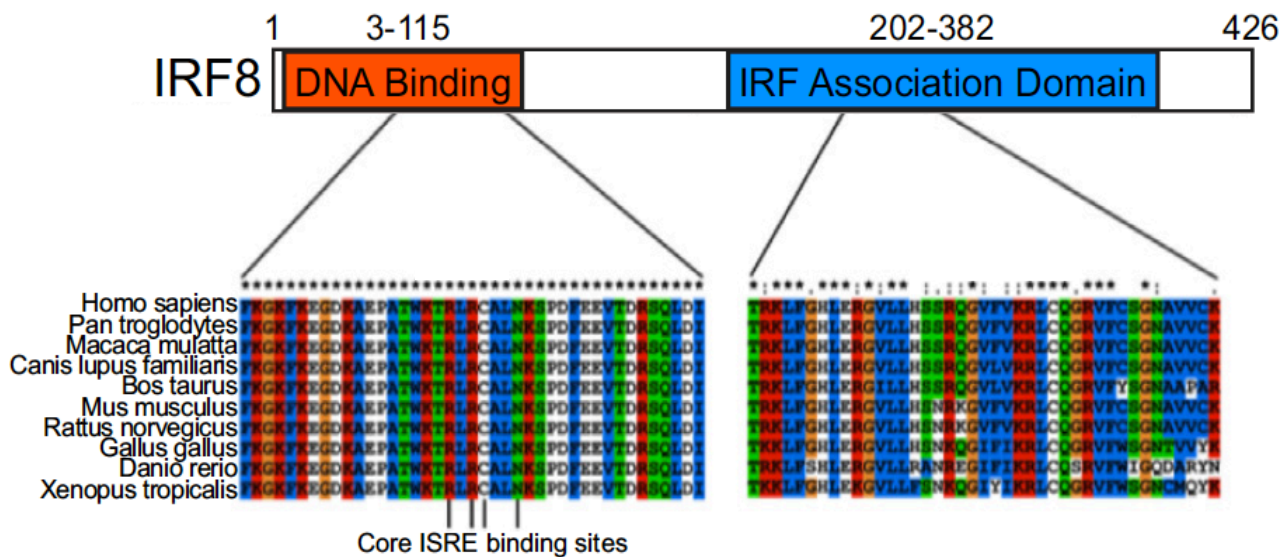


Figure 1.6. Schematic of the IRF8 protein and multiple sequence alignment of IRF8 orthologues, illustrating conserved coding sequences across multiple species (adapted from Bigley et al., 2018).

IRF8 contains a conserved N-terminal region with a DNA-binding domain (DBD, depicted in orange). Core ISRE binding residues are marked. The *IRF8* C-terminus region (also referred to as the IRF association domain, IAD, shown in blue) is less well conserved and is thought to mediate the protein–protein interaction with other transcription factors.

Most of the knowledge about IRF8 has been derived from murine studies. The role played by this transcription factor in human has only recently been highlighted by the discovery of patients with bi-allelic *IRF8* mutations, experiencing a loss of all the monocyte and DC subsets (Bigley et al., 2018; Hambleton et al., 2011; Salem et al., 2014), along with reduced numbers of mature NK cells (Mace et al., 2017). A more restricted loss of cDC1 and pDC subsets was observed in mono-allelic *IRF8* mutation (Cylak and Resteu et al., 2020). Studies in mouse showed that within the progenitor compartment, IRF8 inhibits CEBPA to limit granulocyte production in favour of monocyte/DC differentiation (Kurotaki, et al, 2014). It is then required to maintain the identity of terminally differentiated cDC1s and control lineage survival (Grajales-Reyes et al., 2015). Recent lineage-specific conditional knockout models suggest

that IRF8 is primarily required for pDC function rather than development (Sichien et al., 2016).

1.7.3 Notch signaling pathway and dendritic cell development

The Notch signaling pathway is highly conserved across multicellular organisms, and it is involved in both developmental and homeostatic processes. The mammalian Notch signaling pathway consists of membrane-bound ligands of the Delta-like (DL) and Jagged families and four receptors: NOTCH 1-4. The signaling is initiated via the binding of Notch ligands on an adjacent cell (Figure 1.7), leading to the release of intracellular segment of Notch, which migrates to the nucleus, where it interacts with the transcription factor CSL (also called RBPJ), converting it from a repressor to an activator (Kramer, 2015). Coactivators of the Mastermind family are also recruited to activate Notch-dependent gene expression programmes (McElhinny et al., 2008).

During haematopoiesis, Notch is one of the main signalling pathways engaged in the direct interaction of progenitors and bone marrow stromal cells, and has long been established to play a crucial role in the development of T and B lymphocytes (reviewed by Cheng et al., 2010). In human, mutations in Notch signaling pathway genes have been reported to cause developmental phenotypes affecting an array of organs and are linked to conditions such as T cell acute lymphoblastic leukaemia (Penton et al., 2012; Elisen et al., 1991). The role of Notch signaling in DC development was first highlighted by studies in mice, which identified that Notch2 controls differentiation of splenic cDC1s and cDC2s (Lewis et al., 2011). Recent *in vitro* experiments performed by the Human Dendritic Cell Lab and overseas collaborators, established the role DL1-Notch2 signaling in the generation of authentic cDC1s in mouse and revealed that DL1 signaling facilitates the generation of human cDC1s (Kirkling and Cytlak et al., 2018). Chapter 3 of this thesis further elaborates on a novel culture system, able to generate large number of cDC1s with the aid of Notch signaling.

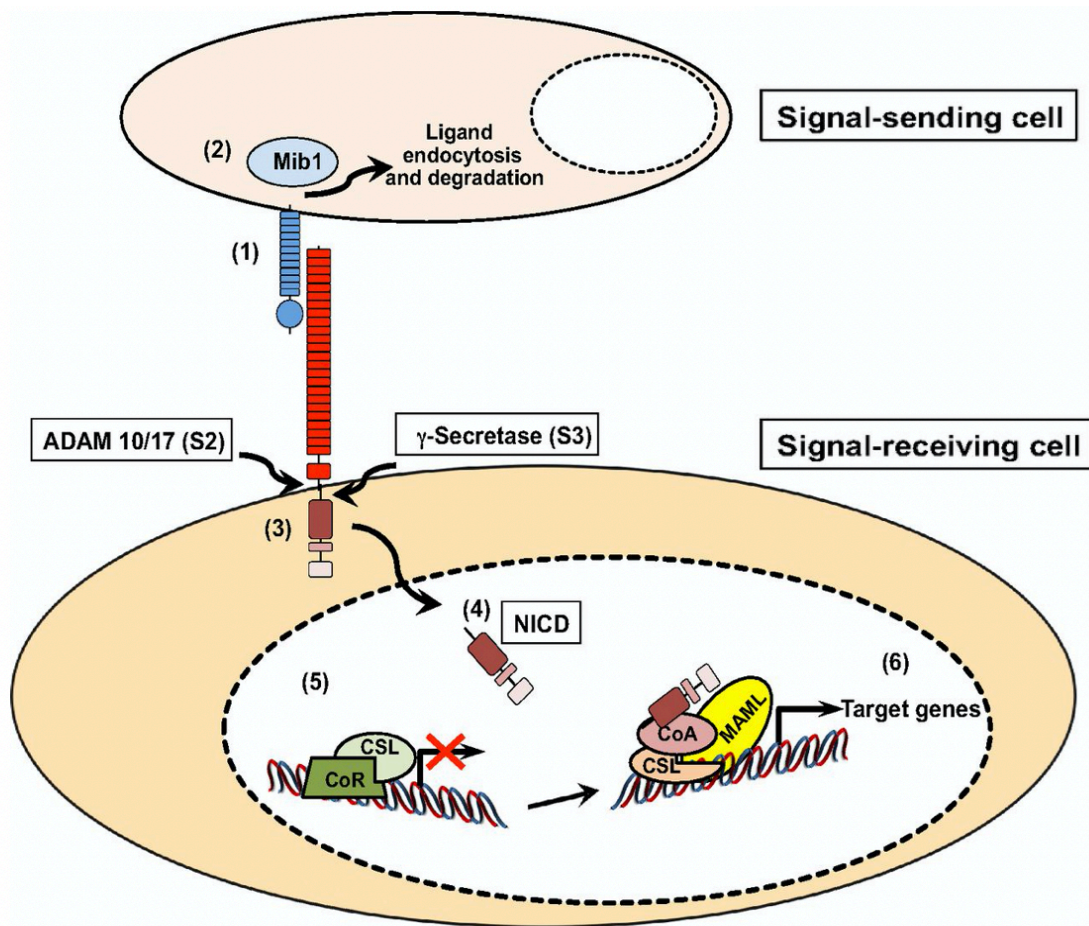


Figure 1.7. Canonical Notch signaling (Koch et al., 2013).

A Notch ligand expressed on the surface of a signal-sending cell interacts with a Notch receptor on the signal-receiving cell, triggering Notch receptor activation (1). The E3 ubiquitin ligase mindbomb 1 (Mib1) promotes ligand endocytosis and is required for efficient Notch receptor activation (2). The trans-interaction between ligand and receptor induces two consecutive proteolytic cleavages of the heterodimeric Notch receptor (3). The first cleavage is mediated by the metalloprotease ADAM10/17 (S2 cleavage), followed by a second cleavage through the γ -secretase complex (S3 cleavage). These cleavages lead to the generation of a free intracellular domain (NICD), which translocates to the nucleus of the signal-receiving cell (4). In the absence of NICD (5), a transcriptional repressor complex composed of CSL and additional co-repressors (CoR), such as N-CoR, keeps Notch target genes silent. The interaction of NICD with CSL (6) dissociates the repressor complex and leads to the recruitment of MAML and additional co-activators (CoA, e.g. p300) to the complex. The assembly of this transcriptional activation complex on the promoter regions of Notch target genes leads to an upregulation of gene expression.

1.7.4 Growth factor requirements for DC development

External factors driving the development of dendritic cells include a set of lineage-specific cytokines or growth factors. These small secreted proteins interact with their cognate receptors on the cell surface, activating signaling pathways that result in alterations in gene expression and a biological response.

1.7.4.1 FLT3L

Fms-related tyrosine kinase 3 ligand (FLT3L) is a potent dendritic cell stimulator, with the ability to expand DC populations *in vivo* and *in vitro* (Dong et al., 2002). Administration of FLT3L *in vivo* dramatically increases the number of DCs in both mice and humans (Maraskovsky et al., 1996; Maraskovsky et al., 2000), while incorporation of FLT3L in *in vitro* culture of bone marrow progenitors allows for the generation of diverse DC subsets (Brasel et al., 2000; Naik et al., 2005). In murine cell culture, Flt3L alone is sufficient to generate DCs from BM, while in human, additional factors are required.

1.7.4.2 GM-CSF

Granulocyte-macrophage colony-stimulating factor (GM-CSF or CSF2) is a cytokine commonly used to generate DCs in both murine and human cell culture systems. It was the first growth factor recognised to promote the *in vitro* development of DC, along with granulocytes and macrophages, in cultures of murine bone marrow progenitors (Inaba et al., 1992). The combined GM-CSF and Flt3L requirement for DC homeostasis *in vitro* was later demonstrated by a double-negative murine model, where mature DCs and pre-DC populations were reduced more significantly in combined deficiency than in Flt3L or GM-CSF deficiency alone (Kingston et al., 2009).

1.7.4.3 SCF

Stem cell factor (SCF) is the ligand of the c-KIT cytokine receptor. SCF is able to regulate DC production in bone marrow progenitor-based *in vitro* systems supplemented with other cytokines, via the recruitment of early progenitors of a high proliferative potential with the capacity to differentiate into erythroid and myeloid cells, as well as into DCs (Saraya et al., 1996).

1.7.4.4 M-CSF

Macrophage colony stimulating factor (M-CSF or CSF-1) is a cytokine known to drive the development of the monocyte/macrophage lineage. Evidence from murine studies suggests that it can also drive the development of cDCs and pDCs, likely because the expression of its receptor is retained through the pre-DC stage (Fancke et al., 2008; Sichen et al., 2017).

1.8 Technologies employed in DC research

Technological advances of the last decades in the fields of cytometry, genomics, and transcriptomics enabled great progress to be made in dendritic cell research.

1.8.1 Cytometry

1.8.1.1 Flow cytometry

Flow cytometry is a single cell technique widely used in immunophenotyping, due to its ability to perform accurate enumeration and identification of immune cells. This versatile technique has a vast number of applications and is fundamental for dendritic cell research, as it offers a time- and cost-effective means to enumerate *ex vivo*- and *in vitro*-derived DCs and discern between the dendritic cell subsets in these samples. This technique is indispensable for the interrogation of the output of *in vitro* culture systems, and for phenotyping and identifying missing immune cell populations in human DC immunodeficiency. Furthermore, fluorescence activated cell sorting (FACS) enables the sorting of desired cells populations or individual cells for further studies, facilitating the closer inspection of rare populations.

Flow cytometry is a laser-based technique. It operates by passing cells suspended in a liquid stream in front of a laser in order to measure the physical and chemical characteristics of each individual analysed cell. Several detectors, placed around the stream, measure a combination of scattered and fluorescent light, and simultaneously produce readings for up to 20 parameters, also known as channels. Generally, the forward scatter (FSC) channel values reflect the cell size, while the side scatter (SSC) readings provide information on its inner complexity, such as granularity. In addition, fluorescence detectors sense fluorescently labelled molecules present on the cell surface or within the cells. Fluorescently-conjugated antibodies and fluorescence dyes are commonly used to aid the detection of the molecules of interest, and are added to the samples prior to analysis in the form of an

antibody “cocktail”. The signals produced by the detectors are converted to channel values and analysed with specialised software. The cell populations are identified with the help of markers, many of which are named according to the CD nomenclature. The traditional approach to flow cytometry data analysis involves the inspection of series of 2D scatter plots, where cell populations are separated by “gates”. Figure 1.8 illustrates an example of a gating strategy employed in the flow cytometric identification of peripheral blood human DCs and monocytes. Recently, novel computational methods have been developed, aiming to provide more rapid and less biased analysis techniques, and improve data visualisation. Examples of analysis algorithms widely applied to flow cytometry data include t-distributed stochastic neighbour embedding (tSNE) and minimum spanning trees. These computational techniques become more important as the number of channels or parameters increases with advancing flow cytometer capabilities.

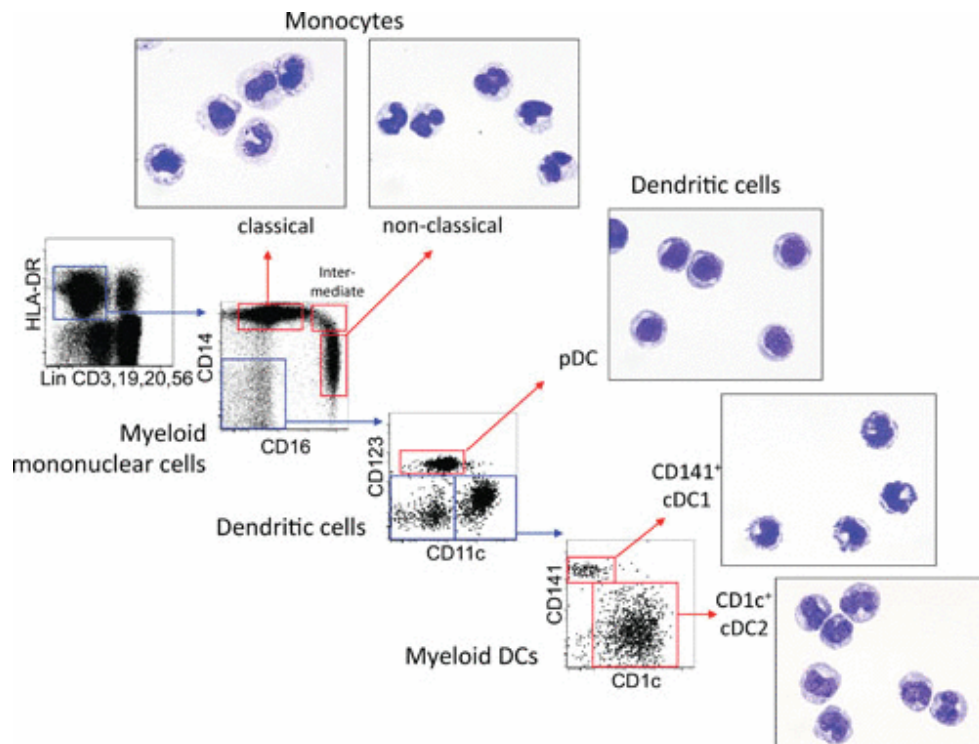


Figure 1.8. Gating strategy for flow cytometric identification of human DCs and monocytes in peripheral blood (Collin and Bigley, 2016).

Monocytes and DCs are found in the HLA-DR positive, and lineage (CD3, 19, 20, and 56) negative compartment. CD14 versus CD16 displays monocyte subsets. The double-negative population contains the DC subsets. In this example, CD123 and CD11c are used to define pDCs and myeloid/classical DCs. cDCs can be separated into cDC1 and cDC2 using CD141 and CD1c, respectively. Lineage – a set of mature cell lineage markers not expressed by DC, including T, B, and NK cell markers.

1.8.1.2 Mass cytometry

Mass cytometry, also known as cytometry by time of flight (CyTOF) is a recently developed next generation flow cytometry platform. Its main advantage is the ability to assess the expression of up to 40 antigens, with further potential for multiplexing. Metal isotopes, rather than fluorophores, are used for labelling the antibodies, allowing to reduce spectral overlaps compared to multi-colour flow cytometry, with little “spillover” between channels. Cells are stained with the metal-labelled antibodies, then loaded into a CyTOF mass cytometer, where they are nebulised, and the readout is produced by time of flight mass spectrometry. The analysis of the mass cytometry data involves a similar approach to that of flow cytometry, and is often performed via the gating of cell populations, or more recently, using computational approaches.

Drawbacks of this method include the inability to retrieve the analysed cells for further analysis, as the cells are nebulised within the instrument in order to be analysed by mass spectrometry. In addition, the throughput of this technique (~ 1,000 cells/s) is typically lower than that of flow cytometry, which is able to analyse tens of thousands of cells per second (Li et al., 2018).

1.8.2 Transcriptomics

1.8.2.1 NanoString gene expression assay

The nCounter Analysis System performs a highly multiplexed direct digital counting of transcripts with no requirement for amplification and is ideal for low amounts of input material, such as RNA derived from DCs, present in blood in low numbers. The accurate detection of transcripts is achieved by the usage of a pair of short probes for the identification of each gene of interest. The reporter probes are approximately 50 base pairs long RNA sequences complementary to target genes of interest. Each reported probe is labelled with a fluorescent barcode, unique for every target. The expression of up to 800 individual genes can be measured in one reaction. The capture probes have a structure similar to the reporter probes, the fluorescent barcode being replaced with a biotin molecule. As a result of a hybridisation step, the capture and reporter probes, along with the input RNA from the sample form a probe-target complex (Figure 1.9). The samples are transferred in the nCounter Prep Station (Figure 1.10), and following the removal of excess probes, this complex bind

to the imaging surface of a NanoString cartridge via the capture probe's biotin molecule. The probes are then aligned on the cartridge, and the fluorescent barcodes are scanned, counted, and assigned to their respective genes by the Digital Analyser (Figure 1.10), producing a Reporter Code Count (RCC) dataset. A pre-built 'Human Immunology V2' panel with 594 genes associated with innate and adaptive immune responses is among the most applicable in an immunology context.

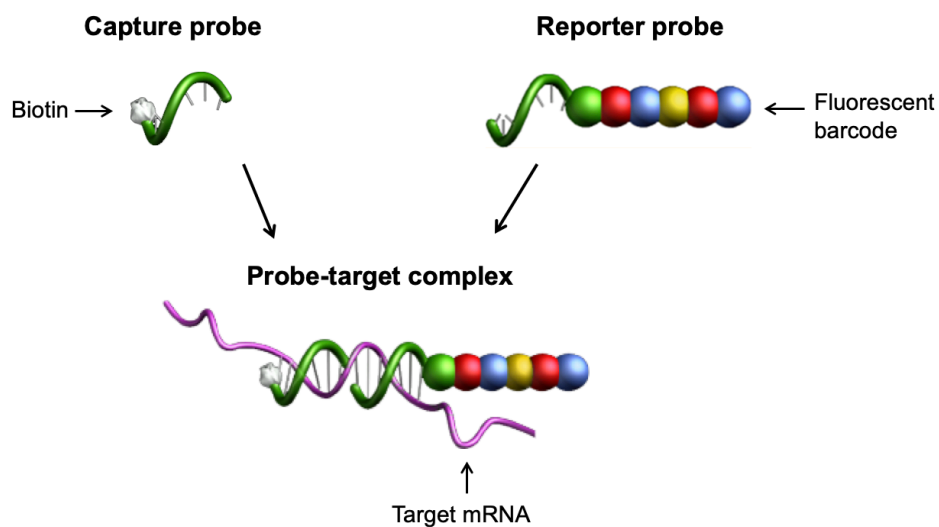


Figure 1.9. Schematic of the NanoString gene expression assay hybridisation chemistry (adapted from NanoString Technologies).

As part of the NanoString assay, the capture and reporter probes are added to purified RNA or cell lysates, then hybridised in a thermocycler, forming the probe-target complex.

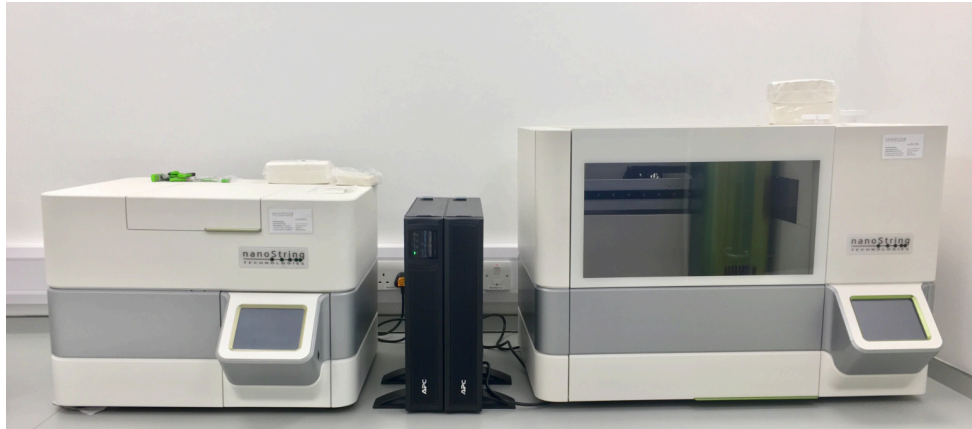


Figure 1.10. The NanoString nCounter FLEX platform.

The NanoString nCounter FLEX system consists of the prep station (right) and the digital analyser (left). The battery pack in the centre ensures the sample processing will not be affected by power outages. The hybridised samples are first placed in the nCounter Prep Station, and following the removal of excess probes, the probe-target complexes bind to the imaging surface of a NanoString cartridge via the capture probe's biotin molecule. The probes are then aligned on the cartridge via an electric current. The cartridge is transferred into the Digital Analyser, where the fluorescent barcodes are scanned, counted, and assigned to their respective genes.

1.8.2.2 Bulk RNA-Seq

As the cost of sequencing has been steadily decreasing in the last two decades, the field of transcriptomics shifted rapidly from probe-based microarrays to bulk RNA-Seq. A standard RNA-Seq workflow begins with the purification or enrichment of the cells populations of interest, often performed via FACS or via magnetic-activated cell sorting. The RNA is then extracted in bulk from each cell type or biological replicate. This is often followed by rRNA depletion or Poly A mRNA enrichment, RNA fragmentation, cDNA synthesis (also called reverse transcription), and finally sequencing library preparation, which involves adapter ligation and polymerase chain reaction (PCR). The libraries are then sequenced on a next generation sequencing platform, such as Roche 454, Illumina, Helicos, or PacBio (Chu et al., 2012).

RNA-Seq is indispensable for the profiling of the whole transcriptome of bulk populations of mature DCs and progenitors. However, its main disadvantage is that

the expression values represent an average of the thousands of cells present in each sample. This renders the method unsuitable for the analysis of DC development, as understanding the mechanisms and patterns of cell decision-making and early lineage bias at molecular level can only be achieved by analysing individual cells (Nimmo et al, 2015).

1.8.2.3 Single cell RNA-Seq

Single cell RNA sequencing (scRNA-Seq) is a powerful technique developed in the last decade, used for characterising genome-wide mRNA expression of individual cells. In the context of dendritic cell research, this technique is vital for studying cell fate decisions of individual DC progenitors or precursors, as bulk RNA sequencing obscures the cellular heterogeneity by averaging the gene expression in the samples. The scRNA-Seq technique has proven useful for unravelling heterogeneity across multiple tissues (Human Cell Atlas, <https://www.humancellatlas.org>; Collin et al., 2019). Using methods such as pseudotime to determine the position of a given cell in a developmental continuum can help explore developmental pathways. In addition, the scRNA-Seq method excels at identifying rare cell populations.

Generally, the workflow for the generation of scRNA-Seq data is similar to that of bulk RNA-Seq. The main difference is the tissue dissociation and the isolation of single cells, rather than of a bulk population, at the start of the scRNA-Seq protocols. Multiple approaches have been developed for this purpose. The first multiplexed scRNA-Seq protocols, able to analyse the expression of multiple cells in parallel, emerged in 2011 and were plate-based (Wu et al., 2018). As part of the plated-based approach, the single cells are isolated into 96 well or 384 well plates (Figure 1.11). A widely-used plate-based method is the Smart-seq2 protocol, which enables full-length coverage across transcripts and is therefore sensitive for gene detection (Picelli et al., 2014). Moreover, using FACS to isolate the cell into plates offers an additional advantage, as the gene expression for each individual cell can be linked with its antigen expression determined by flow cytometry. Droplet based scRNA-seq systems, which rely on capturing each cell in its own microfluidic droplet, are also widely used for the isolation of single cells. Examples include the Drop-seq, inDrop and Chromium 10X technologies. These methods often offer a high cell throughput, however, are also generally more prone to more noise and gene dropouts than plate-based approaches (Wang et al., 2019).

Drawbacks of scRNA-Seq are the technical challenges, the complex and noisy data output due to false-negative errors, and the prohibitive cost of the procedure. The technical challenges include the necessity for standardised sample preparation, reliable isolation of single cells, the limited efficiency of RNA capture and subsequent conversion to cDNA and sequencing (Grun and van Oudenaarden, 2015). In order to assess the sample-to-sample variability of scRNA-Seq data, a set of synthetic control genes are often added to each cell's lysate in a theoretically constant and known amount (Vallejos et al., 2017). However, the commonly used External RNA Control Consortium (ERCC) molecules may have different molecular properties to the molecules of interest (Yuan et al., 2017).

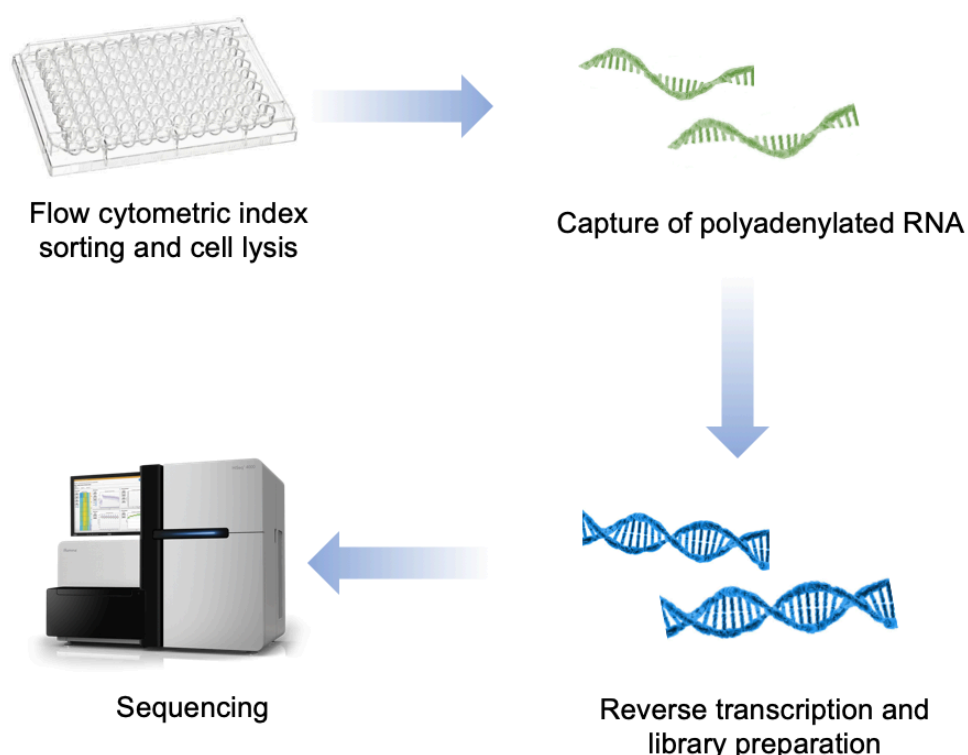


Figure 1.11. Schematic of the plate-based approach to single cell RNA sequencing.

Cells are first isolated in to 96 or 384 well plates, often via fluorescence activated cell sorting. The cells are then lysed, and the mRNA molecules are selected via the capture of the polyadenylated mRNA tails by complementary poly[T] primers. Next, depending on the protocol used, the cells are the barcoded and combined into a single tube for library preparation, or separate sequencing libraries are prepared for each individual cell. The libraries are then sequenced using a next generation sequencing platform.

1.8.3 Genomics

1.8.3.1 ChIP-Seq

Chromatin immunoprecipitation, followed by high-throughput DNA sequencing (ChIP-Seq) is used to examine protein-DNA interactions at molecular level. Studying the DNA binding sites and binding targets of transcription factors is of particular interest in the context of DC research, in order to unveil the roles of transcription factors critical for DC haematopoiesis. Standard ChIP-Seq protocols require abundant starting material, in the region of 1-20 million cells (Gilfillan et al, 2012). However, novel ChIP-Seq protocols, designed for a low cell number, have been able to reduce the number of input cells to 50,000-100,000 cells per immunoprecipitation reaction. As the name suggest, chromatin immunoprecipitation is performed with an antibody against the transcription factor or histone of interest. Chapter 5 of this thesis is dedicated to the optimisation of the low cell ChIP-Seq protocol for transcription factor IRF8 and describes the ChIP-Seq method in great detail.

1.9 Aims and objectives

This project aimed to investigate multiple aspects of human DCs and their development. First, in order to address the rarity of DC *in vivo* and make them accessible for research, an *in vitro* culture system was developed by members of the Human Dendritic Cell lab. While the system generated DCs with a seemingly appropriate phenotype, as determined by flow cytometry, this work was aimed at confirming the similarities between the culture-derived DCs and their *ex vivo* counterparts at transcriptomic level. An additional focus was the verification of the transcriptomic identities of cDC1s (the rarest DC subset), produced in large numbers in culture via Notch stimulation, against blood and tissue DCs.

The second main objective was to identify distinct DC lineages and their precursors in human bone marrow, and to establish whether there is a developmental basis for the recently described heterogeneity in the cDC2 population. Following the verification of the culture output as part of the first aim, the novel culture method was used to interrogate multiple phenotypic spaces of the bone marrow in order to establish increasingly committed progenitor and precursor populations. This project then sought to examine the transcriptomes of individual DC progenitors or precursors from human bone marrow using single cell RNA-Seq. Analyses of these data were

aimed at inferring the developmental trajectories of the single cells and defining the distinct pathways of DC development. A further objective was to determine the link between the developmental pathways originating in the bone marrow and the mature DC subsets found in peripheral blood.

Following the discovery of two DC developmental pathways marked by the expression of transcription factor IRF8, as part of the previous objective, the final goal of this project was to establish the role of IRF8 in the homeostasis of human cDC1 and pDC. These DC subsets develop through the IRF8 high pathway, retain the expression of this transcription factor as they mature, and are the most affected by the loss of IRF8 in monoallelic *IRF8* mutation in human (Cytlak and Resteu et al., 2020). Of particular interest were the functional modules controlled by IRF8 in cDC1 and pDC and the presence of IRF8 auto-activation, a phenomenon previously described in murine studies (Grajales-Reyes et al., 2015). This objective was set to be achieved via ChIP-Seq, a gold-standard technique for interrogating transcription factor binding sites. Previously, this study was not possible in human due to the rarity of DC *in vivo*, however this was overcome via the use of a novel culture system, designed to produce large numbers of *bona fide* DC subsets *in vitro*.

As the ChIP-Seq assay relies heavily on the quality of the used antibody, it was imperative to identify and test ChIP-compatible IRF8 antibodies available on the market in order to select an antibody with appropriate sensitivity and specificity prior to performing IRF8 ChIP-Seq on rare human cells. Further aspects of this assay that needed to be tested and optimised include the sonication settings and input cell number.

Chapter 2. General materials and methods

2.1 Buffers and reagents

2.1.1 Lymphoprep Solution

Lymphoprep (STEMCELL Technologies) is a density gradient solution used for the isolation of human peripheral blood, cord blood, and bone marrow mononuclear cells. During centrifugation, granulocytes and erythrocytes sediment through the Lymphoprep medium due to their higher density, while the mononuclear cells form a layer on top of the solution (Figure 2.1). Lymphoprep contains sodium diatrizoate (9.1% w/v) and polysaccharide (5.7% w/v), along with other (unlisted) ingredients, and has a density of 1.077 g/mL.

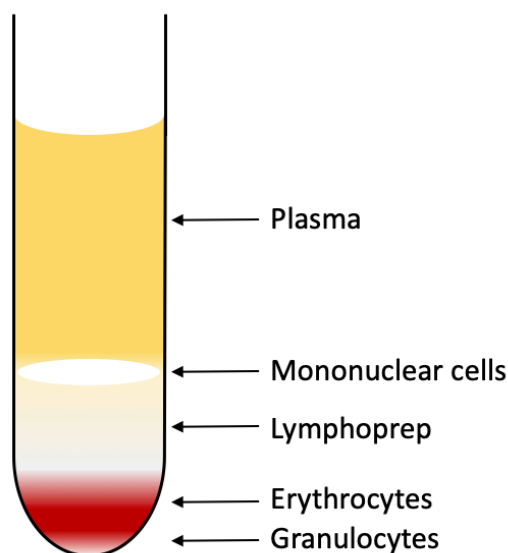


Figure 2.1. Schematic of samples separated using Lymphoprep.

Lymphoprep employs density gradient to separate granulocytes and erythrocytes, which form a pellet at the bottom of the tube, from mononuclear cells, found in a distinct layer above the Lymphoprep solution.

2.1.2 Dulbecco's Phosphate-Buffered Saline

Dulbecco's Phosphate-Buffered Saline (DPBS) is a buffered salt solution used for washing and dilution of mammalian cells. The pH of the solution is maintained within the range of 7.1 - 7.5 by phosphate buffering. DPBS is water-based solution, and each litre of it contains 8g sodium chloride, 0.2g potassium phosphate, monobasic, 1.15g sodium phosphate, dibasic, and 0.2g potassium chloride. The Sigma-Aldrich

formulation, used in this work, does not contain calcium and magnesium ions, which can cause cell clumping.

2.1.3 Foetal calf serum

Foetal calf serum (FCS) is a media supplement, containing essential nutrients and growth factors for cell culture. Heat inactivation at 56°C for 30 minutes is a common treatment of FCS, performed to disable the complement system and any potential inhibitors of cell growth in culture.

2.1.4 Flow and sort buffer

Flow and sort buffer, used to dilute the cells in flow cytometry assays and FACS, was made up of DPBS, 0.1%–2% Heat Inactivated Foetal Calf Serum (HI-FCS, Gibco) and 0.4% EDTA (Sigma-Aldrich). The role of the FCS is to reduce the non-specific antibody binding, while the EDTA acts as chelating agent to reduce cell clumping.

2.1.5 Culture media

All culture media were supplemented with HI-FCS and penicillin/streptomycin (Sigma). The role of FCS is described above in subsection 2.1.3. Penicillin-streptomycin is a mix of antibiotics, used to control bacterial contamination in cell culture. All media were supplemented with L-glutamine (Sigma). This amino acid supports the growth of cells that have high energy demands and serves as an alternative energy source for rapidly dividing cells.

2.1.6 Minimum Essential Medium α

Minimum Essential Medium α (MEM α) is used for the suspension and adherent culture of mammalian cells. It is a modification of MEM, containing non-essential amino acids, sodium pyruvate, lipoic acid, vitamin B12, biotin, and ascorbic acid. In addition, the used formulation contained L-glutamine. MEM α was supplemented with 1% penicillin/streptomycin, 1% L-glutamine, and 10% HI-FCS.

2.1.7 Dulbecco's Modified Eagle Medium

Dulbecco's Modified Eagle Medium (DMEM) is a medium broadly suitable for the culture of adherent cells. The Gibco formulation, used for this work, contained high glucose, pyruvate, and no glutamine. DMEM contains no proteins, lipids, or growth

factors. For cell culture, the medium was supplemented with 10% HI-FCS 1% penicillin-streptomycin, 1% L-glutamine, and 1% Fungizone (Gibco).

2.1.8 RF-10

RF-10 was prepared from Roswell Park Memorial Institute (RPMI) 1640 medium (Sigma), with the addition of 10% FCS, 1% penicillin-streptomycin, and 1% L-glutamine. RPMI 1640 medium is suitable for a variety of mammalian cells and contains biotin, vitamin B12, para-aminobenzoic acid, inositol and choline.

2.1.9 RNA Lysis Buffers

RNA Lysis Buffers RLT and RLT plus (QIAGEN) are used for cell lysis prior to RNA isolation. RLT buffer contains a high concentration of guanidine isothiocyanate, which supports the binding of RNA to the silica membrane in the spin columns used for RNA extraction. RLT plus has the same composition, supplemented with a proprietary blend of detergents. 1% β -mercaptoethanol (Sigma) was added to both buffers before use in order to effectively inactivate RNAses in the cell lysate.

2.1.10 Freezing solution

Freezing solution was prepared from 90% HI-FCS (Gibco) and 10% dimethyl sulfoxide (NBS Biologicals), which acts as a cryoprotectant.

2.1.11 NanoString nCounter reagents

The NanoString nCounter kits for gene expression assays are composed of CodeSets and a master kit. The CodeSets are shipped in two separate vials with reporter and capture probes. The nCounter master kits contain a sodium chloride-based hybridisation buffer, prep plates, cartridges, and all plastics necessary for the assay.

2.2 Cell lines

2.2.1 OP9 and OP9-DL1 cell lines

OP9 is a cell line derived from *Csf1*^{-/-} murine bone marrow stromal cells. The cells are adherent and have a fibroblastic appearance. Due to their hematopoietic supportive capacity, OP9 can be used as a feeder layer to co-culture stem cells and induce differentiation into haematocytes. Unlike bone marrow stromal cells, they are

unable to produce M-CSF, as they are derived from the *Csf1* knockout “osteopetrotic” mouse (hence OP). OP9 cell lines transduced with retroviruses encoding green fluorescent protein (GFP) or Notch ligand DL1 (referred to as OP9-DL1 cells) were obtained from Juan Carlos Zúñiga-Pflücker and the Sunnybrook Research Institute.

2.2.1 HEK 293T cell line

HEK 239T is a cell line derived from human embryonic kidney cells. The cells are adherent and have an epithelial morphology. The cell line is an altered version of the HEK 293 cell line, modified to express the SV40 T-antigen, in order to transiently maintain a high copy number of transfected plasmids that carry the SV40 origin of replication. HEK 239T (and its parent HEK293) are highly transfectable, making them suitable for gene expression, protein production, and retroviral production. HEK 239T cells were obtained from American Type Culture Collection.

2.3 Sample collection and processing

This work was performed in accordance with the Declaration of Helsinki. Written informed consent was obtained from participants or their parents. The studies relevant to this thesis were approved by local review board NRES Committee North East-Newcastle and North Tyneside: 08/H0906/72 and REC 14/NE/1136; REC 14/NE/1212, 17/NE/0361.

2.3.1 Isolation of peripheral blood and bone marrow aspirate mononuclear cells

Prior to performing flow cytometry and FACS, as well as before cell storage, mononuclear cells were isolated using density centrifugation. Blood samples, collected in Vacutainer EDTA tubes (BD), were diluted at a 1:1 ratio with DPBS (Sigma-Aldrich). Bone marrow samples were collected from excess donor material from clinical bone marrow transplants or excavated from femoral heads removed in joint replacement surgery (see next section). Cells from bone marrow donations were obtained by flushing the collection bags and tubes with DPBS.

The blood and bone marrow cells were layered on a volume of Lymphoprep equivalent to approximately a quarter of that of the diluted sample in 50mL Falcon tubes. The tubes were spun down at room temperature for 15 minutes at 800g. The cells situated at interphase of the Lymphoprep and plasma layers were aspirated with

a Pasteur pipette and washed twice with cold DPBS for 5 minutes at 800g. An extra wash (7 minutes at 200g) was introduced for the blood samples in order to remove platelets. Cells were then washed with DPBS (5 minutes at 800g), and pelleted for 5 minutes at 800g.

2.3.2 Isolation of cells from hip femoral bone marrow

Femoral heads were obtained from hip replacement surgery donations in sterile DMEM or RPMI 1640 medium. Bone marrow was extracted with bone clippers, then placed on a 100µm cell strainer (Corning), and washed with DPBS. BMMCs were extracted from the DPBS solution containing cells via the Lymphoprep density centrifugation.

2.3.3 Cryopreservation of cells

Cells were frozen in freezing solution, prepared as described in subsection 2.1.10. PBMC and BMMC were stored at a concentration of 6.6 million cells/mL in 1.5 mL freezing solution. OP9 and OP9-DL1 were frozen at a concentration of 0.5 million cells/mL in 1 mL freezing solution. HEK 239T were frozen at a concentration of 3.3 million cells/mL in 1.5mL freezing solution. Cells in freezing solution were stored in Nunc cryovials (Sigma-Aldrich), placed at -80°C, then transferred to -140°C for long-term storage.

For defrosting, vials were removed from -140°C and immediately thawed at 37°C in a water bath. The liquid containing the cells was then transferred in a dropwise manner into a 15mL Falcon tube with pre-warmed medium (RF-10 for PBMC and BMMC, MEM α for OP9 and OP9-DL1, and DMEM for HEK 293T). The tube was then spun down at room temperature for 5 minutes at 800g. The supernatant was discarded, and the cell pellets were resuspended in the appropriate sort/flow buffer or culture media.

2.3.4 Cell counting

A volume of 10µl was pipetted out the cell suspension and mixed with an equivalent volume of trypan blue stain for dead cells (Invitrogen). The cells were then counted using a haemocytometer (Hawksley).

2.3.5 Co-culture of bone marrow progenitors with OP9 and OP9-DL1 cell lines

The OP9 and OP9-DL1 cell lines were defrosted as described in section 2.3.3 and seeded into 24 well plates (Corning) at a density of 25,000 cells/well in 1 mL growth medium or in 96 well plates at a density of 5,000/well in 200 μ l medium. The feeder layer OP9 and OP9-DL1 cells were left to settle for 4-24 hours prior to the addition of FACS-purified BM CD34⁺ progenitors at a density of 15,000 cells/ well for the 24 well plates and up to 3,000 cells/well for the 96 well plates. The cells were co-cultured in MEM α medium supplemented with 10% HI-FCS and 1% penicillin/streptomycin at 37°C in a humidified atmosphere at 5% CO₂. The medium was also supplemented with 20 ng/ml granulocyte-macrophage colony-stimulating factor (GM-CSF, R&D systems), 20ng/ml (used for pDC generation for ChIP-Seq) or 100 ng/ml (all other experiments) FLT3-ligand (Immunotools), and 20 ng/ml stem cell factor (SCF, Immunotools). Half of the volume of MEM α with the same cytokine concentration was replaced weekly. Microscopy images of the OP9-DL1 culture are shown in Appendix A. Cells were harvested on ice at day 14-21 of culture, passed through a 50 μ m filter, then washed with DPBS, and stained for flow cytometric analysis or cell sorting.

2.3.6 General flow cytometry and cell sorting

Fresh or thawed PBMCs or BMMCs, separated by Lymphoprep density centrifugation, were aliquoted into flow tubes at a density of 1–3 million cells per 50 μ l of DPBS with 0.1%–2% fetal calf serum (FCS, Gibco) and 0.4% EDTA. 3-5 μ l of anti-mouse IgG were added to each sample to prevent non-specific antibody binding, and incubated for 10 minutes at room temperature. Only fresh cells were used for bulk and single-cell RNA-Seq experiments. The samples were then stained with fluorescently-conjugated antibodies and incubated for 30 minutes at room temperature in the dark. Cells were washed in flow/sort buffer for 5 minutes at 500g and resuspended in 300-500 μ l flow/sort buffer. Dead cells (typically <5%) were excluded by staining cells with Zombie (Biolegend) amine dyes prior to staining, or 4',6-diamidino-2-phenylindole stain (DAPI, Partec), added immediately before the samples were analysed. Flow cytometry was performed with an LSRFortessa X-20 (BD Biosciences) running BD FACSDIVA™ 8.0.1. Cell sorting was performed using the FACSARIA III sorter (BD Biosciences) running BD FACSDIVA™ 8.0 software. Data were processed with FlowJo 10.4.1 and 10.1r5 (Tree Star, Inc).

2.3.7 NanoString Assay

Ex vivo or culture-generated DCs were FACS purified (> 98% purity) and lysed in RLT buffer containing 1% β -mercaptoethanol, at a concentration of 2,000 cells/ μ l. Samples were analysed on the NanoString nCounter FLEX platform according to manufacturer's instructions. Briefly, 5 μ l of lysate (10,000 cells) were mixed with reporter probes, hybridisation buffer, and capture probes and hybridised at 65°C for 12-30 hours. Samples were then processed on the NanoString Prep station and cartridges were read on the NanoString Digital Analyzer to yield a reporter code count (RCC) dataset. The human Immunology_V2 panel was used, supplemented with the following 30 genes: ASIP, DAXX, MERTK, C19orf59, DBN1, Ki67, CCL17, F13A1, NDRG2, CD1c, FGD6, PACSIN1, CD207, FLT3, PPM1N, CLEC10A, GCSAM, PRAM1, CLEC9A, GGT5, S100A12, CLNK, LPAR2, TMEM14A, COBLL1, LYVE1, UPK3A, CXCL5, MAFF, ZBTB46.

2.4 Bioinformatic techniques and analysis pipelines

2.4.1 Clustering and dimensionality reduction

The majority of the datasets generated and analysed within this thesis (including flow and mass cytometry, NanoString, and bulk and single cell RNA-Seq) contained a large number of measured variables, and are considered to be high dimensional datasets. The data become increasingly more difficult to analyse with the number of dimensions, and dimensionality reduction techniques are often employed to aid the visual exploration and interpretation of high dimensional data. These techniques can be broadly classified into linear and non-linear methods. The linear methods rely on a linear projection of the data on a subspace. The non-linear methods focus on recovering the low-dimensional surface of the underlying manifold (low-dimensional data embedded in higher dimensional space) that the data have been sampled from.

2.4.1.1 Principal component analysis

Principal component analysis (PCA) is linear dimensionality reduction technique. An orthogonal transformation of the data is performed in order to convert potentially correlated variables into a set of linearly uncorrelated variables called principal components. The new variables are ranked in order of the amount of variance explained, principal component 1 (PC1) explaining the most variance. The PCA method can be used on its own or combined with other dimensionality reduction

techniques. An initial PCA step is often performed prior to running non-linear methods, in order to reduce noise and speed up the computation.

2.4.1.2 t-distributed stochastic neighbour embedding

tSNE is a method for data visualisation and exploration suitable for high-dimensional datasets (van der Maaten and Hinton, 2008). It is a non-linear dimensionality reduction technique, relying on the attraction/repulsion balance (Barnes-Hut approximation). tSNE preserves the local structure of the data, and is able to identify patterns and clusters in the data. However, the distances between the clusters and their density are not meaningful.

2.4.1.3 Diffusion maps

Diffusion maps are a non-linear technique, based on a distance metric, known as diffusion distance. Diffusion maps are able to preserve the developmental continuum and can be used to infer cell trajectories. Formulations adapted especially for single cell data analysis have been developed (Haghverdi et al., 2015).

2.4.1.4 Hierarchical clustering

Clustering represents the task of grouping a set of data into clusters. As a result, data objects similar to one another are placed within the same cluster, and dissimilar to the objects are placed in other clusters. Hierarchical clustering, also known as connectivity clustering, is based on connecting the data objects to form clusters based on their distance. Clustering results are often represented as dendrograms. Strategies for clustering include the bottom-up and top-down approaches. The bottom-up approach, also known as agglomerative, starts with single data objects and aggregates them into clusters. The top-down, divisive, approach, starts with the complete dataset and divides it into partitions. In addition, the clustering methods differ by the way the distance is computed. Examples of distance calculation methods include single, complete, average, and Ward linkage.

2.4.2 Analysis of transcriptomic data

Common steps in all used transcriptomic pipelines include quality control (QC) and normalisation. Crucial QC steps are read QC (for sequencing data), as well as sample and gene filtering. Sample filtering (or cell filtering for scRNA-Seq) aims to

identify and remove poor quality samples/cells, often behaving as outliers in the dataset. Gene QC filters out genes with low expression that do not hold any useful information. Normalisation aims to make the levels of gene expression across the samples directly comparable. Raw data are adjusted for factors that prevent this, such as technical variability, differences in input material, and variations in library size (for sequencing data). Depending on the type of data, normalisation is performed based on library size, distribution, and/or controls. Control methods include housekeeping genes and normalisation with ERCC spike-ins (Evans et al., 2018).

To discover differences in expression levels between experimental groups, differential gene expression testing is often employed to perform statistical analysis on normalised gene counts. Statistical testing is used to confirm whether differences in the expression of certain genes are significant enough to not be attributed to random variation. A variety of methods for differential gene expression testing exist, appropriate for certain types of data and experimental design.

Gene Set Enrichment Analysis (GSEA) is an analytical method commonly applied to transcriptomic data. It determines whether a selected set of genes shows correlation with a biological state, in a pairwise comparison manner (Subramanian et al., 2005). Bubble GUM (GSEA Unlimited Map), a computational tool based on GSEA, allows to automatically extract gene signatures based on transcriptomic data by performing multiple GSEA runs in a row (Spinelli et al, 2015).

2.4.3 Analysis of ChIP-Seq data

The ChIP-Seq analysis pipelines commonly include read QC and alignment to the reference genome, followed by filtering of the alignment file to remove PCR duplicates and reads with poor alignment quality. A typical processing step for ChIP-Seq data is peak calling, executed in order to identify the binding sites of the DNA-binding protein of interest. This is commonly performed against a similarly processed reference sample prepared from appropriate control chromatin or a control immunoprecipitation (Landt et al., 2010). Peak calls are then compared between biological groups, as well as within groups. Peaks are annotated, and further analyses often revolve around the functional enrichment analysis of the genes that are found in the proximity of the binding peaks.

Chapter 3. Transcriptional identity of DCs generated in culture

Questions answered in Chapter 3:

1. How can human *bona fide* DCs be generated *in vitro*?
2. Is it possible to produce sufficient cDC1s for research applications and therapeutic approaches in culture?
3. Do the culture-derived cDC1s resemble peripheral blood and/or tissue DCs?

3.1. Introduction

The ability of DCs to initiate and regulate immune responses has made them prime candidates for use in immunotherapy and vaccination. Nevertheless, DCs, and cDC1s in particular, remain relatively inaccessible for therapeutic use, as well as research, due to their rarity *in vivo* (5,000 pDCs, 4,000 cDC2s, and 500 cDC1s per mL of peripheral blood). Due to the relative abundance of peripheral blood monocytes, DC research and therapeutics have largely focused on *in vitro* monocyte-derived DCs. However, natural DCs show functional specialisation, while the role of the inflammatory, monocyte-derived DCs is not well understood. The ability to generate functional human DCs *in vitro*, in a scalable system, could facilitate translational studies to exploit their therapeutic benefit.

Common methods for producing DCs in culture include culture of haematopoietic stem cells and progenitors from bone marrow with the addition of GM-CSF, a cytokine known to stimulate stem cells to differentiate into granulocytes and monocytes. The culture output in these conditions is skewed toward macrophages and cDC2-like cells in mouse (Helft et al., 2015). Supplementation of murine BM with the ligand for the cytokine receptor Flt3 was shown to yield pDCs and cDC2s in culture (Naik et al., 2005). However, the produced cDC1s did not express the appropriate phenotypic markers and were likely immature (Kirkling and Cytlak et al., 2018).

In human, co-culture of HSCs and progenitor cells from cord blood, with a mix of cytokines, including FLT3L and SCF, produced DCs with the appropriate expression profile and function, after an initial progenitor expansion step (Balan et al., 2014). Lee et al. (2015) identified that all three major human DC subtypes can be produced without CD34⁺ progenitor expansion in co-culture with murine stromal cell line MS5 in

the presence of cytokines SCF, FLT3L, and GM-CSF. However, the proportion of generated cDC1s under these conditions was low.

Stromal cell lines producing M-CSF have been shown to induce the differentiation of embryonic stem cells down the monocyte-macrophage route (Nakano et al., 1994). The use of a stromal cell line with no M-CSF expression could have the potential to improve the DC culture output. One such cell line is OP9. It was established from stromal cells derived from a mouse with an M-CSF mutation ($Csf1^{-/-}$), resulting in the lack of functional M-CSF expression. This enables the OP9 stromal cells to promote differentiation of progenitors into haematopoietic lineages other than monocytes and macrophages, such as the erythroid, myeloid and lymphoid lineages (reviewed by Trakarnsanga et al., 2018).

In mouse, Notch2 deficiency results in reduction of splenic $CD8^+$ cDC1s (Lewis et al., 2011). It was therefore hypothesized that Notch signaling could improve human cDC1 differentiation *in vitro*. To optimise DC production, the differentiation of DCs from human $CD34^+$ progenitor cells was studied in liquid media, on the OP9 stromal layer or with OP9 expressing Notch ligand DLL1 (OP9-DL1), supplemented with FLT3L, SCF and GM-CSF (FSGM cocktail) (Figure 3.1). The transcriptomic profile and the surface markers of the *in vitro* generated DC subsets and their *ex vivo* counterparts were then assessed.

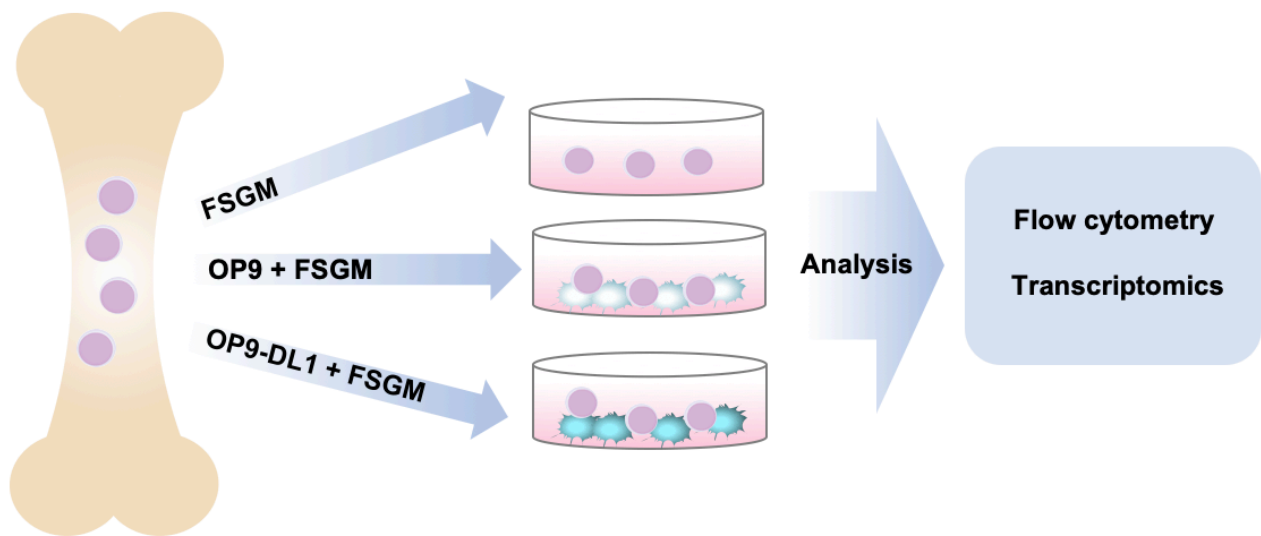


Figure 3.1. Schematic of the culture conditions used for experiments in Chapter 3.

A range of culture conditions were interrogated in order to investigate human DC development. Bone marrow-derived CD34⁺ progenitors were placed in culture media supplemented with a mix of cytokines (FSGM) and in co-culture with the OP9 or OP9-DL1 feeder layers with the addition of the cytokine mix.

FSGM – growth medium supplemented with FLT3 ligand (100ng/ml), SCF (20ng/ml) and GM-CSF (20ng/ml) (FSGM); OP9-DL1 - OP9 expressing Notch ligand DL1.

3.2. Materials and methods

3.2.1. Co-culture of bone marrow progenitors with OP9 and OP9-DL1 cells

The OP9 and OP9-DL1 cell lines were defrosted as described in Chapter 2 subsection 2.3.3 and cultured in 96 well plates as described in Chapter 2 subsection 2.3.5 by Dr Urszula Cytlak.

3.2.2. Flow cytometry and cell sorting

Flow cytometry and fluorescence-activated cell sorting were performed by Dr Urszula Cytlak as described in Chapter 2 subsection 2.2.6, using the flow panel outlined in Table 3.1.

Antigen	Fluorochrome	Clone	Manufacturer
CD11c	BV421/AF700/BV711	B-ly6	BD
CD123	PerCP-Cy5.5/BUV395	7G3	BD
CD14	BV650	M5E2	Biolegend
CD141	APC/BV510	AD5-14H12/1A4	Miltenyi/BD
CD16	FITC	3G8	BD/Biolegend
CD19	FITC	4G7	BD
CD1c	PE-Cy7/PERCPCy5.5	L161	Biolegend
CD2	PE	RPA-2.10	BD
CD20	FITC	L27	BD
CD3	FITC	SK7(Leu4)	BD
CD303	APC/BV605	201A	Biolegend
CD304	APC/BV605	12C2/U21-1283	Biolegend/BD
CD33	BV711	WM53	Biolegend
CD34	BV605/APCCy7	581	Biolegend
CD45	AF700	HI30	Biolegend
CD5	BUV730	UCHT2	BD
CD7	FITC	M-T701	BD
HLA-DR	BV785/V500	L243/G46-6	Biolegend/BD

Table 3.1. Antibodies used for flow cytometry and cell sorting.

