Fakultät für Medizin

II. Medizinische Klinik und Poliklinik

The transcriptional landscape of plasmacytoid dendritic cell differentiation

Andrea Musumeci

Vollständiger Abdruck der von der Fakultät für Medizin der technischen Universität München zur Erlangung des akademischen Grades eines

Doctor of Philosophy (Ph.D.)

genehmigten Dissertation.

Vorsitzsender: Prof. Dr. Jürgen Ruland

Betreuerin: Prof. Dr. Anne Krug

Prüfer der Dissertation:

1. Prof. Dr. Thomas Korn
2. Prof. Dr. Roland Rad

Die Dissertation wurde am 14.06.2017 bei der Fakultät für Medizin der Technischen Universität München eingereicht und durch die fakultät für Medizin am 11.09.2017 angenommen.

Contents

List of Figures

List of Tables

Abbreviations

AIDS	acquired immune deficiency syndrome
ANODEV	analysis of deviance
ANOVA	analysis of variance
APC	antigen presenting cell
Batf	basic leucine zipper transcription factor, ATF-like
Bcl	B cell CLL/lymphoma (zinc finger protein)
BCR	B cell receptor
BDCA	blood dendritic cell antigen
BM	bone marrow
BST2	bone marrow stromal antigen 2
Cbf	core binding factor
CD	cluster of differentiation
cDC	conventional dendritic cell
CDP	common dendritic cell progenitor
CLP	common lymphoid progenitor
cMoP	committed monocyte progenitor
CMP	common myeloid progenitor
CRP	C-reactive protein
CTL	cytotoxic T lymphocytes
DC	dendritic cell
DTR	diphteria toxin receptor
EAE	experimental autoimmune encephalomyelitis
ER	endoplasmic reticulum
FCS	fetal calf serum

Flt3	FMS-like tyrosine kinase 3
GLM	generalized linear model
GM-CSF	granulocyte-macrophage colony-stimulating factor
GMP	granulocyte-macrophage progenitor
HIV	human immunodeficiency virus
HSC	hematopoietic stem cell
Id	inhibitor of DNA binding
IFN	interferon
IL	interleukin
Irf	interferon regulatory factor
LC	Langerhans cell
LN	lymph node
LRT	likelihood ratio test
MBP	mannose binding protein
MCMV	mouse cytomegalovirus
M-CSF	macrophage colony-stimulating factor
M-CSFR	macrophage colony-stimulating factor receptor
MDP	macrophage-DC progenitor
MHCI	major histocompatibility complex, class I
MHCII	major histocompatibility complex, class II
MSigDB	Molecular Signature Database
NEAA	non-essential aminoacids
NET	neutrophil extracellular trap
Nfil3	nuclear factor, interleukin 3, regulated
PBS	phosphate buffered saline
PCA	principal component analysis
pDC	plasmacytoid dendritic cell
PU.1	spleen focus forming virus (SFFV) proviral integration oncogene
Runx	runt related transcription factor
Siglec-H	sialic acid binding Ig-like lectin H
Tcf	transcription factor
TCR	T cell receptor

TF	transcription factor
TLR	toll-like receptor
TNF	tumor necrosis factor
TOM	Topological Overlap Matrix
VSV	vesicular stomatitis virus
WGCNA	Weighted Gene Co-expression Network Analysis
Zbtb	zinc finger and BTB domain containing
Zeb	zinc finger E-box binding homeobox

1 Introduction

1.1 Dendritic Cells and response to pathogens

In mammals and other animals, the response to pathogens is organized in two distinct but strictly interconnected components: an innate response, that arises rapidly and with limited specificity following tissue damage or pathogen encounter, directly in the affected tissue, and an adaptive response, which is organized in secondary lymphoid organs against specific antigens, in order to accurately remove every last trace of the offending pathogen without causing damage to the host.

1.1.1 Innate and adaptive immunity

The innate immune response is mastered by a series of specialized cells and soluble factors that are found both in the blood stream and within the tissues, and exert different functions[1]: Macrophages are professional phagocytes, that are able to actively remove pathogens from the site of infection, inactivate and process them, and can also present antigens on major histocompatibility complex, class I (MHCI) and II, contributing to the activation of effector T cells[2, 3]. Neutrophils (so called because of their typical neutral staining with hematoxylin and eosin (H&E) histological or cytological preparations) are also effective phagocytes and kill internalized bacteria and fungi, but differently from macrophages they are not able to present antigens on MHC class II. Neutrophils are able to extrude their nuclear chromatin in the form of neutrophil extracellular traps (NETs) upon encounter with pathogens, thus physically trapping the microorganisms and limiting their spreading through the organism[4]. The soluble factors are proteins that are able to non-specifically bind and opsonize microorganisms and apoptotic cells, such as complement pro-

teins[5], mannose binding protein (MBP)[6] and C-reactive protein (CRP)[7], as well as chemokines and cytokines which attract and activate innate and adaptive immune cells.