For each antibody, the table lists the target antigen, the conjugated fluorochrome, the clone, and the manufacturer.

3.2.3. NanoString gene expression assay

The NanoString assay was performed as described in Chapter 2 subsection 2.3.7. Counts were normalised within the nSolver 3.0 software (NanoString Technologies). Samples marked with a QC flag by the nSolver software were removed from analysis. Data were normalised in nSolver advanced analysis module version 1.1.4. The module uses the geNorm algorithm to select the best housekeeping genes, which are used for normalisation. The log₂ transformed normalised output data were analysed using R version 3.6.0 (R Core Team, 2019).

Heatmaps and hierarchical clustering of primary DCs or DCs cultured on OP9 or OP9-DL1 were performed based on subset specific surface antigen, TLRs and TFs, as well as chemokine receptor gene expression using the "heatmap.2" function from the pheatmap R package version 1.0.12.

For further analysis, genes that were not expressed at a detectable level and did not reach normalised log₂ expression values of at least 4 in at least one third of the samples were removed from analysis (235 out total 608 endogenous genes were filtered out).

A culture signature was derived by performing pairwise comparisons (two-tailed t-test with Benjamini-Hochberg correction of p-values) of all culture versus all *ex vivo* populations. The results of the differential gene expression testing were displayed as a volcano plot using the EnhancedVolcano version 1.2.0 package in R (Blighe, 2019). Genes with adjusted p-values < 0.05 (the "culture signature") were excluded (95 out of the 373 expressed genes were removed) and the remaining 278 genes were used to construct the combined *ex vivo* and culture-derived cell PCA plot (Figure 3.5 B). For comparison, PCA plot was also constructed using all 608 assayed genes (Figure 3.5 A). Principal component analysis was performed using the "prcomp" function within the stats package version 3.6.0 and visualised using the ggbiplot package version 0.55.

The culture signature was split into genes with higher expression in blood (32/95 genes) and in culture-derived populations (63/95 genes), according to the fold changes in mean expression between the two conditions. The hypergeometric test for association of categories and genes within the Category R package version

2.50.0 was employed to evaluate the overrepresentation for Gene Ontology (GO) terms in culture signature genes with higher expression in blood and in culture. Prior to being used in the hypergeometric test, the gene symbols were converted into Entrez Gene IDs with the “translate” function within the AnnotationFuncs package version 1.34.0. The hypergeometric overrepresentation test was carried out for the GO Biological Processes (BP) category using the two converted gene sets as input. The 608 genes assayed by NanoString gene expression panel were specified as the gene universe. The adjusted p-value cut-off was set at 0.05. The p-values and the OddsRatio statistics for the GO BP terms with more than 5 observed genes counts were visualised as a bar chart with the ggplot2 package version 3.2.1 (Wickham, 2016).

3.2.4. Bulk RNA-Seq

Ex vivo or culture-generated DCs were FACS purified (> 98% purity) and lysed in RLT buffer containing 1% β-mercaptoethanol. RNA was extracted using the QIAGEN RNeasy Mini Kit. The RNA was quantified with the Qubit RNA HS Assay Kit and diluted to 5ng in 10µl. The SMART-Seq v4 protocol was used for cDNA synthesis. Sequencing libraries were prepared with the Nextera XT library prep kit. The Illumina NextSeq 500 platform was employed to generate 75bp paired-end reads. Library preparation and sequencing were performed by the Genomics Core Facility, Newcastle University.

Reads were trimmed based on quality with Trimmomatic v 0.36 (Bolger et al., 2014). Bases with quality scores below Q20 (inferred base call accuracy below 99%) were trimmed and reads shorter than 50bp were dropped. The remaining reads were aligned with the STAR mapping algorithm v 2.4.0 (Dobin and Gingeras, 2015) to the human reference genome version GRCh38.p7 (GENCODE release 25). The files were converted from SAM format to the more compressed BAM format with SAMtools v 1.3 (Li et al., 2009). The count tables were obtained using HTSEQ v 0.6.1 (Anders et al., 2015). ENSEMBL IDs were converted to HGNC gene names using biomaRt v 2.30.0 (Durinck et al., 2015).

Further analyses of the data were undertaken in R v 3.6.0 and Rstudio v 1.0.143. 15,081 protein coding genes with over 50 reads in total were retained for the analysis. The DESeq2 package v 1.24.0 (Love et al., 2014) was used for

normalisation. The top 1,000 genes with highest variance across the samples were selected for the PCA, performed using the "prcomp" function within the stats package version 3.6.0 and visualised using the ggbiplot package version 0.55.

To cluster the samples by similarity, the regularised-logarithm transformation was applied to the data. The Euclidean distances between the samples were then calculated, and used as input for the hierarchical clustering with the complete linkage method. The results were displayed as a heatmap with the pheatmap R package version 1.0.12 (Kolde, 2019).

The blood and tissue cDC1 signatures were computed via the BubbleGUM software. The peripheral blood (4,957 genes) and tissue (8,086 genes) signatures were used as input for the single sample Gene Set Enrichment Analysis (ssGSEA), along with the normalised gene expression values for cDC1s from all 5 sources. The enrichment scores were displayed as a dot plot.

The hypergeometric test for association of categories and genes within the Category R package version 2.50.0 was employed to evaluate the overrepresentation for Gene Ontology terms in the blood and tissue cDC1 signatures. Prior to being used in the hypergeometric test, the gene symbols were converted into Entrez Gene IDs with the "translate" function within the AnnotationFuncs package version 1.34.0.

3.3. Results

3.3.1. Culture output phenotyping and cell enumeration

To optimise the DC production *in vitro*, human CD34⁺ progenitors were FACS-purified and placed in culture under 3 different conditions: (1) liquid culture containing growth medium supplemented with the cytokine mix FSGM, consisting of FLT3 ligand (100ng/ml), SCF (20ng/ml) and GM-CSF (20ng/ml), (2) co-culture with stromal cell line OP9, supplemented with FSGM, and 3) co-culture with OP9-DL1 with the addition of FSGM.

The phenotype of the cells produced at 14 days of culture was assessed using flow cytometry and compared to that of peripheral blood DCs and CD14⁺ monocytes (Figure 3.2). Overall, the culture-derived cells occupied phenotypic spaces

corresponding to blood DCs and monocytes, and were gated in a similar manner. The surface markers were expressed appropriately, classical monocytes showing high CD14 and CD11c expression, cDC1s were marked by CD141 and CLEC9A, cDC2s – by CD1c and CD11c, and pDCs – by CD123 and CD303/304. Most of the APCs, distinguished by high HLA-DR expression, were represented by classical monocytes in blood, as well as in liquid culture. The APC gate contained significantly lower proportions of classical monocytes in co-culture with feeder layers OP9 and OP9-DL1.

The culture output was normalised per input progenitor placed in culture to directly compare the efficiency of generating each of the DC subsets under different conditions (Figure 3.3). The culture in the liquid medium with FSGM supported only a minor expansion of the DC subsets, creating a small proportion of cDC2-like cells, and no pDCs or cDC1s were observed. The addition of the OP9 feeder layer saw an increase in DC output, the majority of DCs produced still being cDC2s, with little pDCs and cDC1s generated. The incorporation of Notch ligand DLL1 in the co-culture with the use of OP9 cells resulted in the selective expansion of cDC1, with an 11-fold increase in cDC1 output per progenitor compared to the OP9 co-culture.

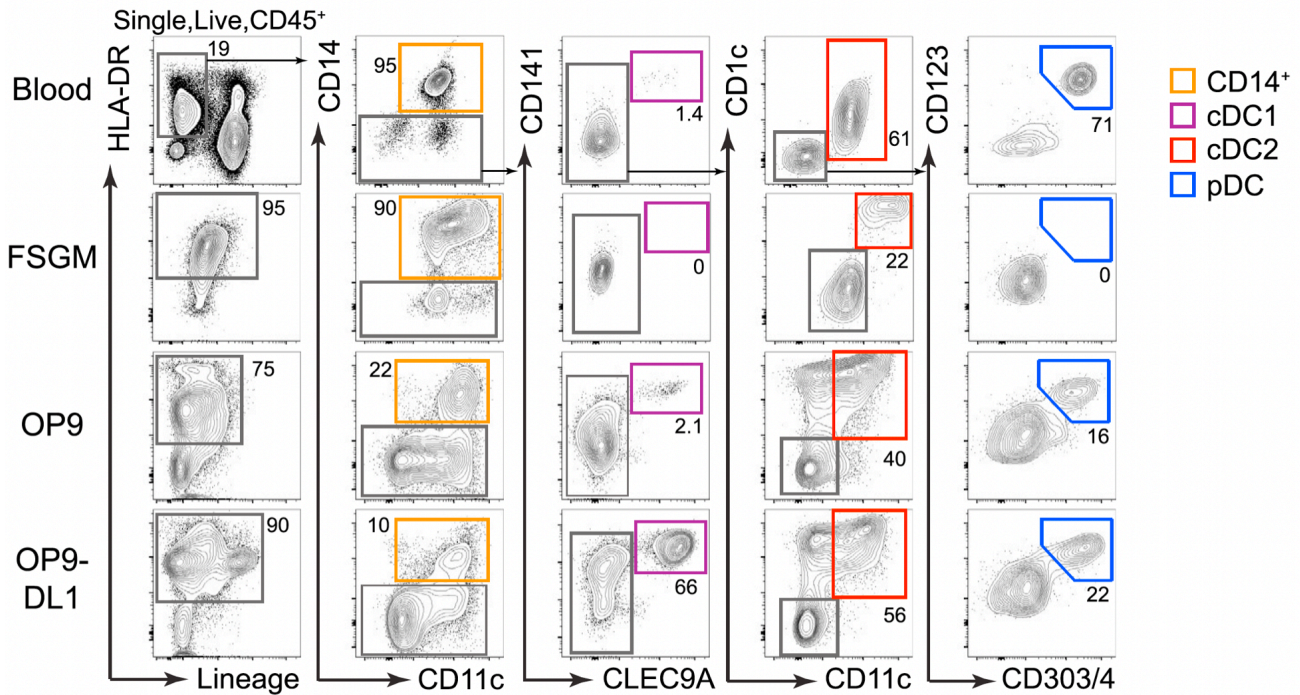


Figure 3.2. Analysis of primary peripheral blood and in vitro-derived cells via flow cytometry (Dr Urszula Cytlak and Dr Venetia Bigley).

Sorted $CD34^+$ progenitor cells purified from human BM were cultured for 14 days in liquid culture in the presence of the cytokine cocktail FSGM and on monolayers of OP9 or OP9-DL1 cells supplemented with the same cytokine mix. Single, live, and $CD45^+$ cells were gated for the analysis of the culture output. Bivariate flow plots show expression of cell markers: HLA-DR for APCs, CD14 and CD11c for classical monocytes, CD141 and CLEC9A for cDC1, CD1c and CD11c for cDC2, and CD123 and CD303/304 for pDC. Lineage markers include B, T and NK cell markers CD3, 19, 20, and 56, and cells expressing them were gated out.

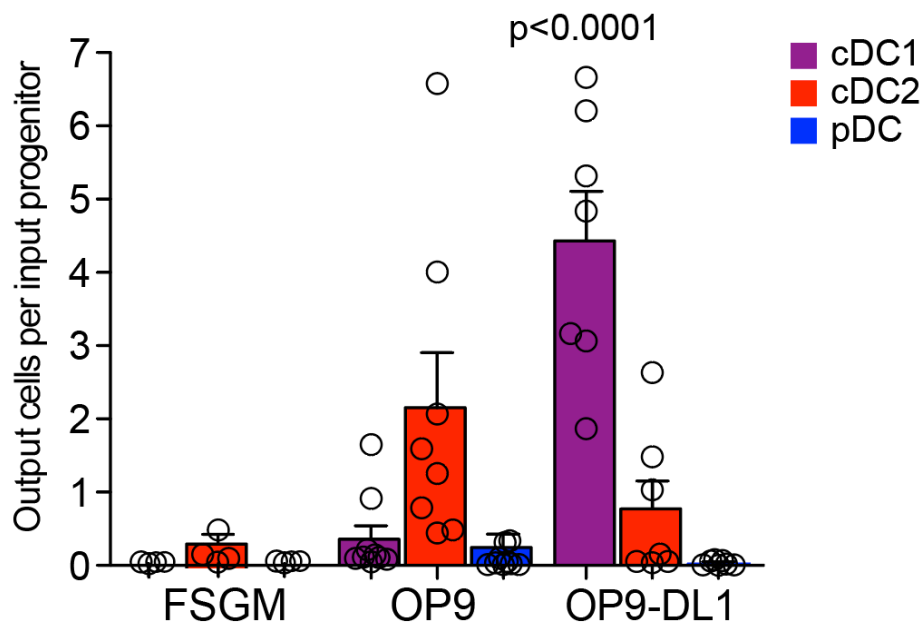


Figure 3.3. The number of output DCs generated per input progenitor cell in culture media supplemented with a mix of cytokines (FSGM) and in co-culture with the OP9 or OP9-DL1 feeder layers (Kirkling and Cytlak et al., 2018).

Data points represent values in 14 days BM cultures from different donors ($n=4$ for FSGM, 8 for OP9, and 7 for OP9-DL1); bars represent mean with SEM. Indicated p -values were derived by unpaired two-tailed Student's t -test conducted on cDC1 proportions derived from OP9 and OP9-DL1.

3.3.2. Confirmation of cultured cell identities via NanoString analysis

Having identified a culture system able to enrich the output in cDC1s and generate all DCs subsets with a seemingly appropriate phenotype, transcriptomic approaches were sought to validate the identities of the produced cells. The NanoString gene expression assay was selected for this analysis, as it provides robust measurements of gene expression from low input material and requires no amplification steps.

All 3 DC subsets from peripheral blood and OP9, along with cDC1s derived from OP9-DL1 were subjected to NanoString analysis using the NanoString pre-built human Immunology V2 panel. The panel contains 578 immunology-related genes and 15 housekeeping genes, and was supplemented with 30 DC-related genes.

To confirm the transcriptomic identity of the cells generated in culture, the expression of DC subset-specific genes and chemokine receptors was compared between the *ex vivo*-derived and cultured cells. All culture-derived cells resembled their *ex vivo* counterparts, and clustered closely with them based on the expression of DC subset-specific genes (hierarchical clustering displayed as a dendrogram above the heatmap in Figure 3.4 A). Crucially, genes encoding subset-specific transcription factors, surface markers, and TLRs were expressed faithfully. The marker genes included *CLEC9A*, *XCR1*, *IRF8*, *BATF3*, and *TLR3* in cDC1s, *CD1C*, *CD2*, *IRF4*, *TLR2*, *TLR8* in cDC2, and *IRF8*, *IRF4*, *TLR7*, *TLR9* in pDCs (Figure 3.4 A). Higher *CD1C* expression was noted on the culture-derived cDC1s. Likewise, the expression of chemokine receptors was split by subset (Figure 3.4 B), however a higher expression of *CCR7* in OP9-derived cDC1s and cDC2s was observed.

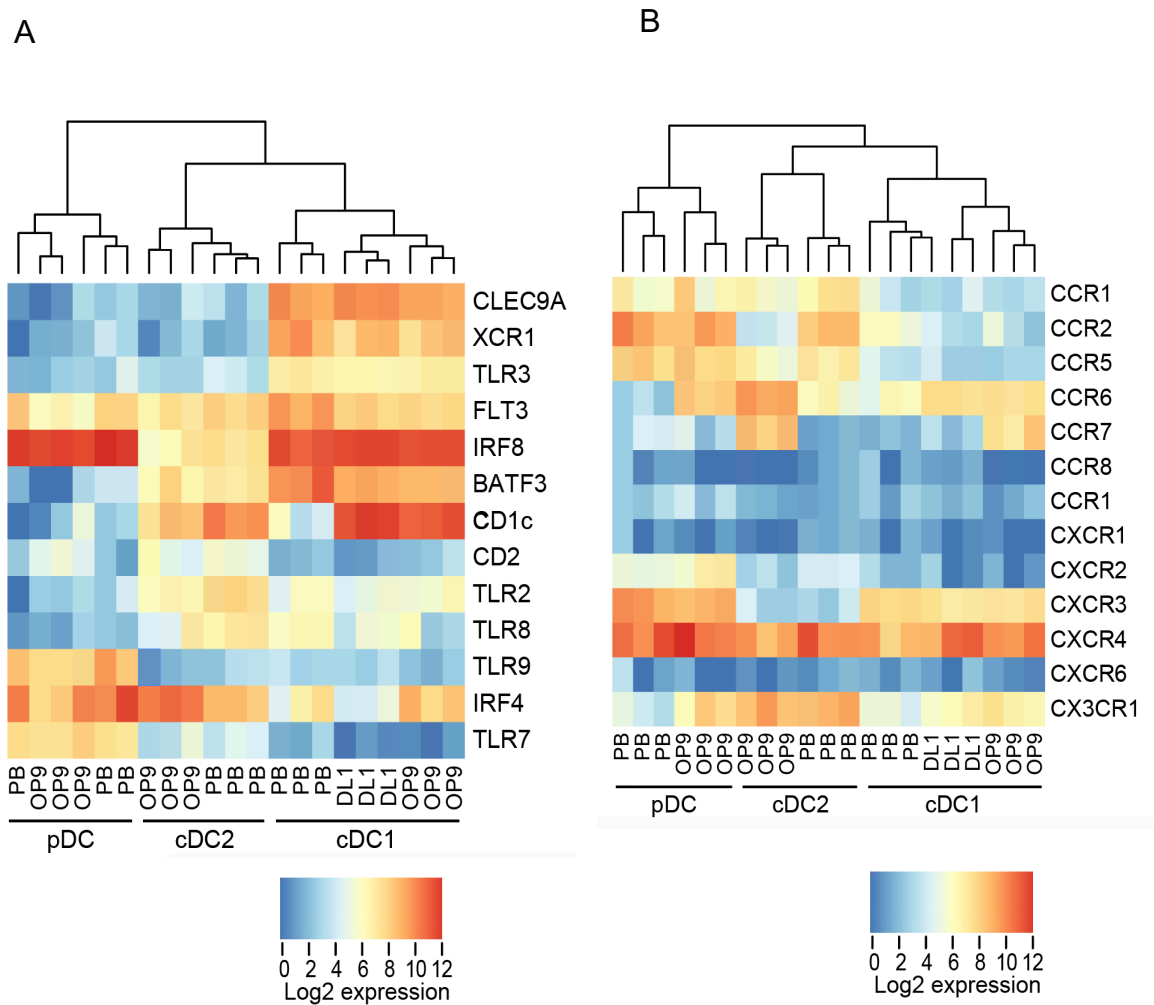


Figure 3.4. Heatmaps of DC subset-specific genes encoding surface markers, transcription factors, and TLRs (A) and chemokine receptors (B), as determined by NanoString nCounter analysis.

The dendrograms above the heatmaps represent hierarchical clustering based on the log₂ gene expression of the listed genes.

Next, principal component analysis was employed to obtain a summary of the dataset. In this analysis, all cultured subsets were shifted in a similar direction and by a similar distance both on PC1 and PC2 in relation to their *ex vivo* counterparts, implying that a particular set of genes, conserved between all DC subsets, was differentially expressed in all blood versus all cultured samples (Figure 3.5 A). Removing this set of differentially expressed genes, also referred to as “culture signature”, resulted with the perfect alignment of primary and *in vitro*-generated subsets, PC1 splitting cDCs from pDCs, and PC2 splitting cDC1s and cDC2s, irrespective of their source (Figure 3.5 B). Notably, both cDC1s generated on OP9 and on OP9-DL1 aligned closely with the peripheral blood cDC1s.

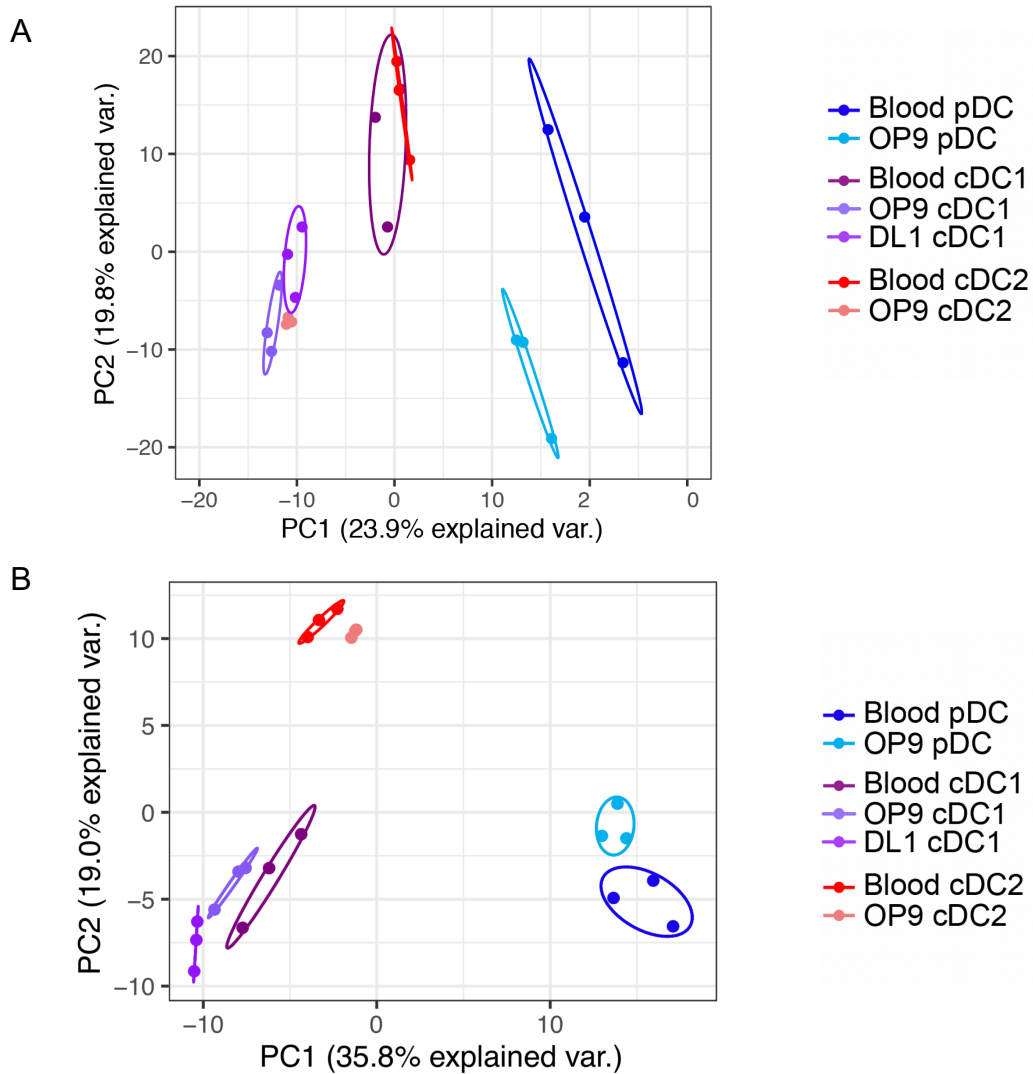


Figure 3.5. Principal component analysis of mRNA expression from FACS-purified primary and in vitro-derived DC subsets using the NanoString Human Immunology V2 panel plus 30 custom DC-related genes.

A. PCA performed based on all 608 genes assayed by the NanoString panel.

B. Analysis based on the expression of 278 genes, performed after the removal of genes with low expression (235/608) and of a “culture signature” (95/608) derived by pairwise comparison of all culture-generated versus all primary cells via a two-tailed *t*-test with Benjamini-Hochberg correction of *p*-values.

Further investigation of the culture signature revealed that it contained 95 genes, 32 of which had higher expression in blood DCs, and 63 were expressed higher in culture.

Among the culture signature components with higher expression in culture (Figure 3.6) were genes encoding the complement subunits C1QA and C1QB, costimulatory molecules CD80 and CD276, chemokines CXCL12, CCL17, and CCL22, transmembrane protein CD1A, highly expressed in tissue DCs, dendritic cell lysosomal associated membrane glycoprotein LAMP3, phosphatase DUSP4, associated with the negative regulation of the cellular proliferation and differentiation, as well as LAG3, a protein involved in the maturation and activation of dendritic cells (Andreae et al., 2002).

The top genes with higher expression in blood included *PRAM1*, expressed during normal myelopoiesis (National Centre for Biotechnology Information, <https://www.ncbi.nlm.nih.gov/>) and the receptor for granulocyte colony-stimulating factor *CSF3R* (Figure 3.6).

In order to gain functional insight into the lists of differentially expressed genes, an overrepresentation test for Gene Ontology terms from the Biological Processes category was performed for both gene sets (Figure 3.7). The set of genes with higher expression in culture was enriched in GO terms relating to the regulation of apoptosis, response to DNA damage, metabolism, cell cycle, and cell growth. The biological processes significantly overrepresented in blood included actin cytoskeleton organisation and regulated exocytosis.

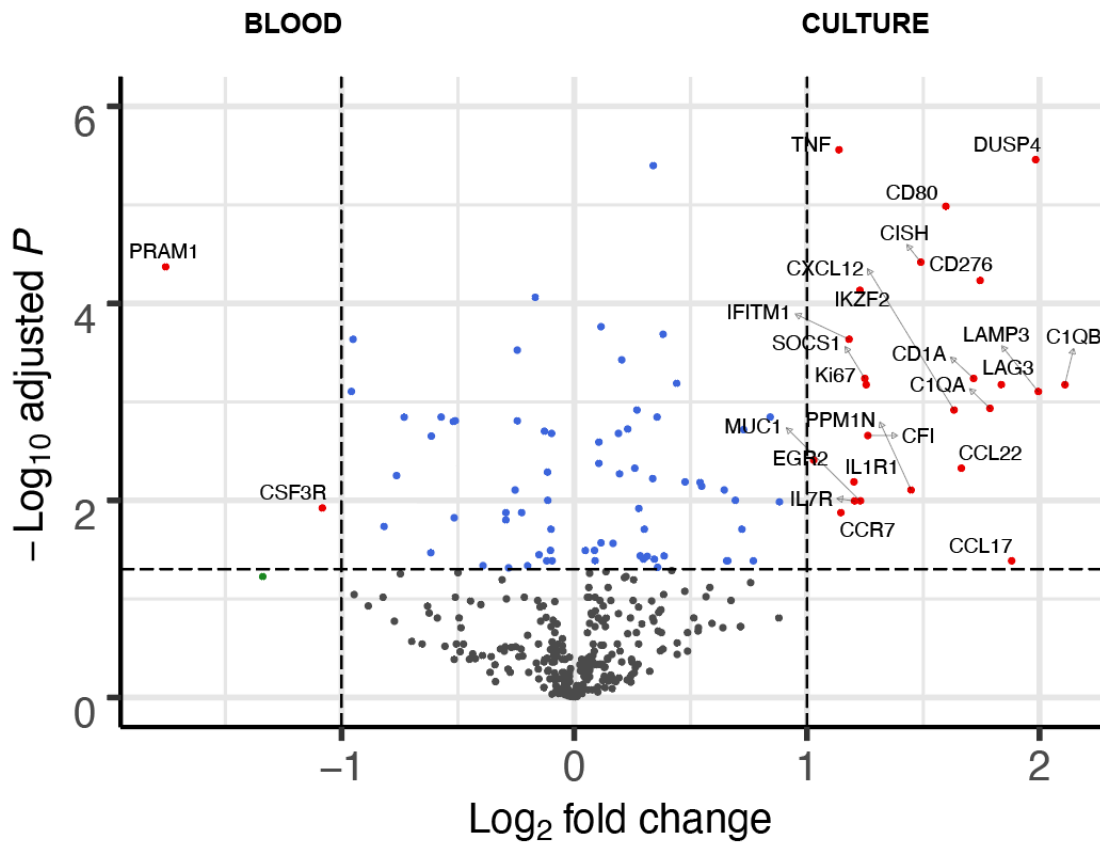


Figure 3.6. Differential gene expression analysis between all culture-derived versus all blood DC subsets, as determined by NanoString nCounter analysis. Genes with higher expression in culture are shown on the right, and genes with higher expression in blood are displayed on the left. The volcano plot displays the log₂ fold change and the log₁₀ adjusted p-values obtained via a paired t-test with Benjamini-Hochberg correction. Names are shown for top differentially expressed genes with adjusted p-values below 0.05 and log₂ fold change above 1 and below -1 (equivalent to linear fold change of 2 and 0.5, respectively). All genes passing the adjusted p-value threshold of 0.05 were regarded as the “culture signature”.

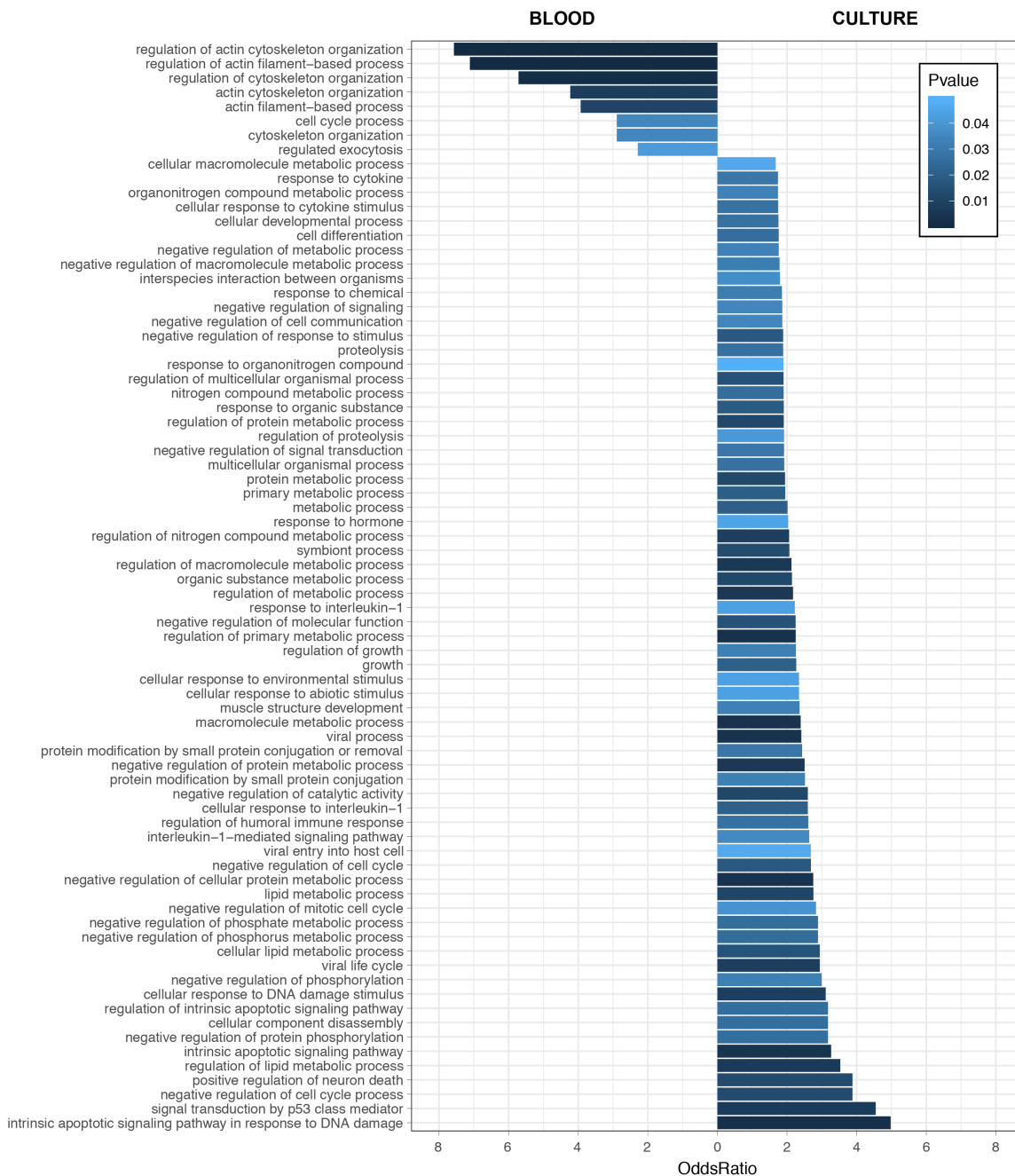


Figure 3.7. Bar chart depicting the overrepresentation of Gene Ontology terms among the culture signature genes with higher expression in blood and in culture, as determined by NanoString nCounter analysis.

The significantly overrepresented GO Biological Processes (p value < 0.05) are displayed on the y axis. The x axis and the length of the bars correspond to the OddsRatio, derived from the observed and expected gene counts for each GO term. The OddsRatio values increase toward the right for terms overrepresented in culture, and toward the left for terms overrepresented in blood. The bars are coloured according to the p -values resulting from the hypergeometric test for over or under-representation of each Biological Process term among the specified gene set.

The presence of marker CD1A as a top differentially expressed gene with higher expression in culture compared to blood suggested that the transcriptome of cultured cells could more closely resemble tissue than blood DCs. Typically, CD1A is abundantly expressed by LCs (reviewed by Brigl and Brenner, 2004), as well as by dermal and other tissue DCs (Haniffa et al., 2012).

3.3.3. Transcriptome analysis of cDC1s from blood, tissue, and culture

To analyse the relationship between culture-derived cDC1s, grown with OP9 and OP9-DL1, and their *ex vivo* counterparts from human spleen, bone marrow, and peripheral blood at the whole transcriptome level, bulk RNA-Seq was undertaken. PCA, performed using top 1,000 variable genes showed 3 separate clusters that contained; (1) PB cDC1s, (2) spleen and bone marrow cells, and (3) culture-derived cells (Figure 3.8). Akin to the NanoString analysis, the cDC1s cultured with OP9 and OP9-DL1 clustered very closely.

Next, hierarchical clustering was employed to further assess the overall similarity between the cultured and *ex vivo*-derived cDC1s (Figure 3.9). The initial division of the populations in the clustering hierarchy was between all culture derived cells (OP9 and OP9-DL1) and all *ex vivo* cells, suggestive of the presence of a set of genes specific to cultured cells. Peripheral blood cDC1s were the most dissimilar *ex vivo* compartment to the cultured cells. Interestingly, both the PCA and the clustering analyses showed no significant differences between the spleen and BM cDC1s. These samples were treated as one group titled “tissue cDC1s” in further analyses.

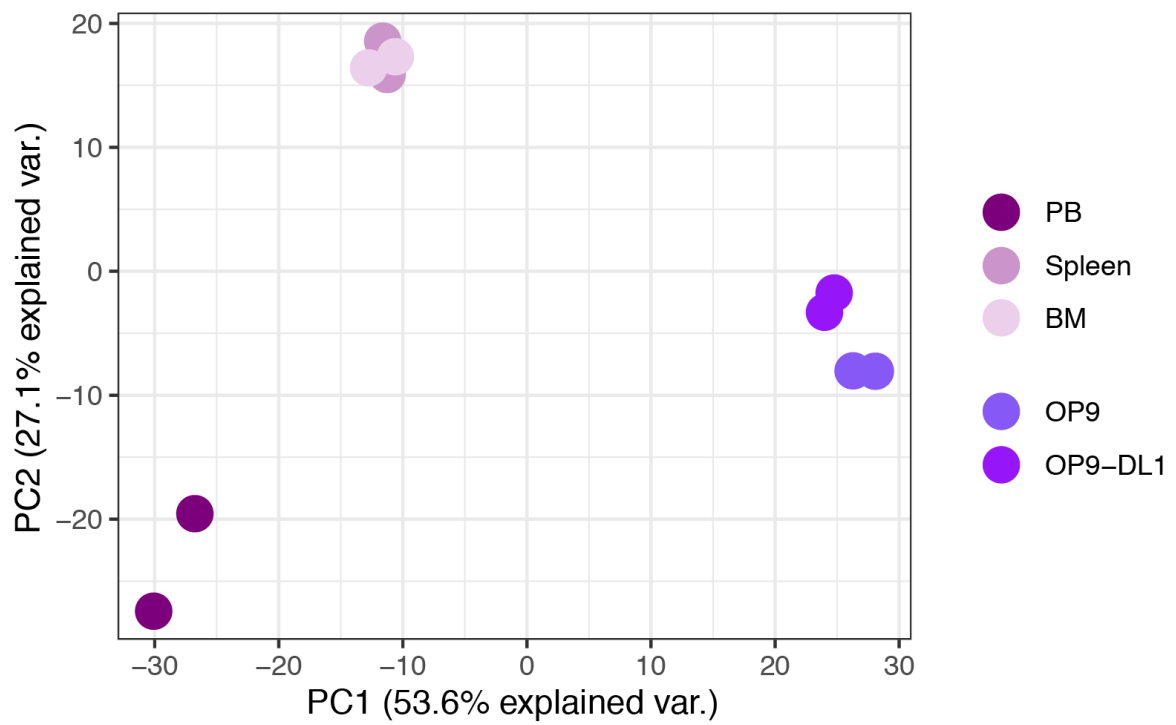


Figure 3.8. Principal component analysis of primary cDC1s from peripheral blood (PB), bone marrow (BM), and spleen, and cDC1s co-cultured with OP9 and OP9-DL1, as determined by bulk RNA-Seq.

Top 1,000 genes with highest variance across the samples were selected for the analysis.

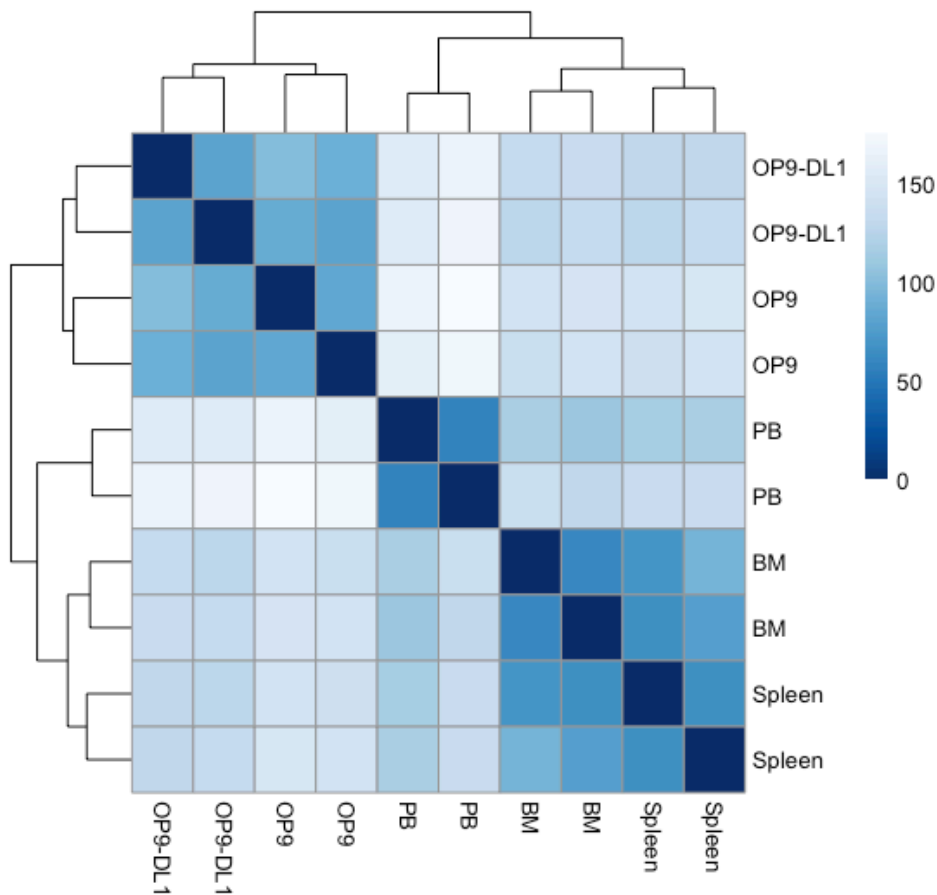


Figure 3.9. Heatmap and clustering of primary cDC1s from peripheral blood (PB), bone marrow (BM), and spleen, and cDC1s co-cultured with OP9 and OP9-DL1, as determined by bulk RNA-Seq.

The colours of the heatmap represent Euclidean distances between the samples, the dark blue indicating similarity and lighter blue showing dissimilarity between the samples. The mirrored dendrogram depicts the hierarchical clustering using the complete linkage method based on the Euclidean distances.

In order to further investigate whether the cultured cells aligned more closely with peripheral blood or with spleen and bone marrow cDC1s, the blood and tissue signatures from the RNA-Seq dataset were computed via the BubbleGUM software. The enrichment scores for the two signatures were then calculated using the single sample Gene Set Enrichment Analysis (ssGSEA) method (Figures 3.10 A and B). As expected, the peripheral blood samples had the highest “blood signature” scores, while both groups of *ex vivo*-derived cells had high scores for the “tissue signature”. Keeping with their previously observed similarity to tissue cells, culture-derived populations were similarly enriched for the “tissue signature” genes. The blood samples displayed the least “tissue signature” enrichment.

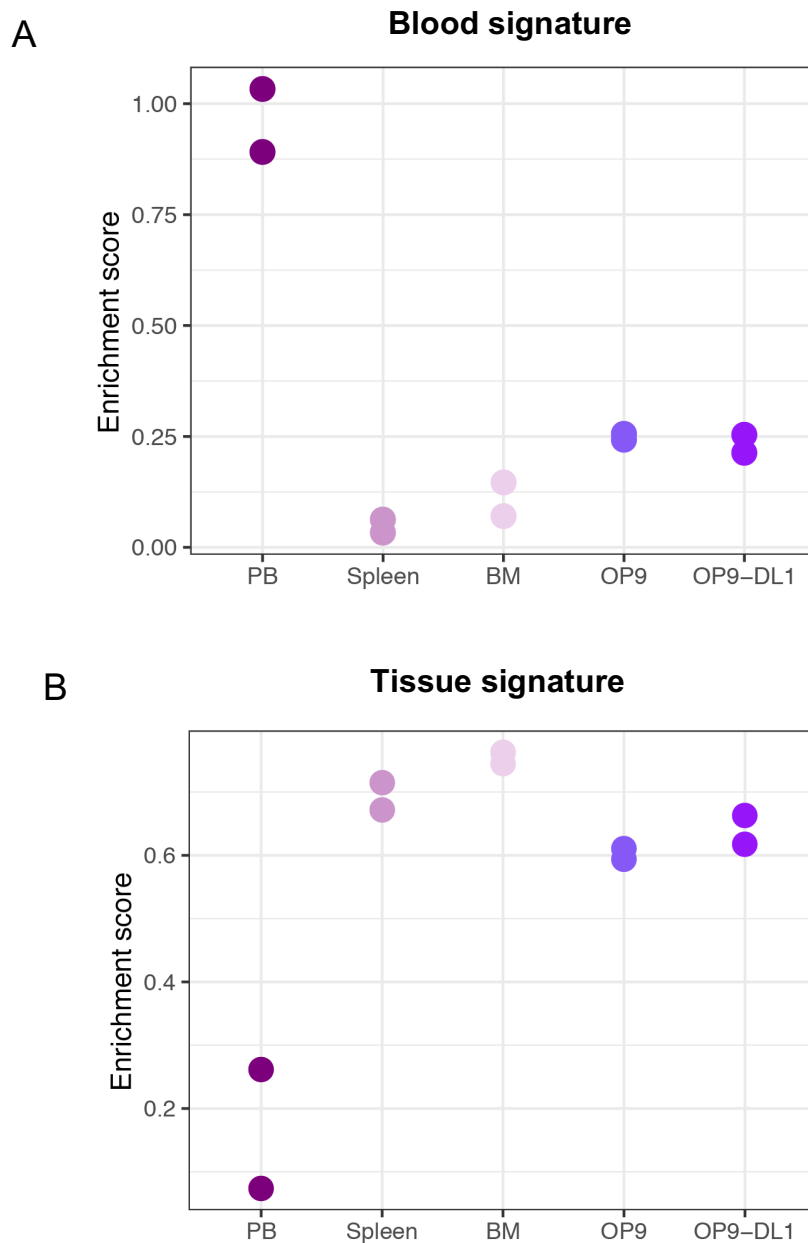


Figure 3.10. Enrichment scores of gene signatures across primary cDC1s from peripheral blood (PB), bone marrow (BM), and spleen, and cDC1s co-cultured with OP9 and OP9-DL1, as determined by bulk RNA-Seq.

The gene signatures for peripheral blood (A) and tissues (B), derived using the BubbleGUM software (Spinelli et al., 2015), were used as input for the ssGSEA enrichment method, along with the normalised gene expression values for all cDC1 conditions.

Next, the tissue and blood signatures were explored with an overrepresentation test for Gene Ontology terms (Figure 3.11). The tissue signature was enriched in terms relating to metabolism, response to DNA damage, cell cycle and division, as well as in mitochondrial gene expression. A large proportion of the GO terms overlapped with the terms enriched in culture, as determined by the NanoString analysis (Figure 3.7). The GO terms overrepresented in the peripheral blood cDC1s centred around leukocyte activation, immune processes, localisation, transport and exocytosis. A fraction of these terms was also observed in the list of terms enriched in blood compared to culture, as determined by the NanoString assay.

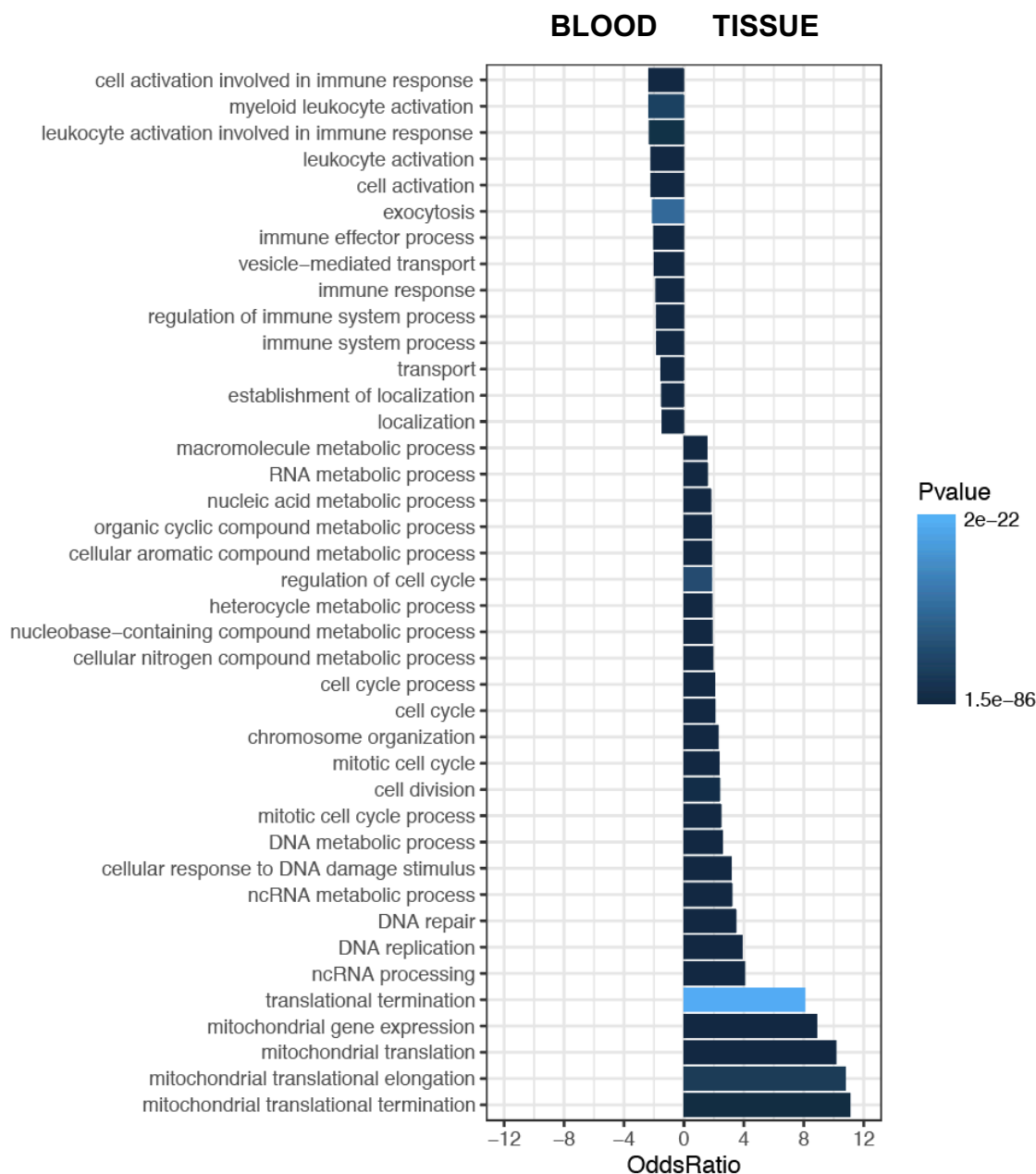


Figure 3.11. Bar chart depicting the overrepresentation of Gene Ontology terms among the BubbleGUM signatures for the blood cDC1s and tissue cDC1s, as determined by bulk RNA-Seq.

The top 40 significantly overrepresented GO Biological Process are displayed on the y axis. The x axis and the length of the bars correspond to the OddsRatio, derived from the observed and expected gene counts for each GO term. The OddsRatio values increase toward the right for terms overrepresented in tissue. The OddsRatio values increase toward the left for terms overrepresented in blood and are denoted with a negative sign. The bars are coloured according to the p-values resulting from the hypergeometric test for over or under-representation of each Biological Process term among the specified gene set.

3.4. Discussion

3.4.1. The production of human DC subsets *in vitro*

To address the requirement of large-scale DC production techniques, a method enabling the generation of all DC subsets in culture has been developed in the Human Dendritic Cell laboratory. The OP9 culture system is able to produce large numbers of cDC1s, cDC2s, and pDCs, in proportions of similar to those seen in peripheral blood and bone marrow. Together with the cytokines FLT3L, SCF, and GM-CSF, the OP9 feeder layer facilitates cell proliferation and differentiation *in vitro*, without the requirement for a progenitor expansion step, used in the past for DC production (Naik et al., 2005). Evidence that OP9 support CD34⁺ cells comes from previous experiments which were able to create CD34⁺ progenitors from human embryonic stem cells in co-culture with OP9 (Vodyanik et al., 2005). In these experiments, the OP9 cell line proved to be superior at growing progenitors to other murine stromal cell lines, such as MS5. This could be attributed to the fact that OP9 lack M-CSF expression, which inhibits the differentiation of progenitors into haematopoietic cells and drives the development of the monocyte/macrophage lineage (Nakano et al., 1994).

3.4.2. Confirmation of culture-derived cell identity

The authenticity of cDC1s, cDC2s, and pDCs produced in co-culture with OP9 and OP9-DL1 (for cDC1s) was verified via both proteomic and transcriptomic approaches, including gene expression analysis on the NanoString platform and surface phenotype analysis using flow cytometry.

The NanoString assay was used to investigate the expression of 578 immunology-related genes and that of 30 DC-related genes. All DC subsets faithfully expressed subset-specific transcription factors, surface markers, and TLRs, previously reported as DC markers in literature (Bigley, et al., 2016; Collin and Bigley, 2018; Schlitzer et al., 2018). Both peripheral blood and culture-derived cDC1s expressed high levels of key markers including *CLEC9A*, *XCR1*, *IRF8*, *BATF3*, and *TLR3*. cDC2s were marked with *CD1C*, *CD2*, *IRF4*, *TLR2*, and *TLR8*, and pDCs exhibited appropriate expression of *IRF8*, *IRF4*, *TLR7*, *TLR9*. The summary of the data based on all genes provided by the PCA revealed the presence of a gene set with altered expression in

culture. The identification and removal of these genes from the analysis resulted in the perfect alignment of the DC subsets with their *ex vivo* counterparts.

Phenotypic investigation of the OP9 and OP9-DL1-derived cells via flow cytometry aligned them closely with their *ex vivo* counterparts from peripheral blood, showing that the cultured and the primary cells occupied the same phenotypic spaces when defined by at least two subset-specific surface markers. The gates were determined based on the *ex vivo* populations, with the exception of the APC gate, marked by high HLA-DR and low lineage marker expression. The APC gate had to be adjusted for the cultured samples, as unlike the peripheral blood, they contained no lineage positive cells. This suggests that culture system does not support the growth of cell types with lineage markers, such as non-classical monocytes, and B, T, and NK cells. Co-expression of two established markers was used to identify each of the DC subsets and the classical monocytes. This practice is highly recommended for flow cytometry analysis, as using several antibodies coupled with different fluorochromes generally improves the reliability of the data.

In addition, the functional integrity of the culture-derived DCs was confirmed via assays performed by members of the group showed appropriate cytokine production in response to TLR stimulation by culture-derived cDC1s, cDC2s, and pDCs (Kirkling and Cytlak et. al, 2018).

3.4.3. The cells produced in culture bear close transcriptional resemblance to tissue DCs

As revealed by the NanoString analysis, the cultured DCs resemble the peripheral blood subsets based on markers defined in the literature. However, a set of genes was differentially expressed between all cultured versus all peripheral blood DCs (95 genes out of 608 assayed genes). Disparities in the cell microenvironment in peripheral blood and culture are the likely cause of the observed differences in gene expression. The factors include cell interactions, soluble factors, haemodynamic differences, and epigenetics.

The genes with higher expression in culture were linked to apoptosis, cell growth, and cell cycle. Complement component C1Q, upregulated in culture, has previously been shown to facilitate apoptotic cell clearance (reviewed by van Cooten et al.,

2008), which aligns with the apoptotic signature of the culture. The cultured cells are likely more metabolically active, divide more rapidly, and are more prone to apoptosis.

Higher *CD1C* expression was noted on the culture-derived cDC1s. The elevated expression of this marker has previously been observed in tissue DCs (Haniffa et al., 2012). LC marker *CD1A* also had higher expression in culture than in blood. These observations are in agreement with the cDC1 RNA-Seq analysis results, which show an enrichment of the “tissue signature” in the OP9 and OP9-DL1-derived cDC1s. Together, the NanoString and RNA-Seq analyses reveal an overlap of the tissue and culture signatures. This could be due to the similarity of the two environments, in particular the contact with stromal cells and lack of consistent shear forces in both the culture and the tissue settings.

The peripheral blood DCs had higher expression of genes regulating actin cytoskeleton organisation and regulated exocytosis compared to cultured cells. As DCs pass through blood vessel endothelium on their journey from bone marrow into peripheral blood, they might involve actin rearrangement when moving through tight spaces. Peripheral blood, being subject to shear forces, may also require additional cytoskeletal activity. Furthermore, blood DCs could be in preparation to migrate through endothelium once more in order to enter peripheral or lymphoid tissues.

The upregulation of exocytosis-regulated genes in blood can be attributed to the fact that DCs undergo major endocytic processes, such as phagocytosis. Large fractions of the plasma membrane are transferred to the cytoplasm, requiring non-secretory exocytosis to take place in order to compensate for the internalised volume (Cocucci and Meldolesi, 2013). Moreover, the blood DCs appear more activated than their tissue and culture counterparts. The analysis of the differentially expressed genes between blood and tissue (spleen and bone marrow) cDC1s, showed that the cells in blood are marked by the expression of genes related to leukocyte activation and immune responses.

3.4.4. Notch signaling facilitates *in vitro* generation of human cDC1s

Together, the results of the transcriptomic validation and the surface marker evaluation show that through Notch stimulation, OP9-DL1 system produced large

numbers of functionally proficient cDC1s with an appropriate gene expression and phenotype. Co-culture of human BM CD34⁺ progenitors with OP9 generated all DC subsets in similar proportions to peripheral blood, while addition of Notch Delta like Ligand 1 resulted in an eleven-fold increase of CLEC9A⁺ CD141⁺ cDC1 output per progenitor cell.

The addition of Notch also improved cDC1 functionality and gene expression profile. Assays performed by members of the group (Kirkling and Cytlak et. al, 2018) demonstrated enhanced CD4 and CD8 T cell activation ability of OP9-DL1 cDC1 compared to OP9-derived cells and similar to that of *ex vivo* cDC1s from peripheral blood *in vitro*.

The OP9-DL1 culture system is able to produce large numbers of cDC1 *in vitro*, making them more accessible for therapeutic use and research. A 4mL bone marrow aspirate (equivalent to the volume taken in diagnostic procedures) typically yields sufficient CD34⁺ progenitors to generate 1.5-3 million cDC1s on OP9-DL1 in 96 well plates. This output is equivalent to the number of cDC1s in 3L whole blood (which contains approximately 500 cDC1s per mL), providing sufficient numbers of functional cDC1s for further biological study and translational applications. Furthermore, the culture is highly scalable, and large numbers of cDC1s have also been produced in 24 well plates and successfully used for the ChIP-Seq experiment in Chapter 5. The DC output per primary CD34⁺ stem/progenitor cell may be further increased with the addition of a CD34⁺ cell expansion step prior to DC differentiation. The use of moDCs in clinical trials for DC vaccines are often listed as a reason of failure. The use of defined, patient-derived DC subsets for next-generation DC vaccines could offer better vaccine functionality. cDC1s are particularly important for cancer vaccines, as they excel at activating critical effector cell types in antitumor immunity, such as cytotoxic T lymphocytes, NK cells, and NKT cells (reviewed by Cancel et al., 2019). In addition, cDC1s have superior cross presentation abilities *in vivo* (reviewed by Gutiérrez-Martínez et al., 2015), offering a promising prospect in therapy as the most potent anti-cancer DCs.

3.5. Summary and further work

Experiments undertaken as part of this chapter showed that co-culture of human bone marrow CD34⁺ progenitors with OP9 cells generates all DC subsets. The

resulting pDCs, cDC1s, and cDC2s aligned closely with their *ex vivo* counterparts by gene expression analysis and exhibited an appropriate phenotype. However, the proportion of generated cDC1s was low. Addition of Notch ligand DLL1 resulted in an eleven-fold increase of CLEC9A⁺ CD141⁺ cDC1 output per progenitor cell. The resulting cDC1s showed equivalent T cell stimulatory capacity to *ex vivo* derived cells in further experiments (Kirkling and Cytlak et al., 2018). Interrogation of their transcriptome revealed that the cells bear a striking similarity to the tissue cDC1 compartment. The OP9-DL1 culture system addresses one major hurdle for the utility of human cDC1s for vaccination – the rarity of cDC1s *in vivo*. The system allows the production of sufficient numbers of cells for functional study and therapeutic applications (Figure 3.12).