Adaptive immune responses are orchestrated in secondary lymphoid organs, and require receptors that are selected for reactivity with specific antigens (T cell receptors (TCRs) and B cell receptors (BCRs)). Antigens processed from the offending pathogens are carried to lymph nodes (LNs) and spleen by antigen presenting cells (APCs) and presented as peptides on MHC class I and II to naive T cells. T cells that specifically recognize the presented antigen are then activated and differentiate into effector T cells, acquiring the ability to directly kill infected cells (cluster of differentiation (CD)8^+ cytotoxic T lymphocytes (CTL)) or to help other effector T and B cells in their function by producing cytokines and growth factors ($CD4^+$ T helper cells).

The connecting elements of these responses are several and still not fully elucidated, but a major contribution is given by professional APCs, specialized cells that are able to recognize pathogens and efficiently process and present specific antigens to T and B cells, orchestrating the adaptive response.

1.1.2 Dendritic cells and subsets

Professional APCs include macrophages and dendritic cells (DCs), that often act in concert to present antigens and organize T cell responses[8]; however, macrophage's antigen presentation capacity is limited, while DCs possess highly specialized structures and mechanisms for antigen acquisition, processing and presentation, and for regulating activation and function of effector cells, by means of co-stimulatory molecules and cytokine production.

DCs were first identified in the 70s by Ralph Steinman [9–12]; their name is due to the characteristic morphology they exhibited, with numerous dynamic dendrites protruding from the cell body, continuously sampling the surrounding tissues. Since the discovery, a number of phenotypically and functionally different DC subsets have been identified, both in lymphoid and non-lymphoid tissues.

The human equivalents of all murine DC subpopulations have been identified. They resemble closely their murine counterparts in ontogeny and function and also

share some but not all of the phenotypic markers of murine DC subsets. In this chapter I will describe the murine DC subpopulations.

Lymphoid tissue DC subsets

DCs can be distinguished in two major subsets, with important phenotypic and functional differences: the conventional DCs (cDCs), which comprise several other subsets described in the next paragraphs, and the plasmacytoid DCs (pDCs), which are described in detail in the next section.

All mature DCs in mouse are characterized by expression of the integrin CD11c and of MHC class II; several other surface markers are used today to easily discriminate the different subsets: CD8α, CD4, CD11b, CD103, DCIR2, CD205, XCR1 and Sirpα.

In the spleen of wild type, healthy mice, where DCs were originally identified, two major subsets can be found: the $CD8^+$ $CD205^+$ cDC1, localized in the marginal zone and T cell zone of the follicles, in the white pulp[13], and the $CD8^-$ $CD11b^+$ $DCIR2^+$ cDC2 subset, residing in the red pulp. Functionally, these subsets are distinguished in their ability to present antigens: $CD8^+$ cDC1 are unique in their ability to capture extracellular antigens and cross-present them to $CD8^+$ T cells on MHC class I, thus eliciting a CTL response. In contrast, $CD8^-$ cDC2 in the spleen are more efficient in processing antigens and presenting them on MHCII, eliciting $CD4^+$ T cell responses[14].

In other lymphoid tissues, such as the lymph nodes, subsets equivalent to both splenic $CD8^+$ cDC1 and $CD8^-$ cDC2 can be found, expressing the same surface markers[15]. In addition, lymph nodes continuously receive non-lymphoid resident DCs (migratory) from peripheral tissues, through the afferent lymphatic vessels[16]. These cells are characterized by a higher MHCII and lower CD11c expression on the surface.

Non-lymphoid tissue DC subsets

Langerhans cells (LCs) were identified in the skin long before their immunogenic properties were recognized in 1985[17]. These cells reside in the epidermal layer of mammalian skin, continuously sampling the environment by extending and

retracting long processes (dendrites) between epidermal cells. They are thus able to readily sense external pathogens breaching the protective skin layer, and quickly acquire an activated DC morphology and functions, such as the ability to present antigens on MHCII and activate T cells. In the steady state, LCs are characterized by surface expression of intermediate levels of CD11c, low MHCII, and high levels of Langerin. They also express CD11b and F4/80, but lack CX_3CR1 expression[18].

In the dermis, two major DC subsets can be found: $CD103^+$ $CD11b^{low}$ $Langerin^+$ and $CD103^-$ $CD11b^{high}$ $Langerin^-$ DCs. While it has been reported that the first are able to migrate to skin draining lymph nodes and cross-present antigens, and are thus functional equivalents of cDC1s in lymphoid tissues, the functions of the latter subset are not yet clear[19].

Three subsets of DCs have been identified in the intestine, distinguishable by surface expression of CD103 and CD11b: $CD103^+$ $CD11b^-$, phenotypically and functionally similar to lymphoid resident $CD8\alpha^+$ cDC1s and capable of cross-presentation; $CD103^+$ $CD11b^+$ DCs, which can take up bacteria from the intestinal lumen and transport them to mesenteric lymph nodes[20, 21]; and a $CD103^-$ $CD11b^{high}$ population, which is heterogeneous and dependent on both Flt3L and M-CSFR for development[22]. $CD103^+$ $CD11b^-$ cDC1 and $CD11b^+$ cDC2, and $CD103^-$ $CD11b^{high}$ cDC2, can be also found in other non-lymphoid tissues such as lung, liver and kidney[22].

The two major branches of cDCs can also be distinguished by the mutually exclusive expression of XCR1 (only on cDC1s) and Sirpα (only on cDC2s), in the intestine[23] as well as in other lymphoid and non lymphoid organs[24].

Table 1.1 summarizes the surface phenotype of murine DC subsets.

cDC1s	**cDC2s**	**LCs**	**pDCs**
$Langerin^+$	$CD4^+$ (LT)	$F4/80^+$	$Siglec\text{-}H^+$
$CD8\alpha^+$ (LT)	$CD11b^+$	$CD205^+$	$Bst2^+$
$CD103^+$ (NLT)	$XCR1^-$	$CD103^-$	$Ly6C^+$
$CD205^+$	$Sirp\alpha^+$	$Langerin^+$	$B220^+$
$CD24^+$		$CD11b^+$	$XCR1^-$
$XCR1^+$		$Sirp\alpha^+$	$Sirp\alpha^+$
$Clec9A^+$		$XCR1^-$	$CD14^+$
$Sirp\alpha^-$		CX_3CR1^+	$CD45RA^+$

Table 1.1: Phenotype of dendritic cell subsets. LT, lymphoid tissue; NLT, non-lymphoid tissue. Table adapted from Murphy et al., 2016[25].

1.1.3 Plasmacytoid dendritic cells

All the DC subsets mentioned above make up the so called conventional dendritic cell (cDC) population. In 1989, Facchetti et al. [26] identified a novel type of monocyte with plasmacytoid morphology in different human histopathological samples. In parallel, natural interferon-producing cells were identified in peripheral blood mononuclear cells by Fitzgerald-Bocarsly et al.[27] Later, these rare cells were both characterized to be functionally very similar to DCs, presenting antigens and producing inflammatory cytokines mainly in response to viral stimulation. Further investigations confirmed that they were indeed a peculiar subset of DCs, with a resting morphology resembling plasma cells, but upon activation by viral antigens they would acquire a distinct dendritic phenotype, with a lower capacity to present antigens and activate T cells, and a unique ability to produce vast amounts of Type I interferons (IFNs) and therefore activate antiviral response pathways in bystander cells, as well as orchestrate B cell activation and maturation[28].

Plasmacytoid dendritic cells (pDCs) are rare cells that can be found mainly in murine bone marrow (BM), spleen and lymph nodes, and in low numbers in peripheral blood and non-lymphoid tissues. They show a similar distribution in humans, with a slightly higher percentage in the peripheral blood. They circulate mainly through blood vessels, and can enter lymphoid organs via high endothelial venules.

Morphologically, resting pDCs resemble antibody-producing plasma cells, with oval shape, eccentric nucleus and abundant endoplasmic reticulum (ER). Upon activation, they acquire a canonical dendritic cell morphology, with dendrites protruding from a central body and smaller nucleus and ER.

Phenotypically, murine pDCs can be distinguished from cDCs by their lower expression of CD11c and MHCII, and by the expression of B220, Ly6C, sialic acid binding Ig-like lectin H (Siglec-H) and bone marrow stromal antigen 2 (BST2) on the surface. Other surface markers such as the CC-chemokine receptor 9 (CCR9), Ly49Q and Sca-1 are expressed by the majority of pDCs in peripheral tissues, however they can be used to distinguish functionally and developmentally different subsets in the BM[29].