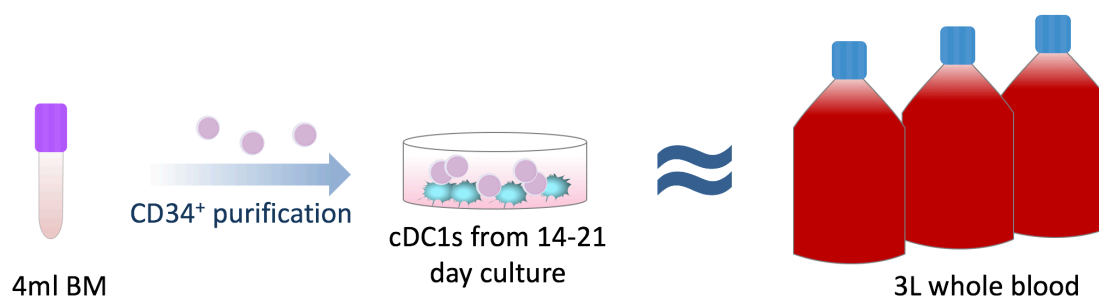


Figure 3.12: Schematic of OP9-DL1 co-culture process and its cDC1 output.

Approximately 4mL of bone marrow aspirate are required for culture. Once the CD34⁺ progenitors are FACS-purified, they are co-cultured with the OP9-DL1 stromal cell line for 14-21 days in the presence of a cytokine cocktail. The cDC1 output of the culture is equivalent to the cDC1 numbers in 3 litres of whole blood.

Future studies, necessary for the use of *in vitro*-derived DCs in therapy include adaptation of culture system to be Good Manufacturing Practice-compliant by removing animal-derived products from the culture system. The utilised animal-derived products include the foetal calf serum, used as a supplement in culture, as well as the murine OP9 feeder layer itself. Initial studies will determine whether cell-cell interactions, soluble factors or both are necessary for DC generation on OP9. For Notch stimulation, it is possible that the ligand signal (DLL1) could be provided by coating the culture vessel or beads with DLL1. In-depth proteomics of the factors produced by OP9 cells could help narrow down the proteins strictly necessary for DC production, which could be then used to supplement feeder layer-free cultures. The

contribution of autologous or haematopoietic cell derived factors may also be important in the generation of a mixed DC output. Practical factors will also be considered, including whether autologous DCs would be necessary or whether “off-the-shelf” third party DC products could be used.

Chapter 4. The two pathways of DC development defined by *IRF8* expression

Questions to be answered in Chapter 4:

1. Can distinct DC lineages and their precursors be identified in human bone marrow via single cell transcriptomics?
2. Is there a developmental basis for the recently described heterogeneity in the cDC2 population?
3. Can the developmental pathways originating in the bone marrow be linked to peripheral blood DC subsets?

4.1 Introduction

Human DC subsets develop in the bone marrow under the control of specific transcription factors, mutation of which can result in dendritic cell immunodeficiency. In humans, important roles are played by transcription factors *IRF8* (Hambleton et al., 2011; Bigley et al., 2018), *GATA2* (Dickinson et al., 2014), and *IKZF1* (Cytlak et al., 2018). DCs can be divided into at least 3 subsets, by transcription factor requirement, phenotype, and function (Bigley et al., 2016). The subset specialisation in DCs results directly from haematopoiesis (Lee et al., 2017; See et al., 2017; Villani et al., 2017), however, the pathways giving rise to cDC1, cDC2, and pDC in the bone marrow are not well mapped.

DCs have an apparent “dual” lympho-myeloid origin and, during their development, traverse the phenotypic spaces of hematopoietic stem cells (HSC), multipotent progenitors (MPP), common myeloid progenitors (CMP), lymphoid-primed multipotent progenitors (LMPP) and granulocyte-macrophage progenitors (GMP) (Doulatov et al., 2010; Lee et al., 2015; Lee et al., 2017; Helft et al., 2017). The CD123⁺ region of GMP has been demonstrated to contain potential to generate human cDC1s, cDC2s, and pDCs (Lee et al., 2015; Helft et al., 2017). However, the phenotypic identities of the unipotent DC progenitors are currently unknown.

Additional complexity is added by the recently described heterogeneity in the cDC2 compartment. The bulk of cDC2s are comprised of two subpopulations, one closer in gene expression and function to cDC1, referred to as DC2, and the other more similar to monocytes, referred to as DC3 (Villani et al., 2017; Dutertre, et al. 2019).

As this is a novel finding, earlier studies investigating DC development did not explore the source of cDC2 heterogeneity. It is currently unknown whether both subtypes of cDC2 arise from distinct lineage trajectories, regulated by different transcription factors, whether some of them are monocyte-derived, or whether they represent two transcriptional states of a common lineage originating from the CD123^{high} GMP (Figure 4.1).

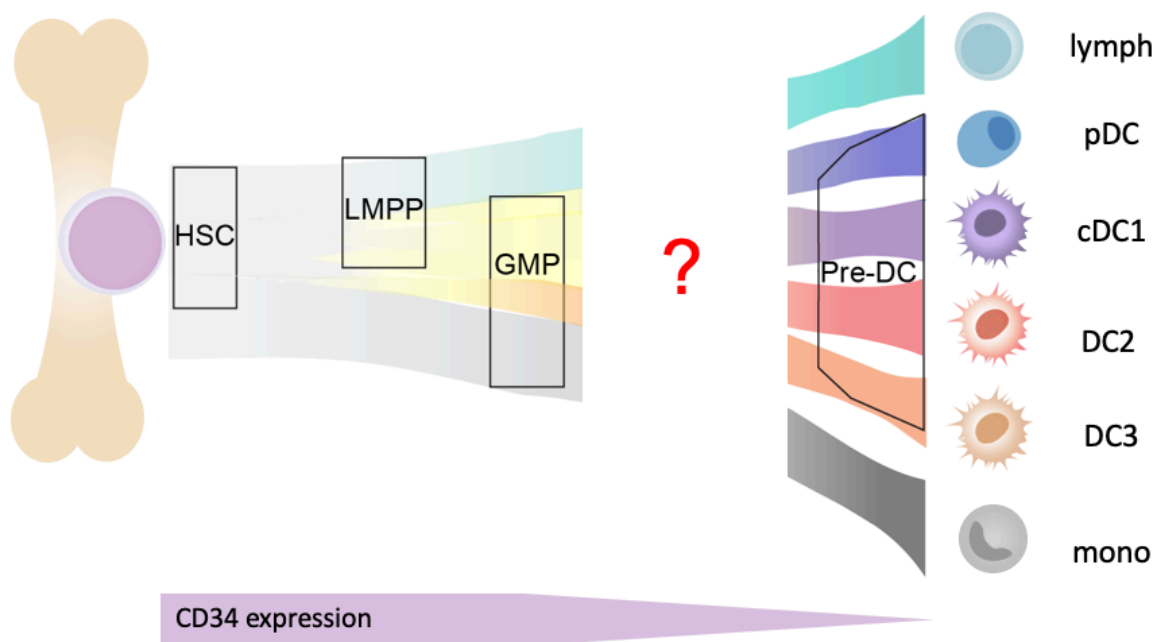


Figure 4.1. Schematic of the development of haematopoietic lineages from bone marrow HSCs, illustrating the unknown phenotypic identities of the DC progenitors.

During their development, the DC subsets traverse the phenotypic spaces of hematopoietic stem cells (HSC), multipotent progenitors (MPP), common myeloid progenitors (CMP), lymphoid-primed multipotent progenitors (LMPP) and granulocyte-macrophage progenitors (GMP). Early haematopoietic cells show high expression of marker CD34 (depicted in pink). As the cells develop, the CD34 expression is gradually lost. In this work, the term “progenitor” was used for cells expressing marker CD34, and “precursor” - for cells with intermediate or negative CD34 expression and without mature DC markers.

The apparent dual myeloid and lymphoid origin of DCs may be explained by early lineage commitment, not detectable by analysis of a few surface antigens, as used in traditional progenitor cell classification by flow cytometry. The work in this chapter aimed to employ single cell transcriptomics and mass cytometry, combined with phenotyping and *in vitro* culture, to define distinct pathways of DC development. This work was undertaken in close collaboration with other members of the Human DC Lab. In particular, Dr Urszula Cytlak performed the *in vitro* culture and phenotypic analysis and Sarah Pagan undertook the CyTOF experiments. The majority of this work has been published as a joint first author manuscript in Immunity (Cytlak and Resteu et al., 2020).

4.2 Materials and methods

4.2.1 Contributions

The following members of the HuDC Lab contributed to the generation and analysis of the data presented within this chapter:

- Urszula Cytlak: Phenotyping analysis, cell purification and *in vitro* differentiation assays (sections 4.2.2 and 4.2.3)
- Sarah Pagan: Antibody conjugation and cell preparation for CyTOF analysis (section 4.2.10)
- Venetia Bigley: pre-processing of mass cytometry data with the FlowJo software (part of section 4.2.11)
- Anastasia Resteu: visualisation of flow cytometry data to generate a 3D plot of antigen expression, generation and analysis of NanoString gene expression data, including normalisation, gene filtering, and principal component analysis (section 4.2.4), processing and analysis of single cell RNA-Seq data, specifically read QC, trimming, and alignment to the reference genome, counting of reads, gene and cell QC, filtering, normalisation, hierarchical clustering, dimensionality reduction using tSNE, lineage tracing with diffusion maps, and pseudotime analysis (sections 4.2.6-4.2.9), computational analysis of mass cytometry data, including visualisation of tSNE embeddings, lineage reconstruction with diffusion maps (section 4.2.11)

4.2.2 Co-culture of bone marrow progenitors with OP9 cells

The culture was undertaken as described in Chapter 2 subsection 2.3.5. Briefly, human bone marrow CD34⁺ progenitors, progenitor subsets or pre-DC were FACS-purified and placed into 96 well U-bottomed plates (Corning) with pre-seeded OP9 stromal cells (5,000/well) in 200µl alpha-MEM (αMEM, Gibco) supplemented with 1% penicillin/streptomycin (Sigma), 10% FCS, 20ng/ml granulocyte-macrophage colony-stimulating factor (GM-CSF, R&D systems), 100ng/ml Flt3-ligand (FLT3, Immunotools), 20ng/ml stem cell factor (SCF, Immunotools).

4.2.3 Flow cytometry

The culture output was assessed via flow cytometry, as described in Chapter 2 subsection 2.3.6, using fluorescently-conjugated antibodies listed in Appendix B.

4.2.4 Analysis of NanoString gene expression data

The NanoString assay was performed as described in Chapter 2 subsection 2.3.7. Data were normalised in nSolver advanced analysis module version 1.1.4. The log₂ transformed normalised output data were analysed using R version 3.6.0 (R Core Team, 2019). For the peripheral blood PCA (Figure 4.3), genes that did not reach above-background counts (normalised log₂ expression values of at least 4) in at least half of the samples were removed (293 out total 608 endogenous genes were filtered out). The remaining 315 genes were used to construct the PCA plot. Principal component analysis was performed using the "prcomp" function within the stats package version 3.6.0 and visualised using the ggbiplot package version 0.55. For the combined *ex vivo*- and culture-derived cell PCA, a culture signature was derived by performing pairwise comparisons (two-tailed t-test with Benjamini-Hochberg correction of p-values) of all culture versus all *ex vivo* populations. 110 genes with adjusted p values <0.05 (the 'culture signature') were excluded from further analysis. The remaining 210 genes were used to construct the PCA plot (Figure 4.5).

4.2.5 Generation of single cell RNA-Seq data

Single human PBMC or BMDC were index-sorted into 96 well round-bottom plates containing 2µl cold RNA lysis buffer (RNase-free water, 2U/µl RNase inhibitor and 0.2% Triton X-100, Sigma) (BM progenitor plates) or SMARTer Dilution buffer (SMARTer Kit, Fluidigm) with the addition of 2U/µl RNase inhibitor (BM precursor

and DC plates). Plates were immediately centrifugated at 500xg for 1 minute, frozen on dry ice then stored at -80°C. Each plate included 2 controls; one blank and one well containing purified mouse RNA. The reverse transcription was performed using an adapted Smart-seq2 protocol (Picelli et al., 2014). Briefly, modifications included 21 PCR cycles and duplicate Ampure clean-up steps, following cDNA generation. The library prep was performed using the Nextera XT DNA Library Prep Kit. The Illumina HiSeq 4000 platform was employed to generate paired-end reads (75bp x 2). An average of 1.5 million reads were retrieved for each cell from the BM CD34⁺ progenitors and 2.3 million per cell for the BM CD34^{med} precursors and DC.

4.2.6 Processing of single cell RNA-Seq data

Reads were trimmed based on quality with Trimmomatic v 0.36 (Bolger et al., 2014). Bases with quality scores below Q10 (inferred base call accuracy below 90%) were trimmed and reads shorter than 60bp were dropped. The remaining reads were aligned with the STAR mapping algorithm v 2.4.0 (Dobin and Gingeras, 2015) to the human reference genome version GRCh38.p7 (GENCODE release 25), supplemented with External RNA Controls Consortium (ERCC) spike-in controls. The files were converted from SAM format to the more compressed BAM format with SAMtools v 1.3 (Li et al., 2009). The count tables were obtained using HTSEQ v 0.6.1 (Anders et al., 2015). ENSEMBL IDs were converted to HGNC gene names using biomaRt v 2.30.0 (Durinck et al., 2015).

4.2.7 QC and normalisation of scRNA-Seq data

Further analysis of the data was undertaken in R v 3.3.3 and Rstudio v 1.0.143. The scater R package v 1.2.0 was used to perform cell and gene QC and filtering (McCarthy et al., 2017). To remove technical outliers with poor coverage, genes with less than 2 counts in at least 2 cells were filtered out. Outlier cells were further removed based on number of total features, total counts and percentage of counts derived from ERCC spike-ins and mitochondrial genes. The normalisation was performed with the RUVg method (Risso et al., 2014) combined with counts per million (CPM) adjustment for library size and log transformation [$\log_2(\text{CPM}+1)$] for all downstream analyses. Only the genes annotated as protein coding in the “gene_type” column of the GENCODE reference genome GTF file were retained. To minimise the effect of cell-division cycle on the clustering performed in future steps,

the genes identified to play a role in cell cycle were downloaded from the supplementary materials provided by Macosko et al. (2015) and removed from all our analyses.

4.2.8 Clustering of scRNA-Seq data

Clustering was performed on the remaining genes using the Single-Cell Consensus Clustering (SC3) R package v 1.3.18 (Kiselev et al., 2017). The SC3 tool requires the k number of number of clusters to be specified by the user. A range of clusters (2 to 15) were visualised and studied for each of the datasets. The output from the “sc3_estimate_k” function guided the minimum number of clusters to be considered for each of the datasets.

Heatmaps with marker genes were generated with SC3. The area under the receiver operating characteristic curve (AUROC) and p-values assigned by a Wilcoxon signed rank test and corrected using the Holm method were used to define the marker genes (thresholds for statistics are stated in the figure legends). Clusters were annotated based on the top statistically significant marker genes from the SC3 output, as well as FACS phenotype and culture output.

4.2.9 Dimensionality reduction of scRNA-Seq data

The tSNE technique for dimensionality reduction was used to visualise the clusters. First, SC3 gene filter was applied to further remove genes with low expression, as well as ubiquitously expressed genes that do not contribute to clustering. The remaining genes were used for tSNE analysis with the Rtsne package v 0.13. An initial PCA step was introduced to reduce dimensionality and eliminate noise. Top principal components accounting for most variance (25-35%) were retained for the tSNE algorithm (the number of PCs is stated in the legend for each plot). Graphics were generated with the ggplot2 package v 3.0.0.

Diffusion maps were used to infer a pseudo-temporal ordering and reconstruct lineage branching. All protein coding genes that are not known to play a role in cell cycle were used in the diffusion map calculation with the destiny tool v 2.14.0 (Angerer et al., 2016) in R v 3.6.0. An initial PCA step was employed to reduce noise, and PCs accounting for most variance (total of approximately 40% for both datasets)

were retained for destiny. PCA was performed using the `prcomp` and `princomp` functions from the `stats` R package. Diffusion components 1-3 were used for trajectory tracing with `slingshot` v 1.2.0 (Street et al., 2018) and visualized on 3D plots. Graphics were generated with the `rgl` package v 0.100.19.

4.2.10 Generation of mass cytometry data

Pre-conjugated antibodies (Fluidigm), purified antibodies conjugated to their respective lanthanide metals using the Maxpar antibody labelling kit (as per manufacturer's instructions; DVS Sciences) or fluorophore-conjugated primary with anti-fluorophore metal-conjugated secondary antibodies were used for surface or intracellular staining (Appendix B).

Healthy control CD45⁺Lineage⁻ (CD3,19,20,56,161) PBMC (3×10^6 cells) or BMMC (1.5×10^6) were FACS purified into 1mL CyTOF staining buffer (PBS plus 2% FCS). Cell staining was performed at room temperature in a final staining volume of 100 μ l. Centrifugation was performed at 500xg for 5 minutes. 'Barcoding' of PBMC and BMMC samples was achieved by staining with 0.5 μ g anti-CD45-Irr115 or anti-CD45-89Y, respectively, (30mins) in CyTOF staining buffer before washing twice in PBS. Barcoded PBMC and BMMC were combined before addition of 2.5 μ M cisplatin for 5 minutes in PBS for live/dead cell discrimination, then washed promptly in CyTOF staining buffer. Successive primary and secondary surface staining was performed using approximately 0.5 μ g of each antibody in CyTOF staining buffer (30mins) before washing twice with PBS. The cells were fixed in 500ml eBioscience fixation buffer (eBioscience FoxP3 fix perm kit) with the addition of 500 μ l of 3.2% formaldehyde (final concentration 1.6%) and incubated for 30 minutes, before washing twice with eBioscience perm buffer. Cells were stained successively in perm buffer for 1hr each with intracellular primary and secondary antibodies then washed twice with PBS. Cells were resuspended in 500 μ l 250nM Iridium in PBS (final concentration 125nM) and 500 μ l 3.2% formaldehyde (final concentration 1.6%) and incubated for 1hr, before centrifugation and resuspension in 500 μ l CyTOF wash buffer for overnight storage at 4 °C. Prior to CyTOF acquisition, cells were washed twice in 200 μ l MilliQ water (800xg for 8 minutes), counted, diluted to a maximum final concentration of 0.55×10^6 /mL in MilliQ water and filtered through a 40 μ m filter (BD). EQ beads were

added (10% by volume) and 1.5×10^6 cells were acquired on the Helios mass cytometer running CyTOF software v 6.7.1014.

4.2.11 Mass cytometry data analysis

Within the CyTOF software, the resultant flow cytometry file (.fcs) was normalised against the EQ bead signals and randomised for a uniform negative distribution. FlowJo software was used to deconvolute live, lineage(CD3,19,20,56)⁻HLA-DR⁺ PB or BM cells by manual gating. For diffusion maps and lineage tracing (Figure 4.14), cells were down-sampled using random sampling within FlowJo, to select a total of 14,000 cells consisting of up to 500 or 1,000 cells per progenitor/precursor or mature cell population, respectively: CD33⁺GMP (300), CD33⁻GMP (200), CD123^{low/med}GMP (298), CD123^{hi}303/4^{low} (499), CD2+pDC (490), pDC (490), early pre-DC2 (498), pre-DC2 (491), CD5⁻DC2 (498), CD5⁺DC2 (800), early pre-DC1 (500), pre-DC1 (254), cDC1 (800), pre-DC3/mono (500), pre-DC3 (298), CD14⁻DC3 (498), CD14⁺DC3 (1000), pre-mono (500), mono (999). Further analysis was undertaken in R version 3.6.0. Diffusion map calculation was performed with the destiny tool v 2.14.0 (Angerer et al., 2016) using log₂-transformed values for the following CD markers: 14, 16, 123, 11b, 116, 303, 304, 2, 38, 10, 33, 11c, 90, 141, 34, 88, 117, 1c, 5, 15 and CLEC9A, AXL, SIGLEC6, SIRPA, IRF4, IRF8, FCER1A, BTLA and FLT3. 3D graphics were produced with the rgl package v 0.100.30.

For combined PB and BM progenitor, pre-DC and DC/monocyte analysis (Figure 4.15), combined lineage-HLA-DR⁺ cells were down-sampled to select 75,000 cells consisting of 20,000 CD11b⁺CD14⁺ monocytes, 4,000 CD11b⁺CD16⁺ monocytes and 50,000 non-monocyte cells. The concatenated .fcs file was subjected to tSNE dimensionality reduction with perplexity 30 from 1000 iterations using CD markers 14, 16, 123, 11b, 116, 303, 304, 2, 38, 10, 33, 11c, 90, 141, 34, 88, 117, 1c, 5 and CLEC9A, AXL, SIGLEC6, SIRPA, IRF4, IRF8, FCER1A, and BTLA. tSNE plots and marker expression heat plots were generated in ggplot2 R package using tSNE coordinates exported from FlowJo.

4.3 Results

4.3.1 Evaluation of mature cDC2 heterogeneity in human blood

cDC2 heterogeneity in human has been recently described in the literature, however, specific antigens for the isolation of the blood cDC2 subsets were unknown at the start of this project. Flow cytometry experiments undertaken by Dr Urszula Cytlak (Human DC Lab) identified markers able to split the cDC2 compartment in blood. The bulk of cDC2s were marked by high expression of CD2 and CD1c, distinct from CD141^{high} cDC1s, and CD123⁺ and CD303/304⁺ pDC. The bulk of cDC2 was split into subsets using markers BTLA, CD5, and CD14 (Figure 4.2 A). BTLA divided the blood cDC2 population into BTLA⁺ DC2 and BTLA⁻ DC3. BTLA⁺ DC2 was further split into CD5⁺ and CD5⁻ populations, while BTLA⁻ DC3 were subdivided into CD14⁺ and CD14⁻ fractions. Expression of CD163 (a marker for the monocyte/macrophage lineage), negatively correlated with that of BTLA (Figure 4.2 B). With this in mind, the two cDC2 subtypes in peripheral blood were defined as CD163⁻ BTLA⁺ CD5^{+/-} DC2 and CD163⁺ BTLA⁻ CD14^{+/-} DC3 (Figure 4.2 C) with CD5 and CD14 marking the poles of the phenotypic continuum.

NanoString gene expression analysis of 10,000 *ex vivo*, FACS-sorted cDC1s, pDCs, cDC2s, and classical monocytes, confirmed the presence of cDC2 heterogeneity. Mirroring that seen in phenotypic analysis, a transcriptional continuum was evident with CD5⁺ DC2s, clustering near the cDC1s, and CD14⁺ DC3s, located closer to the classical monocytes (Figure 4.3).

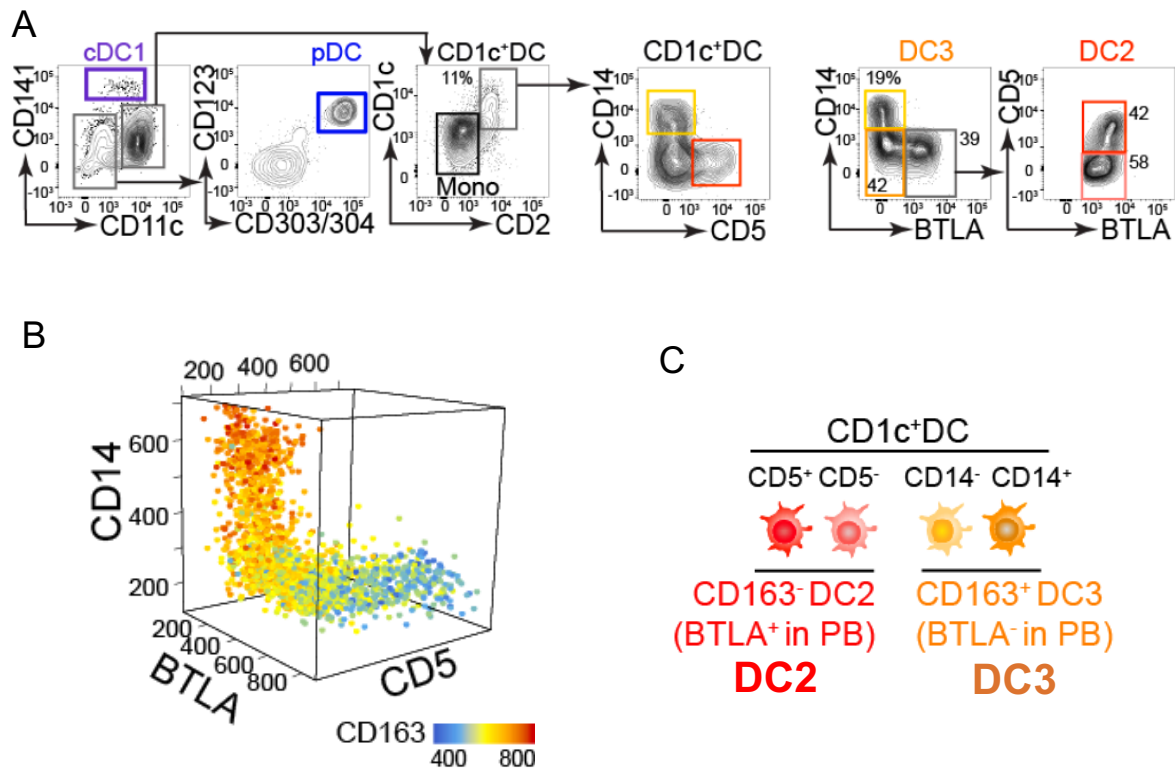


Figure 4.2. Phenotyping of DC populations in human peripheral blood (Cytlak and Resteu, et al., 2020).

A. Gating strategy for flow cytometry analysis. HLA-DR positive and lineage (CD3,16,19,20,34,7) negative cells were used for analysis. cDC2 were defined by CD2 and CD1c expression, distinct from CD141⁺ cDC1, CD123⁺CD303/304⁺ pDC and monocytes (mono). BTLA expression categorised DC2 (BTLA⁺) with (red) or without (pink) CD5; and DC3 (BTLA⁻) with (yellow) or without (orange) CD14. Numbers on the plots represent percentage of the parent gate.

B. 3D plot of the expression of CD14, CD5 and BTLA across the cDC2 (CD1c⁺DC) compartment, as measured by flow cytometry. Cell colours represent CD163 antigen expression.

C. Definition of cDC2 subpopulation phenotypes: CD163⁻CD5^{+/-}(BTLA⁺ in blood) DC2 and CD163⁺CD14^{+/-}(BTLA⁻ in blood) DC3.

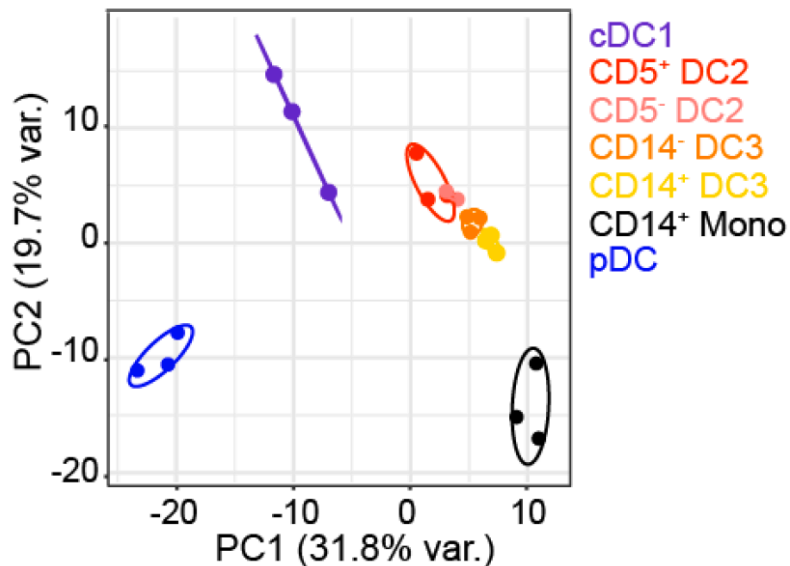


Figure 4.3. Principal component analysis of mRNA expression from FACS-purified cDC1s, cDC2s (DC2 and DC3), pDCs and classical monocytes from peripheral blood, as determined by the NanoString assay.

PCA was performed based on 315 genes with expression levels detectable via the NanoString Human Immunology V2 panel plus 30 custom DC-related genes.

4.3.2 Interrogation of DC haematopoiesis *in vitro*

Following the successful identification and isolation of cDC1, pDC, DC2, and DC3 in human peripheral blood, this project aimed to identify the developmental pathways giving rise to the different cDC2 subpopulations. The *in vitro* culture system described in Chapter 3 was used to interrogate the progenitors and precursors for their potential to generate DC subsets. Firstly, to identify DC2 and DC3 in culture output, the phenotyping of culture-derived DCs was performed via conventional flow cytometry, as for the peripheral analysis described above. Due to the lack of BTLA protein expression in culture, the gating strategy was adjusted (Figure 4.4). CD2 and CD1c defined the bulk of the cDC2s. Within this compartment, CD163 was exclusively expressed by CD14⁺ cells, while the CD5⁺ DC2 population was contained within the CD14⁻ gate. Marker CD14 was therefore used to split cDC2s into (CD14⁻) DC2 and (CD14⁺) DC3. This gating was reproducible for cultured, blood and bone marrow cells (Figure 4.4). Akin to the cultured cells, the lack of BTLA expression was also noted in human bone marrow DCs.

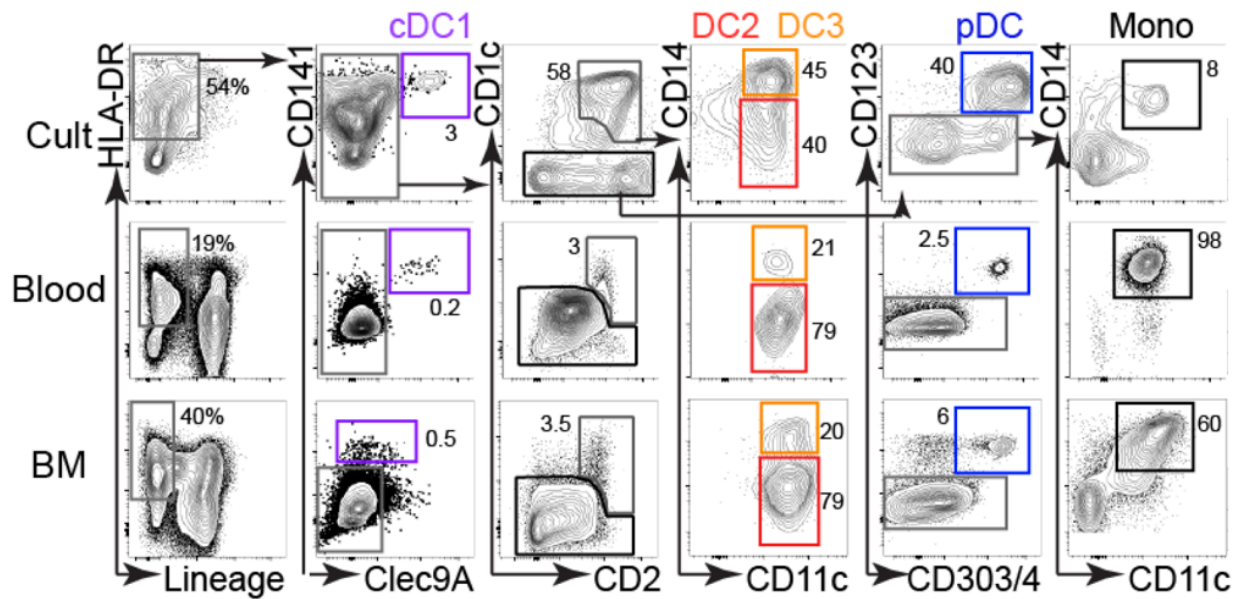


Figure 4.4. Gating strategy used to identify in vitro-derived DCs and monocytes generated from BM CD34⁺ progenitors, shown with blood and bone marrow (BM) for comparison (Dr Venetia Bigley and Dr Urszula Cytlak).

A minimum of two antigens were used to define each of the populations:

CD141⁺CLEC9A⁺ cDC1, CD2⁺CD1c⁺ cDC2 divided as CD14⁻ DC2 and CD14⁺ DC3, CD123⁺CD303⁺CD304⁺ pDC and CD14⁺CD11c⁺CD1c⁻ CD2⁻ monocytes. Numbers on the plots represent percentage of the parent gate.

NanoString gene expression analysis was used to confirm the identity of the culture-derived DC2s and DC3s defined by CD14 expression. Figure 4.5 shows the joint analysis of the culture-derived and peripheral blood DCs and classical monocytes. All cultured subsets clustered close to their *ex vivo* equivalents. The cultured DC2 and DC3 subsets were more polarised towards, but remained distinct from cDC1s or monocytes, respectively, compared to their blood counterparts.

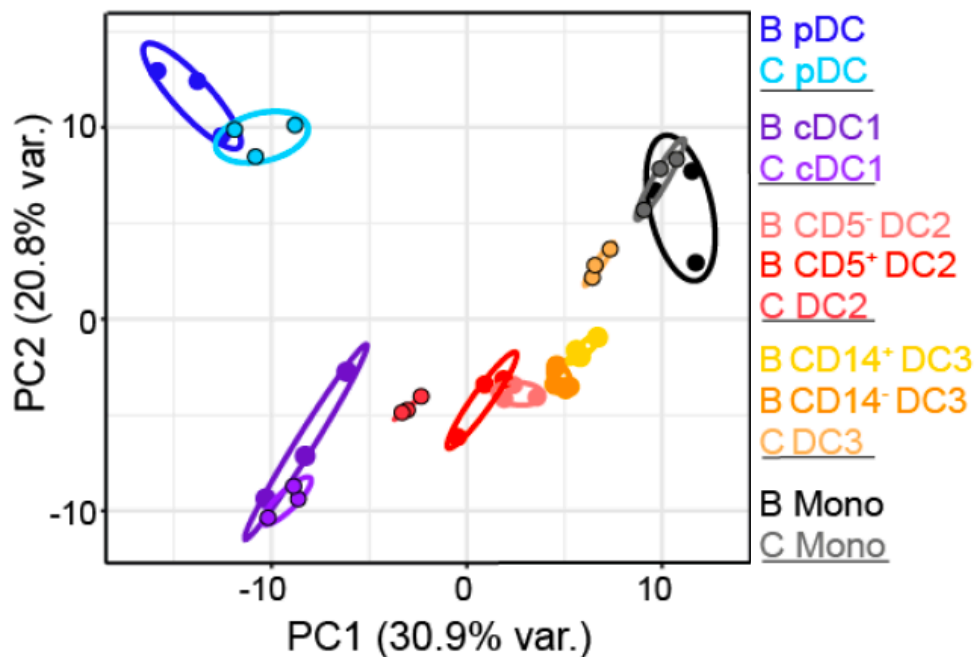


Figure 4.5. Principal component analysis of mRNA expression from FACS-purified primary cDC1, cDC2 (DC2 and DC3), pDC and classical monocytes from human peripheral blood and in vitro culture.

The blood subsets are displayed in darker colours and the in vitro-generated samples are shown in lighter colours marked with a black outline. NanoString Human Immunology V2 panel supplemented with 30 custom DC-related genes was employed to profile the gene expression. PCA was performed based on all genes with detectable expression after removal of a culture signature (210 remaining genes).

Next, the culture system, able to produce bona fide cDC2 subsets, was used to interrogate DC haematopoiesis *in vitro*. Cells originating from the phenotypic spaces known or hypothesized to contain progenitors were FACS-purified and placed in the OP9 culture system. The output was assessed via flow cytometry, following 14 days of culture. Figure 4.6 shows a schematic of haematopoiesis, overlaid by the sampled phenotypic spaces and their output in culture. In culture, HSCs produced DC2s and DC3s at a similar ratio to that observed in blood. The DC2 potential was identifiable in the HSC, LMPP, CD123⁺ GMP, and was highest in the gate identified as pre-DC2. DC3 were progressively enriched through the HSC, CD33⁺ GMP, and the pre-DC3 gates. The HSC, LMPP, CD123⁺ GMP also gave rise to a significant number cDC1s

and pDCs. Despite the culture conditions being designed to inhibit monocyte generation, some monocyte potential was observed. This was seen in the populations giving rise to DC3, but most enriched in a phenotypically distinct precursor population giving rise primarily to monocytes. Detailed gating strategies and culture output results are shown in Appendix C (CD34⁺ progenitors) and Appendix D (CD34^{med/-} precursors and DCs/monocytes).

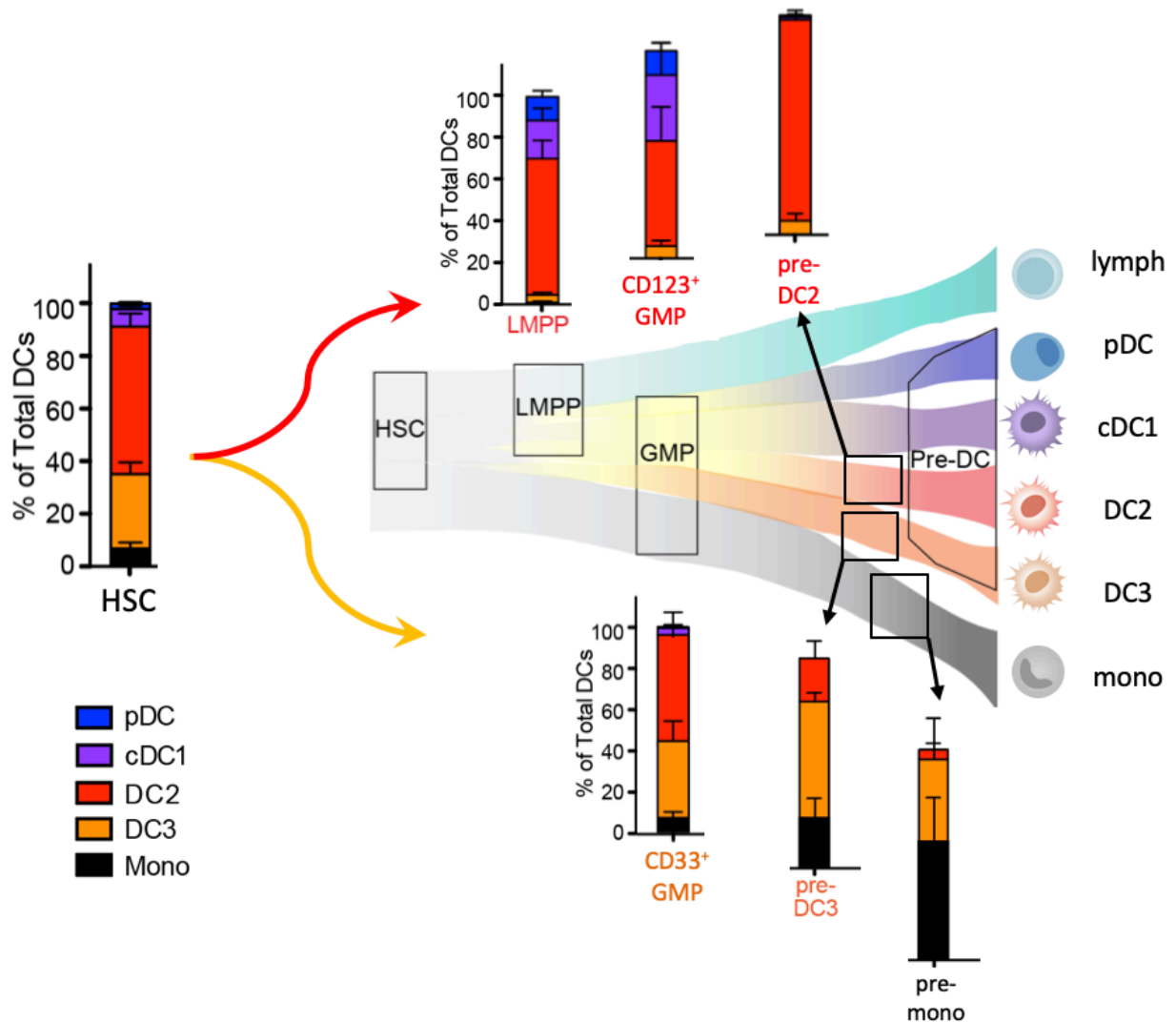


Figure 4.6. Schematic of the development of haematopoietic lineages from bone marrow, showing the DC and monocyte culture output of the phenotypic spaces as bar charts.

The bulk culture output of FACS-purified bone marrow CD34⁺ progenitors was assessed with flow cytometry following 14 days of culture. The proportion of generated DC subsets and monocytes is expressed as % of the total cells captured by all DC and monocyte gates. $n=3-9$ healthy donors for each population. Bars represent mean+SEM.

4.3.3 Single cell transcriptomics of human bone marrow progenitors reveals two pathways of DC development distinguished by differential *IRF8* expression

The cells from the examined phenotypic spaces in healthy human BM were subjected to scRNA-Seq, in order to examine their transcriptomes at single cell resolution. The first scRNA-Seq dataset (also referred to as the “progenitor” dataset) contained CD34⁺ bone marrow progenitors. These cells were index-sorted directly into 96 well plates as shown in Appendix C, allowing the concurrent interrogation of cell-specific surface phenotype and transcriptomic analysis.

Hierarchical clustering of the single cell transcriptomes alone (without considering surface phenotype) from the CD34⁺ bone marrow compartment revealed the presence of distinct clusters, roughly corresponding to the annotation from FACS-guided index sorting. The clusters were annotated based on marker genes as determined by scRNA-Seq and related to the information from index sorting and culture output. HSCs and MPP were marked by vasopressin (*AVP*) expression, and were index sorted from the CD34⁺ CD45RA⁻ fraction of the bone marrow.

Megakaryocyte/erythroid progenitors (referred to as mega/erythro) were identified as the erythroid transcription factor *GATA1*-expressing cluster and were sorted as CD33⁻ CMP. Early myeloid cells formed a cluster of mixed cells from CD123⁻ CD33⁺ GMP and CD33⁺ CMP sort gates, marked with high *MYC* expression, a gene known to drive cell proliferation. LMPP formed a cluster made up almost exclusively of cells from the FACS-purified LMPP gate. The monocyte and neutrophil progenitors clustered closely and were both marked by myeloid cell associated genes, such as myeloperoxidase (*MPO*) and CD33 antigen expression. The monocyte progenitors, were distinguishable by higher lysozyme (*LYZ*) gene expression and were mainly derived from the CD123⁻ CD33⁺ GMP gate, while neutrophil progenitors expressed neutrophil elastase (*ELANE*) and belonged to the CD33⁺ CMP sort gate. The cluster enriched for monocyte genes also displayed a low/medium level of *IRF8* expression. Progenitors with *in vitro* cDC1, pDC, and DC2 potential, also referred to as DC progenitors, formed a cluster adjacent to the LMPPs. This cluster was marked with high *IRF8* expression and contained cells from the phenotypic spaces of CD123⁻ CD33^{low} GMP, CD123^{low} GMP, CD123^{med} GMP.

tSNE, a dimensionality reduction and clustering technique, was used to visualise the scRNA-Seq dataset (Figure 4.7). The cells on the tSNE plots were annotated using

the FACS gates (Figure 4.8 A), or the 10 clusters produced by hierarchical clustering (Figure 4.8 B). Both annotations correlated with the tSNE clusters. Distinct clusters were noted for HSC/MPP, Mega/erythro, early myeloid cells, adjacent to the monocyte and neutrophil progenitors, and DC progenitors, adjacent to the LMPP cluster. The expression of *IRF8*, a top marker gene for the DC cluster (Figure 4.8 C), defined the DC cluster on the tSNE. The monocyte cluster displayed intermediate *IRF8* expression, lower than that of the DC cluster, and higher than the HSC/MPPs (Figure 4.8 D).

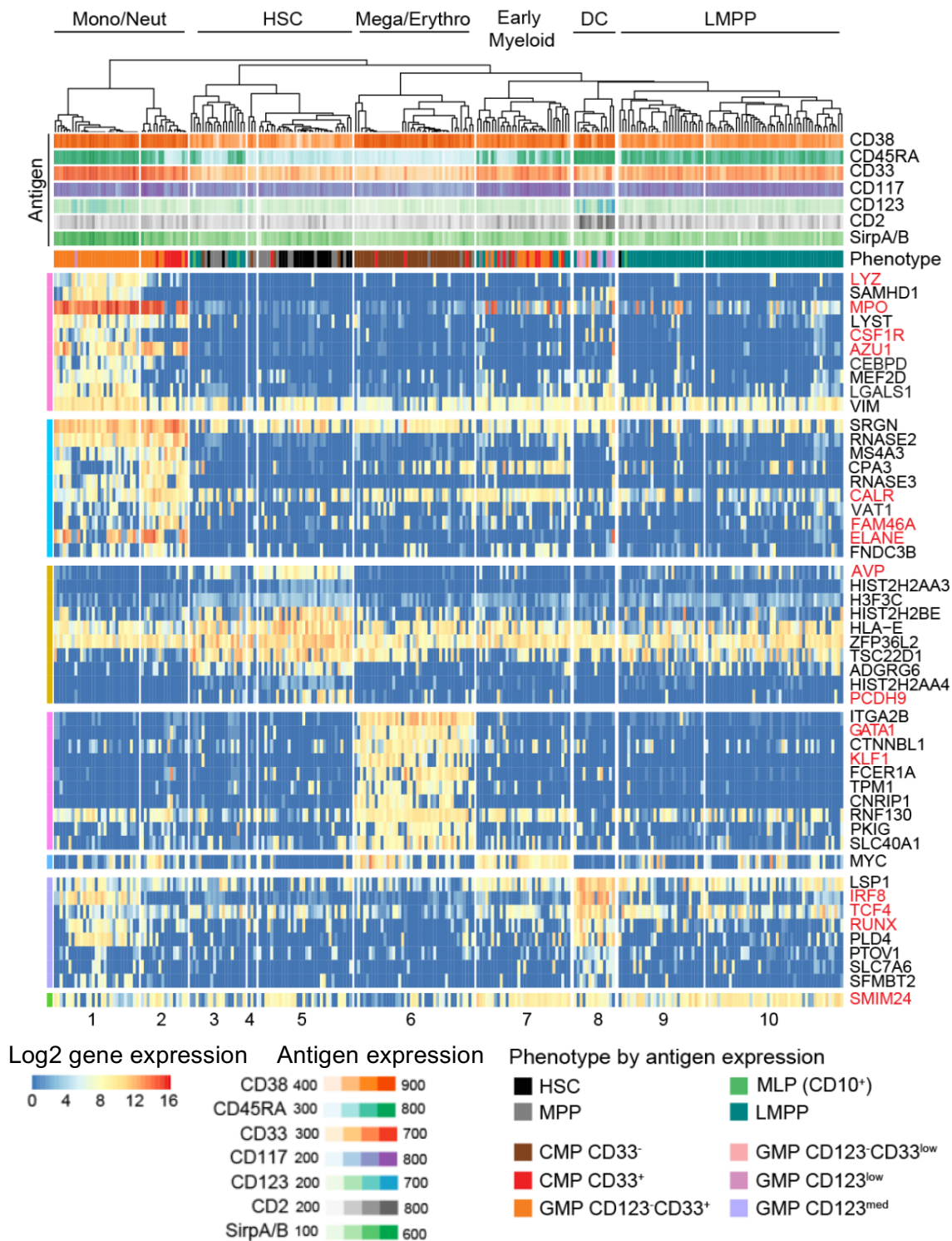


Figure 4.7. Hierarchical clustering of single cell transcriptomes of $CD34^+$ progenitors isolated from BM showing marker genes that identify 10 clusters of all progenitors.

The colours of the heatmap indicate normalised log2 marker gene expression, as determined by scRNA-Seq (p -value < 0.01 , AUROC > 0.6). Antigen expression, as determined by FACS is displayed above the cells, along with the phenotype.

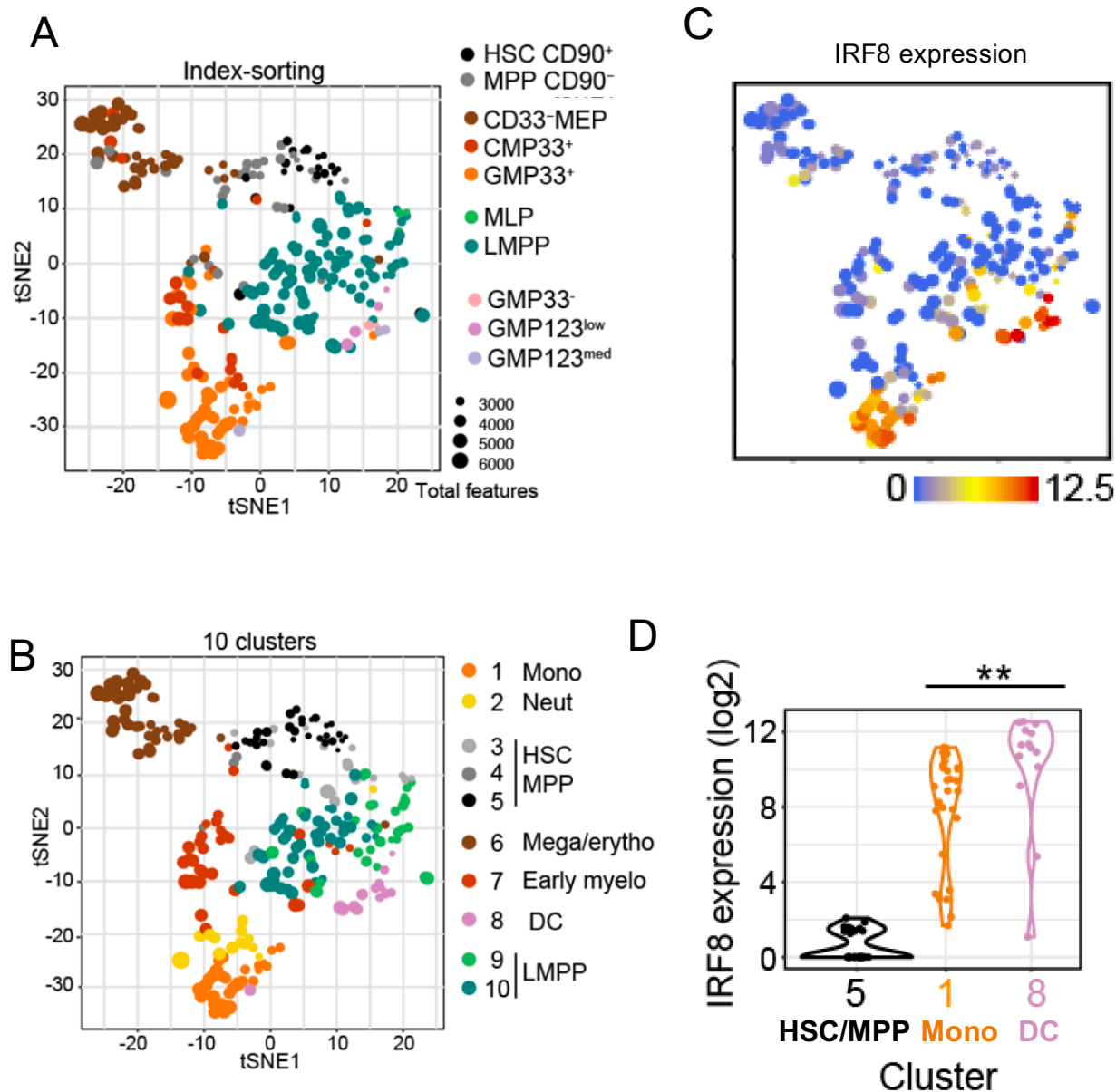


Figure 4.8. Visualisation of single cell transcriptomes of CD34⁺ progenitors.

A - C. tSNE of 262 single cell transcriptomes of CD34⁺ progenitor subsets annotated by the gate of origin from index-linked flow cytometry (A) or 10 clusters from hierarchical clustering (B), or displaying log₂ IRF8 gene expression (C).

D. Violin plot of differential IRF8 expression in clusters 1 (GMP33⁺), 5 (HSC/MPP) and 8 (DC).

Next, diffusion maps, together with Slingshot pseudotime were used to infer the developmental trajectories of the single cells (Figure 4.9 A). HSCs were located at the centre of the diffusion map, as they represented the earliest progenitors, common for all examined cell types. The cells belonging to the DC cluster formed a branch of the diffusion map, rooting from LMPP. The monocyte and neutrophil precursors formed two separate trajectories with early myeloid cells at the base. Mega/erythro progenitors produced a fourth developmental branch.

Together, the *in vitro* culture experiments and the scRNA-Seq data were consistent with a model whereby cDC1, pDC, and DC2 develop through an IRF8^{high} pathway, traversing the phenotypic spaces of LMPP and the CD123⁺ fraction of GMP, while DC3 potential segregates with monocyte potential through a different, CD33⁺ part of GMP, expressing lower IRF8 levels (Figure 4.9 A and B).

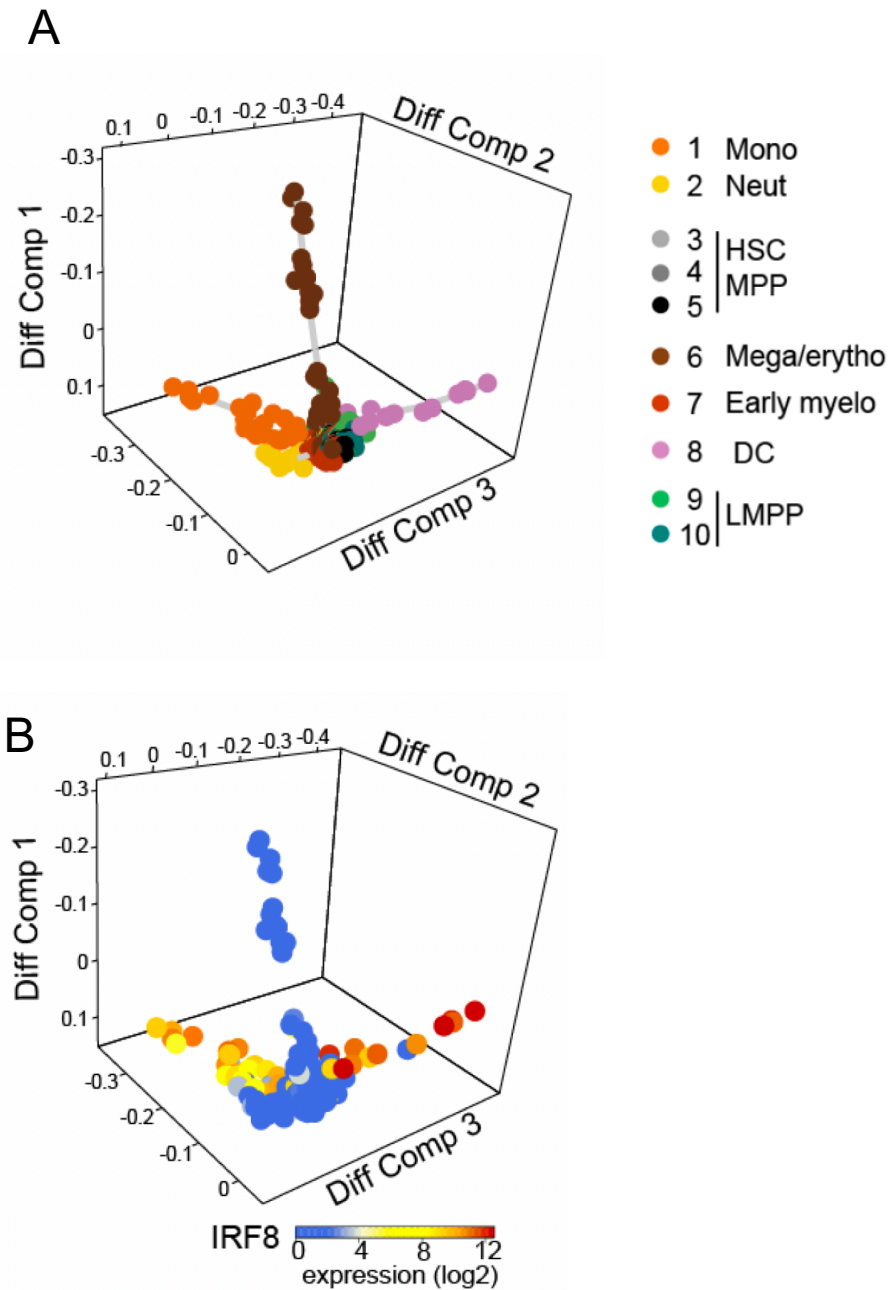


Figure 4.9. 3D diffusion map and lineage tracing of 262 single cell transcriptomes from $CD34^+$ progenitor subsets.

A. Diffusion map showing clusters obtained via hierarchical clustering of the dataset. Top 3 diffusion components, displayed as the x, y and z axes of the 3D plot, were used as input for a pseudotime analysis. Developmental trajectories, inferred with the Slingshot R package (Street et al., 2018), are depicted in grey.

B. Diffusion map displaying \log_2 IRF8 gene expression. Diff Comp, diffusion component.

4.3.4 *IRF8*^{high} and *IRF8*^{low} pathways connect bone marrow progenitors with mature DC subsets

Having identified the two DC developmental pathways within the CD34⁺ fraction of the bone marrow, the analysis was extended to potential DC precursors. To achieve this, a second scRNA-Seq experiment was performed on CD34 intermediate or negative populations defined *in vitro* as DC precursors generating the “precursor” dataset (the detailed gating strategy and culture output of these populations are described in Appendix D). Fully differentiated bone marrow DC subsets and monocytes were also included.

Hierarchical clustering of the transcriptomes of cells within the new dataset identified the presence of 15 clusters (Figure 4.10). The CD34^{med} and mature DC clusters were annotated based on marker genes and related to flow cytometric phenotype and *in vitro* culture potential. The precursors defined in previous experiments clustered adjacent to their mature counterparts. This was observed for all lineages, apart from cDC1, likely due to the rarity of precursors for this population *in vivo*.

Five clusters related to pDCs were identified: two mature pDC clusters and three pre-DC clusters (distinguished by CD34 protein expression). All pDCs clusters displayed high *IRF8* expression and were marked by *GZMB* and *SERPINF1*, identified as pDC markers by Villani et al. (2017).

Two clusters, each containing a subdivision of cDC2 were observed. Both expressed the cDC2 marker *FCER1A*, along with class II MHC genes, such as *HLA-DQB1* and *HLA-DQB2*. Notably, the DC2 cluster had distinguishably higher CD5 antigen expression. A number of early DC2 and early DC3 clusters were observed. One of the early pre-DC2 clusters was marked with *SIGLEC6* expression, a signature gene for the (AXL⁺) AS DC or pre-DC, described by Villani et al. (2017) and See et al, respectively. The early pre-DC3 clusters retained *MPO* expression, while also expressing MHC Class II genes.

Two monocyte-related clusters were found: a mature monocyte cluster with high *CD14* gene expression, and a pre-mono cluster. The pre-mono cluster retained some *MPO* expression, resembling the monocyte progenitors from the CD34⁺ dataset (Figure 4.7).

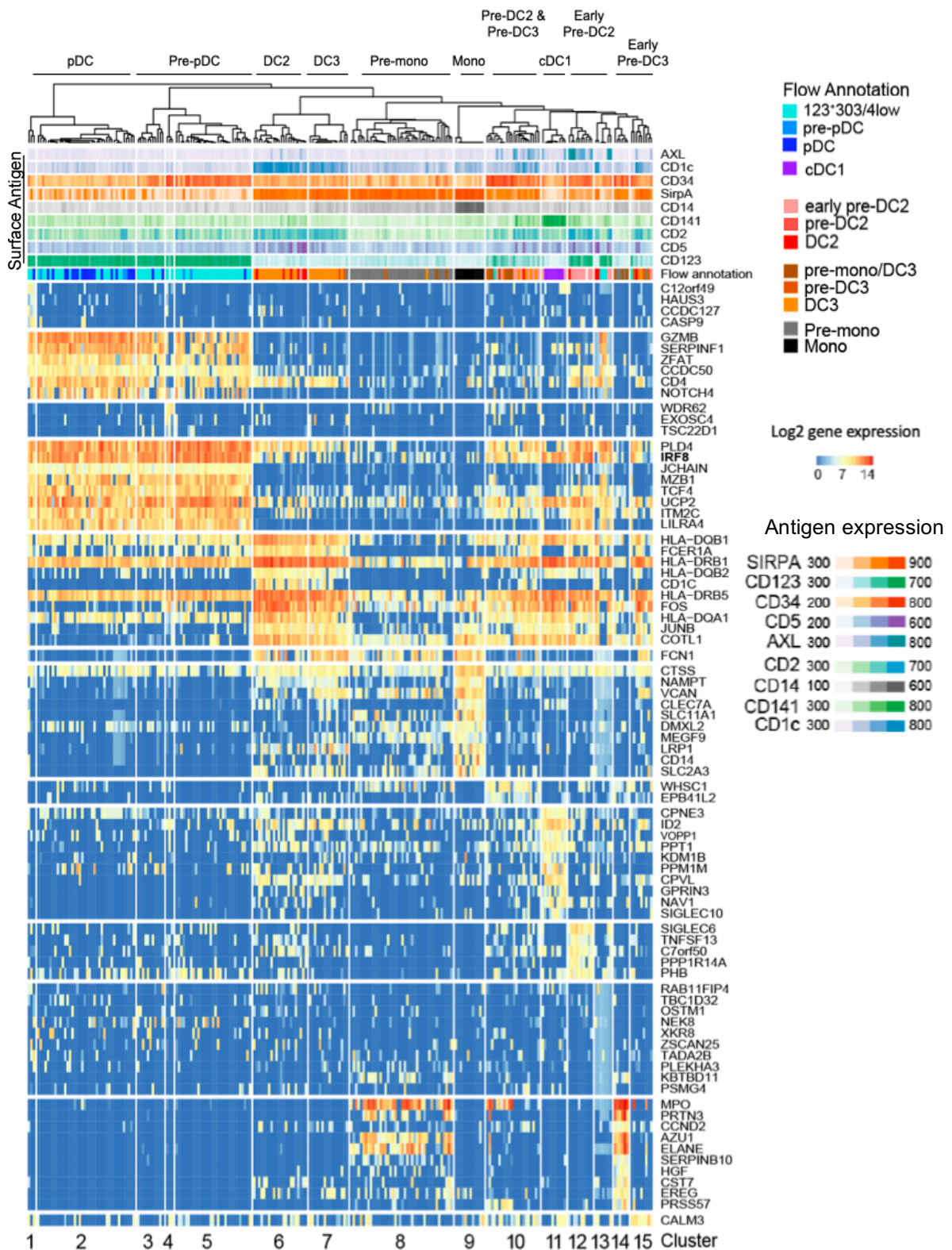


Figure 4.10. Hierarchical clustering of single cell transcriptomes of CD14⁺ monocytes, DCs and precursors isolated from BM, showing signature genes that identify 15 clusters.

The colours of the heatmap indicate normalised log₂ marker gene expression, as determined by scRNA-Seq (p -value < 0.01, AUROC > 0.85). Antigen expression, as determined by FACS is displayed above the cells, along with the phenotype.

tSNE analysis of the single cell transcriptomes identified four poles, corresponding to pDC, cDC1, the two cDC2 subpopulations, and monocytes. This was reflected by both the phenotypic annotation (by flow cytometry) and the hierarchical clustering (Figures 4.11 A and B, respectively). The poles were linked by progenitor populations with preserved CD34 expression (Figure 4.12 A). The two hotspots of CD34 expression corresponded to the early pre-DC2 and early pre-DC3 cluster populations defined by the *in vitro* experiments, as well as hierarchical clustering of the scRNA-Seq data. The early pre-DC2s exhibited high *IRF8* expression and were located in the proximity to the DC populations derived from the IRF8-high pathway, including DC2, cDC1, and pDC (Figure 4.12 B). The early pre-DC3 clustered closely to pre-monocytes and pre-DC3 and displayed medium IRF8 levels (Figure 4.11 C).

Diffusion maps, in combination with pseudotime analysis were used to reconstruct the developmental trajectories of the DC populations (Figure 4.13). The resulting diffusion map showed 4 trajectories, corresponding to pDC, cDC1, monocytes, and cDC2, and resembled the tSNE graph. Consistent with their independent origin, DC2 and DC3 formed two distinct branches of the cDC2 trajectory. The lineage tracing with Slingshot linked early pre-DC2 to the DC2 fraction, and early pre-DC3 to the DC3 with subsequent phenotypic convergence of the trajectories to form a cDC2 population.

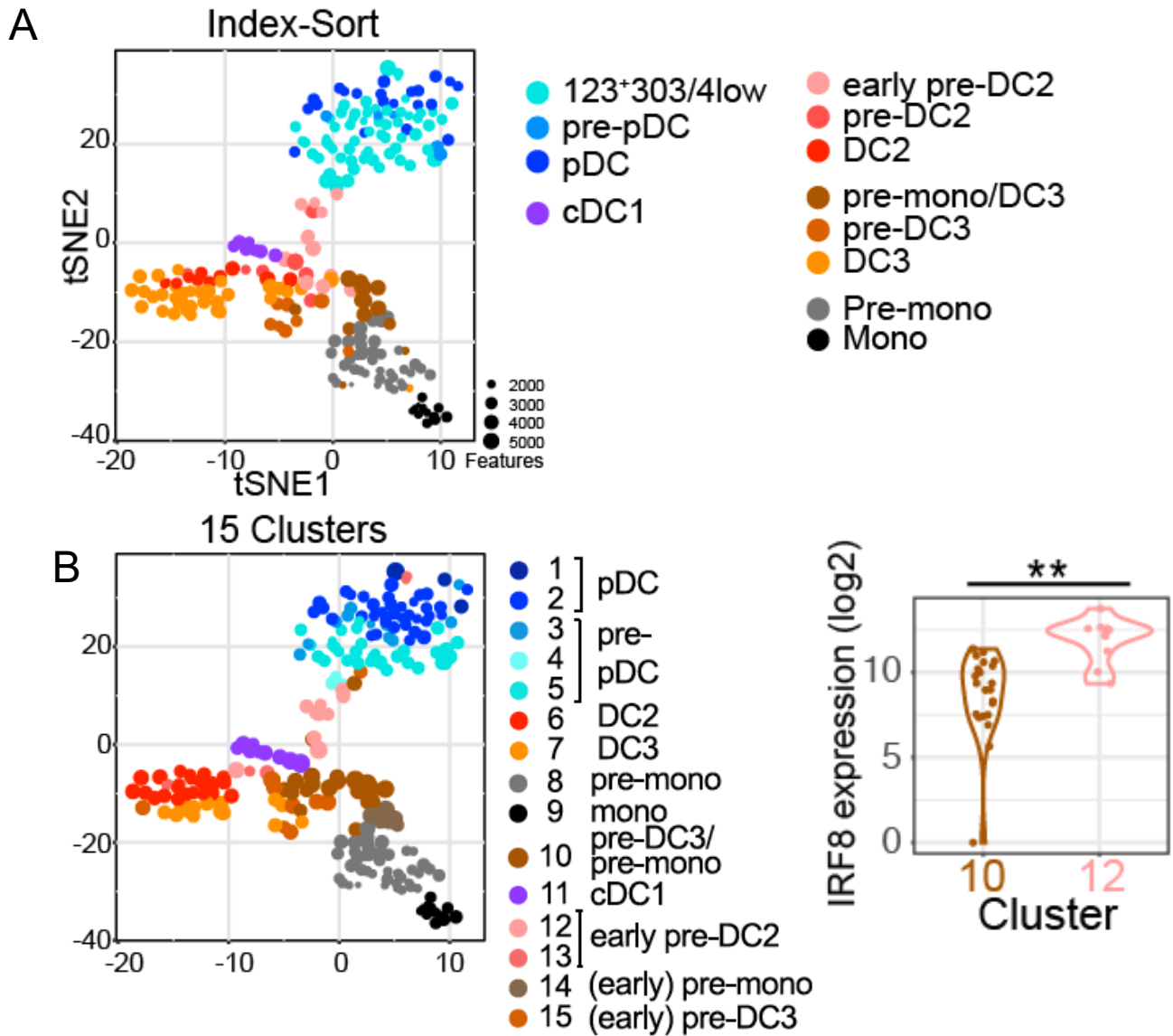


Figure 4.11. tSNE of 244 single cell transcriptomes sampled from mature DCs and classical monocytes and pre-DC populations of human BM.

A. Single cells annotated by the gate of origin from index-linked flow cytometry.

B. Cells annotated by 15 clusters generated from hierarchical clustering of the transcriptomes.

C. Violin plot of differential IRF8 expression in clusters 10 (pre-DC3/mono) and 12 (early pre-DC2), as determined by hierarchical clustering of single cell transcriptomics of pre-DC populations.

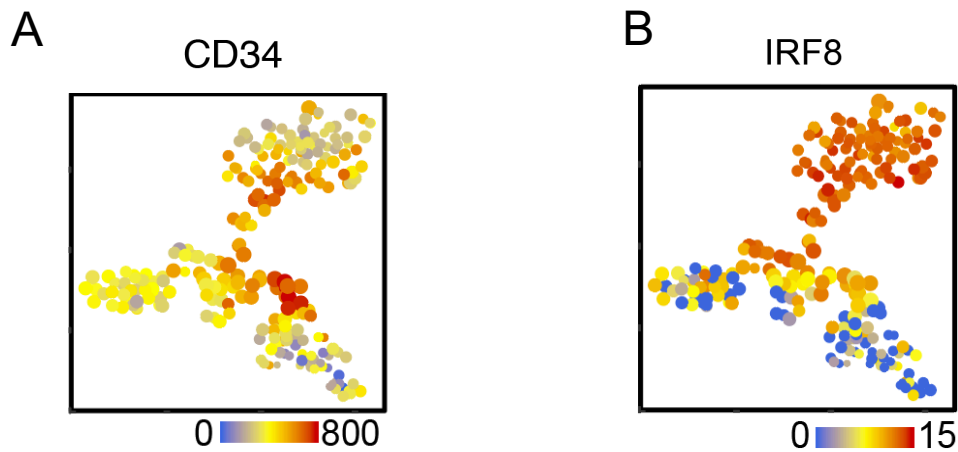


Figure 4.12. tSNE of 244 single cell transcriptomes sampled from pre-DC populations of human BM.

The heatmap shows CD34 antigen expression, as detected by FACS (A) and log₂ IRF8 gene expression, as determined by scRNA-Seq (B).

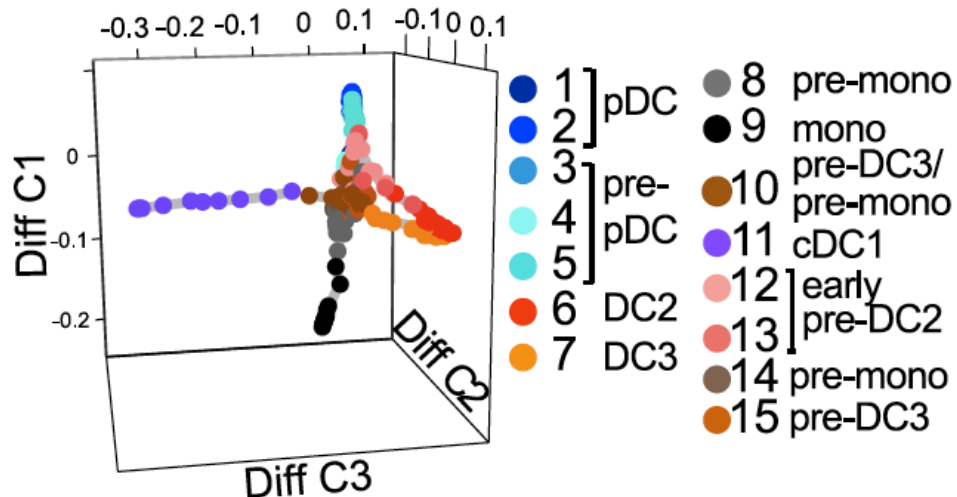


Figure 4.13. 3D diffusion map and lineage tracing of 244 single cell transcriptomes pre-DC populations of human BM annotated by 15 clusters from hierarchical clustering.

Diffusion map was performed following a principal component analysis, aimed at reducing noise. Top 3 diffusion components, displayed on the plot, were used as input for the pseudotime analysis. The inferred trajectories, originating from HSCs, are depicted in grey. Diff Comp, diffusion component.

4.3.5 Coupling the two pathways in bone marrow and peripheral blood via mass cytometry

Following the detection of the two DC pathways in the CD34^{high} and CD34^{med} compartments of the human bone marrow via scRNA-Seq, the next steps were to validate the link between these compartments, and couple the two pathways with mature cells in peripheral blood. Mass cytometry (CyTOF) was chosen for these experiments to allow the simultaneous analysis of blood and bone marrow with sufficient parameter capacity to analyse the phenotype of progenitors, precursors and mature subsets across the same platform.

To this end, a panel of 33 markers was designed, consisting of mature DC and monocyte markers, early DC lineage markers (including AXL, SIGLEC6, CD123, CD2, CD33, SIRPA), progenitor markers (including CD34 and CD117) and transcription factors IRF4 and IRF8, found intracellularly. PB and BM cells were stained with distinct CD45 antibody conjugates so they could be distinguished in subsequent analyses. Cells were then combined for subsequent experimental steps. Including initially only BM derived cells, diffusion maps, employed to infer pseudo-temporal ordering of cells and reconstruct lineage branching, showed four branches dominated by different cell types: GMP, cDC1, pDC, and monocytes (Figure 4.14 A). The cells showed appropriate antigen expression, cDC1 being marked by CLEC9A, pDC – by CD123, cDC2 subsets – by FCER1, SIRPA and CD2, with mutually exclusive expression of CD14 and CD5 (Figure 4.14 B). The earliest progenitors were concentrated within the uppermost points of the diffusion map (marked by CD34 expression, Figure 4.14 B) and were linked to the peripheral DC populations via arms of precursor cells. The DC3 trajectory ran parallel with and in close proximity to the branch dominated by monocytes, originating from CD33⁺ GMP. DC2s were located between the other DC subsets descending from CD123^{low/med} IRF8^{high} GMP and the IRF8^{low} DC3/monocytes. However, unlike mature pDCs and cDC1s, which retained IRF8 expression, differentiated DC2s downregulated this transcription factor (Figure 4.14 B).

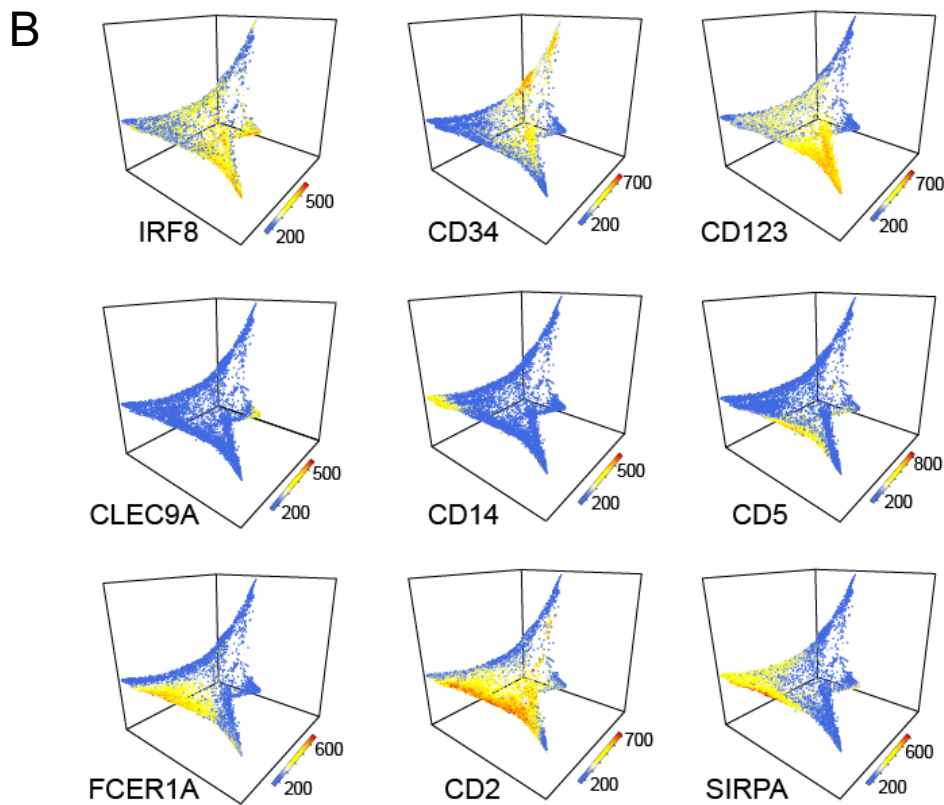
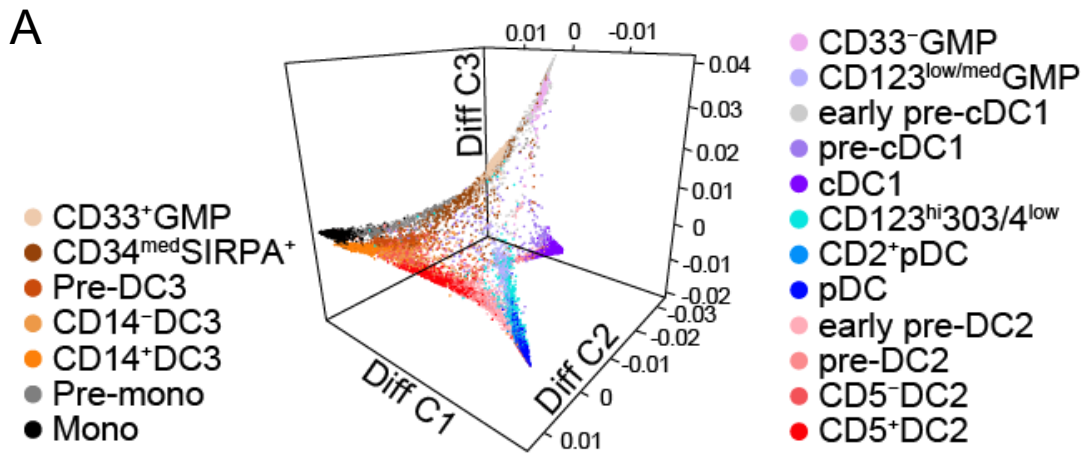


Figure 4.14. 3D diffusion map generated with mass cytometry data for 14,000 granulocyte-monocyte progenitors (GMP), precursor and mature DC/monocyte cells from human bone marrow.

A. Diffusion map showing cells color-coded according to the gating strategies depicted in Appendix Figures C (progenitors) and D (precursors and DC/mono). Diff C, diffusion component.

B. Heatmaps showing the expression of key markers across the diffusion map trajectories.

Next, to validate the link between the two pathways originating in bone marrow and peripheral blood DC/mono populations, BMMC and PBMC compartments were analysed together. The tSNE dimensionality reduction technique was employed to visualise the dataset. On the resulting tSNE plot, most bone marrow cells occupied a central position, while the cells derived from blood were found at the peripheries (Figure 4.15 A). A cluster containing predominantly CD34⁺ progenitors was formed of almost exclusively bone marrow cells. In addition to that, the BMMC sample contained precursors and most mature populations, with the exception of CD16⁺ monocytes (Figure 4.15 B). In keeping with flow cytometric analysis, the majority of the cells in blood were mature DC and monocytes, with the exception of AXL⁺ cells and a few progenitors. The two DC development pathways were apparent on the joint BMMC and PBMC tSNE plot, when guided by the IRF8 expression (Figure 4.15 C). The IRF8^{high} pathway progenitors, such as cells from the LMPP and CD33^{low} and CD123^{med} GMP compartments were close to mature pDC, CD5⁺ DC2, and were adjacent to the cDC1 cluster in tSNE space. In contrast, IRF8^{low} progenitors were located in the proximity of classical monocytes and the CD14⁺ DC3 (Figure 4.15 B).

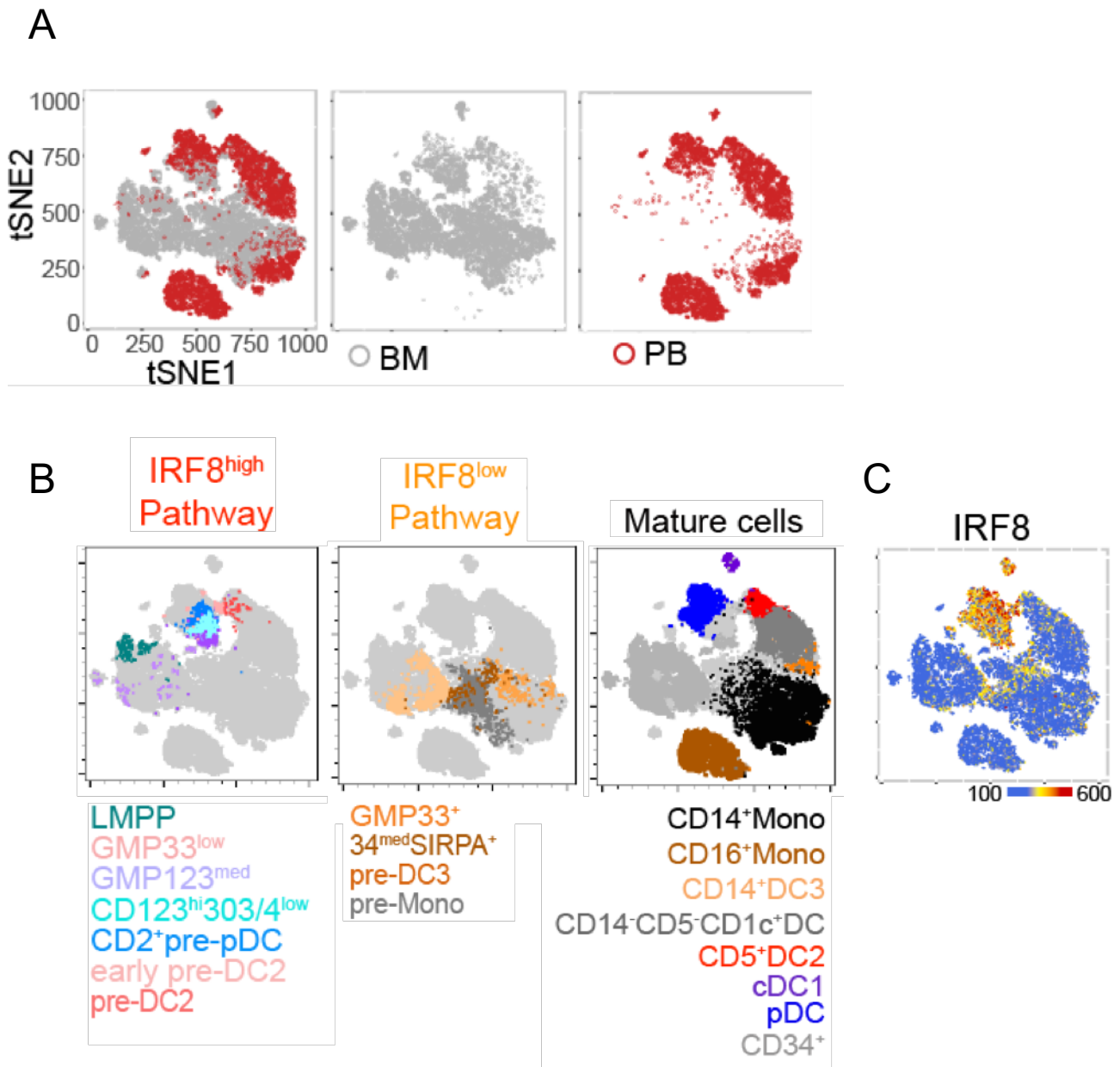


Figure 4.15. tSNE analysis of mass cytometry data for a total of 75,000 progenitors, precursors and mature DCs and monocytes from human PBMC and BMMC.

A. Plot depicting cells derived from bone marrow (grey) and peripheral blood (red), barcoded and combined prior to CyTOF analysis.

B. Progenitor and precursor cells from the IRF8^{high} and IRF8^{low} pathways, as well as mature DCs and monocytes, highlighted on the tSNE plot.

C. Heatmap of IRF8 expression superimposed on the tSNE.

4.4 Discussion

4.4.1 Transcription factor IRF8 defines two DC developmental pathways in human

Recent literature showed that DCs, originate from bone marrow progenitors and precursors with increasingly restricted potential. However, the identity of the pre-DC was unknown prior to the start of this project. The combination of several approaches employed in this chapter the mapping of two pathways of DC development. A culture system, able to support the growth of DCs and monocytes, was used to single out increasingly committed progenitor and precursor populations. Single cell transcriptomic analyses enabled the reconstruction of lineage branching within the progenitor and precursor populations to infer a pseudotemporal ordering of the cells, linking them to mature DC subsets. And finally, computational analyses of mass cytometry data connected the bone marrow compartment to peripheral blood.

Together, the work undertaken within this chapter revealed the presence of two distinct DC development pathways in human bone marrow, defined by expression levels of transcription factor IRF8. High IRF8 expression marked a CD123⁺ developmental pathway that gave rise to pDC, cDC1 and DC2, while the IRF8^{low} pathway produced DC3, along with classical monocytes. The discovery of the IRF8^{high} CD123⁺ pathway is in line with previous studies reporting the presence of unipotent progenitors of cDC1, cDC2, and pDC within the CD123⁺ compartment of GMP (Lee et al., 2017). However, earlier studies did not account for the presence of cDC2 subtypes, recently described in literature, and investigated the bulk CD1c⁺ cDC2 population.

This chapter shows that the two developmental pathways are the source of the blood cDC2s, as the DC2 and DC3 subsets develop via different routes. DC2, the subpopulation bearing higher resemblance to cDC1, originates within the IRF8^{high} pathway, while monocyte-resembling DC3 develops through the IRF8^{low} pathway. Importantly, the data support a model whereby DC3 develop independently from monocytes. While the two cell types shared a “pre-mono/DC3” phenotypic gate, separate immediate precursors for these lineages were captured within the bone marrow compartment, and linked to either mature BM monocytes and DC3 via lineage branching reconstruction based on scRNA-Seq. In addition, DC3 cells emerged before monocytes in progenitor cell culture experiments (data available within the following publication: Cylak and Resteu et al., 2020).

The different levels of IRF8 requirement exhibited by cells of the IRF8^{hi} and IRF8^{low} pathways are congruent with the phenotypes exhibited by patients with primary immunodeficiency caused by mutations in *IRF8*. The bi-allelic loss of IRF8 causes absence of all monocytes and DC subsets (Hambleton et al., 2011; Bigley et al., 2018), while *IRF8* haploinsufficiency primarily affects the cells developing via the IRF8^{hi} pathway, which have greater requirements for this transcription factor (Cytlak and Resteu et al., 2020).

As DCs are relatively short-lived, small numbers of bone marrow-derived cells constantly replenish the DC pool, traveling through blood to peripheral organs (Collin and Bigley, 2018). Mass cytometry analyses aligned the two pathways originating in bone marrow with bone marrow mature DC subsets, as well as their peripheral blood counterparts. The segregation of DC potential at an early stage, noted in this chapter, is consistent with the early lineage priming model for haematopoietic lineages in human (Notta et al., 2016; Velten et al., 2017).

4.4.2 DC2 and DC3 markers support their functional specialisation

A crucial part of this project was the identification of markers able to discern between the cDC2 subpopulations *in vivo* and *in vitro*. This was necessary in order to correctly assess the output of the progenitor cultures. The identified markers include lymphoid-associated antigens CD5 and BTLA for DC2 and monocyte-associated markers CD14 and CD163 for DC3. B and T lymphocyte attenuator BTLA, was originally identified as a DC2 marker in peripheral blood, but its use was hindered by the lack of expression in culture. The absence of BTLA expression was also noted on bone marrow DC, in agreement with the observation that culture produces tissue-like cells, made in Chapter 3. The use of a second DC2 marker, CD5, which was ubiquitously expressed in blood, bone marrow, and culture, greatly aided the separation of *ex vivo*-derived cDC2 subsets. This surface marker suggests that cDC2 subtypes might play different roles in immunity, as CD5 expression on DCs has a regulatory effect on their activity to stimulate T cells and inhibits the production of pro-inflammatory cytokines (Li et al., 2019).

Both DC3 markers identified in this project have been previously reported as markers of the monocyte-macrophage lineage. Pathogen recognition receptor CD14 is well known as a marker for classical monocytes as was shown at single cell level by Villani

et al. (2017). Villani and colleagues also noted that DC3 showed higher CD14 expression than DC2. The expression of scavenging receptor and bacterial sensor CD163 has been previously reported mainly in human monocytes and macrophages, as well as on a fraction of peripheral blood DC, where it is thought to induce an immunostimulatory response (Maniecki et al., 2006). Overall, the markers are in agreement with previous findings that cDC2 subpopulations exhibit functional specialisation into anti-inflammatory DC2 and pro-inflammatory DC3 (Yin et al., 2017; Villani et al., 2017; Brown et al., 2019; Dutertre et al., 2019; Bourdely et al., 2020).

4.4.3 Advances in single cell technologies help reveal the fate of DCs in human

The investigation of DC development in the human was made possible by recent technological advances and novel analysis algorithms. In the last decade, the field of transcriptomics has moved rapidly from microarrays to bulk RNA-Seq, and more recently to single cell RNA-Seq. Employing single cell transcriptomics and phenotyping in this work allowed the characterisation of the heterogeneity of bulk progenitor and precursor populations at single cell level in the BM CD34⁺ and CD34^{med} compartments, and established differential IRF8 expression as a defining feature of the two pathways of DC development. The use of a plate-based index-sorted approach in conjunction with the Smart-seq2 protocol (Picelli et al., 2014) enabled the coupling of transcriptomics data with cell-specific antigen expression parameters by FACS. The main benefit of this approach is the traceability of the flow gates for every analysed cell, which can be used to enrich the desired DC or pre-DC populations in the future via flow cytometry. In addition, antigen expression data were used in the analysis, where gene expression values were missing due to a phenomenon called “dropout”, characteristic of scRNA-Seq data. Dropouts occur due to low RNA input of these experiments and the failure of mRNAs to be reversed transcribed (Andrews and Hemberg, 2019).

The chapter also employed novel tools designed specifically for scRNA-Seq data. Clustering is a common approach for the identification of groups of cells or samples. However, this can be challenging due to high levels of noise specific to scRNA-Seq, combined with the high-dimensionality of the transcriptome. Recent methods, such as the Single-Cell Consensus Clustering (Kiselev et al., 2017), have been developed with this in mind. SC3 is a method for unsupervised clustering, based on techniques such as PCA and k-means, and functions by combining multiple clustering solutions through

a consensus approach. Other common computational approaches specific to scRNA-Seq aim to perform trajectory inference in order to arrange the single cells in an order that represents their developmental trajectories. Over 70 tools for pseudotemporal ordering have been developed in the last years (Saelens et al., 2019). The choice of software for the analysis of DC haematopoiesis was determined by the ability to recreate multiple lineage branching events. This was crucial for the very complex human bone marrow scRNA-Seq datasets, as they encompassed a multitude of closely related lineages, including at least 3 DC subsets. The Slingshot software, designed for inferring continuous, branching lineage structures (Street et al., 2018), was selected for this analysis. The software is flexible enough to handle arbitrarily many branching events and allows for the incorporation of prior knowledge, such as clustering information (Human Cell Atlas, <https://www.humancellatlas.org>). In addition, it was highly ranked by a study aiming to benchmark single-cell trajectory inference methods based on cellular ordering, topology, scalability and usability (Saelens et al., 2019).

The experiments in this chapter were also made possible by recent advances in cytometry. Fluorescence flow cytometry is the most commonly used platform for identifying human DC, and has been invaluable for analysing surface antigens and intracellular molecules in cells derived from human peripheral blood, as well as single-cell suspensions of tissues including skin, lung, intestine, liver, and body fluids (Collin and Bigley, 2018). However, flow cytometry is currently limited to detecting 15-18 antigens. This technology uses fluorophores with overlapping emission spectra, which must be mathematically compensated, limiting the number of parameters that can be assessed simultaneously (Gadalla et al., 2019). Mass cytometry, a next generation single-cell proteomic analysis technique, utilises rare metal isotopes which have relatively little overlap (<2%), as each atom's time of flight is determined by its mass, in order to overcome the limit of multiplexing capability of flow cytometry (Gadalla et al., 2019). CyTOF allows the simultaneously assess the expression of up to 40 antigens, facilitating the analysis of complex cell populations.

In this chapter, the mass cytometry technique enabled the use of a panel of 33 antigens to align the DC developmental pathways in bone marrow and peripheral blood. However, the mass cytometry technique has a number of limitations, including a lower throughput than flow cytometry. It is also a destructive method, and the cells

used as input are nebulised, then analysed by mass spectrometry, and cannot be retrieved for further experiments, such as cell culture or sequencing.

As many single cell methods produce high dimensional data, a number of algorithms have been developed to reduce dimensionality and produce two- or three-dimensional visualisations where similar cells are grouped together. t-Distributed stochastic neighbour embedding is a non-linear dimensionality reduction technique, particularly suitable for the visualisation of large high dimensional datasets (van der Maaten, 2008). It was indispensable for the analysis of the scRNA-Seq, in order to obtain a visualisation of the dataset and the grouping defined by hierarchical clustering within the SC3 software. tSNE was also used for the analysis of the CyTOF data, and helped directly compare the groups of cells present in the bone marrow and in peripheral blood. A different approach was undertaken to map the cellular differentiation in these data, as during development cells follow continuous branching lineages, instead of forming distinct clusters. Diffusion maps were the technique of choice, as they deal with the problem of defining differentiation trajectories, and preserve the global relations between data points, are robust to noise, and are insensitive to the sampling density, aiding the detection of rare cell populations (Angerer et al., 2015).

4.5 Summary and further work

The experiments undertaken in this chapter support the existence of two developmental pathways giving rise to dendritic cells in human bone marrow (Figure 4.16). High IRF8 expression defines a CD123⁺ DC developmental pathway giving rise to pDC, cDC1 and DC2. In contrast, DC3 arise through an IRF8^{low} pathway with precursors independent from monocytes.

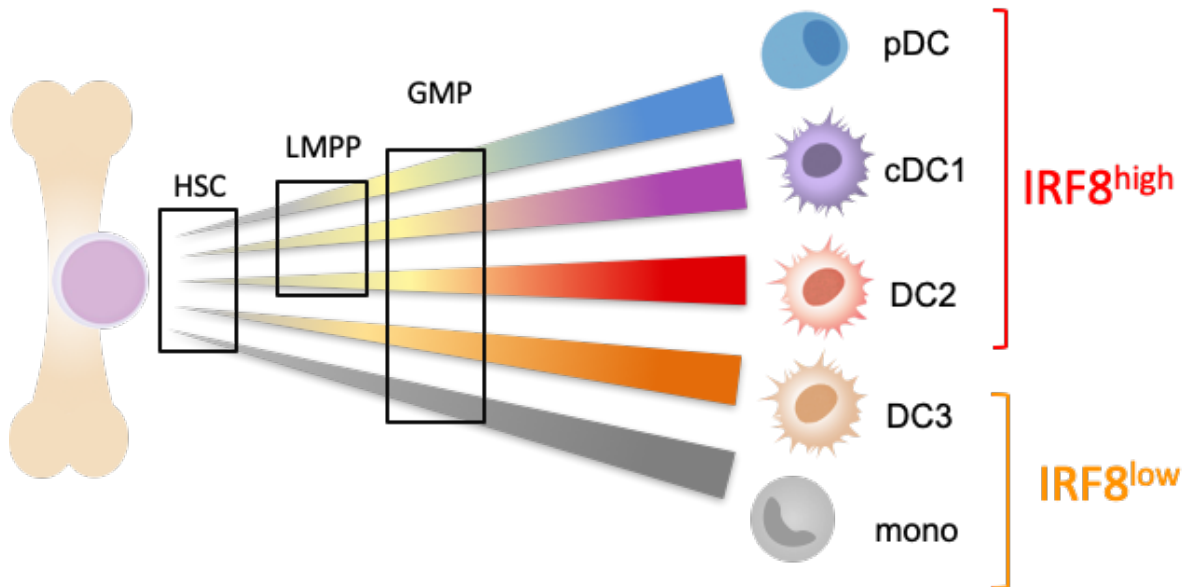


Figure 4.16. Schematic summary of the main findings from Chapter 4.

Human DCs subsets develop in the bone marrow via two pathways marked by *IRF8* expression. The *IRF8*^{high} pathway gives rise to pDC, cDC1 and DC2, while the *IRF8*^{low} pathway produces DC3 and classical monocytes. The traditional progenitor cell classification by flow cytometry is displayed as black rectangles and includes haematopoietic stem cells (HSC), lymphoid-primed multipotent progenitors (LMPP) and granulocyte-macrophage progenitors (GMP).

To derive further insights into the role of *IRF8* in DC development, future work will involve the analysis of a scRNA-Seq dataset comprising of cells derived from the bone marrow of patients with *IRF8* mutations. This will allow the detailed mapping of pathways and cells affected by *IRF8* deficiency.

To further investigate the early lineage priming model in human, supported by this work, future efforts should focus on back-tracing the developmental pathways in primitive populations of the bone marrow. This could be achieved via single cell culture of human HSC/MPP.

Finally, the two pathways of DC development and the cells they give rise to must be explored in other settings (e.g. inflammation and disease) and expanded to other tissues.

Chapter 5. Optimisation of the low cell ChIP-Seq protocol

Questions answered in this chapter:

1. Can a ChIP-compatible IRF8 antibody be identified?
2. What are the optimal sonication conditions for performing low cell IRF8 ChIP-Seq on culture-derived DCs?
3. Does cell number impact the quality of the resulting ChIP-Seq libraries?

5.1 Introduction

Protein–DNA interactions take place when a protein binds a DNA molecule and often result in the regulation of the biological function of the DNA. DNA-binding proteins include transcription factors, involved in the process of transcribing DNA into RNA, and histones, able to control DNA accessibility. Chromatin immunoprecipitation, followed by high-throughput sequencing analysis is the gold-standard technique for examining the distribution of transcription factors and histone modifications in a genome-wide manner (Kidder et al., 2011). The ChIP-Seq method is indispensable for studying the multitude of biological processes that depend on protein-DNA interactions, including cell differentiation and function, cell cycle progression, DNA replication, recombination, repair, gene expression, chromosome stability, and epigenetic silencing (Mundade et al., 2014).

IRF8, also known as interferon consensus sequence-binding protein (ICSBP), is an important immune transcription factor, playing a critical role in the development and homeostasis of DC subsets. Most studies of IRF8 have been performed in mice and its role in humans is not well understood. Applying the chromatin immunoprecipitation (ChIP) technique to its study could help gain significant insights into human immune biology and transcriptional programmes and signaling pathways regulated by IRF8 in human. First, this method can be applied to study the role of IRF8 in mature DC, an experiment that has not previously been performed in human due to the rarity of DC. This issue has been recently addressed via the development of a novel culture system, confirmed to produce large numbers of *bona fide* DC in Chapter 3 of this work. Next, the study could be extended to understand the behaviour of IRF8 during DC differentiation and in immune responses, e.g. to interferon stimulation.

5.1.1 The ChIP-Seq method

ChIP-Seq is an established and powerful technique for the analysis of protein-DNA interactions. However, to obtain accurate and reproducible results, a number of steps in the ChIP-Seq assay require careful optimisation.

At the start of most ChIP-Seq protocols (outlined in Figure 5.1), the proteins are crosslinked to the DNA with formaldehyde. Formaldehyde readily permeates cell membranes and acts in several steps, which result in the formation of covalent bonds between macromolecules. Small proteins, such as Tris or glycine, interact with formaldehyde, and are used to quench the crosslinking reaction (Hoffman et al., 2015). Alternative crosslinking methods have also been described (Zeng et al., 2006). Chromatin shearing is then performed in order to obtain smaller DNA fragments, optimal for immunoprecipitation (IP) and sequencing. DNA shearing is often achieved via mechanical methods, such as acoustic sonication, or enzymatic digestion. Sonication is the preferred approach, as it produces evenly distributed fragments, while the enzymes used for digestion often have a preference for specific sites, without regard for the distance between them, and may introduce bias into the experiment. Sonication consists of several ON/OFF cycles, aiming to produce 200-1200bp DNA fragments. Typically, samples are immersed in a cold water bath (4°C) during sonication. The shearing is achieved during the ON setting. Throughout the OFF setting, the sonication is paused and the samples are cooled down to prevent DNA degradation. Next, the DNA-binding protein, attached to its specific DNA fragments, is immunoprecipitated using a specific ChIP-grade antibody. The antibody is incubated with the lysate, along with agarose or sepharose beads covered in an immunoglobulin-binding protein, such as Protein A, G, A/G, or L. The protein-DNA complexes are then eluted off the beads and descrosslinked, commonly via a digestion at 60-65°C, lasting a few hours to overnight. The protein-free DNA is extracted and purified, and sequencing libraries are prepared. The DNA concentration and fragment size distribution are assessed, and the libraries are sequenced.

The ChIP-Seq assay can be difficult to optimise, as multiple aspects, such as the duration of the crosslinking, the optimal number and length of sonication cycles, the number of PCR amplification rounds applied during preparation of the sequencing library, as well as the sequencing depth have a major impact on the quality of the

ChIP-Seq dataset (Mendoza-Parra and Gronemeyer, 2014). In addition, the following steps of the ChIP-Seq experimentation may give rise to artefacts: (1) DNA shearing: open chromatin regions are easier to shear than closed chromatin regions and produce higher background signals; (2) antibody cross reactivity during IP; (3) base-composition bias during sequencing (Kidder et al., 2011). Two types of controls are commonly used in ChIP-Seq experiments: input chromatin and samples immunoprecipitated with nonspecific IgG antibodies. The use of input chromatin is generally recommended, as it provides enough DNA for a more complex sequencing library and requires less input material (Kidder et al., 2011), which is ideal when studying rare cell types, such as DC. However, the controls do not account for the antibody cross-reactivity, and antibody validation must be performed through independent experiments.

The Low Cell ChIP-Seq protocol (Active Motif, Catalog No. 53084) was selected for the ChIP-Seq experiments in this chapter, as it provided reagents for a complete ChIP-Seq workflow including chromatin preparation, immunoprecipitation and Illumina-compatible next generation sequencing library preparation. In addition, the protocol has been designed specifically for the use on limited cell numbers of 50,000 and above. However, the number and length of the sonication cycles and the IRF8 antibody required further testing and optimisation.

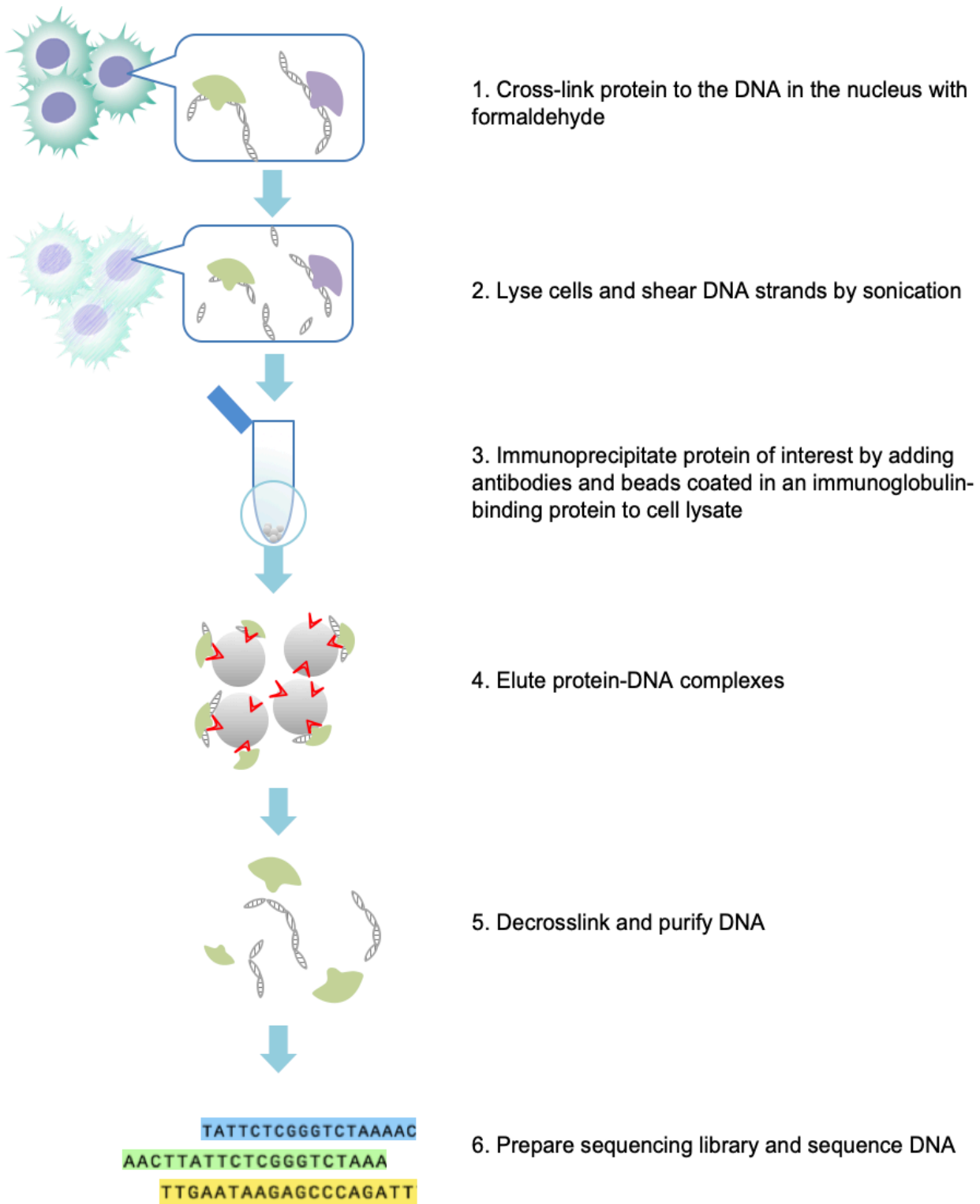


Figure 5.1. Schematic of the ChIP-Seq workflow.

The schematic outlines the common steps in most ChIP-Seq protocols: formaldehyde cross-linking, cell lysis, sonication, IP, elution and decrosslinking of protein-DNA complexes, DNA extraction, library preparation, and sequencing.

5.1.2 Guidelines for antibody testing

The success of the ChIP-Seq assay relies heavily on the quality of the selected antibody. To ensure the validity of the results, a ChIP-grade antibody, previously validated for this application by the vendor, must be utilised for the ChIP-Seq assay. Alternatively, antibody testing for ChIP suitability can be performed by the user.

The Encyclopaedia of DNA Elements Consortium (ENCODE; <https://www.encodeproject.org>) is an ongoing international collaboration of research groups set up with the goal to build a comprehensive parts list of functional elements in the human genome. ENCODE release curated, uniformly processed and validated experiments to the scientific community and provide strict guidelines to ensure high quality data standards are met. According to the ENCODE consortium, antibody deficiencies are of two types: poor reactivity against the intended target (low sensitivity) and cross-reactivity with other DNA-associated proteins (poor specificity) (Landt et al., 2012). The ENCODE consortium guidance recommends performing both primary and secondary characterisation for antibodies against transcription factors to ensure that they are suitable for ChIP-Seq experiments. The suggested characterisation workflow is outlined in Figure 5.2. The starting method recommended for primary characterisation is western blotting. The ENCODE consortium state that in order to pass primary characterisation, the antibody must detect more than 50% of bands on a western blot. The size of the detected protein must be within 20% of the size predicted from amino acid sequence. If the antibody fails the test, it can be rescued by an immunofluorescence assay. Multiple options are provided for secondary characterisation, including small interfering RNA knockdown, IP/mass spectrometry, IP with epitope-tagged version of target, and motif enrichment.

A high quality IRF8 antibody (ICSBP Antibody (C-19): sc-6058, Santa Cruz Biotechnology, Inc) has been previously successfully used for ChIP-Seq experiments in human (Shin et al, 2011; Mohaghegh et al., 2019). Unfortunately, this antibody has been discontinued and was therefore unavailable for this project. In addition, the ENCODE database did not contain any validated IRF8 antibodies suitable for this work. Consequently, a new ChIP-grade antibody had to be identified prior to performing the IRF8 ChIP-Seq assay.

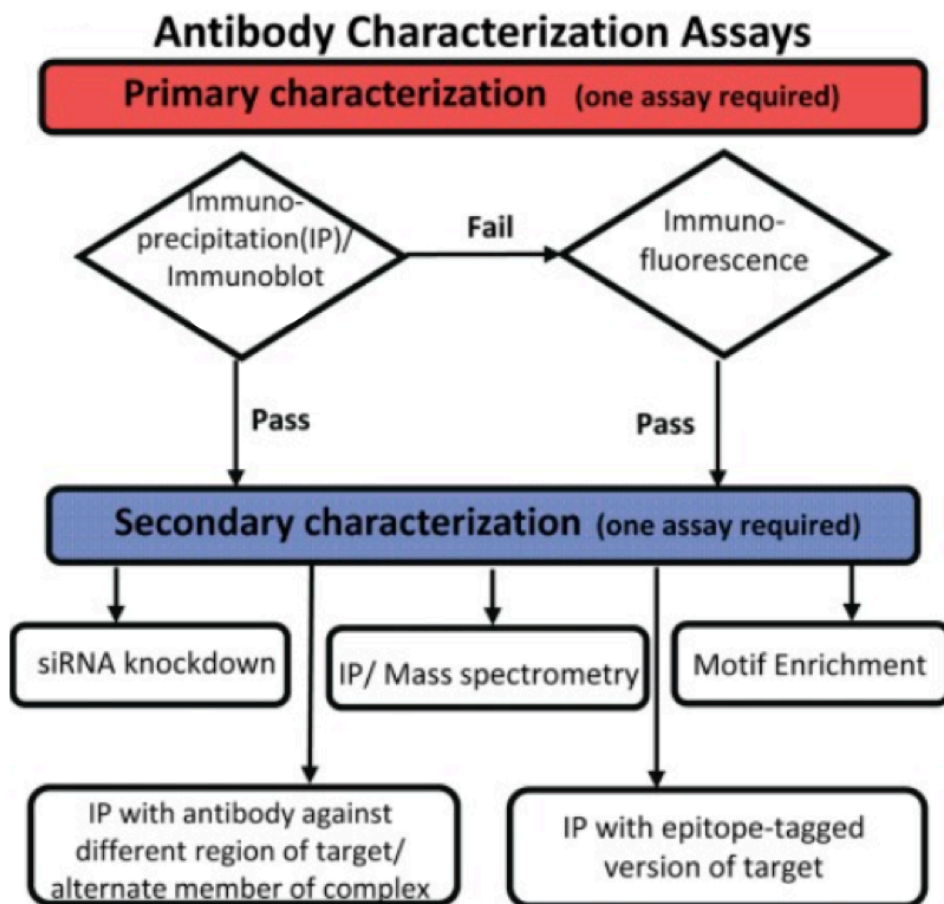


Figure 5.2. Flowchart of antibody characterisation assays, as outlined by ENCODE (adapted from Landt et al., 2012).

One assay is required for primary characterisation. Immunoprecipitation or western blotting are suggested as starting experiments. In case of a fail, the characterisation can be saved via a successful immunofluorescence experiment. Once the primary characterisation is completed, the secondary characterisation may be performed via one of the following approaches: siRNA knockdown, IP against different region of target protein or complex, IP coupled with mass spectrometry, IP with epitope tagged version of protein, or motif enrichment.

5.2 Materials and methods

5.2.1 Shipment and rehydration of expression vectors

The vectors pIRES2-EGFP-HA-IRF8 (IRF8 vector) and pIRES2-EGFP (empty vector) were kindly provided by Dr Gina Doody from the School of Medicine at the University of Leeds. The plasmid DNA was shipped and Whatman paper at room temperature. On receipt, the 1-2µg plasmid DNA, were placed in 500µl TB buffer (Invitrogen) for 6 hours at 4°C. After a brief centrifugation, 5µl of the supernatant (equivalent to 10-20ng DNA) were used for transformation into competent *E. coli* cells (Promega).

5.2.2 Preparation of LB medium supplemented with kanamycin

For the preparation of LB medium for bacterial culture, 20.6g Lennox LB Broth (Sigma) were stirred with a magnetic stirrer until suspended in 1L water, then autoclaved at 121°C for 15 minutes to sterilise. The solution was allowed to cool down before the addition of antibiotic kanamycin (Sigma) to obtain a working concentration of 50µg/mL, and stored at 4°C.

5.2.3 Preparation of LB agar plates supplemented with kanamycin

For the preparation of agar plates for bacterial culture, 30.5g of Luria low salt LB Broth with agar (Sigma) were heated while stirring until dissolved in 1L water, then autoclaved at 121°C for 15 minutes to sterilise. The solution was allowed to cool slightly before the addition of antibiotic kanamycin (Sigma) to obtain a working concentration of 50µg/mL, then poured into petri dishes (Sigma) and allowed to solidify. The plates were stored upside down at 4°C for up to 2 weeks.

5.2.4 Bacterial culture and transformation

The Promega “single-use cells” protocol was used for the bacterial transformation. Briefly, the competent cells were thawed and mixed with 10-20ng plasmid DNA and placed on ice for 30 minutes. Cells were then heat-shocked in a water bath at 42°C for 15 seconds, and placed on ice for 2 minutes. 450µl room temperature SOC medium (New England Biolabs) were added to the cells, and incubated at 37°C for 1 hour with shaking. 100µl and 50µl undiluted cells were plated on LB agar plates, supplemented with 50mg/ml Kanamycin and grown overnight at 37°C. The next day, colonies were picked and grown overnight at 37°C with shaking in LB medium (Sigma) supplemented with 50mg/ml Kanamycin.

5.2.5 Plasmid DNA extraction and quantification

The plasmid DNA was extracted the next day after transformation, using the QIAprep Spin kit (Qiagen), then quantified with a NanoDrop spectrophotometer (Thermo Scientific).

5.2.6 Restriction enzyme digestion and gel electrophoresis

Restriction enzyme digestion was performed using restriction enzymes EcoRI, BglII, and BamHI (Promega), following manufacturer's instructions. Gel electrophoresis was performed on a 1% agarose gel with the addition of 0.01% GelRed nucleic acid stain (Biotium) at 80V for 60 minutes. 300-1,500ng undigested plasmid DNA were loaded as control. The gel was visualised using the Odyssey Imaging System (LICOR Biosciences).

5.2.7 Sanger sequencing of vectors

Vectors were sent to GATC Biotech, Germany for Sanger sequencing. 5µl purified plasmid DNA at a concentration of 80-100ng/µl were aliquoted in 1.5ml Eppendorf tubes (Sigma) and shipped at room temperature.

5.2.18 Culture and transfection of HEK293T cells

HEK293T cells were defrosted as described in Chapter 2 section 2.3.3 and placed in culture in a T25 flask (Sigma) with 5mL DMEM medium, prepared as described in section Chapter 2 section 2.1.7. The cells were passaged twice into T75 flasks (Sigma) when they reached 70-80% confluency. The cells were then counted using a haemocytometer (Hawksley) and seeded into 6 well plates (Costar) to a density of 300,000 cells per well in 2ml DMEM culture medium (Gibco). The seeded cells were incubated overnight at 37°C in a humidified atmosphere at 5% CO₂.

For each of the vectors (pIRES2-EGFP-HA-IRF8 and pIRES2-EGFP), 6µg plasmid DNA were added to 610µl unsupplemented DMEM medium, and vortexed at full speed for 5 seconds. 30µl of 1% Polyethyleneimine (Sigma-Aldrich) were added to the transfection mixes. The mixes were then vortexed at full speed for 10 seconds, and incubated for 20 minutes at room temperature. Following an overnight incubation of the seeded cells at 37°C and 5% CO₂, 200µl of the transfection mixes

were added to each well of the 6 well plates (Costar) in triplicates for each of the vectors.

Fluorescence and light microscopy were performed 23 hours post-transfection using EVOS FL Cell Imaging System (Life Technologies). The cells were lysed on ice 27 hours post-transfection with 200µl RIPA cell lysis buffer (Cell Signaling Technology) per well, supplemented with cOMplete EDTA-free Protease Inhibitor Cocktail (Roche). The triplicates were combined for each vector, yielding lysates containing approximately 1.8 million cells, considering the 24-30 hours doubling time of the HEK293T cell line (DSMZ; <https://www.dsmz.de>). The lysates were spun down at full speed (21,000g) for 10 minutes at 4°C. The pellets were discarded and the clarified lysates were retained. 30µl of each of the lysates were frozen at -20°C for western blotting the following day, and the rest of the supernatant was used for immunoprecipitation immediately.

5.2.9 Immunoprecipitation

For each IP reaction, 1µg IRF8 antibody (Santa Cruz (E-9): sc-365042) or 1 µg Purified Mouse IgG2b, κ isotype control antibody (Biolegend), were added to 570µl lysate, and incubated on a rocker (Bibby Stuart) at 4°C overnight. The following day, 20µl agarose beads coated with Protein G (Cell Signaling Technology) were added to each of the lysate and antibody mixes and incubated on a rocker (Bibby Stuart) at 4°C for 60 minutes. The lysates were spun down at 500g for 2 minutes, and the supernatant was discarded. The pellets were washed 4 times with 0.8ml RIPA cell lysis buffer (Cell Signaling Technology).

5.2.10 Western blotting

A loading mix was prepared by adding 20µl reducing agent 10X NuPAGE Sample Reducing Agent (Invitrogen) to 180µl sample loading buffer 4X NuPAGE LDS Sample Buffer (Invitrogen). For lysates, 10µl loading mix was added to 30µl lysate. Considering the HEK293T doubling time of 24-30 hours, the lysates contained approximately 90,000 cells. For IPs, 10µl loading mix and 5µl RIPA cell lysis buffer (Cell Signaling Technology) were added to the washed IP beads prepared in the immunoprecipitation step. All samples were boiled at 95°C for 5 minutes. The lysates were loaded onto a 7.5% Mini-PROTEAN TGX Precast Protein Gel (BIO-RAD). The

IPs were centrifuged at 500g for 2 minutes prior to loading onto the same gel. The gel was run in Tris/Glycine/SDS running buffer (BIO-RAD) at 120V for 15 minutes, then at 150V for 40 minutes. The transfer was performed in Tris/Glycine Buffer transfer buffer (BIO-RAD) at 100V for 90 minutes. The membrane was probed with the following primary antibodies: IRF8 antibody (Santa Cruz (E-9): sc-365042, 1:500 dilution), HA-probe antibody (Santa Cruz (F-7): sc-7392, 1:200 dilution), and Vinculin antibody (Abcam ab129002, 1:10,000 dilution). A secondary anti-mouse κ light chain antibody (Santa Cruz m-IgGk BP-HRP: sc-516102) was used in combination with all primary Santa Cruz antibodies, which were raised in mice. A secondary anti-rabbit light chain antibody (Merck Millipore MAB201P, 1:20,000 dilution) was used with the primary Vinculin antibody (Abcam), raised in rabbits. The visualisation was performed using the chemiluminescent Pierce ECL Substrate (ThermoFisher) and the Odyssey Imaging System (LI-COR Biosciences).

5.2.11 Sorting of bone marrow cells for western blotting and cell culture

Bone marrow cells from hip replacement procedures or haematopoietic stem cell transplantation donors were stored and defrosted as described in Chapter 2 subsection 2.3.3. The cells were stained and FACS-purified as described in Chapter 2 subsection 2.3.6 using fluorescently-conjugated antibodies listed in Table 3.1. CD34⁺ Lineage⁻ cells were retained for culture and CD34⁻ Lineage⁻ cells were lysed with RIPA cell lysis buffer (Cell Signaling Technology) and used for western blotting, as part of the primary characterisation of the Santa Cruz (E-9): sc-365042 antibody, performed in the same manner as for the lysates in section 5.2.10.

5.2.12 Co-culture of bone marrow progenitors with OP9-DL1 cells

CD34⁺ Lineage⁻ progenitors were co-cultured with OP9-DL1 cells for 14-21 days in order to enrich the output in cDC1s, as described in Chapter 2 subsection 2.3.5.

5.2.13 Chromatin immunoprecipitation

The Low Cell ChIP-Seq kit and Next Gen DNA Library Kit (Active Motif) were used for cell fixation, immunoprecipitation, DNA extraction, and sequencing library preparation, according to manufacturer's instructions. The workflow of the protocol is outlined in Figure 5.3. An additional cell sorting step was introduced at beginning of the protocol in order to purify the populations used in the ChIP-Seq experiment.

Cultured cDC1s were FACS-sorted into siliconized tubes (Active Motif) with 300 μ l MEM- α medium (Gibco) and kept on ice. The cells were spun down at RT at 1,300g for 5 minutes and resuspended into 50 μ l PBS (Active Motif). The cells were then fixed for 10 minutes at RT with 5 μ l of Complete Cell Fixation Solution, the 55 μ l fixation reaction containing 1% formaldehyde. The fixation was quenched for 5 minutes with the Stop Solution (Active Motif) at RT. The lysate was transferred into 1.5ml Bioruptor Pico microtubes with caps (Diagenode) and sonicated on a rotating carousel with 6 slots in the Bioruptor Pico device (Diagenode). 2-3 tubes were sonicated at a time, balanced on opposite sides of the carousel. As part of the optimisation, two sonication settings were tested: (1) 10 cycles of 30 seconds ON and 30 seconds OFF, suggested as a starting setting by the Low ell ChIP-Seq kit, and (2) 6 cycles of 30 seconds ON and 90 seconds OFF, suggested by Diagenode for our sample volume and target DNA fragment size. Following sonication, lysates were stored at -80°C. They were thawed on ice prior the pre-clearing step of the protocol. IP was performed with 3 μ g IRF8 antibody (Santa Cruz (E-9): sc-365042). The protein-DNA complexes were then eluted, and decrosslinking was performed overnight at 65°C in a thermocycler (Applied Biosystems). DNA was extracted and stored at -20°C until it was thawed at room temperature and used for sequencing library preparation. A two-sided SPRI clean-up was performed after the PCR to ensure an appropriate size distribution of DNA fragments for sequencing.

5.2.14 Quantification of sequencing libraries

The DNA concentration, molarity, and fragment size distribution of the sequencing libraries were assessed with the Agilent Bioanalyzer 2100 using the high sensitivity DNA kit (Agilent).

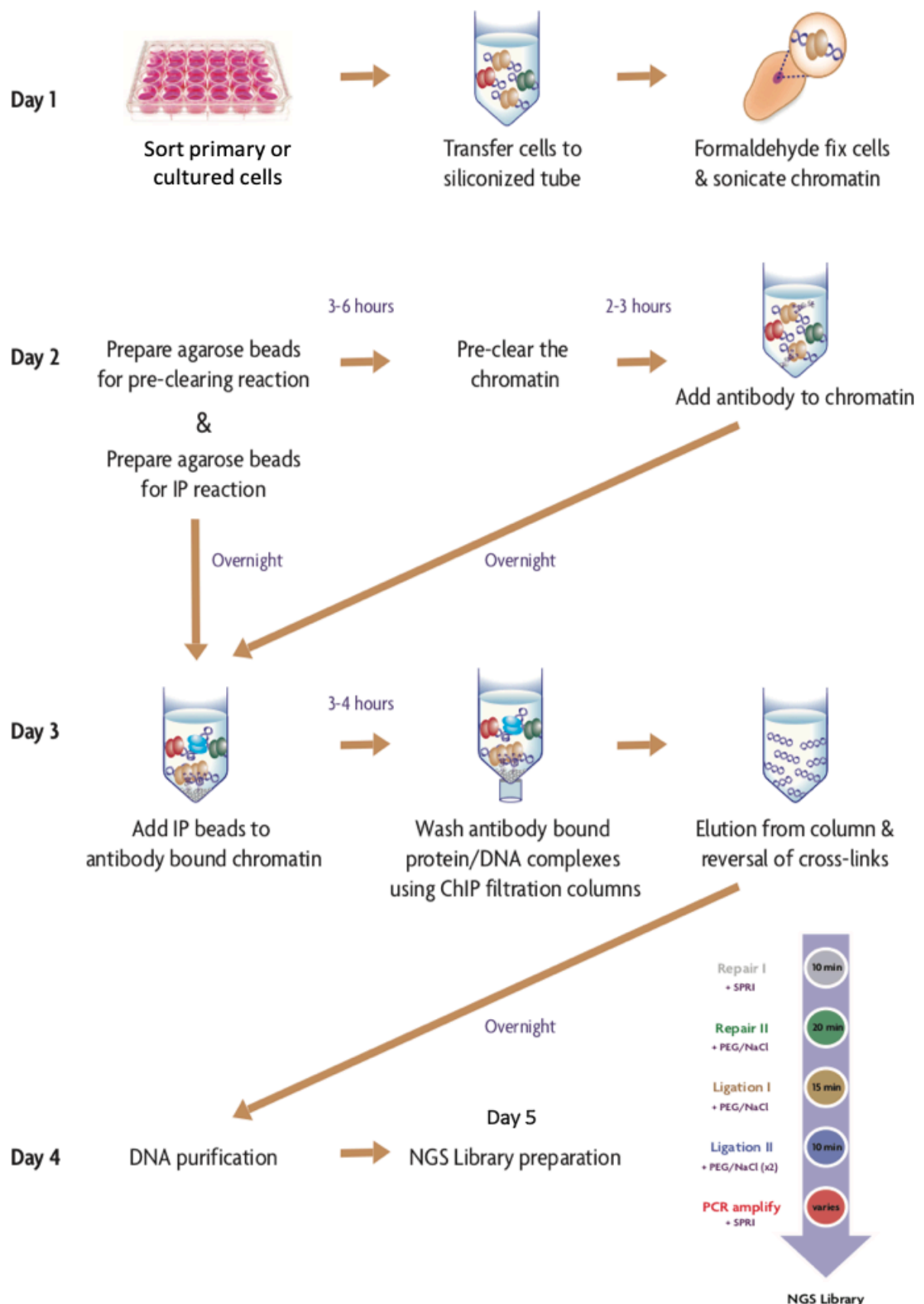


Figure 5.3. Flow chart of the low cell ChIP-Seq process (Active Motif).

The 5-day protocol includes the start-to-end processing of fresh cultured or primary cells up to the sequencing stage. On day 1, cells are collected into siliconized tubes,

fixed with formaldehyde, and sonicated. On day 2, the agarose beads coated in Protein G are prepared for the pre-clearing and IP reactions. The chromatin is then thawed and the pre-clearing reaction is performed, followed by an overnight immunoprecipitation with the desired antibody. On day 3, agarose beads are added to the samples. After 3-4 hours of incubation, and the antibody-bound protein-DNA complexed are eluted off the beads. The cross-linking induced by formaldehyde on day 1 is reversed overnight. The DNA is extracted on day 4, and used for library preparation on day 5. The library construction comprises multiple steps, including two repair steps, two ligation steps, and a PCR amplification step.

5.3 Results

5.3.1 Candidate antibodies for characterisation

Initially, the IRF8 (D20D8) Rabbit mAb #5628 (Cell signaling) was chosen for testing. According to information provided by its manufacturer, this antibody is suitable for the following applications: western blotting, immunoprecipitation, ChIP, ChIP-Seq. However, the Cell signaling IRF8 antibody failed to detect its target by western blotting in the IRF8-expressing lymphoblast-like Raji cell line (Figure 5.4). A band was expected in the western blot analysis at 48kDa, which is the atomic mass of the most common splice variant of the IRF8 protein (Human Protein Atlas; <https://www.proteinatlas.org>). In addition, this antibody lacked specificity, as it detected numerous bands in the HeLa cell line, which does not express IRF8 (Human Protein Atlas; <https://www.proteinatlas.org/>). Therefore, the testing of this antibody was not pursued further.

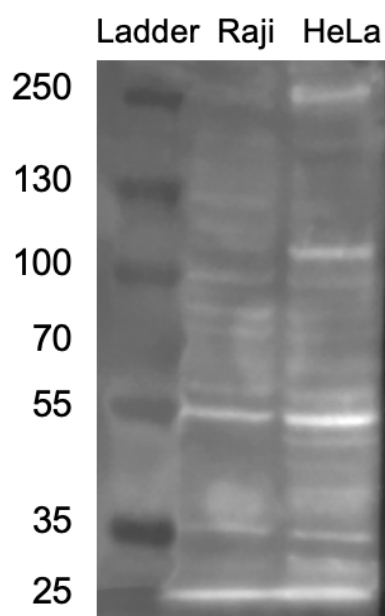


Figure 5.4. Western blot of Raji and HeLa cell lines probed with the primary antibody IRF8 (D20D8) Rabbit mAb #5628 (Cell signaling).

Mouse Anti-Rabbit light chain, HRP conjugate Antibody (Merck Millipore MAB201P) directed against the species of the primary antibody, was used as secondary antibody. Lysates containing 300,000 IRF8-expressing Raji cells and 300,000 IRF8⁻ HeLa cells were used in the analysis. The ladder shows atomic mass of proteins in kDa.

Next, the ICSPB Antibody (E-9): sc-365042 (Santa Cruz Biotechnology) was selected for validation. Its applications recommended by the supplier include IP, western blotting, immunohistochemistry, and immunofluorescence. In addition, the antibody has been successfully used for generating publication-quality data (Arifuzzaman, et al. 2017, Bouamar, et al. 2013). The antibody is a mouse monoclonal IgG_{2b} (kappa light chain), raised against amino acids 357-426 (C-terminus) of the IRF8 protein of human origin. Analysis of the 70-amino acid long sequence targeted by the antibody with the Protein Basic Local Alignment Search Tool (BLASTP) against the NCBI Protein Reference Sequences database, revealed that the sequence shared 100% identity with all three IRF8 isoforms present in the database. Up to 31% identity was also shared with three isoforms of the NLRX1 protein. This minor resemblance did not present a concern for the CHIP-Seq experiment, particularly because the NLRX1 protein does not bind DNA (Xiao et al., 2012). Critically, BLASTP analysis of amino acids 357-426 of the IRF8 protein did not identify any significant alignments to other

DNA-binding members of the IRF family, in particular to IRF4, which shares strong homologies with IRF8 and could therefore contaminate the assay.

5.3.2 Primary characterisation

Primary characterisation of the ICSBP Antibody (E-9): sc-365042 (Santa Cruz Biotechnology) was performed via western blotting of CD34⁻ Lineage⁻ bone marrow cells. This population contained the bulk of DCs, including cDC1s, cDC2s, and pDCs, as well as a small proportion of other immune cells. The CD34⁻ Lineage⁻ cells were expected to express high levels of IRF8, sufficient for detection by western blotting. In addition, testing the bulk of DCs would identify any cross reactivity of the antibody with proteins expressed by any of the DC subsets. Western blotting was performed with biological duplicates, derived from two different bone marrow samples. The western blot was then probed with the candidate ChIP antibody anti-ICSBP (E-9): sc-365042 (Santa Cruz Biotechnology).

A single band was detected in both of the processed biological replicates at around 50kDa (Figure 5.5), corresponding to the atomic mass of the most common splice variant of the IRF8 protein. No other bands were detected on the membrane. This fulfilled the ENCODE criteria for primary characterisation, allowing to proceed with secondary characterisation.

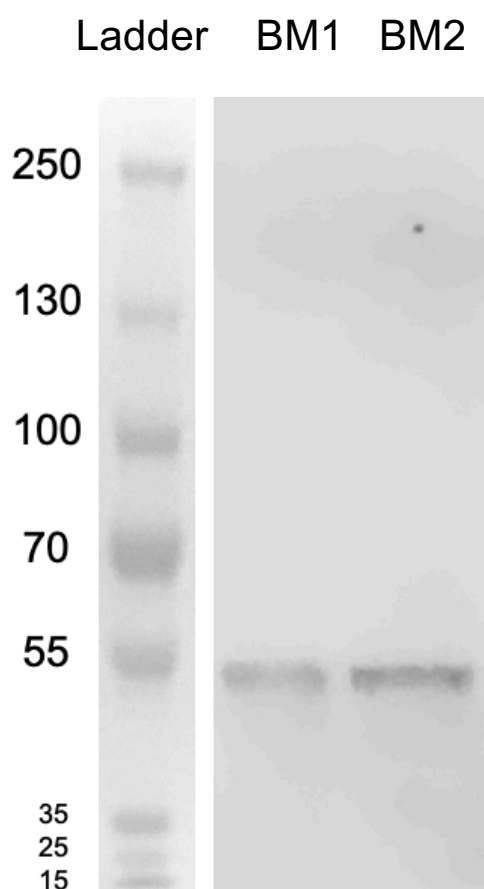


Figure 5.5. Western blot of primary CD34⁺ Lineage⁻ cells from human bone marrow, probed with IRF8 antibody ICSBP (E-9): sc-365042 (Santa Cruz Biotechnology).

The labels above the lanes denote two bone marrow samples derived from different donors. 500,000 cells were lysed and loaded in the lane labelled BM1 and 1,000,000 cells – in BM2. The ladder shows atomic mass of proteins in kDa.

5.3.3 Secondary characterisation

On the successful completion of the primary characterisation, secondary characterisation of the ICSBP Antibody (E-9): sc-365042 (Santa Cruz Biotechnology) was performed via IP and probing of the epitope-tagged version of IRF8. The HEK293T human epithelial cell line was selected for this experiment. The cell line does not express IRF8 (Human Protein Atlas; <https://www.proteinatlas.org>), and would serve as a negative control in IP experiments. HA-tagged IRF8 was intended to be introduced into the HEK293T cells via transfection, following the confirmation of

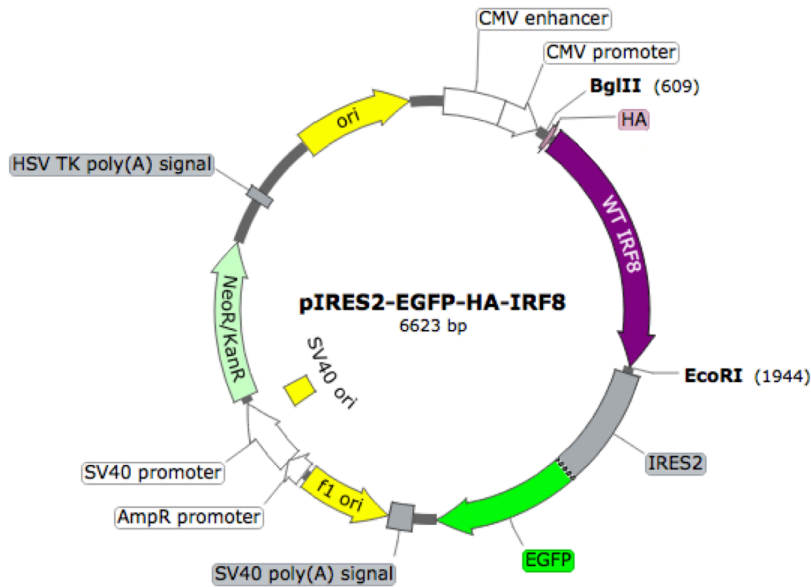
vector identities. The epitope tagging, which is a procedure whereby a short well-known amino acid sequence is attached to a protein under study, would allow the detection of heterologously expressed IRF8 with an HA antibody, as well as with the candidate IRF8 antibody.

5.3.3.1 Confirmation of vector identities

The vectors used in the transfection experiment were gratefully received from Dr Gina Doody (University of Leeds) and propagated in *E. Coli*. The bacteria that incorporated the plasmids were selected using agar plates and medium containing kanamycin, as the vectors featured an antibiotic resistance gene. The IRF8-containing vector (pIRES2-EGFP-HA-IRF8) used for transfection is described in Figure 5.6 A. This vector was sequenced to confirm that the insert maintained the correct orientation and had an appropriate size (Figure 5.6 B). The “empty” vector with no gene insert (pIRES2-EGFP, Clontech), was used to control for factors such as transfection-induced apoptosis and is described in Figure 5.7 A. The multiple cloning site (MCS) of the “empty” vector, where the IRF8 gene and the HA-tag were inserted to create the pIRES2-EGFP-HA-IRF8 vector, is displayed in Figure 5.7 B.

Restriction enzyme digestion, followed by gel electrophoresis were performed in triplicates in order to further confirm the identities of the two vectors (Figure 5.8). Undigested plasmids were loaded alongside vectors digested using restriction enzymes BglII and EcoRI. For the “empty” vector, the digestion was expected to produce fragments similar in size to a 5.3kb linearised plasmid, as the two restriction recognition sites are located in close proximity to each other (Figure 5.7 B). For the *IRF8*-containing vector, the restriction digestion was expected to excise the DNA fragment encoding the HA-tag and the *IRF8* gene with the combined length of 1,335bp and produce a second, 5,288bp-long DNA fragment, containing the rest of the plasmid (Figure 5.6 A). As expected, a single DNA fragment was produced as a result of the “empty” vector digestion and two fragments were produced when digesting the IRF8 vector (wells labelled with “D” in Figure 5.8). The produced fragments corresponded to the predicted size for all replicates for both the *IRF8*-containing (1.3kb and 5.3kb) and the empty vector (5.3kb). All undigested plasmids (“U”) formed multiple bands, representing the different plasmid forms, such as nicked and supercoiled.

A



B

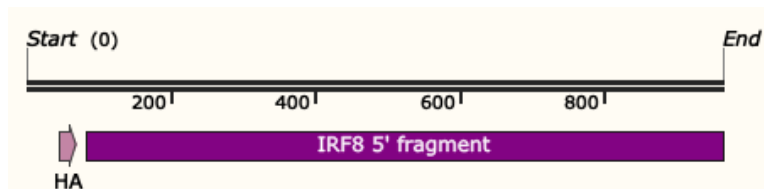


Figure 5.6. The pIRES2-EGFP-HA-IRF8 expression construct.

A. pIRES2-EGFP-HA-IRF8 vector map, displaying restriction recognition sites used for restriction enzyme digestion in this chapter. Unique cutters are shown in bold. The vector contains features similar to those of the pIRES2-EGFP vector (Figure 5.7), with the exception of the multiple cloning site. The MCS was exploited for inserting the IRF8 gene into the “empty” vector.

B. Sanger sequencing results for the pIRES2-EGFP-HA-IRF8 vector using the CMV promoter forward primer. The 963bp-long read contained the sequence encoding the HA-tag, as well as the 5' end of the IRF8 gene.

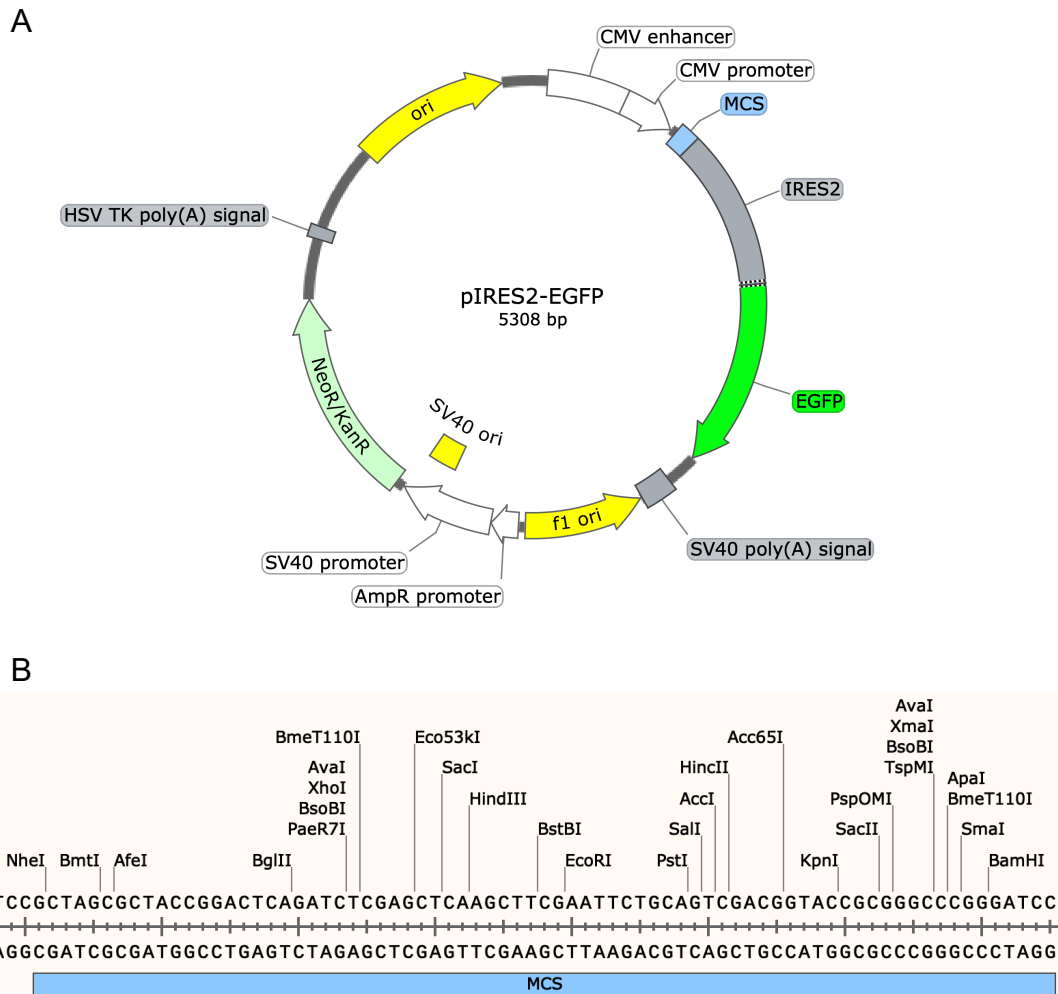


Figure 5.7. The pIRES2-EGFP expression construct.

A. Vector map of pIRES2-EGFP. The plasmid contains the following features at the specified locations: human cytomegalovirus (CMV) immediate early promoter: 1–589, multiple cloning site (MCS): 591–665, internal ribosome entry site (IRES) sequence: 666–1250, enhanced green fluorescent protein (EGFP) gene: 1254–1973, SV40 early mRNA polyadenylation signal: 2096–2217, f1 single-strand DNA origin: 2224–2679, bacterial promoter for expression of KanR gene 2706–2810, SV40 origin of replication: 3020–3155, SV40 early promoter/enhancer: 2812–3169, kanamycin/neomycin resistance gene: 3204–3998, herpes simplex virus (HSV) thymidine kinase (TK) polyadenylation signals: 4234–4252, pUC E. coli plasmid replication origin: 4583–5226.

B. Enlarged view of the pIRES2-EGFP multiple cloning site. The 75bp-long MCS contains restriction recognition sites for numerous commercially-available restriction enzymes.

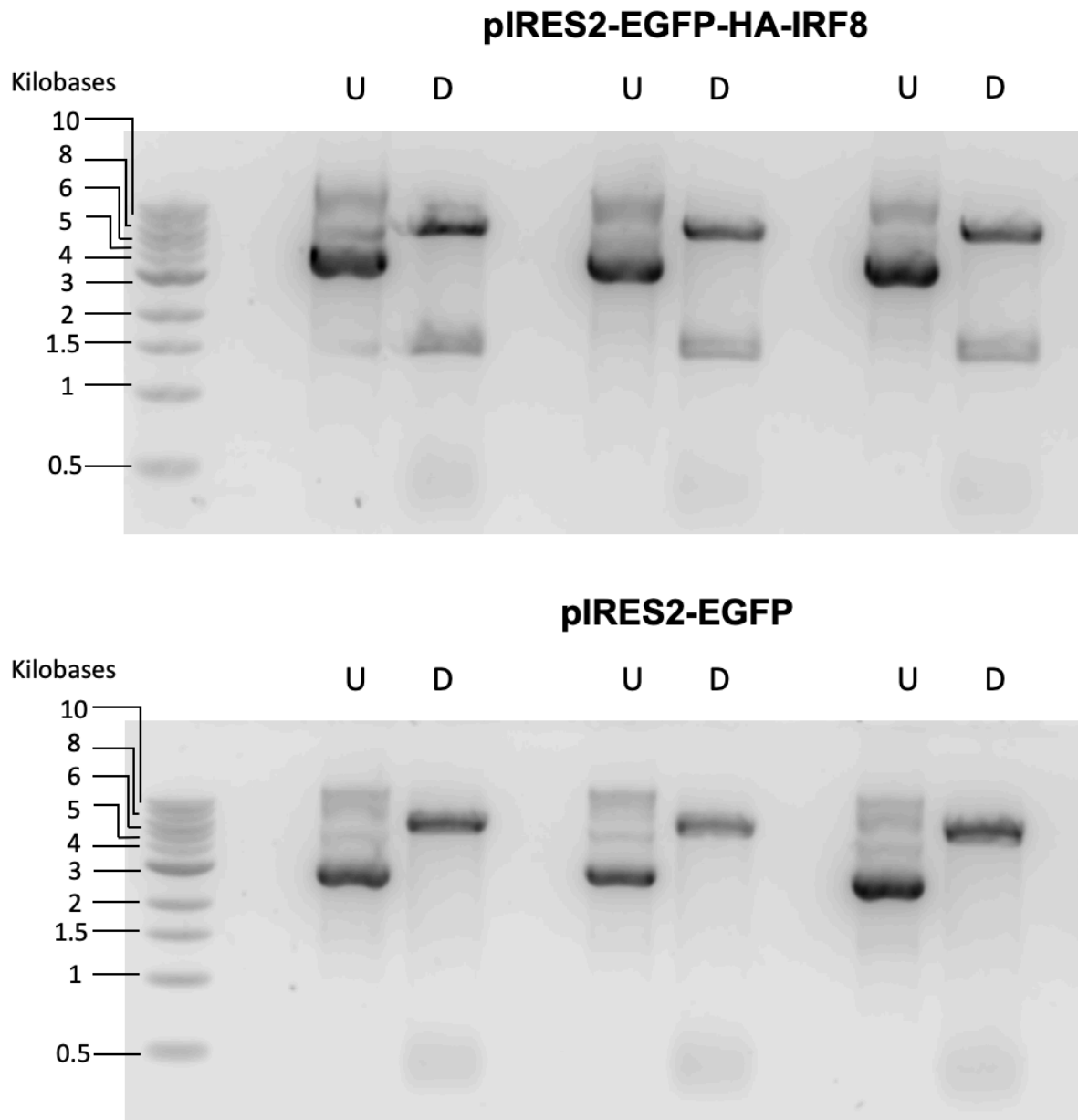


Figure 5.8. Gel electrophoresis of the IRF8-containing vector pIRES2-EGFP-HA-IRF8 (top) and the empty vector pIRES2-EGFP (bottom).

The experiment was performed in triplicates picked from different bacterial colonies. The wells labelled with “U” were loaded with undigested vectors. Plasmids digested with restriction enzymes *Bgl*III and *Eco*RI were loaded in the wells labelled with “D”. The size of the fragments in kilobases is depicted on the left by the well containing the ladder.

5.3.3.2 Immunoprecipitation and western blotting of epitope-tagged IRF8 protein

Following the confirmation of the vector identities, the plasmids were used for transfection into HEK293T cells. The success of the transfection was assessed with fluorescence microscopy. This was made possible by the presence of a gene encoding the Enhanced Green Fluorescent Protein (EGFP) in both vectors. Microscopy confirmed that the transfection was successful for both vectors (Figure 5.9).

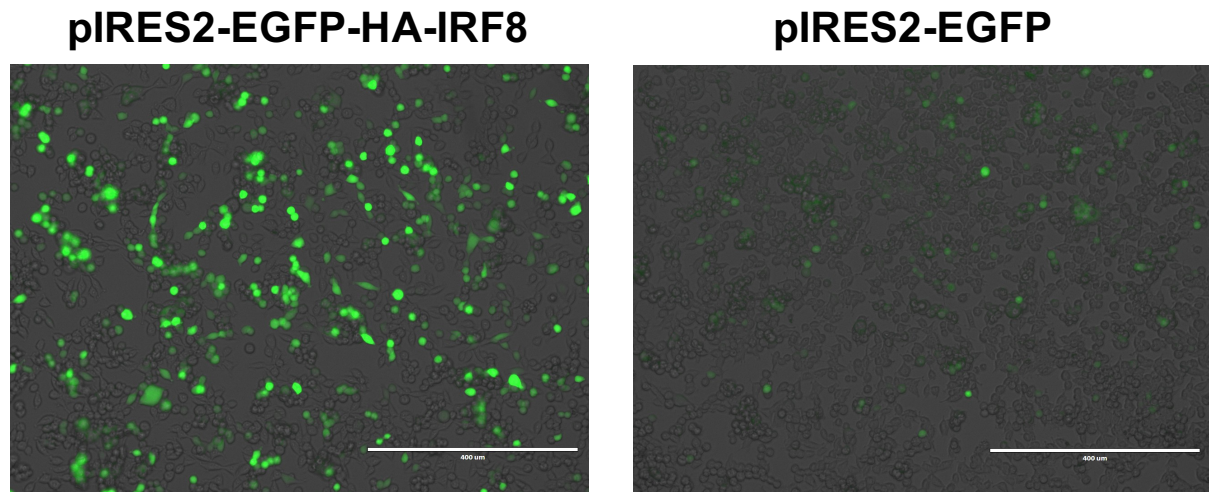


Figure 5.9. Overlay of light and fluorescence microscopy images for HEK293T cells transfected with the IRF8-containing vector (left) and the empty vector (right).

Light microscopy and fluorescence microscopy in the GFP channel were performed 23 hours post-transfection with Polyethylenimine. The same microscopy settings were used to produce both images. The size of the bars is 400 microns.

Once the success of the transfection was confirmed via fluorescence microscopy, the cells were lysed and subjected to IP followed by western blotting, or to western blotting alone (Figure 5.10 A). Lysates were probed with the candidate IRF8 antibody from Santa Cruz and with an HA-probe antibody. A single band was detected at approximately 50kDa by both antibodies in the IRF8-containing lysates. This corresponds to the atomic mass of the most common splice variant of the IRF8 protein – 48kDa.

IP was performed with the Santa Cruz IRF8 antibody, as well as with a mouse IgG2b (kappa light chain) antibody, of the same isotype (class) as the IRF8 antibody. The

isotype control was necessary to identify nonspecific bands resulting from subjecting proteins to immunoprecipitation. All IPs were probed with the Santa Cruz IRF8 antibody being tested. The IRF8 protein was detected at 48kDa in the lysate of cells transfected with the IRF8 vector but not in the “empty” vector-transfected cell lysate. Further bands were observed at low molecular weights in all IP wells. They can be attributed to the IgG light chains (25kDa), originating from the antibodies used for immunoprecipitation and detected by the secondary antibody. The heavy chains of the IP antibodies (50kDa) were not detected by the anti-light chain secondary antibody, which was used specifically to avoid obscuring the IRF8 protein at 48kDa.

An additional control was used to ensure similar amounts of protein were loaded in all wells (Figure 5.10 B). Vinculin, a 117-kDa ubiquitously expressed cytoskeletal protein was selected as “loading” control. The bands in all lysates had a consistent size, for both the *IRF8*-containing and “empty” vector-transfected cells. The Vinculin antibody showed no bands in the IP reactions, as the IRF8 and the mouse IgG2b antibodies did not immunoprecipitate the Vinculin protein.

The Santa Cruz IRF8 antibody proved to be specific, as it detected only one band matching the predicted atomic mass of IRF8. The antibody was also able to immunoprecipitate the IRF8 protein, as evident from the larger size of the 48 kDa band in the IP samples compared to the lysates.

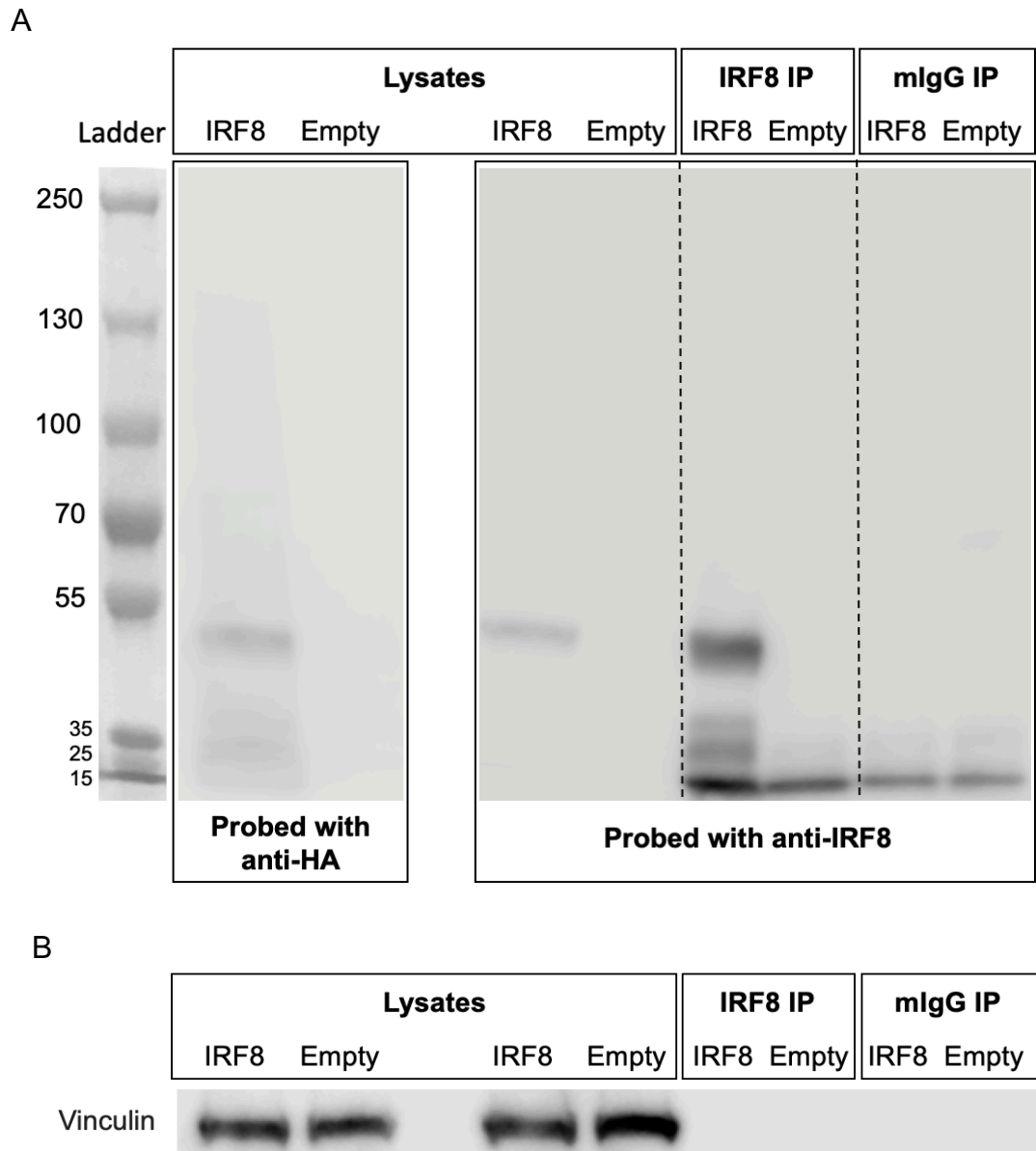


Figure 5.10. Western blot performed on lysates and IP reactions of HEK293T cells transfected with the IRF8-containing vector pIRES2-EGFP-HA-IRF8 and the empty vector pIRES2-EGFP.

A. HEK293T lysates probed with an anti-HA antibody and the Santa Cruz IRF8 antibody ICSP (E-9): sc-365042. Approximately 90,000 cells were loaded in each well labelled as "lysate". IP reactions were performed on 1.7 million transfected cells, using the IRF8 antibody and a mIgG control and were probed with the Santa Cruz IRF8 antibody. Labels above lanes denote the vector used for transfection. The ladder shows atomic mass of proteins in kDa.

B. Probing of all lysate and IP wells with an antibody targeting the 117kDa Vinculin protein.

5.3.4 Optimisation of sonication settings

As a sensitive and specific IRF8 antibody had been identified, the ChIP experiment was able to progress to the next optimisation step – the sonication settings. The testing covered two main aspects of the sonication: the number of cycles and the length of the OFF setting. Two different sonication settings, derived from different sources were used for testing. The protocol for the Active Motif kit, used for the ChIP experiment, recommended a starting setting of 10 cycles of 30 second ON and 30 seconds OFF. The setting suggested by the manufacturer of the Bioruptor Pico sonicator (Diagenode) for our sample was 6 cycles of 30 seconds On and 90 seconds OFF. This recommendation took into account our sample volume (260-300 μ l) and desired fragment length of 400bp (falling within the 200-1200bp range suggested by the manufacturer of the ChIP-Seq kit).

Both suggested settings were tested on the same number of FACS-purified culture-derived cDC1s. The Active Motif Low Cell protocol was used for ChIP and library prep, and the concentration and the fragment size distribution of the resulting libraries were assessed with Bioanalyser (Figure 5.11). A similar fragment size distribution was seen in both settings, most DNA fragments falling between 300 and 450bp in size. However, the library produced from chromatin subjected to 6 cycles of sonication had higher concentration and molarity readings than the library produced following 10 sonication cycles. The readings for DNA concentration were 595.20 pg/ μ l for the 6 cycles setting and 291.27pg/ μ l for 10 cycles. The molarity readings were 2,528.7 pmol/l and 1,278.7 pmol/l, respectively. The fragment size distribution and molarity of libraries from both tested setting were suitable for sequencing. However, peaks containing small DNA fragments below 200bp were detected in the sample subjected to 10 cycles of sonication, but not in the sample subjected to 6. These peaks are undesirable in a sequencing library, as they can undergo preferential amplification during sequencing.

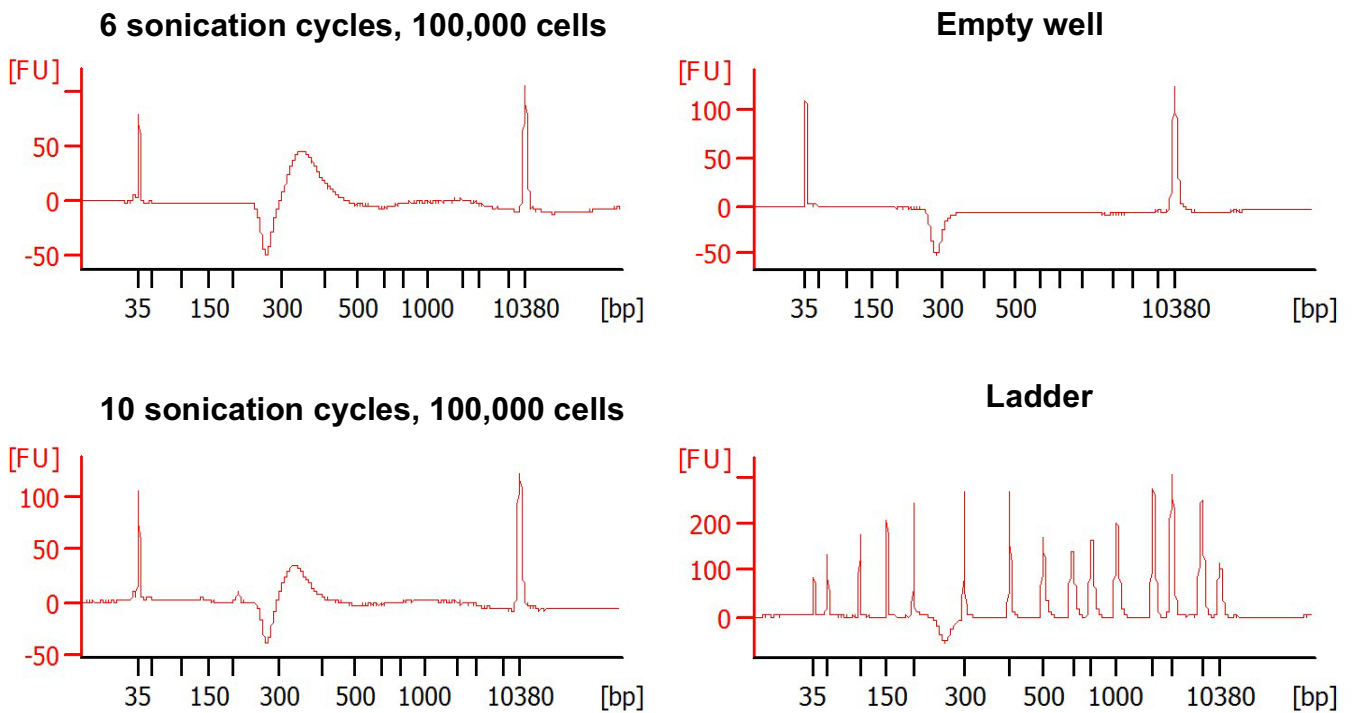


Figure 5.11. Bioanalyser electropherogram for ChIP libraries prepared from 100,000 cultured, sorted, and formaldehyde-fixed cDC1s subjected to different sonication settings and readings for an empty well and the ladder ran on the same Bioanalyser chip.

100,000 cDC1s were produced from CD34⁺ progenitors in OP9-DL1 culture to test each of the sonication conditions, then FACS-purified on day 19. The cells were fixed and subjected to 6 or 10 sonication cycles, then immunoprecipitated with 3 μ g of IRF8 antibody (Santa Cruz Biotechnology) as part of the Low Cell ChIP-Seq protocol (Active Motif). The ChIP DNA was extracted and used for library preparation, quantified with the Bioanalyser high sensitivity DNA kit. The x axis indicates fragment size in base pairs, while the y axis shows signal intensity in fluorescence units. Peaks at 35bp and 10380bp seen on all plots represent lower and upper markers.

Negative baseline dips were detected in all samples, as well as in the empty well and the ladder (Figure 5.11). According to the Agilent 2100 Bioanalyzer System Maintenance and Troubleshooting Guide (Agilent Technologies), this feature is caused by residual RNaseZap on the Bioanalyser instrument electrodes. However, this did not interfere with the interpretation of results, as none of the wells had a QC flag, and all ladder markers were detected. Based on the Bioanalyser results, both 6 and 10 cycles of sonication produced useable libraries from 100,000 cDC1. However, the library prepared following 6 sonication cycles had a more appropriate fragment

size distribution and a significantly higher concentration. The 6-cycle setting was therefore selected for further experiments.

Next, the effect of cell number on the quality of the sequencing library was explored by undertaking the same experiment with 6 sonication cycles while using 50,000 cDC1 cells as input. The resulting library displayed a fragment distribution of 300-450bp and had readings for DNA concentration (829.32 pg/ μ l) and molarity (3,384 pmol/l) comparable to the ChIP experiment performed with 100,000 cDC1 and 6 sonication cycles (Figure 5.12). These results suggested that 50,000 cells, which is the minimum recommended by the Low Cell ChIP-Seq protocol (Active Motif), are sufficient for producing a sequencing library that does not differ in quality from a ChIP library prepared from double the cell number.

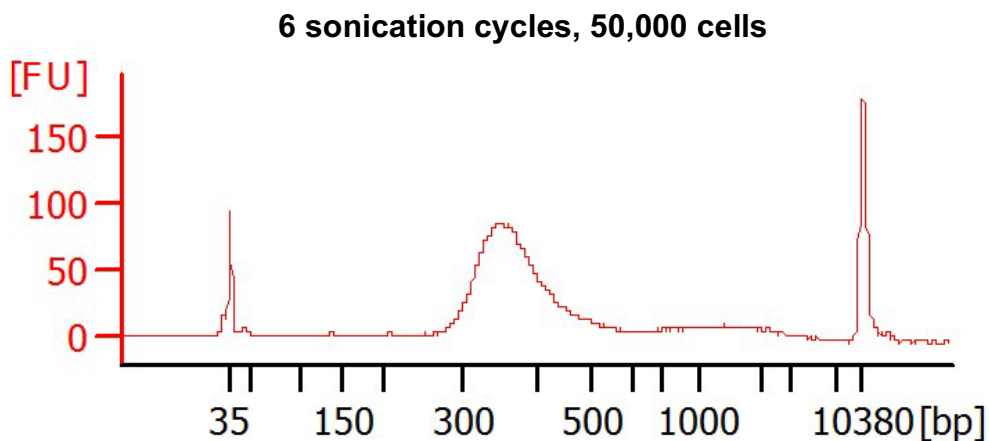


Figure 5.12. Bioanalyser electropherogram for a ChIP library prepared from 50,000 cultured, sorted, and formaldehyde-fixed cDC1s subjected to 6 sonication cycles.

50,000 cDC1s were produced from CD34⁺ progenitors in OP9-DL1 culture, then FACS-purified on day 19. The cells were fixed and subjected to 6 sonication cycles, then immunoprecipitated with 3 μ g of the Santa Cruz IRF8 antibody of IRF8 antibody (Santa Cruz Biotechnology) as part of the Low Cell ChIP-Seq protocol (Active Motif). The ChIP DNA was extracted and used for library preparation, quantified with the Bioanalyser high sensitivity DNA kit. The x axis indicates fragment size in base pairs, while the y axis shows signal intensity in fluorescence units. Peaks at 35bp and 10380bp represent lower and upper markers.

5.4 Discussion

The aim of this chapter was to optimise the IRF8 ChIP-Seq pipeline to probe the transcriptional programs and signaling pathways regulated by IRF8 in human IRF8-expressing DC subsets. First, the identification of a ChIP-grade antibody was necessary, as the IRF8 antibody used in previous IRF8 ChIP studies had been discontinued. Sonication settings and effect of cell number were also explored for the low cell ChIP-Seq protocol.

Performing antibody validation is critical before undertaking ChIP-Seq experiments, to ensure accurate and reproducible data, to maximise the financial efficiency and avoid wastage of limited human samples. When published, it also avoids duplication of work between labs and helps to ensure data quality, facilitating the collaborative use data across the research community. The ENCODE: Encyclopaedia of DNA Elements outlines rigorous criteria for testing candidate ChIP antibodies. These include primary and secondary characterisation steps. The first antibody selected for testing failed to detect IRF8 in an IRF8-expressing cell line and exhibited very low specificity, highlighting the need of rigorous testing of antibodies prior to performing the ChIP-Seq assay. The second antibody selected for testing was a mouse monoclonal antibody manufactured by Santa Cruz Biotechnology, raised against the C-terminus of the human IRF8 protein. This antibody passed its primary characterisation testing, as it detected a single band corresponding to the atomic mass of IRF8 in 500,000 - 1,000,000 bone marrow CD34⁺ cells. The antibody was also able to immunoprecipitate the IRF8 protein and showed high specificity in IP followed by western blotting.

The factors considered when optimising sonication settings included maximising DNA yield and generating appropriate sized DNA fragments. These criteria were best fulfilled by subjecting cells to 6 sonication cycles, compared to 10. The effect of cell number was also explored, and comparable sequencing libraries were produced using 50,000 and 100,000 cells. These data demonstrated the feasibility of performing IRF8 ChIP-Seq in low numbers of primary human DCs using the Santa Cruz IRF8 antibody and the Active Motif Low Cell ChIP-Seq protocol and optimised the pipeline for use in future experiments.

The use of limited human samples in the production of DCs for the optimisation experiment, as well as the high cost and the labour-intensive nature of the work (up to 3 weeks of cell culture, followed by cell sorting and a 5-day long ChIP-Seq protocol) limited the number of replicates and conditions studied during optimisation, representing a weakness of this study. Nevertheless, all tested conditions produced high quality libraries, suitable for sequencing on an Illumina instrument, which will ultimately be employed for the generation of ChIP-Seq data.

5.5 Summary and further work

Through the work in this chapter, a specific and sensitive IRF8 antibody was identified. The Santa Cruz IRF8 antibody met the criteria outlined by the ENCODE consortium and was able to detect IRF8 in human primary CD34⁺ Lineage⁻ bone marrow cells, as well as immunoprecipitate the heterologously expressed IRF8 protein in HEK293T cells. The antibody was successfully used for ChIP experiments aimed at the optimisation of sonication settings for chromatin derived from culture-derived DC.

Future efforts should aim at including the Santa Cruz IRF8 antibody in the ENCODE database as a ChIP-validated antibody. Currently (2020), ENCODE lists 5 IRF8 antibodies, of which 2 are partially characterised and none are fully characterised.

Chapter 6. The role of IRF8 in cDC1 and pDC homeostasis

Questions answered in this chapter:

1. How do the functional modules controlled by IRF8 compare in human cDC1s and pDCs?
2. What motifs are bound by IRF8 in cDC1s and pDCs?
3. Does the auto-activation of IRF8 occur in mature human DC?

6.1 Introduction

IRF8 is a transcription factor of the interferon regulatory factor family, expressed by hematopoietic lineages, including DC, B cells, monocytes, and NK cells. Akin to many other members of the IRF family, IRF8 plays several crucial, yet diverse roles, in the lineage determination of immune cells, directing innate immune responses (including regulation of type I IFNs and IFN-inducible genes), controlling cell growth and survival, as well as in oncogenesis (Tamura et al., 2008). As a transcription factor, IRF8 is able to regulate gene transcription in a positive or negative manner (Huang et al., 2007; Nelson et al., 1993). Regulation of the ability of IRF8 to activate or repress gene transcription is achieved via the association of IRF8 with different binding partners, in order to bind to certain promoter elements (Figure 6.1). IRF8 associates with other IRFs (such as IRF1 and IRF2) to bind the IFN-stimulated response element (ISRE). In this case, IRF8 predominantly represses the transcription of type I IFNs and IFN-stimulated genes (Tamura et al, 2015). IRF8 activates transcription via the association with Ets transcription factors (e.g. PU.1/SPI1 or SPIB) to bind the Ets–IRF composite element (EICE) or the IRF–Ets composite sequence (IECS). IRF8 also promotes gene activation via the formation of the IRF8–BATF3–JUN complex, which binds to activating protein-1 (AP-1)–IRF composite element (AICE) motifs (Murphy et al, 2013).




Molecules	Motif name	Function
IRF8-PU.1 	EICE, IECS	Activation
IRF8-AP1 	AICE	Activation
IRF8-IRF1,2 	ISRE	Repression?

Figure 6.1. Transcriptional regulation by IRF8 and its interacting proteins (Tamura et al, 2015).

Transcription factor IRF8 is able to regulate or repress gene expression by associating with different binding partners. Association of IRF8 with Ets family transcription factors (e.g. PU.1/SPI1) or transcription factor AP-1 (formed by the association of bZIP domain containing proteins BATF or BATF3 and JUN) leads to transcription activation. The interaction of IRF8 with other members of the IRF family often leads to negative regulation of transcription.

From murine studies, IRF8 is known to be vital for the development of mononuclear phagocytes and crucial for driving lineage specification. The *Irf8*^{-/-} mouse displays impaired immunity against viral infections, and has absent cDC1s and monocytes, reduced pDCs, expanded granulocyte precursors, and unaffected cDC2s (Turcotte et al., 2004; Sichien et al., 2016). cDC1s are the only cells sensitive to IRF8 haploinsufficiency, as one copy of *Irf8* is sufficient to support monocyte and pDC development (Sichien et al., 2016). IRF8 is required at several steps of cDC1 development and is a terminal selector for this DC subset, maintaining end stage differentiation through its interaction with BATF3 (Grajales-Reyes et al., 2015). IRF8 was shown not to be essential for pDC development, however it controls several functional modules in differentiated pDCs, such as production of type I IFNs and antigen presentation (Sichien et al., 2016). At the progenitor stage, IRF8 cooperates with PU.1/SPI1 to induce the expression of monocyte-related genes including the critical transcription factor KLF4 (Kurotaki et al, 2013). Crucially, IRF8 inhibits CEBPA promoter activity to block the neutrophil differentiation programme (Kurotaki et al,

2014). However, IRF8 is not required to maintain terminally differentiated monocytes (Sichien et al., 2016).

The mechanisms of induction of *Irf8* expression driving DC lineage specification have recently been described in mouse. Two distinct enhancers were identified within the *Irf8* super-enhancer, located at 32 kb and 41 kb downstream of the *Irf8* transcriptional start site (TSS) (Grajales-Reyes et al., 2015). The +32-kb enhancer is selectively active in mature cDC1s and contains several AICE motifs, bound by *Irf8* and *Batf3*, suggesting that this enhancer might support *Irf8* expression through auto-activation (Durai et al., 2019). The +41-kb *Irf8* enhancer is active in differentiated pDCs, and is transiently accessible in cDC1 progenitors. Several E-box motifs are found within the +41-kb *Irf8* enhancer, suggesting that E-proteins (e.g. E2-2/TCF4, a TF essential for pDC development, Cisse et al, 2008), could use this enhancer to drive *Irf8* expression in pDCs (Durai et al., 2019). Both enhancers are required at different stages of cDC1 development, and the switch between them is crucial for this lineage. Bagadia et al. (2019) proposed that this switch is controlled by Zeb2-Id2-Nfil3 interactions that facilitate the development of cDC1s and maintain *Irf8* expression.

The majority of the insights into the role of IRF8 *in vivo* have been derived from murine studies, as genetically modified mouse models are powerful tools for exploring DC development and function, while human studies rely primarily on *in vitro* systems. The recent discovery of patients with primary immunodeficiency, caused solely by mutations in the *IRF8* gene (Hambleton et al., 2011; Bigley et al. 2018; Cytlak and Resteu, 2020), shed some light on the role of IRF8 in human and revealed a phenotype broadly reminiscent of *Irf8* deficient mice. In humans, biallelic *IRF8* mutations cause complex immunodeficiency with myeloproliferation, absence of all monocytes and DCs, B and T cell defects, leading to susceptibility to viral and intracellular infections (Hambleton et al., 2011; Bigley et al, 2018). As in murine systems, the effects of IRF8 in human are gene dose-dependent, and heterozygous *IRF8* mutations lead to cDC1 and often pDC deficiency, mild-moderate CD14⁺ monocytosis and NK cell defects, causing potential susceptibility to Epstein Barr Virus and intracellular organisms (Bigley and Collin, 2020; Cytlak and Resteu, 2020).

Work by the Human DC Lab has previously described two patients with bi-allelic *IRF8* mutations (homozygous *K108E* mutation and compound heterozygous

R83C/R291Q mutation) with a complete lack of monocytes and DCs (Hambleton et al., 2011; Bigley et al., 2018). *K108E* mutation results in loss of nuclear localization and transcriptional activity, as well as decreased protein stability (Salem et al., 2014). The *R291Q* alteration is orthologous to *R294*, mutated in the BXH2 mouse, which exhibits an immunodeficient phenotype similar to the IRF8 knockout mouse. *R83C* shows reduced nuclear translocation, and neither *R291Q* nor *R83C* mutant was able to regulate the Ets/IRF composite element or ISRE, while *R291Q* retained BATF-JUN interactions *in vitro* (Bigley et al., 2018). Together, the heterozygous parents of these patients, and a newly described kindred with an autosomal dominant phenotype due to a dominant negative *IRF8*^{V426fs} mutation, represent an allelic series of IRF8 activity (Figure 6.2).

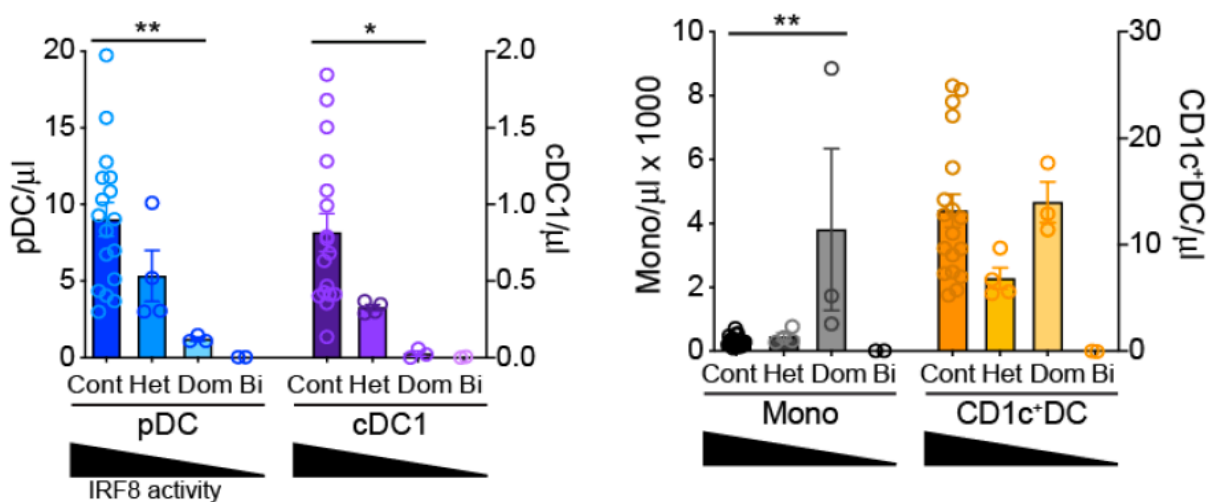


Figure 6.2. Quantification of blood DC and monocyte populations of subjects carrying IRF8 mutations (Cytlik and Resteu et al., 2020).

Cont n=25, Het n=4 (*IRF8*^{R83C}, *IRF8*^{R291Q} and two subjects carrying heterozygous *IRF8*^{K108E} mutations), Dom n=3 (*IRF8*^{V426fs}), Bi n=2 (*IRF8*^{R83C/R291Q} and *IRF8*^{K108E/K108E}). Bars show mean±/ standard error of the mean, circles represent individual subjects. P values derived from two tailed Mann Whitney U testing (*p<0.05, **p<0.01). Cont – healthy controls; Het – patients with heterozygous IRF8 mutation; Dom – patients with heterozygous dominant IRF8 mutation; Bi – patients with homozygous IRF8 mutation.

This chapter aimed to elucidate the role of IRF8 in the homeostasis of human pDCs and cDC1s by analysing its DNA binding patterns in mature cells. Limited by technical restraints related to cell number, this was made possible by the generation of DC from primary stem/progenitor cells *in vitro* using the OP9/OP9-DL1 systems, described in detail in Chapter 3 of this thesis and by Kirkling and Cylak et al. (2018). This approach was employed to produce sufficient numbers of *bona fide* mature cDC1s and pDCs in order to perform chromatin immunoprecipitation followed by high-throughput DNA sequencing. Identification of IRF8 binding sites, related DNA motifs and genes in close proximity drove the exploration of pathways regulated by IRF8 and its binding partners in human.

6.2 Materials and methods

6.2.1 In vitro generation of cDC1s and pDCs from human bone marrow progenitors

Bone marrow cells from hip replacement procedures were collected, stored and defrosted as described in Chapter 2 subsections 2.3.2 and 2.3.3. The BMDC were stained and FACS-purified as described in Chapter 2 subsections 2.3.6 using antibodies listed in Table 3.1. CD34⁺ Lineage⁻ cells were retained for culture. To produce mature cDC1, FACS-purified CD34⁺ Lineage⁻ progenitors were co-cultured with OP9-DL1 cells for 21 days, supplemented with 20 ng/ml granulocyte-macrophage colony-stimulating factor (GM-CSF, R&D systems), 100 ng/ml FLT3-ligand (Immunotools), and 20 ng/ml stem cell factor (SCF, Immunotools), as described in Chapter 2 subsection 2.3.5. For the production of pDC, FACS-purified CD34⁺ Lineage⁻ progenitors were co-cultured with OP9 for 21 days, supplemented with a modified cytokine cocktail, as data from members of the lab (not shown) showed that lower FLT-ligand concentrations promoted pDC production. Therefore, for the generation of pDC, the following cytokine cocktail was used: 20 ng/ml granulocyte-macrophage colony-stimulating factor, 20 ng/ml FLT3-ligand and 20 ng/ml stem cell factor.

6.2.2 Chromatin immunoprecipitation, DNA extraction, and sequencing library preparation

After 21 days of culture, OP9-DL1-derived cDC1s and OP9-derived pDCs were FACS-purified, as described in Chapter 3, and fixed with a formaldehyde mix, as

described in the methods for Chapter 5. Sonication of lysates following fixation was performed with the Bioruptor Pico (Diagenode) using the 6 cycles of 30 seconds ON and 90 seconds OFF setting, established to be best in the optimisation experiments. The total volume for each lysate (200 μ l) contained an average of 93,000 cells per replicate. Of this, 190 μ l were used for immunoprecipitation, and 10 μ l were used as input chromatin control.

The Low Cell ChIP-Seq kit and Next Gen DNA Library Kit (Active motif) were used for cell fixation, immunoprecipitation, DNA extraction, and sequencing library preparation, according to manufacturer's instructions and as described in detail in Chapter 5. 3 μ g IRF8 antibody (Santa Cruz (E-9): sc-365042) were used for immunoprecipitation. The input chromatin controls for each sample were prepared by extracting DNA from 10 μ l of sonicated lysate, as suggested by Dr Stefan Dillinger (Active motif). Lysates were transferred to PCR tubes and mixed with 70 μ l TE pH 8.0 (Active motif) and 0.4 μ l RNase A (Invitrogen) via vortexing. The mix was incubated in a thermocycler at 37°C for 30 minutes. 0.8 μ l Proteinase K (Active motif) were added to each tube. Following a vortexing step, the tubes were incubated in a thermocycler at 55°C for 30 minutes. The temperature was then increased to 80°C for 2 hours. Each chromatin input was transferred into a 1.5ml microcentrifuge tube and mixed with 33 μ l of 5M ammonium acetate (Invitrogen), 0.8 μ l carrier (Active motif), and 300 μ l absolute ethanol via vortexing. Tubes were chilled at -80°C for 30 minutes, then spun at 4°C in a microcentrifuge at 21,000xg for 15 minutes. The supernatant was carefully removed, and the pellet was washed with 500 μ l 70% ethanol and spun at 4°C in a microcentrifuge at 21,000xg for 5 minutes. The supernatant was removed, taking care not to disturb the pellet. Residual ethanol was removed with a P10 pipette. Tubes were left uncapped and air dried for 15 minutes. When the pellets dried, 40 μ l Low EDTA TE (Active motif) were added to each tube and incubated at room temperature for 10 minutes, then vortexed to ensure the pellet was completely resuspended. The solution, containing the input chromatin, was used for library preparation with the Next Gen DNA Library Kit (Active motif). Following library preparation for the immunoprecipitation reactions and the controls, the DNA concentration, molarity, and fragment size distribution of the sequencing libraries was assessed with the Agilent Bioanalyzer 2100 using the high sensitivity DNA kit (Agilent).

6.2.3 Generation and processing of ChIP-Seq data

The Illumina NextSeq 500 platform at the Genomics Core Facility, Newcastle University, was employed to generate 150bp paired-end reads, averaging 27.7 million per sample. The reads were demultiplexed with bcl2fastq version 2.19.0.316 (Illumina). Adapter trimming, removal of low quality bases, and exclusion of short reads (below 70bp) were performed with Cutadapt version 1.9.1 (Martin, 2011) and FastQC version 0.11.2 (Simon Andrews, Babraham Institute) via the Trim Galore wrapper version 0.4.3 (Felix Krueger, Babraham Institute). Reads were aligned to the human reference genome GRCh38p12 (release29) with the BWA-MEM algorithm as part of the BWA software version 0.7.15 (Li, 2013). The SAM to BAM conversion was achieved with samtools version 1.3 (Li et al., 2009). PCR duplicates were removed using a Perl script provided by Active Motif, as part of the Low ChIP-Seq protocol. Peak calling was performed with MACS version 2.1.0.20150731 (Zhang et al., 2008) against input chromatin controls, and employed a q-val cut-off of 0.05. Additionally, all peaks identified on the mitochondrial chromosome were filtered out, as the IRF8 protein is localised exclusively in the nucleoplasm (The Human Protein Atlas, <https://www.proteinatlas.org>), and therefore is not expected to directly regulate mitochondrial gene expression. Peaks mapped to chromosome Y were also removed from analysis, as all bone marrow donors used in this experiment were female. Further analysis was undertaken in R version 3.6 (R Core Team, 2019).

6.2.4 Differential binding analysis

The DiffBind R package version 2.12.0 (Ross-Innes et al., 2012) was used for counting reads within each peak (peak width was set to 500bp) and determining the consensus peaks present in at least two biological replicates of cDC1s and pDCs, as well as for differential ChIP-Seq analysis, including PCA and hierarchical clustering using the complete linkage method based on the Pearson's distances. The Venn diagram displaying the overlap between cDC1s and pDCs consensus peaks was produced with the VennDiagram package version 1.6.20 (Chen, 2018).

6.2.5 Peak annotation and pathway enrichment analysis

The consensus peaks were annotated with the ChIPseeker package version 1.20.0 (Yu et al., 2015). Gene symbols were converted into Entrez Gene IDs with the "translate" function from the AnnotationFuncs package version 1.34.0. Pathway

enrichment analysis for genes regulated by IRF8 in cDC1s and pDCs (with IRF8 peaks located within the promoter region, +/-3kb from TSS), was performed with ClusterProfiler version 3.12.0 (Yu et al., 2012) against the Reactome biological pathways (Jassal et al., 2020).

6.2.6 Motif analysis

The GimmeMotif software (Bruse and van Heeringen, 2018) was used for motif enrichment analysis and for known motif scanning. The motif enrichment analysis in cDC1s and pDCs was undertaken with the “gimme motifs” command, using 500bp-wide consensus peaks located within the promoter region of genes. The enrichment was performed against the GimmeMotifs database of known vertebrate motifs and transcription factors. The p-values for each motif were determined with the hypergeometric/Fisher’s exact test. The motif scan for the consensus peaks identified in cDC1s within +/-100kb of the IRF8 TSS was performed via the “gimme scan” command. The search was performed against the GimmeMotifs database and displays known bZIP, Ets, and IRF motifs identified within the 500bp-wide peaks.

6.3 Results

6.3.1 Overview of the IRF8 ChIP-Seq dataset

To interrogate the IRF8 binding sites in human, an IRF8 ChIP-Seq experiment (optimised in Chapter 5) was performed using *in vitro*-derived cDC1s and pDCs as described in Chapter 3. Based on all IRF8 binding sites detected via ChIP-Seq, the two cell types showed a clear separation. This was confirmed both by hierarchical clustering and PCA (Figures 6.3 A and 6.3 B, respectively). These analyses also suggested that the cDC1 biological replicates bear more similarity to each other than the pDC, as they have higher positive correlation between the samples, and cluster more tightly together on the PCA plot.

In order to further explore the differences, the number of peaks shared by the biological replicates of the two cell types were interrogated (Figure 6.4). Overall, a much higher number of peaks was noted in cDC1, with over 11,000 peaks present in all cDC1 replicates, in contrast to only 212 peaks shared by all pDC. In addition, a higher degree of overlap was evident in cDC1, which shared the majority of their peaks, while most pDC replicates contained mainly sample-specific peaks.

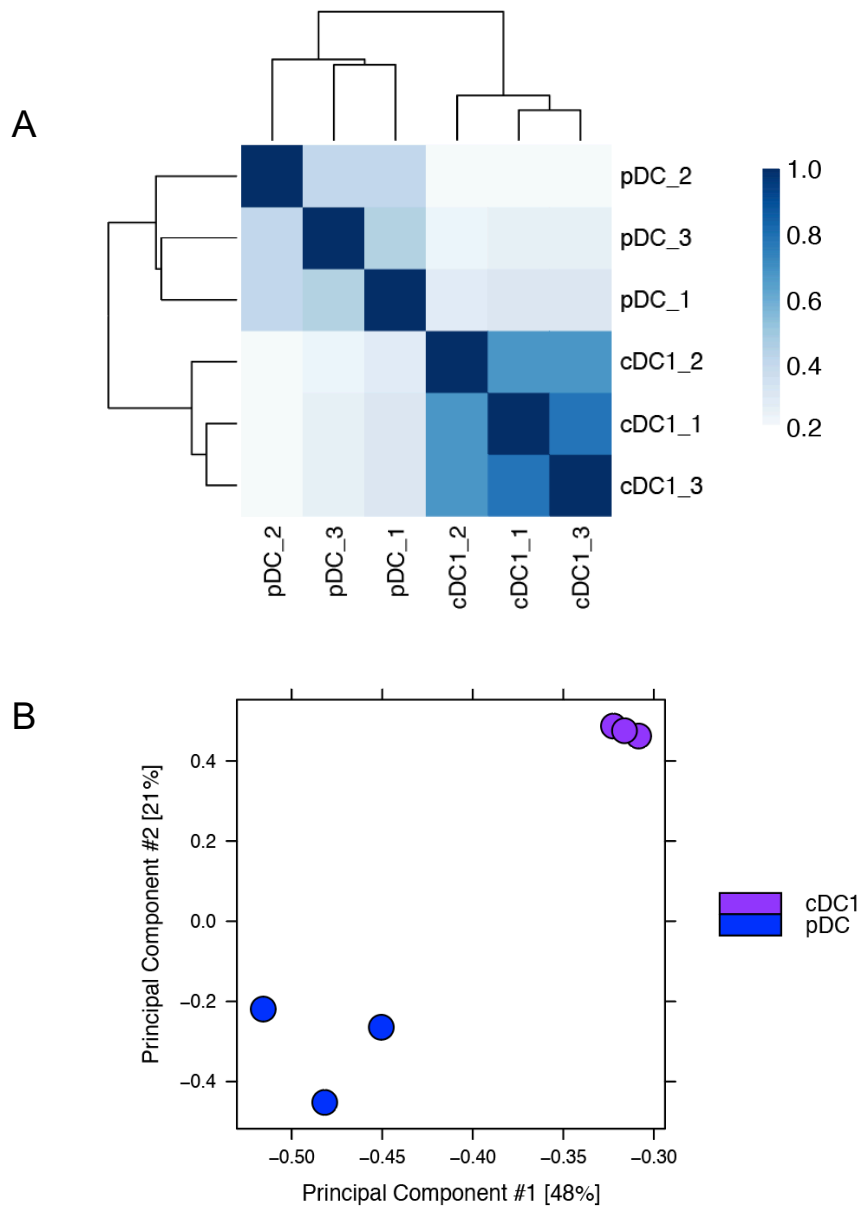


Figure 6.3. Overview of the generated IRF8 ChIP-Seq dataset comprised of culture-derived cDC1s and pDCs with 3 biological replicates per cell type.

A. Heatmap and clustering of the ChIP-Seq samples based on the read counts at all sites. The dendrogram positions together the samples with similar counts at binding sites. The colours of the heatmap represent the correlation between the samples, dark blue indicating high positive correlation, and light blue to white - low correlation.

B. Principal component analysis based on the counts at all binding sites. The percent variance explained by each principal component is indicated in square brackets.

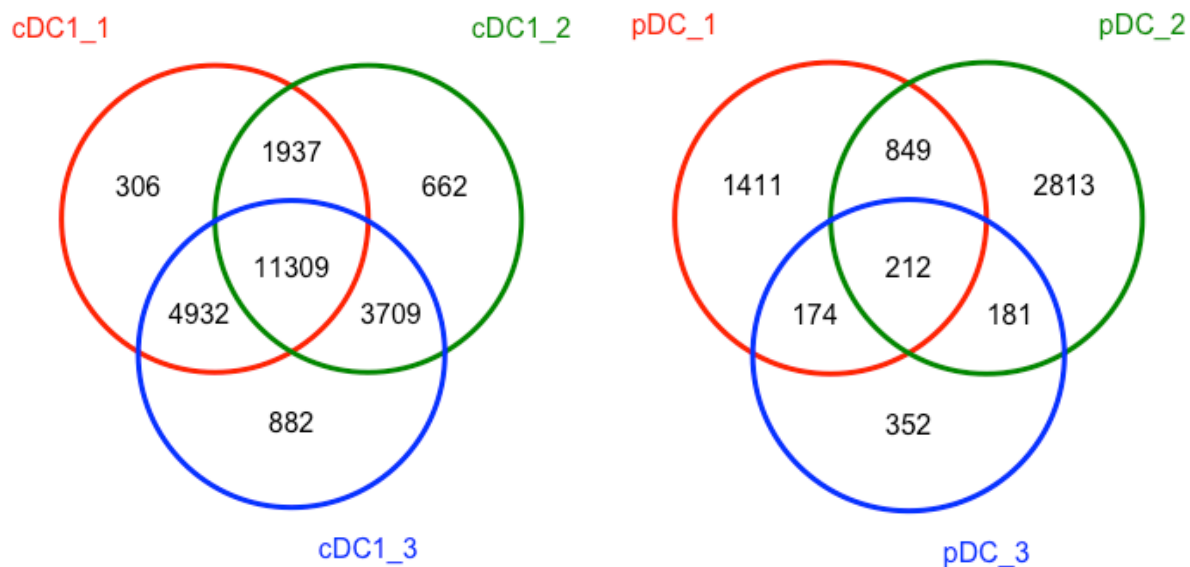


Figure 6.4. Venn diagrams showing IRF8 binding site overlaps in biological replicates of cDC1 (left) and pDC (right).

The peaks were called with MACS2 (Zhang et al., 2008) against input chromatin controls corresponding to each sample. Peaks with q -values below 0.01 were processed, counted, and visualised within the DiffBind package in R (Ross-Innes et al., 2012).

To focus on the high confidence data, peaks called with MACS2 ($qval < 0.05$) and present in at least two out of the three biological replicates, also known as consensus peaks, were used for further analysis. Over 20,000 peaks were identified in at least two of the cDC1 samples, 1,027 of which were shared with pDC. The number of pDC-specific peaks was lower, as only 389 peaks were identified in at least two of the pDC replicates (Figure 6.5).

The consensus peaks identified in the IRF8 ChIP-Seq dataset were annotated with the ChIPseeker R package (Yu et al., 2015). Peak annotation bar charts (Figure 6.6) were used to explore the distribution of peaks. In both cDC1s and pDCs, more than half of the peaks were located in enhancer regions (distal intergenic and intronic regions), and around 20% of the peaks were found in promoter regions, a profile common in ChIP-Seq and ATAC-Seq datasets (Yan et al., 2020).

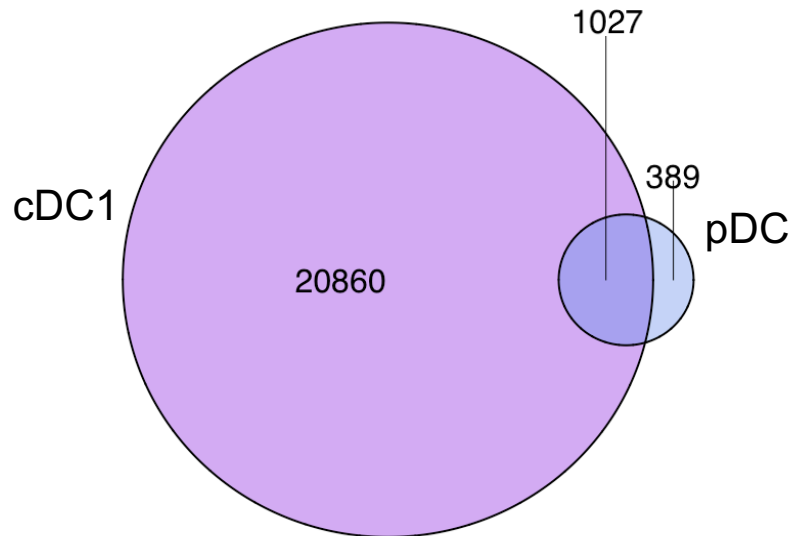


Figure 6.5. Venn diagram displaying the overlap between consensus peak regions in cDC1s and pDCs.

The diagram was constructed using peaks called with MACS2 ($qval < 0.05$) in at least 2 of the 3 biological replicates for each of the cell types. 73% of pDC peaks were shared with cDC1, while 5% of cDC1 peaks were shared with pDCs.

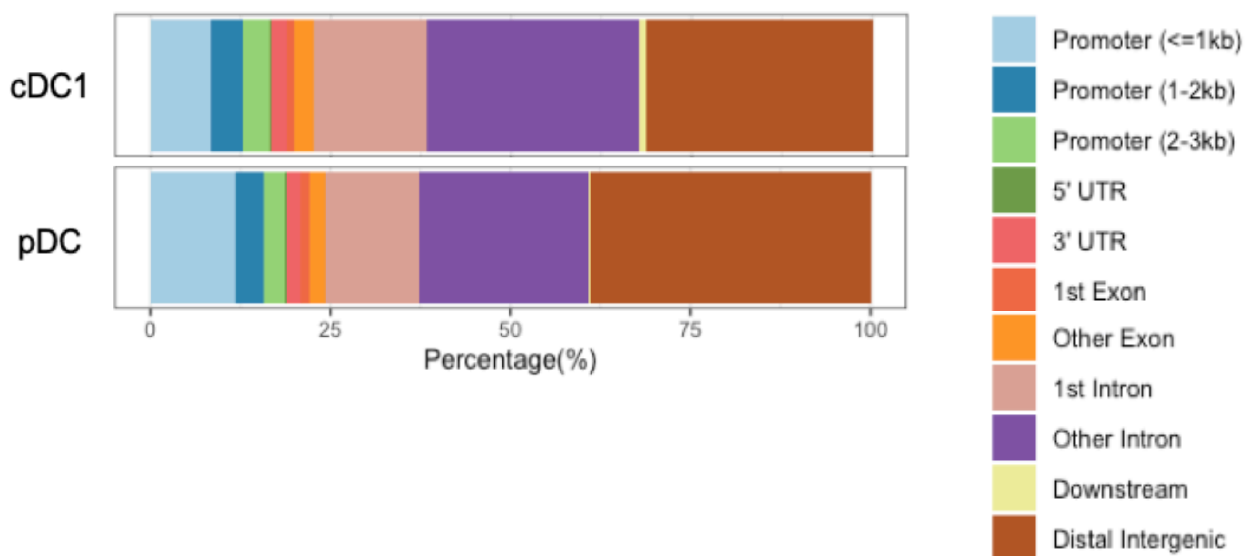


Figure 6.6. Visualisation of the genomic annotation for the IRF8 binding peaks in cDC1s and pDCs.

Annotation was performed with the ChiPseeker R package (Yu et al., 2015) for peaks identified in two out of the three biological replicates for each of the cell types.

6.3.2 IRF8 differentially regulates immune pathways in mature cDC1s and pDCs

To determine the pathways regulated by IRF8 and mature cDC1s and pDCs, the genes with IRF8 peaks in the promoter region (+/-3kb from TSS) were selected for pathway enrichment analysis. The analysis revealed an array of functional modules regulated by IRF8 in cDC1s and pDCs (Figure 6.7 and Table 6.1). The vast majority of the pathways were linked to the function of the immune system. Neutrophil degranulation was the most significant among the commonly regulated pathways, both cell types containing a large number of genes from this pathway with IRF8 peaks in the promoter region. Interferon signaling was also shared by both cell types, and interferon gamma signaling in particular was present in both cell types. Interferon alpha/beta signaling, however, emerged to be the only pDC-specific pathway regulated by IRF8. cDC1-specific pathways included numerous adaptive immune system features, such as MHC II antigen presentation, costimulation by the CD28 family, and immunoregulatory interactions between a lymphoid and non-lymphoid cell, as well as innate immune functions, such as pattern recognition through NLRs, CLRs, and TLRs (within the MyD88 deficiency pathway). Remarkably, IRF8 itself was contained within the genes it regulates in cDC1s (Table 6.1), suggesting potential for IRF8 auto-activation in these cells, but not in pDC.

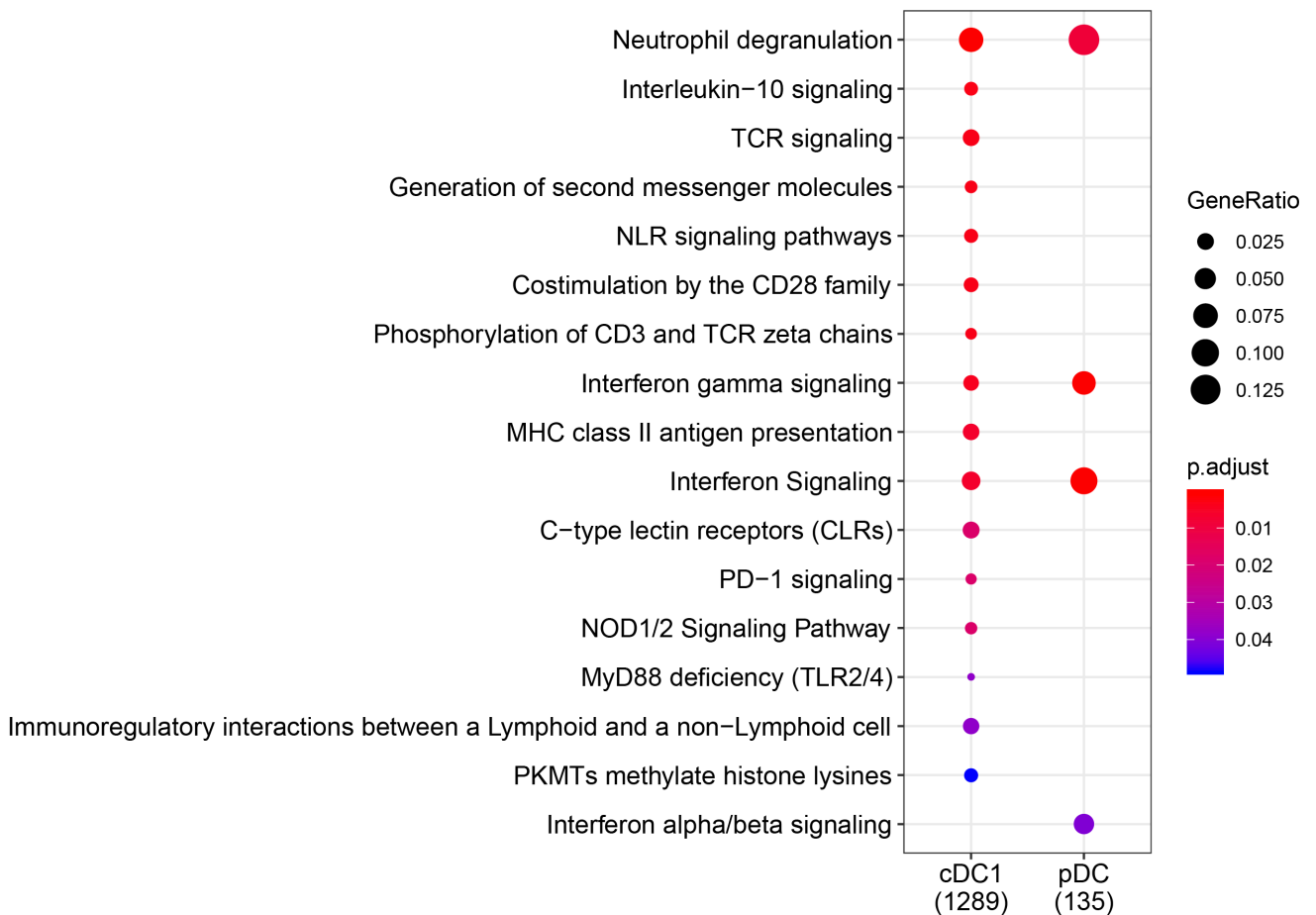


Figure 6.7. Pathway enrichment analysis for genes regulated by IRF8 in cDC1s and pDCs.

The dot plot displays the Reactome biological pathways enriched in one or both cell types (adjusted p -value < 0.05). The pathway enrichment analysis was performed using genes that have at least one peak identified in the promoter region (\pm 3kb from the TSS). The number of identified genes in each category is displayed in parentheses. The size of the dots represents the GeneRatio (the overlap between genes in each list and the genes associated with a Reactome term), and the colour represents the adjusted p -values (calculated based on the hypergeometric model and adjusted using the Benjamini-Hochberg method).

Group	Pathway	Genes with IRF8 peaks near promoter
cDC1	Neutrophil degranulation	CD53, FCER1G, SELL, CHIT1, NFASC, PTPRJ, CCT2, APAF1, SLC15A4, PECAM1, CD300A, XRCC5, PSMD1, GM2A, HK3, PSMC2, DDX3X, PSMA5, MNDA, NCSTN, TMEM63A, PLAUI, CAT, CD44, CRACR2A, C3AR1, METTL7A, GNS, LYZ, MLEC, PNP, LGALS3, ATP6V1D, ATP8B4, ITGAL, IST1, ATP6V0A1, CEACAM6, CEACAM3, FPR2, TMBIM1, CTSZ, ADA2, TOM1, RAB7A, ATP8A1, NFKB1, IQGAP2, CD14, RAB24, VNN1, GUSB, MGAM, GSN, CYBB, RHOG, GSTP1, GLIPR1, SERPINA1, CXCR1, GLB1, GYG1, DIAPH1, MAPK14, SLC2A5, CAP1, CTSS, FCGR2A, PTPRC, CD55, CD59, COMMD9, CTSC, CLEC12A, LRMP, NCKAP1L, RNASE3, SNAP23, CTSH, HMOX2, CD68, GRN, SERPINB10, ADGRE5, SIRPB1, CCT8, CD47, ACPP, TMED7-TICAM2, ENPP4, RNASET2, FGL2, CD36, DEFA4, PRSS3
cDC1	Interleukin-10 signaling	IL18, CCL4L1, IL1R2, IL1RN, CXCL8, IL6, JAK1, CCR1, CCR5, CD80, CD86, IL10, CCL3L1, IL1R1, IL1B, CCR2, TNF, CCL19
cDC1	TCR signaling	PTPRJ, PAK1, CD4, PSMD9, CSK, PSMD1, PSMC2, BCL10, PSMA5, PTPN22, PSMA3, EVL, PDPK1, PSMA8, NFKB1, LCP2, HLA-DRA, HLA-DPB1, CD101, MALT1, GRAP2, FBXW11, PTPRC, UBC, PSMC6, VASP, FYB1, HLA-DRB5, HLA-DRB1, HLA-DQA1, HLA-DPA1, TAB2
cDC1	Generation of second messenger molecules	PAK1, CD4, EVL, LCP2, HLA-DRA, HLA-DPB1, CD101, GRAP2, VASP, FYB1, HLA-DRB5, HLA-DRB1, HLA-DQA1, HLA-DPA1
cDC1	Nucleotide-binding domain, leucine rich repeat containing receptor (NLR) signaling pathways	CASP1, NLRP1, CASP8, NLRP3, MAP2K6, NFKB1, TAB3, BCL2, MAPK14, TXNIP, AIM2, BIRC3, CASP4, UBC, NOD2, CYLD, TNFAIP3, TAB2, NOD1
cDC1	Costimulation by the CD28 family	PAK1, CD4, CSK, MAP3K14, FYN, PPP2R5C, PDPK1, RICTOR, HLA-DRA, HLA-DPB1, GRAP2, CD80, CD86, PPP2R5A, CTLA4, HLA-DRB5, HLA-DRB1, HLA-DQA1, HLA-DPA1, CD274, PDCD1LG2, MAPKAP1
cDC1	Phosphorylation of CD3 and TCR zeta chains	PTPRJ, CD4, CSK, PTPN22, HLA-DRA, HLA-DPB1, PTPRC, HLA-DRB5, HLA-DRB1, HLA-DQA1, HLA-DPA1
cDC1	Interferon gamma signaling	IRF9, CIITA, PTPN2, TRIM5, CD44, SP100, IFNGR2, IRF2, HLA-DRA, HLA-DPB1, IRF5, JAK1, TRIM68, OAS3, JAK2, VCAM1, CAMK2G, TRIM22, OAS1, IRF8, TRIM38, TRIM10, HLA-DRB5, HLA-DRB1, HLA-DQA1, HLA-DPA1
cDC1	MHC class II antigen presentation	CTSE, ARF1, AP2B1, DYNC1I2, CD74, HLA-DMB, DCTN2, TUBB6, RAB7A, SEC31A, SEC24D, HLA-DRA, HLA-DPB1, AP1B1, DCTN4, CAPZB, CTSS, CTSC, KIF5A, LGMN, KIF23, CTSH, AP1S3, KIF3B, KIF20A, NUP160, IRF9, CIITA, IFI35, PTPN2, EIF4G3, TRIM5, CD44, ARIH1, NUP88, SP100, SAMHD1, IFNGR2, NUP50, IRF2, HLA-DRA, HLA-DPB1, IRF5, JAK1, IFIT3, IFIT1, TRIM68, OAS3, JAK2, IFN1, VCAM1, ADAR, RNASEL, CAMK2G, IFIT2, TRIM22, AAAS, OAS1, UBC, IRF8, MX2, TRIM38, TRIM10, HLA-DRB5, HLA-DRB1, HLA-DQA1, HLA-DPA1, POM121, DDX58
cDC1	C-type lectin receptors (CLRs)	FCER1G, PAK1, PSMD9, NFATC3, MAP3K14, ICAM2, CASP8, PSMD1, FYN, PSMC2, PRKACB, BCL10, PSMA5, CLEC4A, PSMA3, PDPK1, PSMA8, NFATC2, ITPR1, NFKB1, TAB3, MALT1, FBXW11, CLEC6A, CD4, CSK, HLA-DRA, HLA-DPB1, HLA-DRB5, HLA-DRB1, HLA-DQA1, HLA-DPA1, CD274, PDCD1LG2
cDC1	PD-1 signaling	CD4, CSK, HLA-DRA, HLA-DPB1, HLA-DRB5, HLA-DRB1, HLA-DQA1, HLA-DPA1, CD274, PDCD1LG2
cDC1	NOD1/2 Signaling Pathway	CASP1, CASP8, MAP2K6, TAB3, MAPK14, BIRC3, CASP4, UBC, NOD2, CYLD, TNFAIP3, TAB2, NOD1
cDC1	MyD88 deficiency (TLR2/4)	TLR6, CD14, TLR1, CD36, LY96, TLR4
cDC1	Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	SLAMF7, SELL, ITGB7, ICAM2, CD300A, CD300LB, LILRA4, COL2A1, ITGAL, CD300E, CD300LF, LILRA2, LILRB4, CD1C, JAML, KLRD1, VCAM1, SLAMF6, FCGR2B, ITGB1, CRTAM, KLRG1, CLEC2B, KLRC4-KLRK1, CD300LD, CD226, SIGLEC7, CD96, CD200R1, TREML2
cDC1	PKMTs methylate histone lysines	KMT5B, ATF7IP, NSD2, KMT2C, NFKB1, KMT2E, EHMT1, RBBP5, SMYD3, SUV39H2, SETD6, SETD2, HIST1H4B, HIST1H4C, HIST1H4D, HIST1H4E, HIST1H4H, HIST1H4L, ASH2L
pDC	Interferon Signaling	NUP205, TRIM22, OASL, GBP3, CAMK2G, IFIT3, OAS1, OAS3, TRIM38, HLA-DRA, IFNA16, IFIT2, PTPN2
pDC	Interferon gamma signaling	TRIM22, OASL, GBP3, CAMK2G, OAS1, OAS3, TRIM38, HLA-DRA, PTPN2
pDC	Neutrophil degranulation	PTPRC, VAMP8, NFKB1, GM2A, EEF1A1, CAP1, LRMP, CTSS, SLC15A4, IST1, SIGLEC9, XRCC5, SIRPB1, ADA2, TMED7-TICAM2, METTL7A, GNS, CD36
pDC	Interferon alpha/beta signaling	OASL, IFIT3, OAS1, OAS3, IFNA16, IFIT2

Table 6.1. Biological pathways enriched in cDC1s and pDCs (adjusted p-value<0.05).

The table displays cells type, name of pathway in the Reactome database, as well as a list of genes from each pathway with IRF8 peaks within +/-3kb of TSS.

Next, motif analysis was undertaken to investigate the role of IRF8 co-operating factors and binding partners in cDC1 and pDC homeostasis. The majority of the sequences bound by IRF8 in both cell types were Ets and IRF motifs (Table 6.2 and Table 6.3) and a very similar top motif was shared. This Ets/C2H2 motif, most significantly enriched in both cell types, is known to be bound by IRF8 and IRF4 (a member of the IRF family), as well as their binding partners from the Ets family: SPI1 and SPIB. Further motifs bound by IRF8 (motif GM.5.0.IRF.0013), and other members of the IRF family (GM.5.0.IRF.0010 and GM.5.0.IRF.0020) were identified in both cell types. In addition, multiple Ets motifs, bound by transcription factors ETS1, ETS2, ELF1, ELK1, ETV6 and ETV7, were shared. A number of differentially bound sequences were identified, including an IRF motif thought to be bound by IRF4 (GM.5.0.IRF.0017), strongly enriched in pDCs but not observed in cDC1s, in keeping with the known expression of IRF4 in pDCs but its absence in cDC1. bZIP motifs, bound by AP-1 forming transcription factors BATF3, BATF and JUN, were found exclusively in cDC1. In mice, the cooperation of Batf3 and Jun with IRF8 is crucial for maintaining the expression of *Irf8* by auto-activation during the development of cDC1s (Grajales-Reyes et al., 2015). This prompted further investigation of the IRF8 auto-activation in human cDC1s.

	direct	indirect or predicted	logo	log10 P-value
GM.5.0.Ets.0010	SPIB,Sfpi1,SPI1,Spi1,Spib	Spi1,SPIC,Spib		152.736
GM.5.0.MADS_box.0005		MEF2C,PAX5,BCL11A,MEF2A,RXRA		144.741
GM.5.0.bZIP.0021	BATF3,Batf3	JUND		68.557
GM.5.0.IRF.0013	IRF8			53.990
GM.5.0.Ets.0039	ELF1			52.852
GM.5.0.Ets.0011	ETS2,ETS1			41.452
GM.5.0.bZIP.0010	BATF,JUN	IRF4,TCF12,RXRA		32.401
GM.5.0.Ets.0017	ETV7	ETV7		28.278
GM.5.0.C2H2_ZF.0085	PRDM1	PRDM1		19.181
GM.5.0.bZIP.0058	NFE2L2			14.408
GM.5.0.Ets.0007	ELK1,ETV4,Etv4	Fev,Erf,Etv2,Ets2		13.561
GM.5.0.bZIP.0086	MAF			12.164
GM.5.0.IRF.0020	IRF5	IRF5		10.919
GM.5.0.Paired_box.0005		PAX5		10.369
GM.5.0.Ets.0021	ELF1	Elf1,ELF1		8.893
GM.5.0.Mixed.0020		IRF4,BATF		7.496

Table 6.2. Known motifs enriched (adjusted p-value <0.005) in cDC1s within the IRF8 binding peaks.

Peaks located in promoter regions were used for analysis. The table shows the motif name from the GimmeMotifs database of known vertebrate motifs and transcription factors known to bind them directly, and indirect or predicted factors. The p-value for the motif, determined with the hypergeometric/Fisher's exact test, and its sequence logo are shown. The height of the characters in the logo indicates their conservation.

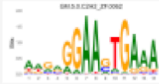

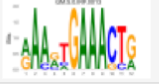
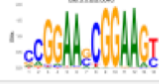
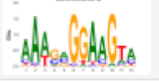


	direct	indirect or predicted	logo	log10 P-value
GM.5.0.C2H2_ZF.0062	BCL11A,Irf4,Irf8,IRF8,IRF4	Bcl11b,Bcl11a,BCL11B		92.937
GM.5.0.IRF.0017		IRF4		25.203
GM.5.0.IRF.0013	IRF8			24.142
GM.5.0.Ets.0040	ETV6	ETV6		8.244
GM.5.0.Ets.0042	SPIC	SPIC,Spib		8.244
GM.5.0.IRF.0010	IRF5	Irf6,Irf4,IRF5,Irf5		5.259
GM.5.0.Ets.0007	ETV4,Etv4,ELK1	Fev,Erf,Etv2,Ets2		3.863

Table 6.3. Known motifs enriched (adjusted p-value <0.005) in pDCs within the IRF8 binding peaks.

Peaks located in promoter regions were used for analysis. The table shows the motif name from the GimmeMotifs database of known vertebrate motifs and transcription factors known to bind them directly, and indirect or predicted factors. The p-value for the motif, determined with the hypergeometric/Fisher's exact test, and its sequence logo are shown. The height of the characters in the logo indicates their conservation.

6.3.3 cDC1s maintain their cell identity through the auto-activation of IRF8

Having shown the potential of IRF8 in regulating a number of key immune pathways in cDC1, the next step of the analysis focused on the induction and maintenance of IRF8 expression through self-activation. This phenomenon has been previously described in mouse, where enhancer and superhancer regions surrounding the IRF8 gene contain motifs bound by IRF8 and its binding partners (Durai et al., 2019). Since enhancers can be located at great distances upstream or downstream away from the transcription start site of the genes they are regulating, all peaks called in the +/-100 kb region around the IRF8 TSS were visualised (Figure 6.8). No evidence of IRF8 auto-activation was found in mature pDC, as no common peaks were identified across the replicates. In contrast, multitude of peaks were present consistently across at least two replicates in cDC1. Three of the consensus peaks

were located upstream of the IRF8 TSS at -32kb, -25kb, -16kb, one near the 3rd exon at +10kb, and three downstream at +37kb, +59kb, and +68kb.

To further explore auto-activation of IRF8 in cDC1, a motif scan of the consensus peaks for known sites bound by IRF8 and its binding partners was undertaken. The scan included the transcription-activating bZIP (bound by JUN and BATF, forming AP-1) and Ets motifs (bound by ETS family TFs, such as SPI1), as well as IRF motifs (repressive if bound by an IRF8-IRF1,2 complex). Table 6.4 highlights the motifs identified in each of the peaks, together with the sequence bound, as well as transcription factors that are known or predicted to bind them. The motif scan identified multiple sites with potential to act as enhancers, and all seven of the consensus peaks contained at least one of the motifs. The transcription-activating Ets motifs bound by known IRF8 partner SPI1(PU.1), were enriched in the peak at -25kb, which contained eight Ets motifs, and present in the peaks at +10kb and +37kb, suggesting capacity for IRF8 auto-activation at these sites. Transcription-activating bZIP motifs known to be bound by IRF8 partners BATF3, BATF and JUN were identified in the -16kb and +68kb. Notably, the motifs identified in the latter peak were bound by BATF3, a TF which maintains the expression of IRF8 by auto-activation during development of cDC1s in cooperation with IRF8 and JUN (Grajales-Reyes et al., 2015). Further motifs identified by the scan include a sequence bound by IRF3 within the +37 peak and an E-box canonical motif at +59kb.

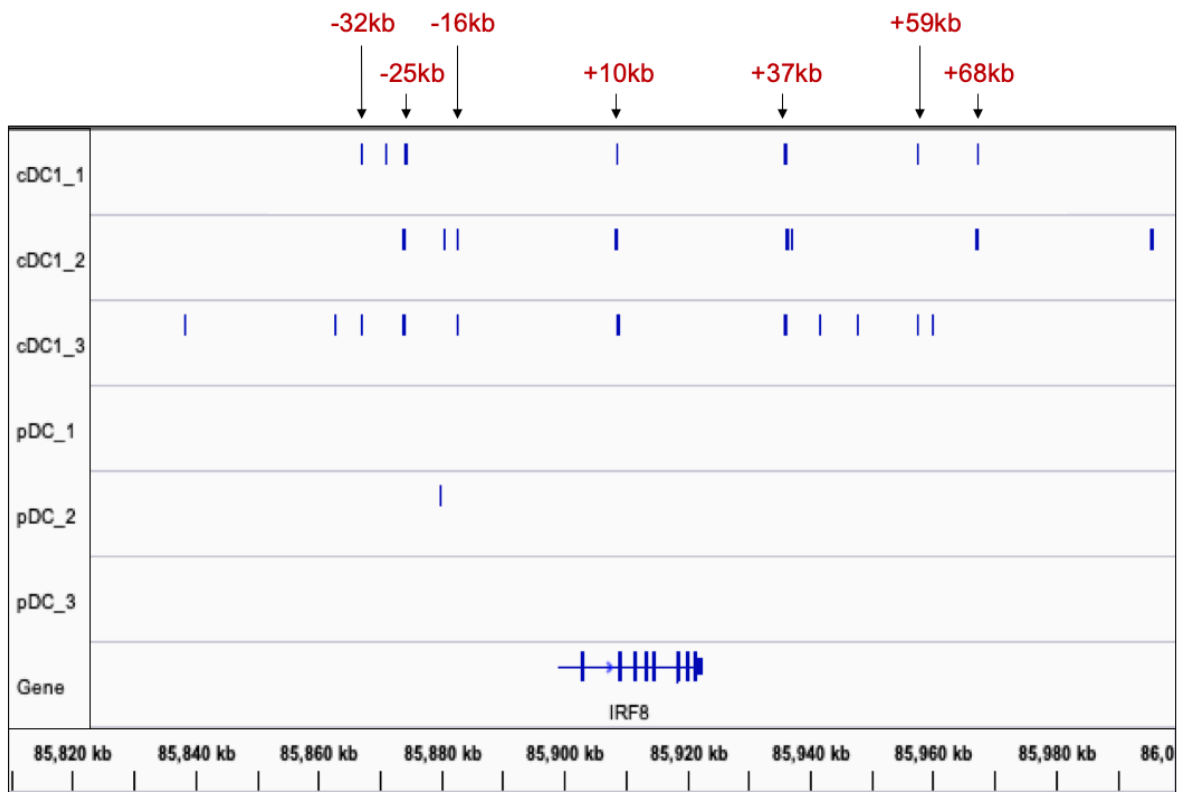


Figure 6.8. Overview of peaks called with MACS2 ($qval < 0.05$) in the region between -50kb and +100kb around the IRF8 TSS in individual cDC1 and pDC samples.

Consensus peaks, present within +/-100kb from IRF8 TSS in at least 2 biological replicates of cDC1s are marked with an arrow and their distance from IRF8 TSS is shown in red. No consensus peaks were observed in cDC1s or pDCs in the region between -100kb and -50kb in relation to the IRF8 TSS (not shown).

Distance from IRF8 TSS	Motif type	Motif instance	Known or predicted factor	Note
-32kb	bZIP	CCACGGCAGCCT	ATF6,CR925796.2,CR753803.1,Atf6,Atf6b,ATF6B,AL662828.6	
		ATTTTTAAATGAG	CEBPG	
-25kb	Ets	CCAAAATACTTCCTCCCA	ELF2,Elf2,Elf1	ETS TF
		ttgacttcctctt	Sfp1,Spi1,SPIC,Spib,SPI1,SPIB	ETS TF
		AAATACTTCCTCCC	ETS1,ETS2	ETS TF
		gacttcctctt	SPI1	ETS TF
		acttcctcttt	SPI1,Spic,SPIB,Spib	ETS TF
		acttcctcttt	ELF1	ETS TF
		gacttcctctt	Spib,SPIC	ETS TF
		ttgacttcctctt	ELK1	ETS TF
-16kb	bZIP	TCTTTTTGATGTCTGCATAGT	JUNB,JUN,JUND	Forms AP-1 TF
		CATGACTTAATG	ATF4	Forms AP-1 TF
	Ets	CTTCTCTTTCC	ELF1,GABPA,ELK4	ETS TF
+10kb	bZIP	TGACTCA	Fos,JUND,Jdp2,Fosb,Junb,Jund,Fos1	Forms AP-1 TF
		CTGACTCAGCT	Bach1,BACH1,Bach2,NFE2,MAFK	
		TTCTGACTCAGCTGTA	Mafg,Maff,Mafk,MAFK,MAFG,MAFF	
		TGACTCAGC	MAF	
		CTGACTCAGCT	NFE2L2	
		CTGACTCAGCT	MAFF,MAFK	
	Ets	TGCAACTTCCTCTTT	Spib,SPI1,SPIB,Sfp1,SPIC,Spi1	ETS TF
		AACTTCCTCTT	SPI1	ETS TF
		ACTTCCTCTTTG	SPI1,Spic,Spib,SPIB	ETS TF
		AACTTCCTCTTT	SPIC,Spib	ETS TF
+37kb	Ets	CAAAGCCACTTCCTCCTG	ELF2,Elf2,Elf1	ETS TF
		CACCTCCTCCT	SPI1	ETS TF
	bZIP	CATCAGTCCAGAGCACCG	ATF3,CREB1	
	IRF	GTTTCCTTTC	Irf3	
+59kb	Ets	CTGGAAGGACACCTG	ETS1	ETS TF
		CCAGGAAATCCTA	Erf,Ets2,Fev,FLI1,Etv2,ERG	ETS TF
		GGTACCAGGAAAT	ELF3	ETS TF
	bZIP	gcacgtggc	ATF3	E-box canonical sequence motif
+68kb	bZIP	AGTAATAAAGAAAC	BATF3,JUND,Batf3	Forms AP-1 TF
		TGAGTAATAAAGAAA	JUND,Batf3,Batf,BATF	Forms AP-1 TF

Table 6.4. Motif scanning results for the 7 consensus peaks from the *cDC1* triplicates, highlighting the bZIP, Ets, and IRF motifs identified with GimmeMotif.

For each peak, the following information is listed: rounded distance in relation to the IRF8 TSS, type of motifs identified by the scan, motif sequence, list of human and murine TF known or predicted to bind the motif, and as well as further notes, highlighting ETS family transcription factors and TFs known to form the AP-1 complex with IRF8, as well as the motif containing the E-box canonical sequence.

The evidence of IRF8 auto-activation, along with the multitude of key immune modules controlled by IRF8 in cDC1, suggested that IRF8 plays a vital role in the homeostasis and function of these cells. To further investigate the role of IRF8 in controlling the genes that define the cDC1 identity, a published scRNA-Seq dataset encompassing mononuclear phagocyte subsets was interrogated. Via single cell transcriptomics, Villani et al. (2017) outlined a list of top six genes that mark the cDC1 lineage. The IRF8 ChIP-Seq dataset helped identify that out of the six marker genes, four genes (including surface markers CLEC9A and XCR1) had IRF8 peaks within the promoter region, and one had multiple intronic and distal intergenic peaks (Figure 6.9). In addition, the genes from an extended list of markers published by Villani et al. (2017) were interrogated in the ChIP-Seq dataset. Out of 112 cDC1 signature genes, the majority of the gene showed potential IRF8 regulation. 47 genes (including multiple HLA class II genes, FLT3, and TLR10) had consensus peaks in the promoter region and a further 30 genes (including BTLA) containing peaks in the intronic or distal intergenic regions. This effect was not observed in pDC, where none of the top markers were controlled by IRF8, and only 15 out of the 390 marker genes from the extended list had peaks in the promoter region.

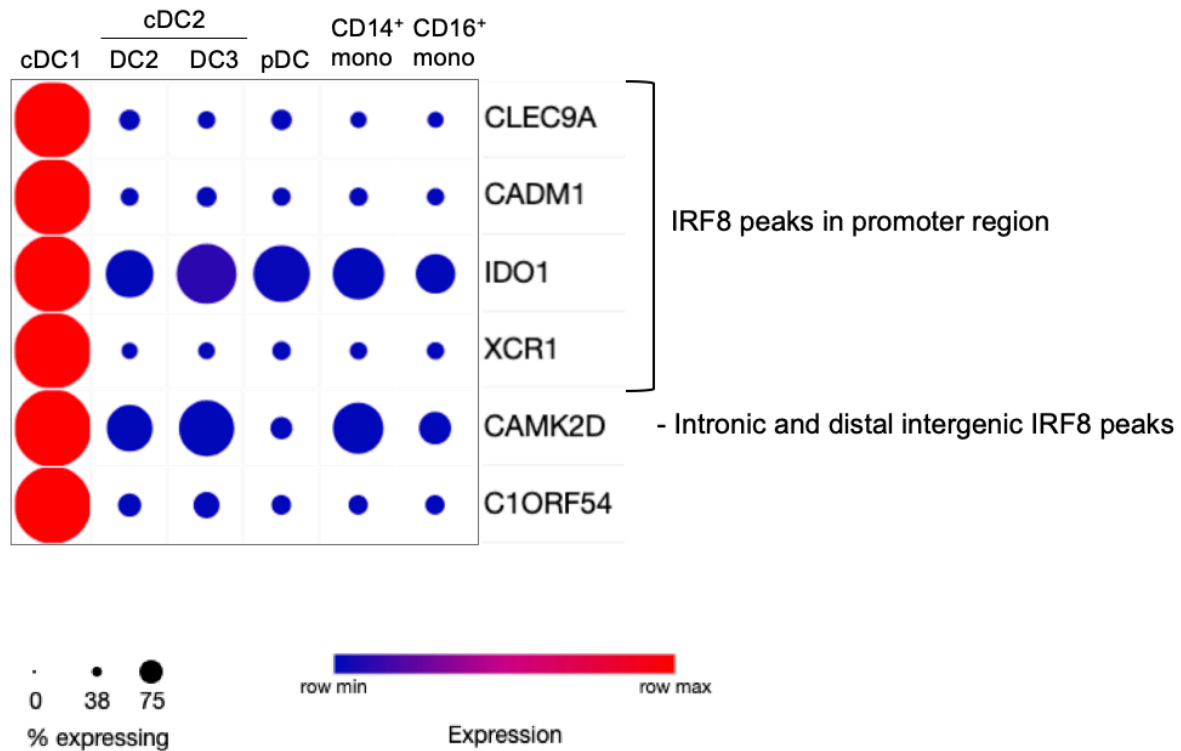


Figure 6.9. Heatmap displaying the expression of cDC1 marker genes identified by Villani et al., 2017, across peripheral blood DC and monocyte subsets.

The heatmap was generated via the Single Cell Portal (The Broad Institute of MIT and Harvard) and displays the conventional notation for the DC and monocyte subsets. For each cDC1 marker gene, information is given on the consensus IRF8 peak location in the cDC1 ChIP-Seq experiment, 4/6 cDC1 signature genes having at least 1 peak in the promoter region, 1/6 having multiple intronic and distal peaks, and 1/6 having no proximal or distal IRF8 peaks.

6.4 Discussion

Transcription factors are able to control cell fate at different phases of the cell lifespan via (1) prompting the specification of precursors toward a given lineage (2) suppressing alternative lineages to preserve the commitment of these precursors, (3) maintaining the cell identity of terminally differentiated cells, or (4) regulating key functional modules (Sichien et al., 2016). This chapter aimed to reveal the role of IRF8 in mature human pDC and cDC1, the two DC subsets marked by the expression of this transcription factor (Collin and Bigley, 2018), allowing indirect inference of its role in their development. ChIP-Seq was employed to assess the

binding sites occupied by IRF8 in these DC lineages. Previously, this study was not possible in human due to the rarity of these cells *in vivo* and the requirement for millions of cells for the chromatin immunoprecipitation technique. The use of a novel OP9/OP9-DL1 culture system, designed to produce large numbers of *bona fide* DC subsets from human bone marrow progenitors, along with advances in the chromatin precipitation protocols, such as the development of the Low Cell ChIP-Seq method (Active Motif), facilitated the success of this experiment. Significant differences between the IRF8 binding sites in pDCs and cDC1s were apparent within the generated ChIP-Seq dataset.

6.4.1 Role of IRF8 in cDC1s

In cDC1, IRF8 regulates multiple immune pathways involved in innate and adaptive immunity, indispensable for their function as antigen-presenting cells. Crucially, among the functional modules controlled by IRF8 is the DC-defining antigen presentation via MHC Class II, which is the first step in CD4 T cell activation. IRF8 also controls the costimulatory signal, mandatory for the activation of T cells. CD80 and CD86, regulated by IRF8, are considered the most important costimulatory molecules, as they serve as very early costimulatory signals and can lead to both the inhibition and activation of T cells (Hubo et al., 2013). The CD80 and CD86 molecules have intermediate expression in immature DCs and are upregulated in terminally differentiated DC, possibly due to the activator effect of transcription factor IRF8. The third signal required for T cell polarisation has also been found to be controlled by IRF8, mainly the production of IL-10, a potent immunomodulatory cytokine. Furthermore, IRF8 is likely to play a role in sensing and elimination of pathogens by cDC1. The transcription factor controls the signaling of three out of the four major pattern recognition receptors families, namely membrane-bound TLRs and CLRs, and cytoplasmic NLRs. Receptors TLR1, 4 and 6, regulated by IRF8, are specialised in the recognition of bacterial lipids. Additionally, TLR stimulation leads to the secretion of pro-inflammatory cytokines such as IL-6, TNF-alpha and IL-12 or anti-inflammatory cytokines such as IL-10, which shape the T-cell responses. CLRs recognise fungal and bacterial glycans, and are involved in the recognition of glycosylated self-antigens, and as adhesion and/or signalling molecules (van Kooyk, 2008). NOD1, a member of the NLR family regulated by IRF8, senses intracellular bacterial peptidoglycan and can trigger signal transduction via NF- κ B and MAPK pathways (Saxena et al., 2014). In addition to controlling antigen sensing,

presentation, and T cell priming, IRF8 was found to control cDC1 cell identity, as peaks were found in the promoter region of the most cDC1-defining marker genes and surface antigens, including the established markers XCR1 and CLEC9A. This finding is in line with the role of IRF8 in mouse cDC1s, where it acts as the terminal selector for this lineage and is therefore required to maintain the identity of terminally differentiated cells (Grajales-Reyes et al., 2015).

Mature cDC1s are likely to maintain the expression of IRF8, critical for their function and phenotype, through auto-activation. Evidence of IRF8 auto-activation was found in the cDC1 ChIP-Seq in the form of peaks surrounding the IRF8 TSS, consistent across biological replicates. Two of the peaks identified upstream of IRF8 TSS closely resembled the peaks detected in murine cDC1s via IRF8 ChIP-Seq at -25kb and -16kb of IRF8 TSS (Grajales-Reyes et al., 2015). The peaks identified downstream of IRF8 TSS fell within the superenhancer region reported in murine cDC1, but were not an exact match with the peaks reported in murine studies (Grajales-Reyes et al., 2015). Notable peaks which contained several motifs bound by Ets factor PU.1/SPI1, a known IRF8 binding partner with an activator role, were located at -25kb, +10kb and +37kb relative to IRF8 TSS. Motifs bound by a second prominent IRF8 partner, BATF3, were found within the peak at +68 kb relative to the IRF8 TSS. These sites are strong candidates for IRF8 auto-activation, as both PU.1/SPI1 and BATF3 known to support IRF8 auto-activation at different stages of cDC1 development in mouse (Grajales-Reyes et al., 2015). In addition to the above-described activating motifs bound by IRF8 and its partners, the scan revealed an E-box canonical sequence motif within the +59kb peak. This advocates for the role for E-proteins in the induction of IRF8 expression in cDC1s, a concept recently proposed in murine studies (Durai et al., 2019). Further insights derived from the analysis include the presence of an IRF motif, known to be bound by Irf3, at +37kb from IRF8 TSS. As cDC1s show IRF3 and IRF8 expression (Single Cell Portal, https://singlecell.broadinstitute.org/single_cell), this could be achieved by the direct interaction of the two transcription factors, previously demonstrated to play a role in transcription activation (Li et al., 2013).

6.4.2 Role of IRF8 in pDCs

The pDC-specific role of IRF8 comprises the control of one of the pDC functional modules - the transcriptional regulation of the IFN α/β signaling. This is a significant role, as pDCs are the main producers of type I IFNs in response to viral infections (Tel et al., 2012). This observation is consistent with mouse models, where IRF8 positively regulates essential functional modules in differentiated pDCs, such as production of type I IFNs, while being dispensable for pDC development (Sichien et al., 2016). Remarkably, no evidence of auto-activation was found in mature pDCs, suggesting that other mechanisms are employed to maintain IRF8 expression in these cells.

pDCs, but not cDC1s, exhibit high expression of a second transcription factor from the IRF family – IRF4. Interestingly, a motif bound by IRF4 was strongly enriched in pDC. This can be explained by the direct interaction of the two transcription factors as cooperative partners (Humblin et al., 2017). Alternatively, the two IRF family transcription factors could bind similar motifs, due to their high homology (Antonczyk et al., 2019), and therefore functioning redundantly in binding certain motifs.

6.4.3 Pathways likely to be downregulated by IRF8

In both cDC1s and pDCs, IRF8 controls pathways relating to two biological terms: neutrophil degranulation and IFN γ signaling, and likely represses the expression of the genes associated with these pathways. IRF8 is known to block the neutrophil differentiation programme to favour the development of DCs and monocytes (Kurotaki et al, 2014), and could act in a similar manner toward other neutrophil-related genes, repressing their expression in mature cDC1s and pDCs. In the case of IFN γ signaling, downregulation of the pathway is the also the most probable outcome, as IFN γ is primarily secreted by activated T cells and NK cells, and cDCs and pDCs show little capacity for its production (Vremec et al., 2007).

6.4.4 IRF8 in immunodeficiency

The findings of this chapter are congruent with the gene dosage-related phenotypes observed in patients with IRF8 deficiency. IRF8 haploinsufficiency manifests as a reduction or depletion of all DC subsets developing through the IRF8^{high} pathway: pDC, cDC1 and DC2 (Cytlak and Resteu, et al., 2020). cDC1s are most sensitive to

the loss of IRF8, their numbers being selectively depleted as a result of the dominant negative *T80A* mutation in IRF8 (Kong et al., 2018). This observation is in line with the critical role IRF8 plays in the multiple aspects of cDC1 function and maintaining the identity of these cells, as determined by IRF8 ChIP-Seq. In heterozygotes, pDCs exhibit an abnormal pDC function due to deficits in IFN α and TNF production (Cytlak and Resteu, et al., 2020). Findings of this chapter revealed that the deficits in IFN α could be the direct cause of a lower IRF8 dosage in these individuals, as peaks corresponding to the transcription factor's binding site were found in the promoter region of type I IFN-related genes, including the Interferon Alpha 16 (IFNA16) gene. An even lower IRF8 dosage is found in biallelic IRF8 deficiency, which leads to the complete loss of all monocyte and DC subsets, including cells developing through both the IRF8^{hi} and IRF8^{low} pathways (Bigley et al., 2018). Therefore, IRF8 plays an important, currently unrevealed role in the development of all monocyte and DC precursors in human.

6.4.5 Technical caveats of the study

This study relied heavily on the FACS-purification of pDCs and cDC1s prior to the ChIP-Seq assay. pDCs were defined as CD123⁺CD303/304⁺ cells, however the expression of the established pDC markers CD303/304 is fickle in culture. In addition, other cells have been shown to express CD123, including the pre-DC2 in Chapter 4 of this thesis. It is therefore likely that the population sorted as pDC is more heterogeneous than the cDC1 population, the markers for which (CLEC9A and CD141) are very faithful in culture.

Other caveats include the interpretation of ChIP-Seq peaks. This work focused on the peaks located in the immediate vicinity of the gene TSS (+/-3kb), as often performed in similar studies. However, the location of the promoter of a gene near a peak does not absolutely guarantee that it is regulated by IRF8. In addition, distal peaks were excluded from analysis, as it was not possible to determine with confidence the genes they regulated. Further experimental data is required to explore the distal peaks and verify the peaks in the promoter regions.

6.5 Summary and further work

Together, the findings of this chapter expand the knowledge about the role of transcription factor IRF8 in human. A unique ChIP-Seq dataset, generated within this project, demonstrated that IRF8 controls different functional modules in the IRF8-expressing human DC subsets. In pDC, IRF8 controls an important functional module, type I IFN signaling. The role of IRF8 is different in cDC1, where it maintains both the function and surface phenotype of these cells. Strikingly, evidence of IRF8 auto-activation was found in mature cDC1, but not pDC.

In order to corroborate the role of IRF8 within the reported biological pathways in cDC1 and pDC, future efforts should focus on generating and analysing an RNA-Seq dataset that complements the ChIP-Seq samples. This will allow to assess the effect of IRF8 on the expression of genes within the pathways deemed significant by the ChIP-Seq analysis. The cells for this analysis have been produced in culture and subjected to FACS purification by the candidate and members of the Human DC Lab. However, the project was paused at the library preparation stage due to the impact of COVID19.

To confirm the location of active transcription sites and define the enhancer and super-enhancer regions controlling the expression of IRF8 in human, further experiments should focus on histone ChIP-Seq in order to identify active promoters via H3K4me3 enrichment, inactive promoters via H3K27me3 enrichment, enhancers via enrichment of H3K4me1 and H3K27ac in regulatory regions, and active gene bodies with H3K36me3 enrichment. In addition, ChIP-Seq of IRF8 binding partners could reveal the interactions that are most crucial in maintaining IRF8 expression through auto-activation. Furthermore, genome-wide chromatin accessibility could be assessed with ATAC-Seq, which can be performed at single cell level, facilitating the use of small numbers of rare cells, such as DC and their precursors and progenitors.

IRF8 is required at multiple stages of cDC1 development in mice, however this requirement has not yet been explained in human DC precursors. To reveal how IRF8 prompts the specification of precursors toward a given lineage, efforts should be made to perform IRF8 ChIP-Seq using recently defined pre-DC populations (Cytlak and Resteu et al., 2020). However, the main limitations to this are the rarity of

these cells *in vivo* and the difficulty of producing sufficient numbers of precursors *in vitro* under the current culture conditions of the OP9/OP9-DL1 systems.

Finally, the role of IRF8 in the control of type I IFN signaling in pDC, will be explored further in the future. Since the produced ChIP-Seq experiment used unstimulated cells, the use of DCs stimulated with TLR ligands to actively produce interferons could benefit further studies wishing to uncover the role of IRF8.

Chapter 7. Overview, discussion, and further work

7.1 Summary of novel findings

This work explored multiple aspects of human dendritic cell development. A combination of approaches, including cytometry, transcriptomics and genomics, enabled the following novel observations to be made:

- 1. Co-culture of human bone marrow CD34⁺ progenitors with feeder layer OP9 in the presence of a cytokine cocktail generates all DC subsets in proportions similar to peripheral blood.** The resulting pDCs, cDC1s, and cDC2s exhibit an appropriate phenotype and align closely with their *ex vivo* counterparts by gene expression analysis.
- 2. Addition of Notch ligand DL1 to the OP9 culture system results in an eleven-fold increase of cDC1 output per progenitor cell (Kirkling and Cytlak et al., 2018).** The OP9-DL1 system, able to produce large numbers of *bona fide* cDC1s *in vitro* makes these cells more accessible for therapeutic use and research.
- 3. cDC1s produced in culture closely resemble their *ex vivo*-derived tissue counterparts.** Comparative transcriptomics revealed that OP9 and OP9-DL1 derived cDC1s bear higher resemblance to spleen and bone marrow cells than to cDC1s from peripheral blood.
- 4. Human DC subsets develop in the bone marrow via two pathways, marked by differential expression of transcription factor IRF8 (Cytlak and Resteu et al., 2020).** A combination of cell culture and cutting-edge single cell analysis techniques facilitated the discovery that high *IRF8* expression defines a developmental pathway giving rise to pDC, cDC1 and DC2. In contrast, the DC3 population arises through an *IRF8*^{low} pathway with precursors independent from monocytes.
- 5. A specific and sensitive IRF8 antibody was validated for ChIP-Seq.** The Santa Cruz (E-9): sc-365042 IRF8 antibody met the characterisation criteria outlined by the ENCODE consortium (Landt et al., 2012). As part of the validation, this antibody was able to detect IRF8 in human *ex vivo*-derived bone marrow cells via western blotting and to immunoprecipitate the IRF8 protein heterologously expressed by the

HEK293T cell line. Following characterisation, the Santa Cruz antibody was successfully used to produce a low cell ChIP-Seq dataset.

6. IRF8 controls different functional modules in the homeostasis of the IRF8-expressing human DC subsets. Analysis of IRF8 ChIP-Seq data showed that this transcription factor is crucial for cDC1, where it maintains both the function and surface phenotype. In pDC, IRF8 controls an important functional module, IFN α/β signaling.

7. Mature cDC1, but not pDC, show evidence of IRF8 auto-activation in human. Visualisation of the ChIP-Seq data revealed the presence of several peaks surrounding the IRF8 transcription start site in cDC1. Motif analysis showed that the majority of these peaks contained motifs bound by IRF8 and its binding partners with an activator role. This was not observed in mature pDC, suggesting that IRF8 expression in these cells is maintained via different mechanisms.

7.2 Overview of the techniques used

The above-listed findings were made possible by employing a combination of novel and established methods and analysis pipelines. Initial work was performed using the NanoString gene expression assay, in order to validate the DC output of a novel *in vitro* culture system against *ex vivo*-derived DCs from peripheral blood. The NanoString system was selected, as it offered a material- and cost-effective method of assaying the expression of hundreds of Immunology-related genes via the use of the pre-built Immunology panel. The use of fluorescence activated sorting allowed for the purification and sorting of the DC subsets into lysis buffer, and due to the specialised chemistry of the NanoString assay, the lysates could be used directly for hybridisation, omitting the RNA isolation step, and therefore minimising the loss of material and cutting down on the cost and time needed for RNA extraction. In addition, the system was available in-house, and the samples were processed on the NanoString system shortly after cell sorting. This allowed to reduce the number of freeze-thaw cycles that cause degradation of RNA during sample shipping. Furthermore, the amount of time from cell sorting to obtaining the data on the NanoString system from as short as 48 hours. The data produced by the NanoString assay is in form of counts, each count being equivalent to one mRNA molecule, greatly simplifying the processing steps of this data compared to sequencing

approaches. In this work, data normalisation was performed with the advanced analysis module of the nSolver software (NanoString Technologies), which selects the housekeeping genes with the most consistent expression across the samples and uses them as reference. Normalisation of NanoString data is a critical step, as the gene counts in each sample are affected by the concentration and quality of input RNA and by chaotropic agents and other contaminants originating from RNA extraction reagents or cell lysis buffers. As the nSolver software is not fully customisable, further data analysis was undertaken in the R environment, where it was possible to perform additional filtering of genes with background-level expression, as well as use a t-test to identify list of differentially expressed genes between all culture versus all *ex vivo*-derived populations, also referred to as “culture signature” (Figure 7.1).

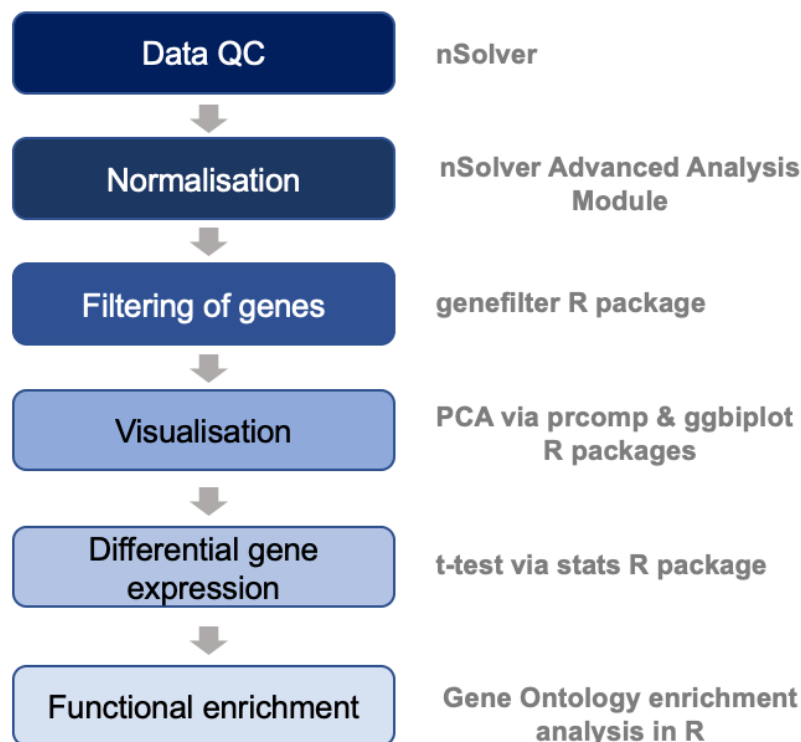


Figure 7.1. Analysis pipeline for NanoString data.

The analysis steps are listed in blue boxes and the tools used at each stage are shown in grey. Data QC and normalisation steps were undertaken within the nSolver software (NanoString Technologies). Further analysis was performed within the R software environment.

The second transcriptomic approach employed in this work was “low input” bulk RNA-Seq. A modified single cell protocol was used for this work, rather than standard RNA-Seq requiring tens of thousands of cells. This whole transcriptome analysis method was used to gain a better understanding of the identity of cDC1s produced in OP9 and OP9-DL1 culture, by comparing the transcriptomes of the *in vitro*-derived population to those of their counterparts from peripheral blood, bone marrow, and spleen. Generally, the production of RNA-Seq data entails a labour-intensive protocol, which involves the isolation or enrichment of cells of interest, RNA extraction, library preparation, then sequencing. In addition, the analysis of RNA-Seq data is more complex and labour-intensive than that of NanoString system output, and requires additional processing steps, often preformed using command line tools. In this work, the processing steps were performed on a computer cluster, due to their memory-intensive nature. An adaptation of analysis scripts provided by the Bioinformatics Support Unit at Newcastle University was used to trim and filter sequencing reads based on quality with Trimmomatic (Bolger et al., 2014), align them with the STAR mapping algorithm (Dobin and Gingeras, 2015), convert the alignment files from SAM format to the more compressed BAM format with SAMtools (Li et al., 2009), and count reads with HTSEQ (Anders et al., 2015) in order to generate the count table, which reports the number of reads assigned to each gene for each sequenced sample. As bulk RNA-Seq is an established technique, a multitude of analysis pipelines designed for performing comparative transcriptomics using count matrices have been made available, primarily in the R software environment. The data analysis in this thesis employed a widely-used pipeline for normalisation and differential expression – the DESeq2 R package (Love et al., 2014). The goal of the RNA-Seq data normalisation is to control for sources of variation, such as differences in the sequencing depth of the samples. The DESeq2 tool also performs the estimation of dispersion values for each gene, followed by the fitting of a generalised linear model to minimise sampling noise. Further “custom” methods that suited the analysis needs the most include the BubbleGUM software (Spinelli et al., 2015), employed to determine the blood and tissue signatures, and the single sample GSEA from the GSVA R package (Barbie et al., 2009; Hänzelmann, et al, 2013), applied to determine the enrichment of these signatures across *ex vivo*-derived and cultured cells (Figure 7.2).

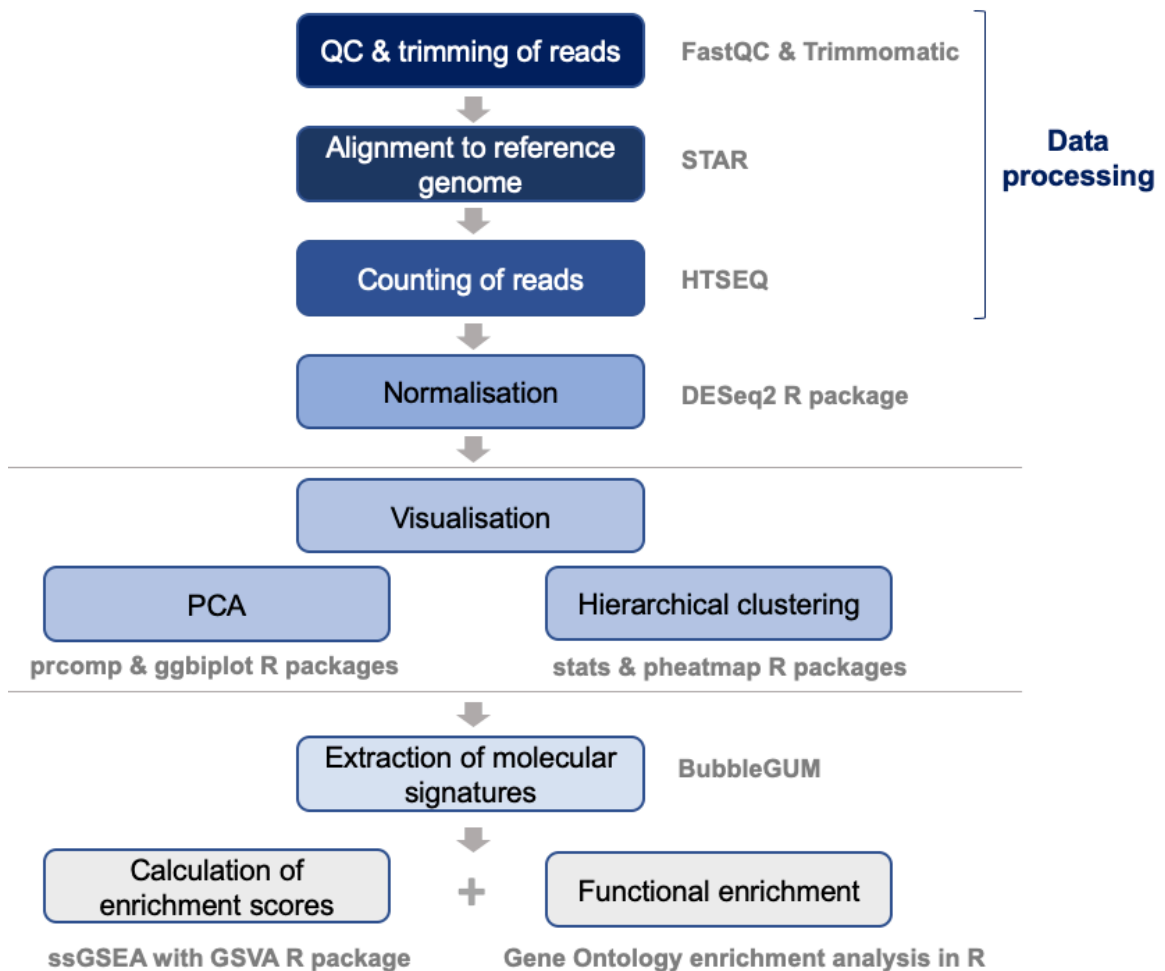


Figure 7.2. Analysis pipeline for bulk RNA-Seq data.

The analysis steps are listed in blue boxes and the software programmes used for each step are shown in grey. The data processing was performed on a computer cluster, using command line tools. Further analysis was undertaken in the R environment by employing the DESeq2 pipeline, especially designed for RNA-Seq experiments. Other R packages, as well as the BubbleGUM software, were incorporated in the analysis pipeline in order to analyse the molecular signatures of tissue and blood DCs.

Recent advances in sequencing lead to an increase in availability and decrease in cost of single cell RNA-Seq, enabling the use of this technology to dissect DC haematopoiesis in Chapter 4 of this thesis. The single cell plate-based approach was adopted, as it allows the index sorting of each cell and is able to generate better quality data and compared to droplet methods, such as 10X. However, this method is more expensive and labour intensive. The workflow for the generation of the scRNA-Seq data generally resembled that of bulk RNA-Seq, the main difference being that

FACS was employed to isolate single cells, rather than bulk cell populations. Pre-processing steps of plate-based scRNA-Seq data often mimic those used for bulk transcriptome analysis, and in this work an almost identical pipeline was employed. However, the two approaches diverge once the count tables are obtained (Figure 7.3). From this point, single cell specific tools are essential for analysis, primarily due to differences in the quality of the data, scRNA-Seq data containing many dropouts and having sparser gene counts. Tools specialised to handle single-cell gene expression data applied in this work include the scater R package (McCarthy et al., 2017), employed to perform cell and gene QC and filtering and the SC3 R package (Kiselev et al., 2017), used to cluster the cells by gene expression and determine the markers defining each cluster of similar cells, and the Slingshot method for inferring cell lineages and pseudotimes (Street et al., 2018). As the scRNA-Seq datasets are often highly multidimensional, containing hundreds of cells and thousands of genes, their visualisation and interpretation can be challenging. Dimensionality reduction techniques, such as tSNE and diffusion maps, allow for low-dimensional representation of the expression data and are therefore indispensable for the single cell transcriptomic analysis (van der Maaten, 2008; Angerer et al., 2016). Transformation of data from a high-dimensional space into a low-dimensional space is also critical for computational analysis and visualising of flow and mass cytometry datasets. In this thesis, dimensionality reduction allowed the representation of all parameters constituting the cytometry datasets as 2D or 3D figures, in contrast to the traditional sequence of multiple 2D scatter plots displaying 2 antigens at a time.

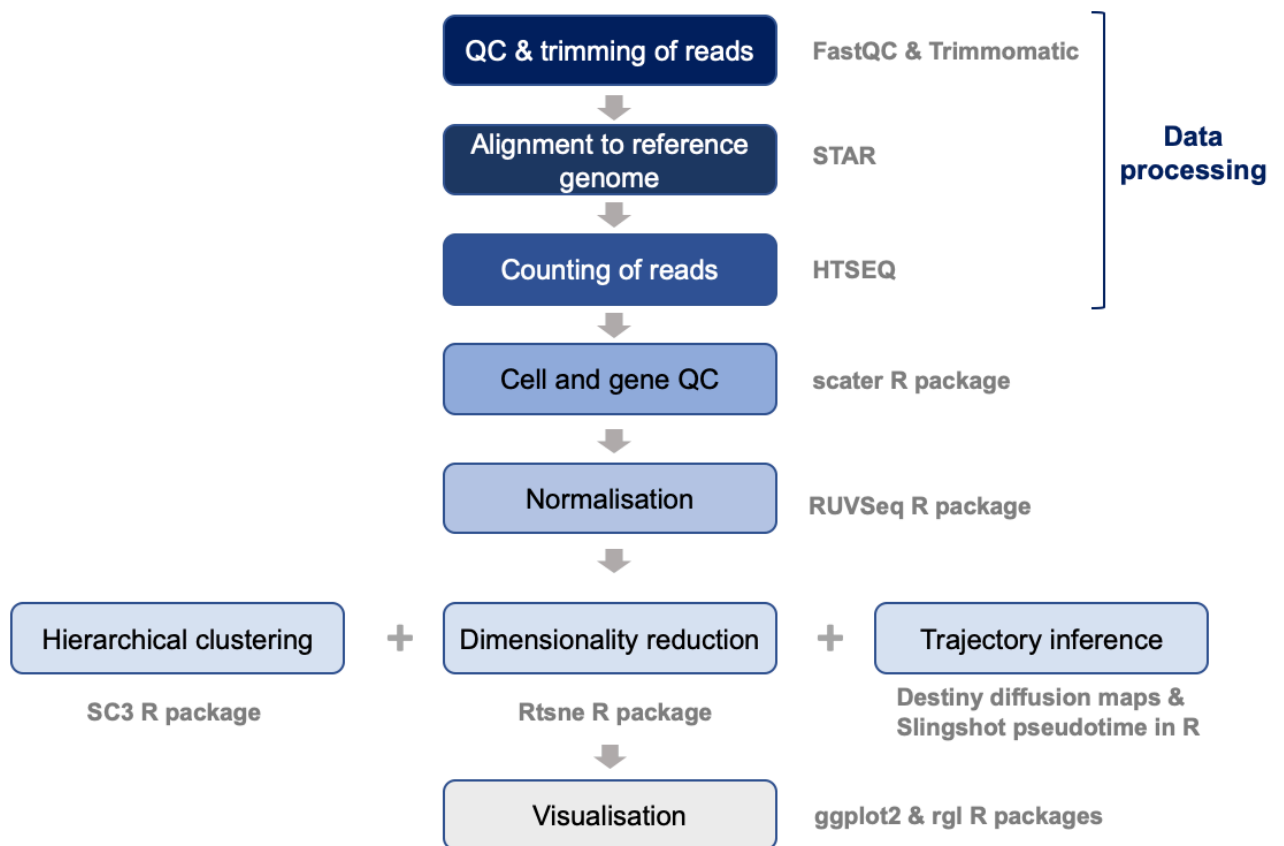


Figure 7.3. Analysis pipeline for scRNA-Seq data.

The analysis steps are listed in blue boxes and the tools used at each stage are shown in grey. The data processing was performed on a computer cluster, using command line tools. The data analysis was undertaken in the R environment and employed packages developed or adapted for scRNA-Seq data.

The final two results chapters of this thesis revolved around the optimisation and analysis of low cell IRF8 ChIP-Seq for the cDC1 and pDC subsets. These experiments were made possible by recent advances in ChIP protocols, which allowed the use of under 100,000 cells for each ChIP reaction, in contrast to over a million cells, previously used in ChIP-Seq experiments. In addition, the rarity of cDC1 *in vivo*, was addressed via a novel culture system, the output of which was verified by two transcriptomic approaches in addition to surface phenotype and functional assays (Kirkling and Cytlak et al., 2018). A hurdle encountered in these experiments was the commercial discontinuation of high-quality ChIP-grade IRF8 antibodies, previously used for most related studies (Langlais et al. 2016, Shin et al, 2011). This was overcome by the validation of a different antibody, following the characterisation criteria outlined by the ENCODE consortium (Landt et al., 2012). Molecular biology

techniques, such as western blotting, bacterial transformation and transfection of mammalian cell lines, as well as immunoprecipitation, were combined to ensure the specificity and sensitivity of the antibody. The generation of sequencing data was guided by the standards outlined by the ENCODE consortium (<https://www.encodeproject.org>) for Transcription Factor ChIP-seq. In accordance to the ENCODE experimental guidelines, each experiment had 3 biological replicates and employed an appropriately characterised antibody, each ChIP-Seq experiment had a corresponding input control experiment with matching run type, read length, and replicate structure, and most replicates contained over 20 million reads. In addition, ENCODE Uniform Processing Pipeline Restrictions were followed to generate 150bp paired-end reads (longer read lengths than 50 base pairs being encouraged). Upon production of the datasets, custom data processing and analysis pipelines were assembled by the candidate to suit the analysis needs (Figure 7.4). For the most part, the data processing pipeline consisted of a set of established genomic tools, such as bcl2fastq (Illumina), used to assign reads to samples following multiplexed sequencing based on the indexes incorporated during library construction, Cutadapt (Martin, 2011) and FastQC (Simon Andrews, Babraham Institute), used for trimming of adapters and low quality bases, BWA-MEM (Li, 2013), a fast and accurate algorithm for mapping of DNA sequences against a large reference genome, and samtools (Li et al., 2009), a tool used for converting SAM files to the more compressed BAM format. To mitigate the effects of PCR amplification bias introduced during library preparation, which can interfere with downstream peak calling, a custom tool developed by Active Motif, Inc was applied to the BAM alignment files. First, a molecular identifier in the form of a 9 base N random sequence was incorporated into each DNA fragment prior to library construction. The custom tool recognised reads with identical molecular identifiers as true PCR duplicates, leading to their removal. Peak calling was undertaken with an improved version of a commonly used tool for identifying transcription factor binding sites, named Model-based Analysis of ChIP-Seq version 2 (MACS2, Zhang et al., 2008). Following the processing steps, performed on a computer cluster due to their prohibitive memory requirements, analysis of called peaks was migrated to the R environment, which contains a multitude of ChIP-Seq-specific packages and analysis pipelines. A blend of R packages was used to create an analysis pipeline that computes differentially bound sites from multiple ChIP-Seq experiments using affinity data with DiffBind (Ross-Innes et al., 2012), annotates the consensus peaks found in

most replicates with ChIPseeker (Yu et al., 2015), and performs pathways analysis with ClusterProfiler (Yu et al., 2012). Motif analysis was migrated back to the command line envired, where a larger number of specialised tools, such as GimmeMotif (Bruse and van Heeringen, 2018) are available. Finally, the incorporation of the external single cell transcriptomic dataset published by Villani and colleagues (2017), was essential to verify and visualise the cDC1 marker genes controlled by transcription factor IRF8.

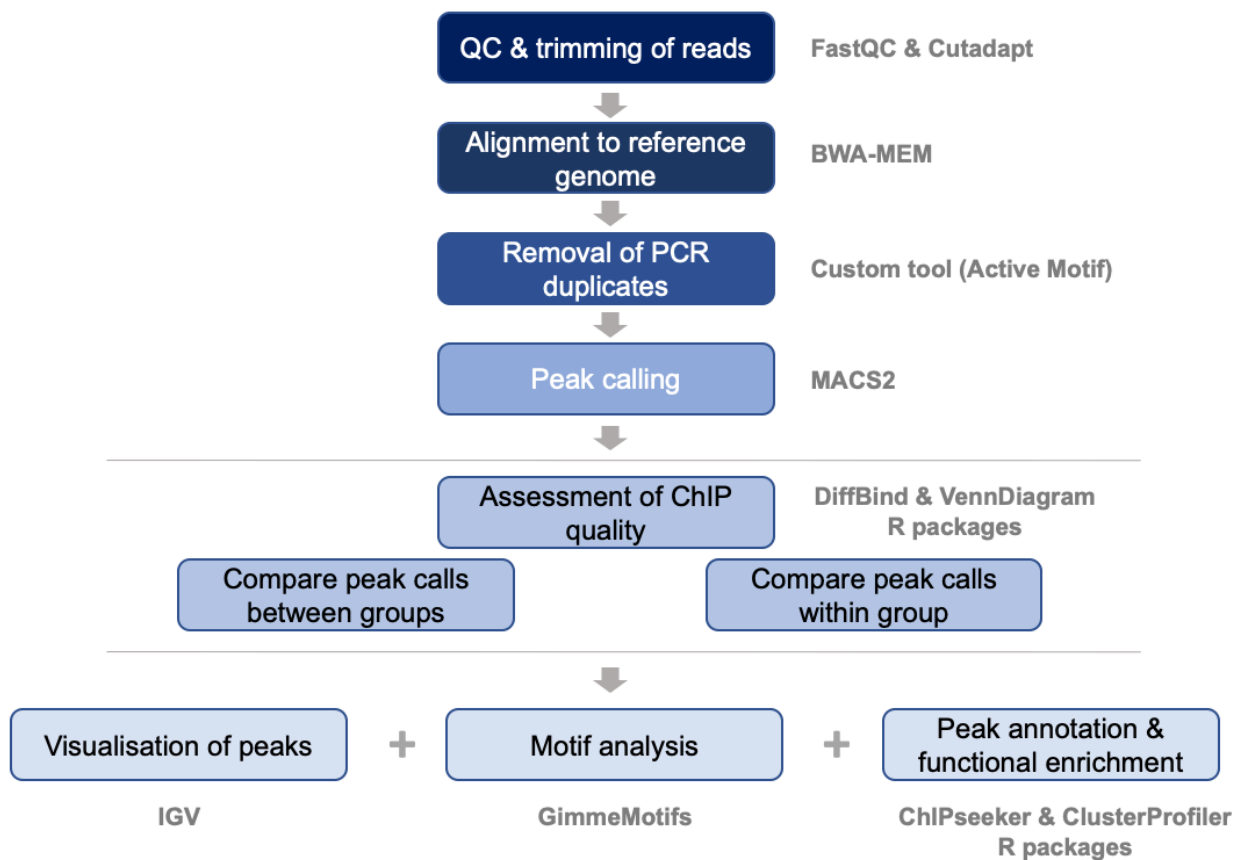


Figure 7.4. Analysis pipeline for ChIP-Seq data.

The analysis steps are listed in blue boxes and the tools used at each stage are shown in grey. Initial data processing and analysis, including peak calling, were performed on a computer cluster using command line tools. Further analysis tools include R packages, as well as the IGV software, used to visualise the genomic locations of the peaks, and the GimmeMotifs command line tool for motif analysis.

7.3 Research impact

As described in the previous sub-chapter, multiple unique datasets have been generated and analysed throughout this project. The biological findings, as well as the data and the analysis pipelines represent valuable contributions to immunology. They pave the way to tackle future research including the leads and further questions raised through the work in this project.

First, this project verified the identity of DCs produced in culture and established the critical role of Notch signaling for human cDC1 differentiation. These results were published (co-author) in collaboration with a research group that determined the crucial role of Notch signaling in the functional maturation of murine cDC1s (Kirkling and Cytlak et al., 2018). The OP9-DL1 culture system, able to produce large numbers of cDC1s *in vitro* via Notch stimulation, renders cDC1, the rarest DC subset *in vivo*, more accessible for therapeutic use and research. A research application has already been found for this culture method in Chapters 5 and 6 of this work, where it was applied to produce sufficient cells for the study of the role of transcription factor IRF8 in the homeostasis of cDC1s. The OP9 system, able to generate DC subsets in proportions similar to peripheral blood, was applied to produce pDCs for the IRF8 ChIP-Seq experiment in Chapter 6 and aided the dissection of DC haematopoiesis in Chapter 4. The R code created for the analysis of NanoString gene expression data from *ex vivo* and culture-derived DCs and monocytes has been used as learning material by BSc and MSc students in the HuDC group. The RNA-Seq data processing and analysis pipelines for cDC1s from blood, tissue, and culture has been adapted for the exploration of other datasets generated within the HuDC Lab, such as the RNA-Seq data created to verify the identity of monocyte-derived Langerhans cells produced in culture via Notch stimulation through DLL4. The findings of transcriptomic analysis of the moLC have recently been published as part of a publication titled “Notch-mediated generation of monocyte-derived Langerhans cells: Phenotype and function” (Bellmann et al., 2020).

Using the OP9 *in vitro* differentiation assay in combination with single cell methods, the research undertaken as part of Chapter 4 of this thesis revealed that two pathways of DC development are present in human bone marrow. Differential expression levels of transcription factor IRF8 marks for the two pathways, high IRF8 developmental pathway giving rise to pDC, cDC1 and DC2, and the IRF8^{low} pathway

producing DC3 and classical monocytes. This has not been previously reported in literature, as most studies of DC haematopoiesis did not account for the cDC2 heterogeneity, which was first described in detail in 2017 by Villani and colleagues. Chapter 4 identifies the two distinct developmental pathways as the source of the cDC2 heterogeneity, as the DC2 and DC3 subsets develop via different routes. Furthermore, work undertaken this chapter revealed the previously unknown phenotypical identities of the progenitor populations with increasingly committed DC potential and determined that DC2s develop through subsets of LMPP and CD123⁻/^{low}CD33⁻ GMP, while DC3 follow a trajectory passing through the CD123^{-/low}CD33⁺ GMP phenotypic space, explaining the apparent dual lympho-myeloid origin of dendritic cells. An important observation made in this thesis was that DC3 are not monocyte-derived, despite of their inflammatory-like expression profile. Instead, the resemblance of DC3 and monocytes can be explained by the similar developmental routes followed by these cell types. The single cell transcriptomics data for primary human bone marrow progenitors and mature DCs and monocytes generated during this project have been deposited within the Gene Expression Omnibus data repository (accession numbers GSE142999 and GSE143002) and are available for use by other researchers.

As part of Chapter 6 of this work, an IRF8 ChIP-Seq dataset for human cDC1s and pDCs has been generated and will be made available to other researchers upon publication of the related manuscript. This importance of this dataset lies in its uniqueness, as only a small number of ChIP-Seq datasets for transcription factor IRF8 in human are currently publicly available and none have been performed using human dendritic cells. Due to the lack of ChIP-validated IRF8 antibodies, extensive laboratory testing was undertaken as part of this project in order to identify a high-quality ChIP-grade antibody. Efforts will be made to include the antibody used in this experiment in the ENCODE database. Collaborators of the HuDC Lab on the IRF8 ChIP-Seq project have been informed about the potential of the newly characterised antibody and are planning to perform further independent validation and use it for ChIP-Seq experiments. Additionally, the analysis of the newly generated ChIP-Seq data expanded the knowledge about transcription factor IRF8 and uncovered its diverse role in the homeostasis of human DC subsets. A pipeline for the analysis of ChIP-Seq data has been compiled by the candidate as part of this project, and can

be applied to any ChIP-Seq datasets that will be generated within the HuDC group in the future.

7.4 Limitations of the project

Together, this work greatly expands the knowledge about the development of human dendritic cells. However, conducting research on this cell type in human is restricted by certain limitations. Dendritic cells are extremely rare *in vivo*, and are difficult to obtain in sufficient numbers for research studies. cDC1 cells are particularly rare, found in very low numbers in peripheral blood (an average of 500 cells per mL of blood, or 0.1% of PBMC; Collin et al 2013). Although they may be more abundant in tissues, for example splenic or lymphoid tissue, this material is less accessible for human research. Typically, the volume of blood drawn from healthy volunteers does not exceed 180mL, giving a total of 90,000 cDC1s in most samples cells. The loss of 30-40% of cells during FACS and subsequent steps further reduces the quantity of available material (Sutermaster et al., 2019), restricting the genomic and transcriptomic methods that can be used to study these cells to low input technologies. The most common methods for cell isolation and purification are FACS and magnetic-activated cell sorting. While magnetic-activated cell sorting is a faster method, the low purity resulting from this protocol renders it suboptimal for DC research. The rarity of dendritic cells was overcome by the use of the scalable OP9 and OP9-DL1 culture systems, able to generate sizable numbers of DC from human bone marrow progenitors. Nevertheless, the scarcity of the samples remains an issue, and in particular the limited availability of healthy donor bone marrow. In this project, bone marrow obtained from hip replacement surgery was used to meet the high demand for CD34⁺ progenitors for cell culture. However, the majority of patients undergoing this type of surgery are of an advanced age, limiting the samples further, if all age groups are to be represented. As the immune system and the process of haematopoiesis undergo changes with aging (Gubbels Bupp et al., 2018), it was imperative to not focus on a certain age group in this project. Therefore, the bone marrow samples used for culture were carefully selected to include range of age groups, and for ChIP-Seq donors were aged between 28 and 50. In addition, only samples derived from female donors were used in this analysis, as differences between male and female immune systems have recently come to light (Moulton, 2018; Griesbeck et al., 2015).

In addition, each of the multitude of techniques employed in this project has its advantages and limitations. As mentioned above, the purification of cell populations for downstream genomic and transcriptomic applications relied heavily on FACS. Drawbacks of this technology include the relatively low number of antigens (up to 18) than can be evaluated in each assay. In addition, FACS can only examine the expression of surface markers and cannot determine the expression of intercellular proteins (including transcription factors) on live cells, as intercellular staining commonly requires a permeabilisation step which causes cell damage and death. Furthermore, cell sorters commonly employ a gating strategy consisting of a sequence of 2D scatter plots. The software used for cell sorting in this project, allowed up to 8 scatter plots, leading to the adoption of alternative gating strategies for the sorting of large numbers of complex populations in order to accommodate this limit.

Further technical limitations stem from the genomic and transcriptomic approaches adopted in this work. While the NanoString assay is a robust alternative for a low input gene expression assay, its multiplexing capability is limited to 800 genes. The pre-built NanoString Immunology panel was the most economical to use for DC research. Nevertheless, many genes present in this panel covered other aspects of the immune system and were not expressed by DCs. This was in part overcome by the addition of 30 custom DC-related genes. However, approximately one third of the 608 total assayed genes displayed background-level counts. The detection of all known DC subsets was still possible based on the expressed genes.

The scRNA-Seq method, while being invaluable for dissecting haematopoiesis in Chapter 4, has a number of limitations. The first challenge of this method is the isolation of single cells, which often leads to cell loss, and depending on the technology used, may cause a high proportion of doublet cells to be sequenced. In addition, a variable proportion of low quality data is generated due to sequencing of broken or dead cells. QC of single cell data is therefore crucial in order to remove poor quality cells. Outlier cells are commonly removed based on the total number of detected features, total gene counts and percentage of counts derived from ERCC spike-ins and mitochondrial genes. These steps were used in this project to filter out doublets with abnormally high numbers of detected features, and remove damaged and apoptotic cells with low total counts, low number of features, a high proportion of

reads derived from spike-ins, rather than endogenous genes, and significant mitochondrial contamination. This QC step relies on the proportion of mitochondrial genes, increased as the cell dies, but doesn't take into account of any lineage differences in expression of mitochondrial genes. Due to low amounts of starting material, the scRNA-Seq data is also prone to dropout events, occurring when specific transcripts cannot be detected and adding substantial challenges for the computational analysis of this data. As the scRNA-Seq datasets are noisier than bulk RNA-Seq, specialised tools must be used for the analysis. The scRNA-Seq methods are continuously evolving, which brings benefits such as decrease in cost and improvement of data quality. However, this constituted a hindrance in this project, due to alterations over the years in the protocol for the generation of the scRNA-Seq data at the Wellcome Centre for Human Genetics, where sequencing was performed. This led to the inability to efficiently remove batch effects in order to combine the datasets containing BM CD34⁺ progenitors and BM CD34^{med} precursors and mature DCs, generated at different points in time. These scRNA-Seq datasets were therefore analysed separately, and mass cytometry data were generated to explore the BM CD34^{high} and CD34^{med} fractions, along with mature cells as a single dataset.

The transcription factor ChIP-Seq method was used study the role of TF IRF8 in the homeostasis of IRF8-expressing DC subsets in human. However, despite being a long-established technique, this assay required extensive optimisation, to which Chapter 5 was dedicated. The quality of the ChIP-Seq data relies heavily on the antibody used for immunoprecipitation, and multiple molecular biology techniques had to be used to ensure the adequate specificity of the employed IRF8 antibody. The number and duration of sonication cycles was one of the aspects that required optimisation. This was a time-consuming experiment, as DCs were first produced in 14-21 days of culture, then fixed and sonicated, and the full 5-day ChIP-Seq protocol, including the library preparation had to be performed in order to assess the effect of different sonication settings. Assessment of fragment size was not possible immediately after sonication, due to the use of a low number of cells, and was performed after PCR amplification as part of library construction protocol. The minimal number of cells required for the low input ChIP-Seq protocol is 50,000 for each replicate, which allowed us to use this technique for mature culture-derived DC. However, this number is still prohibitively high for using this technology on DC

progenitors, which are extremely rare *in vivo* and transient in culture, and on *ex vivo*-derived mature DCs.

7.5 Future research vision

Further work will aim to:

1. **Study and optimise the OP9/OP9-DL1 culture system.** Further optimisation of the culture system will be performed, aiming to enhance it as a research tool and to adapt it for use in DC therapy. To increase the yield of DCs both for research and therapy, introduction of a CD34⁺ progenitor expansion step in the first days of culture will be attempted. Further aspects to be investigated include:
 - a. Examining the effect of exogenous factors: the effect of alternative or additional growth factors or cytokines on the quantity, phenotype or function of the output cells, including factors which activate or tolerise the cells. The culture will also be altered with the goal to produce sufficient numbers of DC progenitors, which can be used for further studies or tolDC therapy.
 - b. Determining the OP9 factors are necessary for human DC development, including cell-cell contact and secreted factors, allowing the design of a mouse-stromal cell free system. This approach is in line with the aim to reduce the use of the animal-derived products in the culture system, such as the feeder layer and the foetal calf serum, in order to allow the adoption of this system for therapeutic use.
 - c. Determining the autocrine factors that influence the culture output. As DCs develop they may secrete factors which inhibit or support other DCs for homeostasis.

2. **Explore the developmental pathways disrupted by mutations in *IRF8*.** This will be achieved via the generation and analysis of scRNA-Seq datasets encompassing cells derived from patients with heterozygous and homozygous *IRF8* mutations. The single cell transcriptomics data will help uncover the dose-dependent effect of *IRF8* on the development of individual cells within heterogeneous DC progenitor populations. Further experiments will also include:

- a. CRISPR-Cas9 techniques to introduce known mutations into primary CD34⁺ cells to explore the phenotypes *in vitro*.
 - b. Creation of induced pluripotent stem cells from patient cells for *in vitro* studies.
 - c. Co-immunoprecipitation experiments to determine whether mutations affect the IRF8 binding partners.
 - d. CHIP-Seq of cells carrying IRF8 mutations to establish the effect of mutations on IRF8 transcriptional regulation – both for cell development and for functional responses, e.g. to IFN stimulation.
3. **Validate the IRF8 ChIP-Seq findings.** IRF8 can act as an activator or repressor of gene expression, and the effect of IRF8 on the biological pathways deemed significant by the ChIP-Seq analysis is presently unknown. Analysis of pDC and cDC1 RNA-Seq data from a complementary experiment will reveal the levels of expression of IRF8-regulated genes and help infer the role of this transcription factor in regulating the expression of individual genes and biological pathways. The active and inactive promoters will be verified via histone ChIP-Seq, and the chromatin accessibility at IRF8 binding sites will be explored with single cell ATAC-Seq on the 10X platform. Further ChIP-Seq experiments involving IRF8 binding partners with an activator role (such as BATF3 and PU.1) will reveal the interactions most crucial for the homeostasis of human DC subsets. The validation of IRF8 auto-activation in cDC1s will also be performed via CRISPR-Cas9 knock out of IRF8 peaks surrounding the *IRF8* gene in order to determine the sites necessary for the different elements of IRF8 function.
4. **Investigate the role of IRF8 in the development of cells emerging from the IRF8^{high} pathway.** This will be explored via an IRF8 ChIP-Seq experiment of increasingly committed progenitor populations, the phenotypic identities of which were revealed in Chapter 4 of this work. However, this can only be achieved via further optimisation of the culture system in order to produce larger number of progenitors. Alternatively, modifications can be applied to the ChIP-Seq protocol with the aim to decrease the number of cells required per IP reaction.

References

Alberts B, Johnson A, Lewis J, et al. *Molecular Biology of the Cell*. 4th edition. New York: Garland Science, 2002.

Alcántara-Hernández M, Leylek R, Wagar LE, et al. High-Dimensional Phenotypic Mapping of Human Dendritic Cells Reveals Interindividual Variation and Tissue Specialization. *Immunity*. 2017;47(6):1037-1050.e6.
doi:10.1016/j.immuni.2017.11.001

Andreae S, Piras F, Burdin N, Triebel F. Maturation and activation of dendritic cells induced by lymphocyte activation gene-3 (CD223). *J Immunol*. 2002;168(8):3874-3880. doi:10.4049/jimmunol.168.8.3874

Andrews TS, Hemberg M. M3Drop: dropout-based feature selection for scRNASeq. *Bioinformatics*. 2019;35(16):2865-2867.
doi:10.1093/bioinformatics/bty1044

Antonczyk A, Krist B, Sajek M, et al. Direct Inhibition of IRF-Dependent Transcriptional Regulatory Mechanisms Associated With Disease. *Front Immunol*. 2019;10:1176. Published 2019 May 24. doi:10.3389/fimmu.2019.01176

Ardouin L, Luche H, Chelbi R, et al. Broad and Largely Concordant Molecular Changes Characterize Tolerogenic and Immunogenic Dendritic Cell Maturation in Thymus and Periphery. *Immunity*. 2016;45(2):305-318.
doi:10.1016/j.immuni.2016.07.019

Arifuzzaman S, Das A, Kim SH, et al. Selective inhibition of EZH2 by a small molecule inhibitor regulates microglial gene expression essential for inflammation. *Biochem Pharmacol*. 2017;137:61-80. doi:10.1016/j.bcp.2017.04.016

Bagadia P, Huang X, Liu TT, et al. An Nfil3-Zeb2-Irf2 pathway imposes Irf8 enhancer switching during cDC1 development. *Nat Immunol*. 2019;20(9):1174-1185.
doi:10.1038/s41590-019-0449-3

Balan S, Ollion V, Colletti N, et al. Human XCR1+ dendritic cells derived in vitro from CD34+ progenitors closely resemble blood dendritic cells, including their adjuvant

responsiveness, contrary to monocyte-derived dendritic cells. *J Immunol.* 2014;193(4):1622-1635. doi:10.4049/jimmunol.1401243

Banchereau J, Palucka AK, Dhodapkar M, et al. Immune and clinical responses in patients with metastatic melanoma to CD34(+) progenitor-derived dendritic cell vaccine. *Cancer Res.* 2001;61(17):6451-6458.

Banchereau J, Steinman RM. Dendritic cells and the control of immunity. *Nature.* 1998;392(6673):245-252. doi:10.1038/32588

Baratin M, Foray C, Demaria O, et al. Homeostatic NF- κ B Signaling in Steady-State Migratory Dendritic Cells Regulates Immune Homeostasis and Tolerance. *Immunity.* 2015;42(4):627-639. doi:10.1016/j.immuni.2015.03.003

Barbie DA, Tamayo P, Boehm JS, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature.* 2009;462(7269):108-112. doi:10.1038/nature08460

Bell GM, Anderson AE, Diboll J, et al. Autologous tolerogenic dendritic cells for rheumatoid and inflammatory arthritis. *Ann Rheum Dis.* 2017;76(1):227-234. doi:10.1136/annrheumdis-2015-208456

Bellmann L, Zelle-Rieser C, Milne P, et al. Notch-mediated generation of monocyte-derived Langerhans cells: Phenotype and function [published online ahead of print, 2020 Jun 6]. *J Invest Dermatol.* 2020;S0022-202X(20)31663-8. doi:10.1016/j.jid.2020.05.098.

Bevan MJ. Cross-priming for a secondary cytotoxic response to minor H antigens with H-2 congenic cells which do not cross-react in the cytotoxic assay. 1976. *J Immunol.* 2010;185(3):1361-1366.

Bigley V, Barge D, Collin M. Dendritic cell analysis in primary immunodeficiency. *Curr Opin Allergy Clin Immunol.* 2016;16(6):530-540. doi:10.1097/ACI.0000000000000322

Bigley V, Collin M. Insights from Patients with Dendritic Cell Immunodeficiency [published online ahead of print, 2020 Apr 25]. *Mol Immunol.* 2020;122:116-123. doi:10.1016/j.molimm.2020.04.019

Bigley V, Maisuria S, Cytlak U, et al. Biallelic interferon regulatory factor 8 mutation: A complex immunodeficiency syndrome with dendritic cell deficiency, monocytopenia, and immune dysregulation. *J Allergy Clin Immunol*. 2018;141(6):2234-2248. doi:10.1016/j.jaci.2017.08.044

Blasius AL, Beutler B. Intracellular toll-like receptors. *Immunity*. 2010;32(3):305-315. doi:10.1016/j.immuni.2010.03.012

Blighe K. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. R package version 1.2.0., 2019. <https://github.com/kevinblighe/EnhancedVolcano>

Blumenreich MS. The White Blood Cell and Differential Count. In: Walker HK, Hall WD, Hurst JW, editors. *Clinical Methods: The History, Physical, and Laboratory Examinations*. 3rd edition. Boston: Butterworths; 1990. Chapter 153. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK261/>

Bouamar H, Abbas S, Lin AP, et al. A capture-sequencing strategy identifies IRF8, EBF1, and APRIL as novel IGH fusion partners in B-cell lymphoma. *Blood*. 2013;122(5):726-733. doi:10.1182/blood-2013-04-495804

Bourdely P, Anselmi G, Vaivode K, et al., Transcriptional and Functional Analysis of CD1c+ Human Dendritic Cells Identifies a CD163+ Subset Priming CD8+CD103+ T Cells, *Immunity* 2020, 53: 1–18, doi:10.1016/j.immuni.2020.06.002

Brasel K, De Smedt T, Smith JL, Maliszewski CR. Generation of murine dendritic cells from flt3-ligand-supplemented bone marrow cultures. *Blood*. 2000;96(9):3029-3039.

Breton G, Lee J, Zhou YJ, et al. Circulating precursors of human CD1c+ and CD141+ dendritic cells. *J Exp Med*. 2015;212(3):401-413. doi:10.1084/jem.20141441

Brigl M, Brenner MB. CD1: antigen presentation and T cell function. *Annu Rev Immunol*. 2004;22:817-890. doi:10.1146/annurev.immunol.22.012703.104608

Brown CC, Gudjonson H, Pritykin Y, et al. Transcriptional Basis of Mouse and Human Dendritic Cell Heterogeneity. *Cell*. 2019;179(4):846-863.e24. doi:10.1016/j.cell.2019.09.035

Bruse N, van Heeringen SJ. GimmeMotifs: an analysis framework for transcription factor motif analysis, bioRxiv; 2018. <https://www.biorxiv.org/content/10.1101/474403v1.full>. doi: 10.1101/474403.

Calzetti F, Tamassia N, Micheletti A, Finotti G, Bianchetto-Aguilera F, Cassatella MA. Human dendritic cell subset 4 (DC4) correlates to a subset of CD14^{dim/-} CD16⁺⁺ monocytes. *J Allergy Clin Immunol*. 2018;141(6):2276-2279.e3. doi:10.1016/j.jaci.2017.12.988

Cancel JC, Crozat K, Dalod M, Mattiuz R. Are Conventional Type 1 Dendritic Cells Critical for Protective Antitumor Immunity and How?. *Front Immunol*. 2019;10:9. Published 2019 Feb 12. doi:10.3389/fimmu.2019.00009

Chen H. VennDiagram: Generate High-Resolution Venn and Euler Plots. R package version 1.6.20, 2018. <https://CRAN.R-project.org/package=VennDiagram>

Cheng P, Gabrilovich D. Notch signaling in differentiation and function of dendritic cells. *Immunol Res*. 2008;41(1):1-14. doi:10.1007/s12026-007-8011-z

Chomarat P, Banchereau J, Davoust J, Palucka AK. IL-6 switches the differentiation of monocytes from dendritic cells to macrophages. *Nat Immunol*. 2000;1(6):510-514. doi:10.1038/82763

Chu Y, Corey DR. RNA sequencing: platform selection, experimental design, and data interpretation. *Nucleic Acid Ther*. 2012;22(4):271-274. doi:10.1089/nat.2012.0367

Cisse B, Caton ML, Lehner M, et al. Transcription factor E2-2 is an essential and specific regulator of plasmacytoid dendritic cell development. *Cell*. 2008;135(1):37-48. doi:10.1016/j.cell.2008.09.016

Cocucci E, Meldolesi J. Nonsecretory, Regulated Exocytosis: A Multifarious Mechanism Employed by Cells to Carry Out a Variety of Functions. In: *Madame Curie Bioscience Database [Internet]*. Austin (TX): Landes Bioscience; 2000-2013. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK6363/>

Collin J, Queen R, Zerti D, et al. Deconstructing Retinal Organoids: Single Cell RNA-Seq Reveals the Cellular Components of Human Pluripotent Stem Cell-Derived Retina. *Stem Cells*. 2019;37(5):593-598. doi:10.1002/stem.2963

Collin M, Bigley V. Human dendritic cell subsets: an update. *Immunology*. 2018;154(1):3-20. doi:10.1111/imm.12888

Collin M, Bigley V. Monocyte, Macrophage, and Dendritic Cell Development: the Human Perspective. *Microbiol Spectr*. 2016;4(5):10.1128/microbiolspec.MCHD-0015-2015. doi:10.1128/microbiolspec.MCHD-0015-2015

Collin M, McGovern N, Haniffa M. Human dendritic cell subsets. *Immunology*. 2013;140(1):22-30. doi:10.1111/imm.12117

Collin M, Milne P. Langerhans cell origin and regulation. *Curr Opin Hematol*. 2016;23(1):28-35. doi:10.1097/MOH.0000000000000202

Colonna M, Trinchieri G, Liu YJ. Plasmacytoid dendritic cells in immunity. *Nat Immunol*. 2004;5(12):1219-1226. doi:10.1038/ni1141

Crozat K, Guiton R, Contreras V, et al. The XC chemokine receptor 1 is a conserved selective marker of mammalian cells homologous to mouse CD8alpha+ dendritic cells. *J Exp Med*. 2010;207(6):1283-1292. doi:10.1084/jem.20100223

Cytlak U, Resteu A, Bogaert D, et al. Ikaros family zinc finger 1 regulates dendritic cell development and function in humans. *Nat Commun*. 2018;9(1):1239. Published 2018 Mar 27. doi:10.1038/s41467-018-02977-8

Cytlak U, Resteu A, Pagan S, et al. Differential IRF8 Transcription Factor Requirement Defines Two Pathways of Dendritic Cell Development in Humans. *Immunity*. 2020;53(2):353-370.e8. doi:10.1016/j.immuni.2020.07.003

Dalod M, Chelbi R, Malissen B, Lawrence T. Dendritic cell maturation: functional specialization through signaling specificity and transcriptional programming. *EMBO J*. 2014;33(10):1104-1116. doi:10.1002/embj.201488027

Dickinson RE, Griffin H, Bigley V, et al. Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. *Blood*. 2011;118(10):2656-2658. doi:10.1182/blood-2011-06-360313

Dickinson RE, Milne P, Jardine L, et al. The evolution of cellular deficiency in GATA2 mutation. *Blood*. 2014;123(6):863-874. doi:10.1182/blood-2013-07-517151

Donev R. *Advances in Protein Chemistry and Structural Biology Volume 107, 2017, Pages 191-222* doi:10.1016/bs.apcsb.2016.11.003

Dong J, McPherson CM, Stambrook PJ. Flt-3 ligand: a potent dendritic cell stimulator and novel antitumor agent. *Cancer Biol Ther*. 2002;1(5):486-489. doi:10.4161/cbt.1.5.161

Doulatov S, Notta F, Eppert K, Nguyen LT, Ohashi PS, Dick JE. Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat Immunol*. 2010;11(7):585-593. doi:10.1038/ni.1889

Dror I, Rohs R, Mandel-Gutfreund Y. How motif environment influences transcription factor search dynamics: Finding a needle in a haystack. *Bioessays*. 2016;38(7):605-612. doi:10.1002/bies.201600005

Durai V, Bagadia P, Granja JM, et al. Cryptic activation of an *Irf8* enhancer governs cDC1 fate specification. *Nat Immunol*. 2019;20(9):1161-1173. doi:10.1038/s41590-019-0450-x

Dutertre CA, Becht E, Irac SE, et al. Single-Cell Analysis of Human Mononuclear Phagocytes Reveals Subset-Defining Markers and Identifies Circulating Inflammatory Dendritic Cells. *Immunity*. 2019;51(3):573-589.e8. doi:10.1016/j.immuni.2019.08.008

Edwards JC, Everett HE, Pedrera M, et al. CD1⁻ and CD1⁺ porcine blood dendritic cells are enriched for the orthologues of the two major mammalian conventional subsets. *Sci Rep*. 2017;7:40942. Published 2017 Jan 20. doi:10.1038/srep40942

Ellisen LW, Bird J, West DC, et al. TAN-1, the human homolog of the *Drosophila* notch gene, is broken by chromosomal translocations in T lymphoblastic neoplasms. *Cell*. 1991;66(4):649-661. doi:10.1016/0092-8674(91)90111-b

- Embgenbroich M, Burgdorf S. Current Concepts of Antigen Cross-Presentation. *Front Immunol.* 2018;9:1643. Published 2018 Jul 16. doi:10.3389/fimmu.2018.01643
- Evans C, Hardin J, Stoebel DM. Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Brief Bioinform.* 2018;19(5):776–792. doi:10.1093/bib/bbx008
- Fancke B, Suter M, Hochrein H, O'Keeffe M. M-CSF: a novel plasmacytoid and conventional dendritic cell poietin. *Blood.* 2008;111(1):150-159. doi:10.1182/blood-2007-05-089292
- Feinberg MW, Wara AK, Cao Z, et al. The Kruppel-like factor KLF4 is a critical regulator of monocyte differentiation. *EMBO J.* 2007;26(18):4138-4148. doi:10.1038/sj.emboj.7601824
- Fujii Y, Shimizu T, Kusumoto M, Kyogoku Y, Taniguchi T, Hakoshima T. Crystal structure of an IRF-DNA complex reveals novel DNA recognition and cooperative binding to a tandem repeat of core sequences. *EMBO J.* 1999;18(18):5028-5041. doi:10.1093/emboj/18.18.5028
- Gadalla R, Noamani B, MacLeod BL, et al. Validation of CyTOF Against Flow Cytometry for Immunological Studies and Monitoring of Human Cancer Clinical Trials. *Front Oncol.* 2019;9:415. Published 2019 May 17. doi:10.3389/fonc.2019.00415
- Gauzzi MC, Del Cornò M, Gessani S. Dissecting TLR3 signalling in dendritic cells. *Immunobiology.* 2010;215(9-10):713-723. doi:10.1016/j.imbio.2010.05.008
- Geissmann F, Manz MG, Jung S, Sieweke MH, Merad M, Ley K. Development of monocytes, macrophages, and dendritic cells [published correction appears in *Science.* 2010 Dec 3;330(6009):1319]. *Science.* 2010;327(5966):656-661. doi:10.1126/science.1178331
- Gieseler R, Heise D, Soruri A, Schwartz P, Peters JH. In-vitro differentiation of mature dendritic cells from human blood monocytes. *Dev Immunol.* 1998;6(1-2):25-39. doi:10.1155/1998/72054

- Gilfillan GD, Hughes T, Sheng Y, et al. Limitations and possibilities of low cell number ChIP-seq. *BMC Genomics*. 2012;13:645. Published 2012 Nov 21. doi:10.1186/1471-2164-13-645
- Gowder S. *Cell Interaction*. IntechOpen. Published: October 10, 2012. doi:10.5772/3194
- Grajales-Reyes GE, Iwata A, Albring J, et al. Batf3 maintains autoactivation of Irf8 for commitment of a CD8 α (+) conventional DC clonogenic progenitor. *Nat Immunol*. 2015;16(7):708-717. doi:10.1038/ni.3197
- Griesbeck M, Ziegler S, Laffont S, et al. Sex Differences in Plasmacytoid Dendritic Cell Levels of IRF5 Drive Higher IFN- α Production in Women. *J Immunol*. 2015;195(11):5327-5336. doi:10.4049/jimmunol.1501684
- Grouard G, Rissoan MC, Filgueira L, Durand I, Banchereau J, Liu YJ. The enigmatic plasmacytoid T cells develop into dendritic cells with interleukin (IL)-3 and CD40-ligand. *J Exp Med*. 1997;185(6):1101-1111. doi:10.1084/jem.185.6.1101
- Grün D, van Oudenaarden A. Design and Analysis of Single-Cell Sequencing Experiments. *Cell*. 2015;163(4):799-810. doi:10.1016/j.cell.2015.10.039
- Gubbels Bupp MR, Potluri T, Fink AL, Klein SL. The Confluence of Sex Hormones and Aging on Immunity. *Front Immunol*. 2018;9:1269. Published 2018 Jun 4. doi:10.3389/fimmu.2018.01269
- Guilliams M, Dutertre CA, Scott CL, et al. Unsupervised High-Dimensional Analysis Aligns Dendritic Cells across Tissues and Species. *Immunity*. 2016;45(3):669-684. doi:10.1016/j.immuni.2016.08.015
- Guilliams M, Ginhoux F, Jakubzick C, et al. Dendritic cells, monocytes and macrophages: a unified nomenclature based on ontogeny. *Nat Rev Immunol*. 2014;14(8):571-578. doi:10.1038/nri3712
- Gupta K, Sari-Ak D, Haffke M, Trowitzsch S, Berger I. Zooming in on Transcription Preinitiation. *J Mol Biol*. 2016;428(12):2581-2591. doi:10.1016/j.jmb.2016.04.003

- Gutiérrez-Martínez E, Planès R, Anselmi G, et al. Cross-Presentation of Cell-Associated Antigens by MHC Class I in Dendritic Cell Subsets. *Front Immunol*. 2015;6:363. Published 2015 Jul 17. doi:10.3389/fimmu.2015.00363
- Haghverdi L, Buettner F, Theis FJ. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics*. 2015;31(18):2989-2998. doi:10.1093/bioinformatics/btv325
- Hambleton S, Salem S, Bustamante J, et al. IRF8 mutations and human dendritic-cell immunodeficiency. *N Engl J Med*. 2011;365(2):127-138. doi:10.1056/NEJMoa1100066
- Hammerstrom AE, Cauley DH, Atkinson BJ, Sharma P. Cancer immunotherapy: sipuleucel-T and beyond. *Pharmacotherapy*. 2011;31(8):813-828. doi:10.1592/phco.31.8.813
- Hampsey M. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev*. 1998;62(2):465-503.
- Haniffa M, Bigley V, Collin M. Human mononuclear phagocyte system reunited. *Semin Cell Dev Biol*. 2015;41:59-69. doi:10.1016/j.semcdb.2015.05.004
- Haniffa M, Shin A, Bigley V, et al. Human tissues contain CD141^{hi} cross-presenting dendritic cells with functional homology to mouse CD103⁺ nonlymphoid dendritic cells. *Immunity*. 2012;37(1):60-73. doi:10.1016/j.immuni.2012.04.012
- Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013;14:7. Published 2013 Jan 16. doi:10.1186/1471-2105-14-7
- Harry RA, Anderson AE, Isaacs JD, Hilken CM. Generation and characterisation of therapeutic tolerogenic dendritic cells for rheumatoid arthritis. *Ann Rheum Dis*. 2010;69(11):2042-2050. doi:10.1136/ard.2009.126383
- Hasegawa H, Matsumoto T. Mechanisms of Tolerance Induction by Dendritic Cells In Vivo. *Front Immunol*. 2018;9:350. Published 2018 Feb 26. doi:10.3389/fimmu.2018.00350

- Helft J, Anjos-Afonso F, van der Veen AG, Chakravarty P, Bonnet D, Reis e Sousa C. Dendritic Cell Lineage Potential in Human Early Hematopoietic Progenitors. *Cell Rep.* 2017;20(3):529-537. doi:10.1016/j.celrep.2017.06.075
- Helft J, Böttcher J, Chakravarty P, et al. GM-CSF Mouse Bone Marrow Cultures Comprise a Heterogeneous Population of CD11c(+)MHCII(+) Macrophages and Dendritic Cells. *Immunity.* 2015;42(6):1197-1211. doi:10.1016/j.immuni.2015.05.018
- Hémont C, Neel A, Heslan M, Braudeau C, Josien R. Human blood mDC subsets exhibit distinct TLR repertoire and responsiveness. *J Leukoc Biol.* 2013;93(4):599-609. doi:10.1189/jlb.0912452
- Hoeffel G, Wang Y, Greter M, et al. Adult Langerhans cells derive predominantly from embryonic fetal liver monocytes with a minor contribution of yolk sac-derived macrophages. *J Exp Med.* 2012;209(6):1167-1181. doi:10.1084/jem.20120340
- Hofer TP, van de Loosdrecht AA, Stahl-Hennig C, Cassatella MA, Ziegler-Heitbrock L. 6-Sulfo LacNAc (Slan) as a Marker for Non-classical Monocytes. *Front Immunol.* 2019;10:2052. Published 2019 Sep 13. doi:10.3389/fimmu.2019.02052
- Hoffman EA, Frey BL, Smith LM, Auble DT. Formaldehyde crosslinking: a tool for the study of chromatin complexes. *J Biol Chem.* 2015;290(44):26404-26411. doi:10.1074/jbc.R115.651679
- Hoffman R, Benz EJ, Silberstein L, et al. *Hematology: Basic Principles and Practice E-Book, Seventh Edition.* Elsevier, 2018 p.252
- Huang W, Horvath E, Eklund EA. PU.1, interferon regulatory factor (IRF) 2, and the interferon consensus sequence-binding protein (ICSBP/IRF8) cooperate to activate NF1 transcription in differentiating myeloid cells. *J Biol Chem.* 2007;282(9):6629-6643. doi:10.1074/jbc.M607760200
- Hubo M, Trinschek B, Kryczanowsky F, Tuettenberg A, Steinbrink K, Jonuleit H. Costimulatory molecules on immunogenic versus tolerogenic human dendritic cells. *Front Immunol.* 2013;4:82. Published 2013 Apr 3. doi:10.3389/fimmu.2013.00082

Humblin E, Thibaudin M, Chalmin F, et al. IRF8-dependent molecular complexes control the Th9 transcriptional program. *Nat Commun.* 2017;8(1):2085. Published 2017 Dec 12. doi:10.1038/s41467-017-01070-w

Inaba K, Inaba M, Romani N, et al. Generation of large numbers of dendritic cells from mouse bone marrow cultures supplemented with granulocyte/macrophage colony-stimulating factor. *J Exp Med.* 1992;176(6):1693-1702. doi:10.1084/jem.176.6.1693

Inaba K, Steinman RM, Van Voorhis WC, Muramatsu S. Dendritic cells are critical accessory cells for thymus-dependent antibody responses in mouse and in man. *Proc Natl Acad Sci U S A.* 1983;80(19):6041-6045. doi:10.1073/pnas.80.19.6041

Ishikawa F, Nihiro H, Iino T, et al. The developmental program of human dendritic cells is operated independently of conventional myeloid and lymphoid pathways. *Blood.* 2007;110(10):3591-3660. doi:10.1182/blood-2007-02-071613

Jagannathan-Bogdan M, Zon LI. Hematopoiesis. *Development.* 2013;140(12):2463-2467. doi:10.1242/dev.083147

Jassal B, Matthews L, Viteri G, et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* 2020;48(D1):D498-D503. doi:10.1093/nar/gkz1031

John LB, Ward AC. The Ikaros gene family: transcriptional regulators of hematopoiesis and immunity. *Mol Immunol.* 2011;48(9-10):1272-1278. doi:10.1016/j.molimm.2011.03.006

Kapsenberg ML. Dendritic-cell control of pathogen-driven T-cell polarization. *Nat Rev Immunol.* 2003;3(12):984-993. doi:10.1038/nri1246

Kassianos AJ, Hardy MY, Ju X, et al. Human CD1c (BDCA-1)+ myeloid dendritic cells secrete IL-10 and display an immuno-regulatory phenotype and function in response to *Escherichia coli*. *Eur J Immunol.* 2012;42(6):1512-1522. doi:10.1002/eji.201142098

Katsnelson A. Kicking off adaptive immunity: the discovery of dendritic cells. *J Exp Med.* 2006;203(7):1622. doi:10.1084/jem.2037fta

Kidder BL, Hu G, Zhao K. ChIP-Seq: technical considerations for obtaining high-quality data. *Nat Immunol.* 2011;12(10):918-922. Published 2011 Sep 20. doi:10.1038/ni.2117

Kingston D, Schmid MA, Onai N, Obata-Onai A, Baumjohann D, Manz MG. The concerted action of GM-CSF and Flt3-ligand on in vivo dendritic cell homeostasis. *Blood.* 2009;114(4):835-843. doi:10.1182/blood-2009-02-206318

Kirkling ME, Cytlak U, Lau CM, et al. Notch Signaling Facilitates In Vitro Generation of Cross-Presenting Classical Dendritic Cells. *Cell Rep.* 2018;23(12):3658-3672.e6. doi:10.1016/j.celrep.2018.05.068

Kobayashi KS, van den Elsen PJ. NLRC5: a key regulator of MHC Class I-dependent immune responses. *Nat Rev Immunol.* 2012;12(12):813-820. doi:10.1038/nri3339

Koch U, Lehal R, Radtke F. Stem cells living with a Notch. *Development.* 2013;140(4):689-704. doi:10.1242/dev.080614

Kolde R. pheatmap: Pretty Heatmaps. R package version 1.0.12., 2019. <https://CRAN.R-project.org/package=pheatmap>

Kong XF, Martinez-Barricarte R, Kennedy J, et al. Disruption of an antimycobacterial circuit between dendritic and helper T cells in human SPPL2a deficiency. *Nat Immunol.* 2018;19(9):973-985. doi:10.1038/s41590-018-0178-z

Kramer IM. *Signal Transduction 3rd Edition*, Academic Press, 2015. Chapter 19 - Cell Fate Determination by Notch, Pages 997-1039

Kurotaki D, Osato N, Nishiyama A, et al. Essential role of the IRF8-KLF4 transcription factor cascade in murine monocyte differentiation. *Blood.* 2013;121(10):1839-1849. doi:10.1182/blood-2012-06-437863

Kurotaki D, Yamamoto M, Nishiyama A, et al. IRF8 inhibits C/EBP α activity to restrain mononuclear phagocyte progenitors from differentiating into neutrophils. *Nat Commun.* 2014;5:4978. Published 2014 Sep 19. doi:10.1038/ncomms5978

Kurts C, Robinson BW, Knolle PA. Cross-priming in health and disease. *Nat Rev Immunol.* 2010;10(6):403-414. doi:10.1038/nri2780

- Lambert SA, Jolma A, Campitelli LF, et al. The Human Transcription Factors [published correction appears in Cell. 2018 Oct 4;175(2):598-599]. Cell. 2018;172(4):650-665. doi:10.1016/j.cell.2018.01.029
- Landt SG, Marinov GK, Kundaje A, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 2012;22(9):1813-1831. doi:10.1101/gr.136184.111
- Langlais D, Barreiro LB, Gros P. The macrophage IRF8/IRF1 regulome is required for protection against infections and is associated with chronic inflammation. J Exp Med. 2016;213(4):585-603. doi:10.1084/jem.20151764
- Laurens van der Maaten, Geoffrey Hinton. Visualizing Data using t-SNE. The Journal of Machine Learning Research. 2008 9(Nov):2579—2605.
- Lee J, Breton G, Oliveira TY, et al. Restricted dendritic cell and monocyte progenitors in human cord blood and bone marrow. J Exp Med. 2015;212(3):385-399. doi:10.1084/jem.20141442
- Lee J, Zhou YJ, Ma W, et al. Lineage specification of human dendritic cells is marked by IRF8 expression in hematopoietic stem cells and multipotent progenitors. Nat Immunol. 2017;18(8):877-888. doi:10.1038/ni.3789
- Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. Cell. 2013;152(6):1237-1251. doi:10.1016/j.cell.2013.02.014
- Lewis KL, Caton ML, Bogunovic M, et al. Notch2 receptor signaling controls functional differentiation of dendritic cells in the spleen and intestine. Immunity. 2011;35(5):780-791. doi:10.1016/j.immuni.2011.08.013
- Lewis KL, Caton ML, Bogunovic M, et al. Notch2 receptor signaling controls functional differentiation of dendritic cells in the spleen and intestine. Immunity. 2011;35(5):780-791. doi:10.1016/j.immuni.2011.08.013
- Li H, Burgueño-Bucio E, Xu S, et al. CD5 on dendritic cells regulates CD4+ and CD8+ T cell activation and induction of immune responses. PLoS One. 2019;14(9):e0222301. Published 2019 Sep 6. doi:10.1371/journal.pone.0222301

Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079. doi:10.1093/bioinformatics/btp352

Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v1 [q-bio.GN]. 2013

Li P, Wong JJ, Sum C, et al. IRF8 and IRF3 cooperatively regulate rapid interferon- β induction in human blood monocytes. *Blood*. 2011;117(10):2847-2854. doi:10.1182/blood-2010-07-294272

Lu L, et al. Chapter Forty-Seven - Natural killer cell induction of tolerance, Natural Killer Cells, Academic Press, 2010, Pages 617-631

Mace EM, Bigley V, Gunesch JT, et al. Biallelic mutations in IRF8 impair human NK cell maturation and function. *J Clin Invest*. 2017;127(1):306-320. doi:10.1172/JCI86276

Macosko EZ, Basu A, Satija R, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*. 2015;161(5):1202-1214. doi:10.1016/j.cell.2015.05.002

Manicassamy S, Reizis B, Ravindran R, et al. Activation of beta-catenin in dendritic cells regulates immunity versus tolerance in the intestine [published correction appears in *Science*. 2011 Nov 4;334(6056):594]. *Science*. 2010;329(5993):849-853. doi:10.1126/science.1188510

Maniecki MB, Møller HJ, Moestrup SK, Møller BK. CD163 positive subsets of blood dendritic cells: the scavenging macrophage receptors CD163 and CD91 are coexpressed on human dendritic cells and monocytes. *Immunobiology*. 2006 ;211(6-8):407-417. DOI: 10.1016/j.imbio.2006.05.019.

Mantegazza AR, Magalhaes JG, Amigorena S, Marks MS. Presentation of phagocytosed antigens by MHC Class I and II. *Traffic*. 2013;14(2):135-152. doi:10.1111/tra.12026

Manz MG. Plasmacytoid dendritic cells: origin matters. *Nat Immunol*. 2018;19(7):652-654. doi:10.1038/s41590-018-0143-x

Maraskovsky E, Brasel K, Teepe M, et al. Dramatic increase in the numbers of functionally mature dendritic cells in Flt3 ligand-treated mice: multiple dendritic cell subpopulations identified. *J Exp Med*. 1996;184(5):1953-1962.

doi:10.1084/jem.184.5.1953

Maraskovsky E, Daro E, Roux E, et al. In vivo generation of human dendritic cell subsets by Flt3 ligand. *Blood*. 2000;96(3):878-884.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads, *EMBnet.journal* 17.1 2011, doi:10.14806/ej.17.1.200

McElhinny AS, Li JL, Wu L. Mastermind-like transcriptional co-activators: emerging roles in regulating cross talk among multiple signaling pathways. *Oncogene*.

2008;27(38):5138-5147. doi:10.1038/onc.2008.228

Mendoza-Parra MA, Gronemeyer H. Assessing quality standards for ChIP-seq and related massive parallel sequencing-generated datasets: When rating goes beyond avoiding the crisis. *Genom Data*. 2014;2:268-273. Published 2014 Aug 14.

doi:10.1016/j.gdata.2014.08.002

Merad M, Sathe P, Helft J, Miller J, Mortha A. The dendritic cell lineage: ontogeny and function of dendritic cells and their subsets in the steady state and the inflamed setting. *Annu Rev Immunol*. 2013;31:563-604. doi:10.1146/annurev-immunol-

020711-074950

Mohaghegh N, Bray D, Keenan J, et al. NextPBM: a platform to study cell-specific transcription factor binding and cooperativity. *Nucleic Acids Res*. 2019;47(6):e31.

doi:10.1093/nar/gkz020

Moulton VR. Sex Hormones in Acquired Immunity and Autoimmune Disease. *Front Immunol*. 2018;9:2279. Published 2018 Oct 4. doi:10.3389/fimmu.2018.02279

Mundade R, Ozer HG, Wei H, Prabhu L, Lu T. Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. *Cell Cycle*. 2014;13(18):2847-2852.

doi:10.4161/15384101.2014.949201

Murphy K, Weaver C. *Janeway's Immunobiology*, Garland Science; 9th edition 2017

Murphy TL, Tussiwand R, Murphy KM. Specificity through cooperation: BATFIRF interactions control immune-regulatory networks. *Nat Rev Immunol* 2013;13:499-509.

Naik SH, Proietto AI, Wilson NS, et al. Cutting edge: generation of splenic CD8⁺ and CD8⁻ dendritic cell equivalents in Fms-like tyrosine kinase 3 ligand bone marrow cultures. *J Immunol*. 2005;174(11):6592-6597. doi:10.4049/jimmunol.174.11.6592

Nakano T, Kodama H, Honjo T. Generation of lymphohematopoietic cells from embryonic stem cells in culture. *Science*. 1994;265(5175):1098-1101. doi:10.1126/science.8066449

Nelson N, Marks MS, Driggers PH, Ozato K. Interferon consensus sequence-binding protein, a member of the interferon regulatory factor family, suppresses interferon-induced gene transcription. *Mol Cell Biol*. 1993;13(1):588-599. doi:10.1128/mcb.13.1.588

Nimmo RA, May GE, Enver T. Primed and ready: understanding lineage commitment through single cell analysis. *Trends Cell Biol*. 2015;25(8):459-467. doi:10.1016/j.tcb.2015.04.004

Notta F, Zandi S, Takayama N, et al. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science*. 2016;351(6269):aab2116. doi:10.1126/science.aab2116

Nussenzweig MC, Steinman RM, Gutchinov B, Cohn ZA. Dendritic cells are accessory cells for the development of anti-trinitrophenyl cytotoxic T lymphocytes. *J Exp Med*. 1980;152(4):1070-1084. doi:10.1084/jem.152.4.1070

Patel AA, Zhang Y, Fullerton JN, et al. The fate and lifespan of human monocyte subsets in steady state and systemic inflammation. *J Exp Med*. 2017;214(7):1913-1923. doi:10.1084/jem.20170355

Penton AL, Leonard LD, Spinner NB. Notch signaling in human development and disease. *Semin Cell Dev Biol*. 2012;23(4):450-457. doi:10.1016/j.semcdb.2012.01.010

Picelli S, Faridani OR, Björklund AK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc.* 2014;9(1):171-181. doi:10.1038/nprot.2014.006

Puhr S, Lee J, Zvezdova E, Zhou YJ, Liu K. Dendritic cell development-History, advances, and open questions. *Semin Immunol.* 2015;27(6):388-396. doi:10.1016/j.smim.2016.03.012

Qu C, Brinck-Jensen NS, Zang M, Chen K. Monocyte-derived dendritic cells: targets as potent antigen-presenting cells for the design of vaccines against infectious diseases. *Int J Infect Dis.* 2014;19:1-5. doi:10.1016/j.ijid.2013.09.023

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2019. URL <https://www.R-project.org/>

Reece JB. Combinatorial control of gene activation. In *Campbell Biology 2011*, 10th ed., pp. 37.

Ross-Innes CS, Stark R, Teschendorff AE, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature.* 2012;481(7381):389-393. Published 2012 Jan 4. doi:10.1038/nature10730

Sadeghzadeh M, Bornehdeli S, Mohahammadrezakhani H, et al. Dendritic cell therapy in cancer treatment; the state-of-the-art. *Life Sci.* 2020;254:117580. doi:10.1016/j.lfs.2020.117580

Salem S, Langlais D, Lefebvre F, et al. Functional characterization of the human dendritic cell immunodeficiency associated with the IRF8(K108E) mutation. *Blood.* 2014;124(12):1894-1904. doi:10.1182/blood-2014-04-570879

Saraya K, Reid CD. Stem cell factor and the regulation of dendritic cell production from CD34+ progenitors in bone marrow and cord blood. *Br J Haematol.* 1996;93(2):258-264. doi:10.1046/j.1365-2141.1996.5131053.x

Saxena M, Yeretssian G. NOD-Like Receptors: Master Regulators of Inflammation and Cancer. *Front Immunol.* 2014;5:327. Published 2014 Jul 14. doi:10.3389/fimmu.2014.00327

Schlitzer A, Zhang W, Song M, Ma X. Recent advances in understanding dendritic cell development, classification, and phenotype. *F1000Res*. 2018;7:F1000 Faculty Rev-1558. Published 2018 Sep 26. doi:10.12688/f1000research.14793.1

See P, Dutertre CA, Chen J, et al. Mapping the human DC lineage through the integration of high-dimensional techniques. *Science*. 2017;356(6342):eaag3009. doi:10.1126/science.aag3009

Segura E, Valladeau-Guilemond J, Donnadieu MH, Sastre-Garau X, Soumelis V, Amigorena S. Characterization of resident and migratory dendritic cells in human lymph nodes. *J Exp Med*. 2012;209(4):653-660. doi:10.1084/jem.20111457

Shin D.-M., Lee C.H., Morse H.C. IRF8 Governs Expression of Genes Involved in Innate and Adaptive Immunity in Human and Mouse Germinal Center B Cells. *PLoS One*. 2011;6(11):e27384. doi: 10.1371/journal.pone.0027384.

Shin KS, Jeon I, Kim BS, et al. Monocyte-Derived Dendritic Cells Dictate the Memory Differentiation of CD8⁺ T Cells During Acute Infection. *Front Immunol*. 2019;10:1887. Published 2019 Aug 16. doi:10.3389/fimmu.2019.01887

Sichien D, Scott CL, Martens L, et al. IRF8 Transcription Factor Controls Survival and Function of Terminally Differentiated Conventional and Plasmacytoid Dendritic Cells, Respectively. *Immunity*. 2016;45(3):626-640. doi:10.1016/j.immuni.2016.08.013

Sichien, D., Lambrecht, B., Guilliams, M. et al. Development of conventional dendritic cells: from common bone marrow progenitors to multiple subsets in peripheral tissues. *Mucosal Immunol* 10, 831–844 (2017). <https://doi.org/10.1038/mi.2017.8>

Sisirak V, Vey N, Vanbervliet B, et al. CCR6/CCR10-mediated plasmacytoid dendritic cell recruitment to inflamed epithelia after instruction in lymphoid tissues. *Blood*. 2011;118(19):5130-5140. doi:10.1182/blood-2010-07-295626

Spinelli L, Carpentier S, Montañana Sanchis F, Dalod M, Vu Manh TP. BubbleGUM: automatic extraction of phenotype molecular signatures and comprehensive visualization of multiple Gene Set Enrichment Analyses. *BMC Genomics*. 2015;16:814. Published 2015 Oct 19. doi:10.1186/s12864-015-2012-4

Steinman RM, Cohn ZA. Identification of a novel cell type in peripheral lymphoid organs of mice. I. Morphology, quantitation, tissue distribution. *J Exp Med*. 1973;137(5):1142-1162. doi:10.1084/jem.137.5.1142

Steinman RM, Hawiger D, Nussenzweig MC. Tolerogenic dendritic cells. *Annu Rev Immunol*. 2003;21:685-711. doi:10.1146/annurev.immunol.21.120601.141040

Steinman RM. Linking innate to adaptive immunity through dendritic cells. *Novartis Found Symp*. 2006;279:101-219.

Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550. doi:10.1073/pnas.0506580102

Sutermaster BA, Darling EM. Considerations for high-yield, high-throughput cell enrichment: fluorescence versus magnetic sorting. *Sci Rep*. 2019;9(1):227. Published 2019 Jan 18. doi:10.1038/s41598-018-36698-1

Tamura T, Kurotaki D, Koizumi S. Regulation of myelopoiesis by the transcription factor IRF8. *Int J Hematol*. 2015;101(4):342-351. doi:10.1007/s12185-015-1761-9

Tamura T, Yanai H, Savitsky D, Taniguchi T. The IRF family transcription factors in immunity and oncogenesis. *Annu Rev Immunol*. 2008;26:535-584. doi:10.1146/annurev.immunol.26.021607.090400

Tavian M, Biasch K, Sinka L, Vallet J, Péault B. Embryonic origin of human hematopoiesis. *Int J Dev Biol*. 2010;54(6-7):1061-1065. doi:10.1387/ijdb.103097mt

Theisen D, Murphy K. The role of cDC1s in vivo: CD8 T cell priming through cross-presentation. *F1000Res*. 2017;6:98. Published 2017 Feb 1. doi:10.12688/f1000research.9997.1

Tibúrcio R, Nunes S, Nunes I, et al. Molecular Aspects of Dendritic Cell Activation in Leishmaniasis: An Immunobiological View. *Front Immunol*. 2019;10:227. Published 2019 Feb 22. doi:10.3389/fimmu.2019.00227

Ting JP, Trowsdale J. Genetic control of MHC Class II expression. *Cell*. 2002;109 Suppl:S21-S33. doi:10.1016/s0092-8674(02)00696-7

Tolouei Semnani R, Moore V, Bennuru S, et al. Human monocyte subsets at homeostasis and their perturbation in numbers and function in filarial infection. *Infect Immun*. 2014;82(11):4438-4446. doi:10.1128/IAI.01973-14

Turcotte K, Gauthier S, Mitsos LM, et al. Genetic control of myeloproliferation in BXH-2 mice. *Blood*. 2004;103(6):2343-2350. doi:10.1182/blood-2003-06-1852

Uhlen M, Oksvold P, Fagerberg L, et al. Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol*. 2010;28(12):1248-1250. doi:10.1038/nbt1210-1248

Vallejos CA, Risso D, Scialdone A, Dudoit S, Marioni JC. Normalizing single-cell RNA sequencing data: challenges and opportunities. *Nat Methods*. 2017;14(6):565-571. doi:10.1038/nmeth.4292

Vander Lugt B, Riddell J, Khan AA, et al. Transcriptional determinants of tolerogenic and immunogenic states during dendritic cell maturation. *J Cell Biol*. 2017;216(3):779-792. doi:10.1083/jcb.201512012

van Kooten C, Fiore N, Trouw LA, et al. Complement production and regulation by dendritic cells: molecular switches between tolerance and immunity. *Mol Immunol*. 2008;45(16):4064-4072. doi:10.1016/j.molimm.2008.07.015

van Kooyk Y. C-type lectins on dendritic cells: key modulators for the induction of immune responses. *Biochem Soc Trans*. 2008;36(Pt 6):1478-1481. doi:10.1042/BST0361478

van Leeuwen-Kerkhoff N, Lundberg K, Westers TM, et al. Transcriptional profiling reveals functional dichotomy between human slan⁺ non-classical monocytes and myeloid dendritic cells. *J Leukoc Biol*. 2017;102(4):1055-1068. doi:10.1189/jlb.3MA0117-037R

van Willigen WW, Bloemendal M, Gerritsen WR, Schreibelt G, de Vries IJM, Bol KF. Dendritic Cell Cancer Therapy: Vaccinating the Right Patient at the Right Time. *Front Immunol*. 2018;9:2265. Published 2018 Oct 1. doi:10.3389/fimmu.2018.02265

- Velten L, Haas SF, Raffel S, et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol.* 2017;19(4):271-281. doi:10.1038/ncb3493
- Villani AC, Satija R, Reynolds G, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science.* 2017;356(6335):eaah4573. doi:10.1126/science.aah4573
- Villard J. Transcription regulation and human diseases. *Swiss Med Wkly.* 2004;134(39-40):571-579.
- Villicaña C, Cruz G, Zurita M. The basal transcription machinery as a target for cancer therapy. *Cancer Cell Int.* 2014;14(1):18. Published 2014 Feb 28. doi:10.1186/1475-2867-14-18
- Vodyanik MA, Bork JA, Thomson JA, Slukvin II. Human embryonic stem cell-derived CD34+ cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential. *Blood.* 2005;105(2):617-626. doi:10.1182/blood-2004-04-1649
- Vremec D, O'Keeffe M, Hochrein H, et al. Production of interferons by dendritic cells, plasmacytoid cells, natural killer cells, and interferon-producing killer dendritic cells. *Blood.* 2007;109(3):1165-1173. doi:10.1182/blood-2006-05-015354
- Vu Manh TP, Bertho N, Hosmalin A, Schwartz-Cornil I, Dalod M. Investigating Evolutionary Conservation of Dendritic Cell Subset Identity and Functions. *Front Immunol.* 2015;6:260. Published 2015 Jun 2. doi:10.3389/fimmu.2015.00260
- Wang JH, Nichogiannopoulou A, Wu L, et al. Selective defects in the development of the fetal and adult lymphoid system in mice with an Ikaros null mutation. *Immunity.* 1996;5(6):537-549. doi:10.1016/s1074-7613(00)80269-1
- Wang, X, He Y, Zhang Q, et al. Direct Comparative Analysis of 10X Genomics Chromium and Smart-seq2 bioRxiv 615013; 2019. doi:https://doi.org/10.1101/615013
- West HC, Bennett CL. Redefining the Role of Langerhans Cells As Immune Regulators within the Skin. *Front Immunol.* 2018;8:1941. Published 2018 Jan 5. doi:10.3389/fimmu.2017.01941

Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Wolf AA, Yáñez A, Barman PK, Goodridge HS. The Ontogeny of Monocyte Subsets. *Front Immunol.* 2019;10:1642. Published 2019 Jul 17.

doi:10.3389/fimmu.2019.01642

Wu X, Yang B, Udo-Inyang I, et al. Research Techniques Made Simple: Single-Cell RNA Sequencing and its Applications in Dermatology. *J Invest Dermatol.*

2018;138(5):1004-1009. doi:10.1016/j.jid.2018.01.026

Xiao TS, Ting JP. NLRX1 has a tail to tell. *Immunity.* 2012;36(3):311-312.

doi:10.1016/j.immuni.2012.03.002

Yan F, Powell DR, Curtis DJ, Wong NC. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biol.* 2020;21(1):22. Published 2020 Feb 3.

doi:10.1186/s13059-020-1929-3

Yanai H, Negishi H, Taniguchi T. The IRF family of transcription factors: Inception, impact and implications in oncogenesis. *Oncoimmunology.* 2012;1(8):1376-1386.

doi:10.4161/onci.22475

Yin X, Yu H, Jin X, et al. Human Blood CD1c+ Dendritic Cells Encompass CD5high and CD5low Subsets That Differ Significantly in Phenotype, Gene Expression, and Functions. *J Immunol.* 2017;198(4):1553-1564. doi:10.4049/jimmunol.1600193

Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284-287.

doi:10.1089/omi.2011.0118

Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics.* 2015;31(14):2382-2383.

doi:10.1093/bioinformatics/btv145

Yuan GC, Cai L, Elowitz M, et al. Challenges and emerging directions in single-cell analysis. *Genome Biol.* 2017;18(1):84. Published 2017 May 8. doi:10.1186/s13059-017-1218-y

Zeng P.-Y. In vivo dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation. *Biotechniques*. 2006 Dec;41(6):694, 696, 698.

Zhang JG, Czabotar PE, Policheni AN, et al. The dendritic cell receptor Clec9A binds damaged cells via exposed actin filaments. *Immunity*. 2012;36(4):646-657. doi:10.1016/j.immuni.2012.03.009

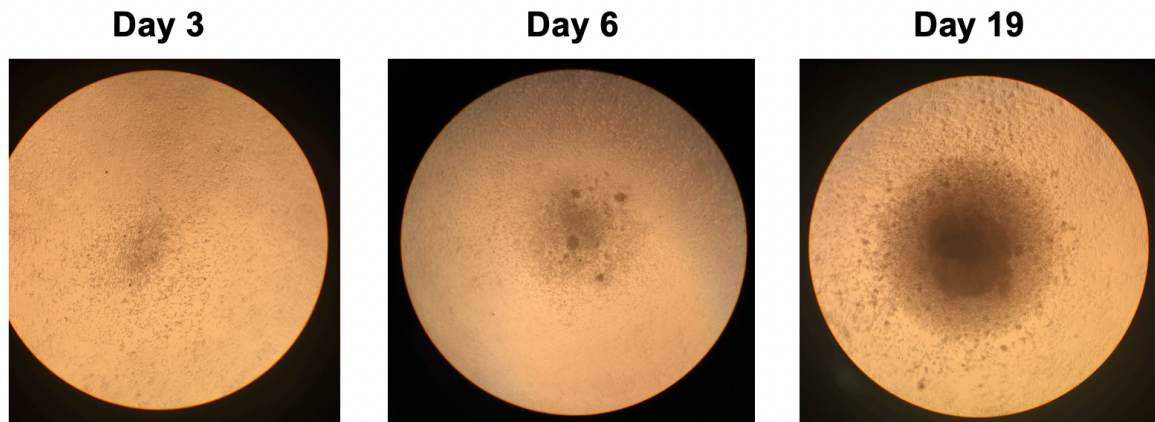
Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137. doi:10.1186/gb-2008-9-9-r137

Zhu H, Wang G, Qian J. Transcription factors as readers and effectors of DNA methylation. *Nat Rev Genet*. 2016;17(9):551-565. doi:10.1038/nrg.2016.83

Ziegler-Heitbrock L, Ancuta P, Crowe S, et al. Nomenclature of monocytes and dendritic cells in blood. *Blood*. 2010;116(16):e74-e80. doi:10.1182/blood-2010-02-258558

Zinkernagel RM, Doherty PC. The discovery of MHC restriction. *Immunol Today*. 1997;18(1):14-17. doi:10.1016/s0167-5699(97)80008-4

Appendix A. Microscopy of cells produced in culture



Microscopy of the cells produced within the OP9-DL1 culture system on days 3, 6, and 19 of culture.

3,000 bone marrow CD34⁺ progenitors were co-cultured with 5,000 OP9-DL1 murine bone marrow stromal cells in the presence of growth factors SCF, GM-CSF, and FLT3L. Microscopy was performed using the Olympus CK2 inverted microscope (100x magnification).

Appendix B. List of antibodies used for flow cytometry, FACS and mass cytometry in Chapter 4

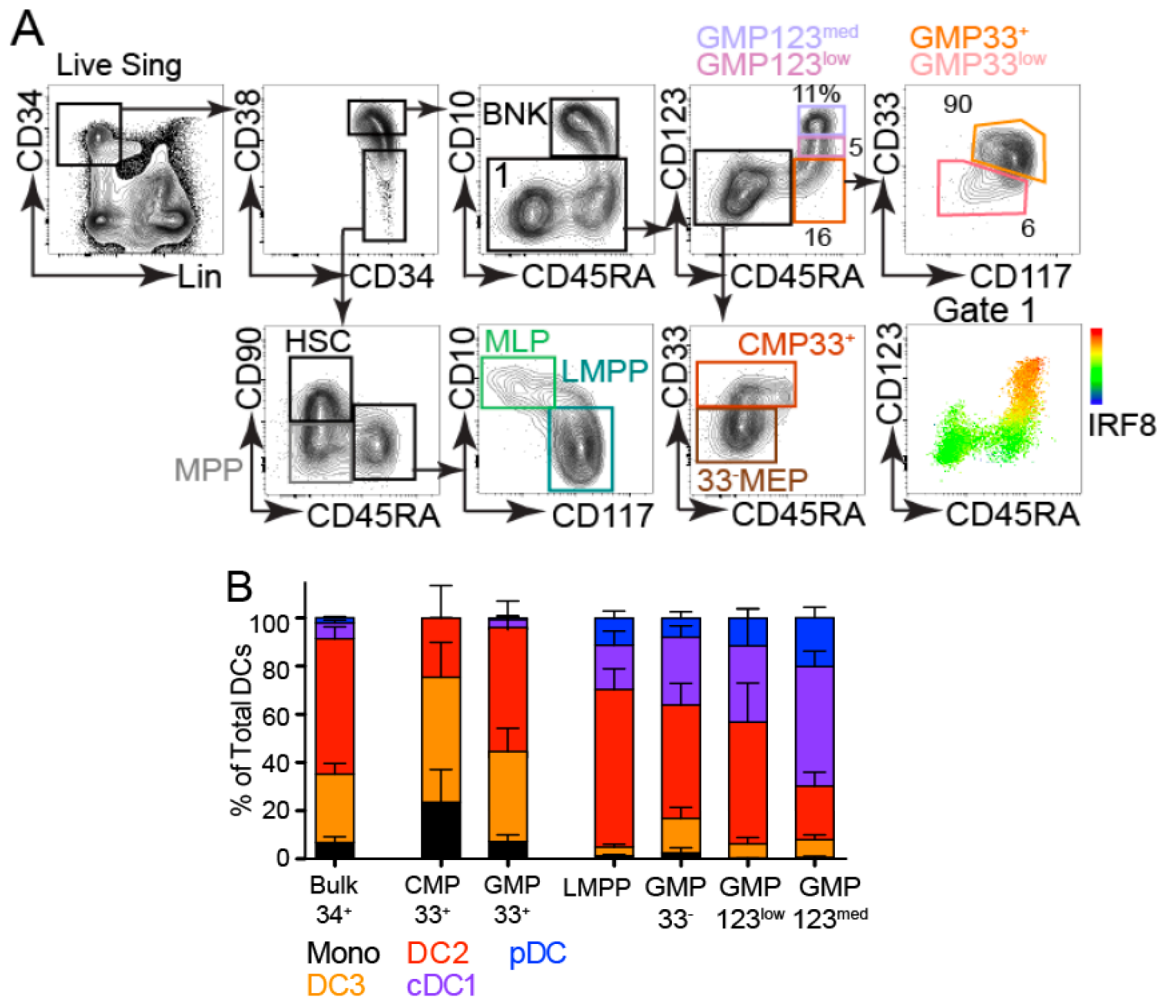
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse anti-APC 176 Yb, clone APC003	Fluidigm	Cat# 3176007B
Mouse anti-human AXL APC, clone 108724	R&D Systems	Cat# FAB154A
Mouse anti-human AXL purified, clone 108724	R&D Systems	Cat# MAB154; RRID:AB_2062558
Mouse anti-human BTLA 163Dy, clone MIH26	Fluidigm	Cat# 3163009B
Mouse anti-human BTLA BV650, clone J168-540	BD Biosciences	Cat# 564803; RRID:AB_2738962
Mouse anti-human CD1c APC-Cy7, clone L161	BioLegend	Cat# 331520; RRID:AB_10644008
Mouse anti-human CD1c PE-Cy7, clone L161	BioLegend	Cat# 331516; RRID:AB_2275574
Mouse anti-human CD1c PerCP-Cy5.5, clone L161	BioLegend	Cat# 331513; RRID:AB_1227536
Mouse anti-human CD1c purified, clone L161	BioLegend	Cat# 331502; RRID:AB_1088995
Mouse anti-human CD2 151Eu, clone TS1/8	Fluidigm	Cat# 3151003B
Mouse anti-human CD2 BV421, clone TS1/8	BioLegend	Cat# 309217; RRID:AB_10915139
Mouse anti-human CD2 PE-CF594, clone RPA-2.10	BD Biosciences	Cat# 562300; RRID:AB_11153492
Mouse anti-human CD3 AF700, clone SK7 (Leu-4)	BioLegend	Cat# 344822; RRID:AB_2563420
Mouse anti-human CD3 FITC, clone SK7(Leu-4)	BD Biosciences	Cat# 345763
Mouse anti-human CD3 PE, clone SK7(Leu9)	BD Biosciences	Cat# 345765
Mouse anti-human CD5 BUV737, clone UCHT2	BD Biosciences	Cat# 564451; RRID:AB_2714177
Mouse anti-human CD5 purified, clone L17F12	BioLegend	Cat# 364002; RRID:AB_2564477
Mouse anti-human CD7 FITC, clone Leu-9	BD Biosciences	Cat# 347483; RRID:AB_400309
Mouse anti-human CD7 PE, clone M-T701	BD Biosciences	Cat# 332774
Mouse anti-human CD10 156Gd, clone HI10a	Fluidigm	Cat# 3156001B
Mouse anti-human CD10 BV650, clone HI10a	BD Biosciences	Cat# 563734; RRID:AB_2738393
Mouse anti-human CD11b 144Nd, clone ICRF44	Fluidigm	Cat# 3144001B
Mouse anti-human CD11c 159Tb, clone Bu15	Fluidigm	Cat# 3159001B
Mouse anti-human CD11c AF700, clone B-ly6	BD Biosciences	Cat# 561352; RRID:AB_10612006
Mouse anti-human CD11c APC-Cy7, clone Bu15	BioLegend	Cat# 337218; RRID:AB_10662746
Mouse anti-human CD11c BV711, clone B-ly6	BioLegend	Cat# 301630; RRID:AB_2562192
Mouse anti-human CD14 BV650, clone M5E2	BioLegend	Cat# 301835; RRID:AB_11204241
Mouse anti-human CD14 FITC, clone M5E2	BD Biosciences	Cat# 555397; RRID:AB_395798
Mouse anti-human CD14 PE, clone M5E2	BD Biosciences	Cat# 555398; RRID:AB_395799
Mouse anti-human CD14 PE-Cy7, clone HCD14	BioLegend	Cat# 325618; RRID:AB_830691
Mouse anti-human CD14 purified, clone M5E2	BioLegend	Cat# 301802; RRID:AB_314184
Mouse anti-human CD15 164Dy, clone W6D3	Fluidigm	Cat# 3164001B
Mouse anti-human CD15 BUV395, clone HI98	BD Biosciences	Cat# 563872; RRID:AB_2738461

Mouse anti-human CD15 BV605, clone W6D3	BD Biosciences	Cat# 562979; RRID:AB_2744292
Mouse anti-human CD16 209Bi, clone 3G8	Fluidigm	Cat# 3209002B
Mouse anti-human CD16 AF700 , clone 3G8	BioLegend	Cat# 302026; RRID:AB_2278418
Mouse anti-human CD16 FITC, clone 3G8	BD Biosciences	Cat# 335035
Mouse anti-human CD16 PE, clone 3G8	BD Biosciences	Cat# 555407; RRID:AB_395807
Mouse anti-human CD16 PE-Dazzle594, clone 3G8	BioLegend	Cat# 302054; RRID:AB_2563639
Mouse anti-human CD19 AF700 , clone 4G7/HIB19	BioLegend	Cat# 302226; RRID:AB_493751
Mouse anti-human CD19 FITC, clone 4G7	BD Biosciences	Cat# 345776
Mouse anti-human CD19 PE, clone HIB19	BD Biosciences	Cat# 555413; RRID:AB_395813
Mouse anti-human CD20 AF700 , clone L27/2H7	BioLegend	Cat# 302322; RRID:AB_493753
Mouse anti-human CD20 FITC, clone L27	BD Biosciences	Cat# 345792
Mouse anti-human CD20 PE, clone L27	BD Biosciences	Cat# 345793
Mouse anti-human CD33 158Gd, clone WM53	Fluidigm	Cat# 3158001B
Mouse anti-human CD33 APC, clone P67.6	BD Biosciences	Cat# 345800
Mouse anti-human CD33 BV711, clone WM53	BD Biosciences	Cat# 563171; RRID:AB_2738045
Mouse anti-human CD34 166Er, clone 581	Fluidigm	Cat# 3166012B
Mouse anti-human CD34 APC-Cy7, clone 581	BioLegend	Cat# 343514; RRID:AB_1877168
Mouse anti-human CD34 BV605, clone 581	BioLegend	Cat# 343529; RRID:AB_2562193
Mouse anti-human CD34 FITC, clone 8G12	BD Biosciences	Cat# 345801
Mouse anti-human CD34 PE-CF594, clone 581	BD Biosciences	Cat# 562383; RRID:AB_11154586
Mouse anti-human CD36 155Gd, clone 5-271	Fluidigm	Cat# 3155012B
Mouse anti-human CD38 PE-Cy7, clone HB7	BD Biosciences	Cat# 335825
Mouse anti-human CD38 purified, clone HB-7	BioLegend	Cat# 356602; RRID:AB_2561794
Mouse anti-human CD45 89Y, clone HI30	Fluidigm	Cat# 3089003B
Mouse anti-human CD45 APC-Cy7, clone 2D1	BD Biosciences	Cat# 557833; RRID:AB_396891
Mouse anti-human CD45 V450, clone 2D1	BD Biosciences	Cat# 642275; RRID:AB_1645755
Mouse anti-human CD45RA 153Eu, clone HI100	Fluidigm	Cat# 3153001B
Mouse anti-human CD45RA BV510, clone HI100	BioLegend	Cat# 304142; RRID:AB_2561947
Rat anti-human CD52 PE, clone YTH34.5	Bio-Rad	Cat# SFL1642PE; RRID:AB_324131
Mouse anti-human CD56 FITC, clone NCAM16.2	BD Biosciences	Cat# 345811
Mouse anti-human CD88 PE, clone S5/1	BioLegend	Cat# 344304; RRID:AB_2067175
Mouse anti-human CD88 purified, clone C5AR	BioLegend	Cat# 344302; RRID:AB_2259318
Mouse anti-human CD90 161Dy, clone 5E10	Fluidigm	Cat# 3161009
Mouse anti-human CD90 AF700, clone 5E10	BioLegend	Cat# 328120; RRID:AB_2203302
Mouse anti-human CD90 PerCP-Cy5.5, clone 5E10	BioLegend	Cat# 328118; RRID:AB_2303335
Human anti-human CD100 APC-Vio770, clone REA316	Miltenyi Biotec	Cat# 130-104-604; RRID:AB_2654328
Mouse anti-human CD100 purified, clone A8	BioLegend	Cat# 328401; RRID:AB_1236386
Mouse anti-human CD115 purified, clone 9-4D2-1E4	BioLegend	Cat# 347302; RRID:AB_2085375

Mouse anti-human CD116 BV421, clone hGMCSFR-M1	BD Biosciences	Cat# 564045; RRID:AB_2738561
Mouse anti-human CD116 BV650, clone hGMCSFR-M1	BD Biosciences	Cat# 564044; RRID:AB_2738560
Mouse anti-human CD116 purified, clone 4H1	BioLegend	Cat# 305902; RRID:AB_314568
Mouse anti-human CD117 BV605, clone 104D2	BD Biosciences	Cat# 562687; RRID:AB_2737721
Mouse anti-human CD117 PE, clone 104D2	BD Biosciences	Cat# 332785
Mouse anti-human CD117 purified, clone 104D2	BioLegend	Cat# 313201; RRID:AB_314980
Mouse anti-human CD123 143Nd, clone 6H6	Fluidigm	Cat# 3143014B
Mouse anti-human CD123 BUV395, clone 7G3	BD Biosciences	Cat# 564195; RRID:AB_2714171
Mouse anti-human CD123 BV421, clone 6H6	BioLegend	Cat# 306018; RRID:AB_10962571
Mouse anti-human CD123 PerCP-Cy5.5, clone 7G3	BD Biosciences	Cat# 558714; RRID:AB_1645547
Mouse anti-human CD135 BV711, clone 4G8	BD Biosciences	Cat# 563908; RRID:AB_2738479
Mouse anti-human CD135 purified, clone BV10A4H2	BioLegend	Cat# 313302; RRID:AB_314987
Mouse anti-human CD141 BV510, clone 1A4	BD Biosciences	Cat# 563298; RRID:AB_2728103
Mouse anti-human CD141 purified, clone M80	BioLegend	Cat# 344102; RRID:AB_2201808
Mouse anti-human CD161 PE-Cy7, clone HP-3G10	Thermo Fisher Scientific	Cat# 25-1619-42; RRID:AB_10807086
Mouse anti-human CD303 147Sm, clone 201A	Fluidigm	Cat# 3147009B
Mouse anti-human CD303 APC, clone 201A	BioLegend	Cat# 354206; RRID:AB_11150412
Mouse anti-human CD303 BV605, clone 201A	BioLegend	Cat# 354224; RRID:AB_2572149
Mouse anti-human CD304 169Tm, clone 12C2	Fluidigm	Cat# 3169018B
Mouse anti-human CD304 APC, clone 12C2	BioLegend	Cat# 354506; RRID:AB_11219600
Mouse anti-human CD304 BV605, clone U21-1283	BD Biosciences	Cat# 743130; RRID:AB_2741297
Mouse anti-human CLEC9A PE, clone 8F9	BioLegend	Cat# 353804; RRID:AB_10965546
Mouse anti-human CLEC9A purified, clone 8F9	BioLegend	Cat# 353802; RRID:AB_10983070
Rat anti-human CX3CR1 APC, clone 2A9-1	BioLegend	Cat# 341610; RRID:AB_2087424
Mouse anti-human FcεRI 150Nd, clone AER-37 (CRA-1)	Fluidigm	Cat# 3150027B
Mouse anti-FITC purified, clone FIT-22	BioLegend	Cat# 408305; RRID:AB_2563769
Mouse anti-human HLA-DR 173Yb, clone L243	Fluidigm	Cat# 3173005B
Mouse anti-human HLA-DR AF700, clone G46-6	BD Biosciences	Cat# 560743; RRID:AB_1727526
Mouse anti-human HLA-DR BV785, clone L243	BioLegend	Cat# 307642; RRID:AB_2563461
Mouse anti-human HLA-DR PerCP-Cy5.5, clone L243	BioLegend	Cat# 307629; RRID:AB_893575
Mouse anti-human ID2 purified, clone 4E12G5	Thermo Fisher Scientific	Cat# MA5-17095; RRID:AB_2538566
Mouse anti-human IFN-α PE, clone LT27:295	Miltenyi Biotec	Cat# 130-092-601; RRID:AB_871560
Rat anti-human IL-10 APC, clone JES3-9D7	BioLegend	Cat# 501410; RRID:AB_315176
Mouse anti-human IL-12p40/p70 BV421, clone C8.6	BD Biosciences	Cat# 565023; RRID:AB_2739045
Mouse anti-human IL-1b FITC, clone JK1B-1	BioLegend	Cat# 508206; RRID:AB_345362

Mouse anti-human IL-8 PE-Cy7, clone E8N1	BioLegend	Cat# 511416; RRID:AB_2565291
Rat anti-human IRF4 PE, clone 3E4	Thermo Fisher Scientific	Cat# 12-9858-80; RRID:AB_10853179
Mouse anti-human IRF4 purified, clone IRF4.3E4	BioLegend	Cat# 646402; RRID:AB_2280462
Mouse anti-human IRF8 efluor710, clone 3GYWCH	Thermo Fisher Scientific	Cat# 46-9852-80; RRID:AB_2573903
Mouse anti-human IRF8 purified, clone GW4CML3	Thermo Fisher Scientific	Cat# 14-7888-82; RRID:AB_2572907
Goat anti-human KLF4 APC, clone POLY	R&D Systems	Cat# IC3640A; RRID:AB_2044690
Mouse anti-PE purified, clone PE001	BioLegend	Cat# 408105; RRID:AB_2563787
Mouse anti-human SIGLEC-6 purified, clone 767329	R&D Systems	Cat# MAB2859
Mouse anti-human SIRPa purified, clone 15-414	BioLegend	Cat# 372102; RRID:AB_2629807
Mouse anti-human SIRPa/b AF700, clone SE5A5	BioLegend	Cat# 323816; RRID:AB_2687275
Mouse anti-human SIRPa/b APC, clone SE5A5	BioLegend	Cat# 323809; RRID:AB_11219399
Mouse anti-human SIRPa/b PE, clone SE5A5	BioLegend	Cat# 323805; RRID:AB_830704
Mouse anti-human SLAN PE, clone DD1	Miltenyi Biotec	Cat# 130-093-029; RRID:AB_871582

Appendix C. Detailed gating strategy for the FACS-purification of bone marrow progenitors for scRNA-Seq

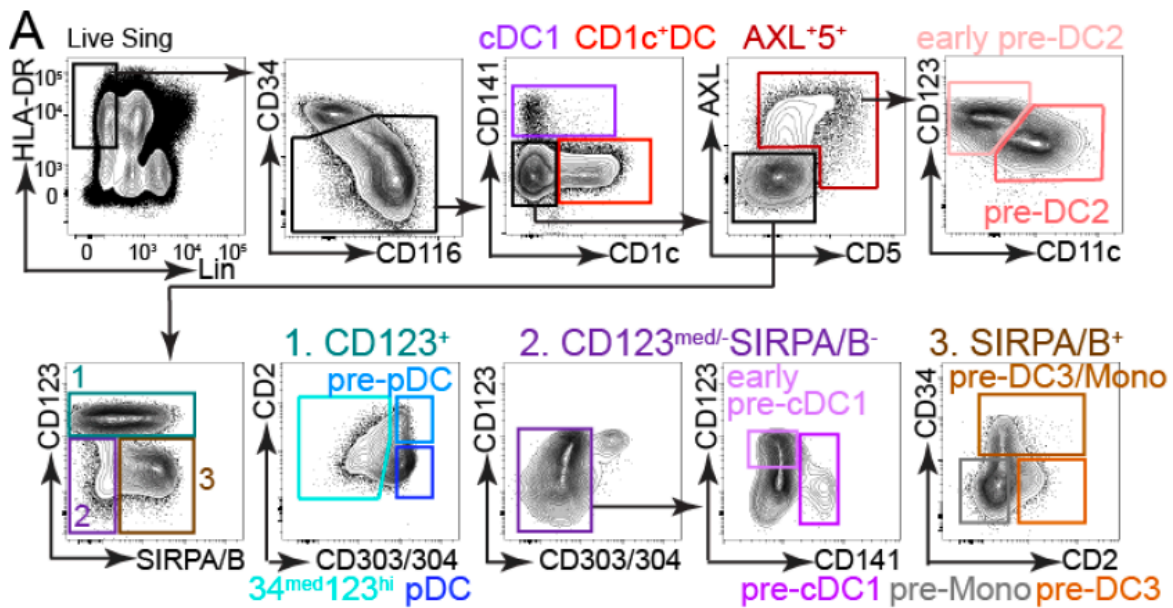


Identification of the components of the CD34⁺ lineage(CD3,14,16,19,20,7)⁻ progenitor compartment of human BM (Cytlak and Resteu et al., 2020).

A. Gating strategy used for the index sorting of bone marrow progenitors subjected to scRNA-Seq, along with IRF8 expression across the GMP region.

B. Culture output of FACS-purified bone marrow progenitors, assessed with flow cytometry following 14 days of culture. The proportion of generated DC subsets and monocytes is expressed as % of the total cells captured by all DC and monocyte gates. $n=3-9$ donors for each population. Bars represent mean+SEM. HSC, hematopoietic stem cell; MPP, multipotent progenitor; MEP, megakaryocyte erythroid progenitor; MLP, multilymphoid progenitor; LMPP, lymphoid primed multipotent progenitor; CMP, common myeloid progenitor; GMP, granulocyte macrophage progenitor.

Appendix D. Detailed gating strategy for the FACS-purification of bone marrow DC and monocyte precursors for scRNA-Seq



The identification of bone marrow pre-DC subjected to scRNA-Seq (Cytlak and Resteu et al., 2020).

A. Gating strategy used for the isolation of cells for scRNA-Seq.

B. Culture output of FACS-purified bone marrow DC precursors, assessed with flow cytometry following 14 days of culture. The proportion of generated DC subsets and monocytes is expressed as % of the total cells captured by all DC and monocyte gates. $n=3-4$ donors for each population. Bars represent mean+SEM.