Functionally, plasmacytoid DCs are the most efficient type I IFN producers;

they express toll-like receptor (TLR)-7 and -9, which are activated by viral nucleic acids, and this leads to secretion of primarily IFN-α and -β, initiating antiviral responses in bystander cells. In addition, they produce interleukin (IL)-6, which, together with the IFNs, contributes to differentiation of B cells into plasma cells[28]. They also produce other pro-inflammatory cytokines, such as tumor necrosis factor (TNF)-α and IL-12, which contribute to their ability to induce T helper cell differentiation[30]. Unlike cDCs, which stabilize peptide-MHCII complexes on their surface for a longer time, pDCs can continuously form new peptide-MHCII complexes and present endogenous and viral antigens following TLR9 activation[31]. Different subsets of pDCs can be distinguished by additional surface markers, that also have important functional differences: CCR9 was discovered to be highly expressed on mature pDCs, and a subset of $CCR9^{low}$ pDC-like cells can be found in the BM and in lymphoid tissues, that is able to respond to TLR stimulation, and produce higher amounts of type I IFN than mature pDCs[32]. A $CD9^{+}$ pDC subset was also identified in the BM that partially overlaps the $CCR9^{low}$ pDC-like cells in their higher type I IFN production following TLR stimulation, while mature $CCR9^{high}$ pDCs are $CD9^{-}$[33]. Sca-1 is differentially expressed on $CCR9^{high}$ pDCs, distinguishing a less mature Sca-1^{low} subset that is more efficient at producing IFNα than Sca-1^{high} pDCs, and that gives rise to the latter following activation[34]. $Ly49Q^{-}$ pDCs are characterized by a lower responsive capacity to RNA viruses than $Ly49Q^{+}$ cells, while they respond equally well to TLR9 stimulation and DNA viruses[35].

Given their central role in organizing immune responses, pDCs are associated with immune tolerance as well as with immunity, both in humans and in mice. For instance, pDCs are pivotal in maintaining tolerance during pregnancy[36] and may directly control Treg proliferation and suppressive activity[37]. They induce Treg mediated tolerance in tumor draining lymph nodes[38] and in experimental autoimmune encephalomyelitis (EAE) models[39]. Alterations in their function are associated with autoimmune diseases, such as psoriasis and systemic lupus erythematosus. In these cases, patients show decreased numbers of pDCs in the circulation and accumulation of IFN producing pDCs in the affected tissues[40].

pDCs produce type I interferon in response to nearly all enveloped viruses and contribute to virus clearance. Their role in antiviral responses was elucidated by

studies on several systemic murine virus infection models[41] and in mucosal viral infections[42]. In addition, specific depletion of pDCs using blood dendritic cell antigen2-diphteria toxin receptor (BDCA2-DTR) transgenic mice led to reduced early IFN-I production, as well as impaired survival and accumulation of $CD8^+$ T cells in mouse cytomegalovirus (MCMV) and vesicular stomatitis virus (VSV) infection[43]. pDCs are infected by human immunodeficiency virus (HIV)-1 and this contributes to the constant immune activation observed in HIV patients, which is associated with faster progression to acquired immune deficiency syndrome (AIDS) and development of comorbidities[44].

1.2 Origin of dendritic cells

With the exception of Langerhans cells, which originate from self-renewing embrionic precursors that have migrated to the skin during early development[45], all DC subsets are generated *in vivo* in the bone marrow from gradually committed progenitors and precursors mainly of the myeloid lineage, although a contribution from lymphoid progenitors to all DC subsets has also been reported[46].

The initiator of the myeloid lineage is the common myeloid progenitor (CMP), characterized by lack of expression of lineage markers (Lin^-) and of Sca-1, high expression of the receptor tyrosine kinase KIT, also known as CD117, and expression of CD34 and CD16/32 (FcγR III and IIb). The CMP gives rise to macrophage-DC progenitors (MDPs), which can generate DCs and monocytes, but lose potential to become neutrophils or other monocytes[47]. The MDPs differ from the CMPs for their lower expression of CD16/32, and expression of the chemokine receptor CX_3CR1 and of the macrophage colony-stimulating factor receptor (M-CSFR), also known as CD115. They also express FMS-like tyrosine kinase 3 (Flt3), also known as CD135, which characterizes all the DC lineage downstream. From the MDPs a committed monocyte progenitor (cMoP) ($CD117^+$ $CD115^+$ $CD135^-$ and $Ly6C^+$) and a common dendritic cell progenitor (CDP) are generated, the latter giving rise to all and only the DC subsets through further sequential differentiation steps[48, 49]. The CDPs are defined as Lin^-, $CD117^{int/low}$, $CD135^+$, $CD115^+$.

cDCs are generated from intermediate precursors that exit the BM, circulate in

the blood and enter peripheral tissue where they complete differentiation. These so-called pre-cDCs were initially identified as a dividing population expressing CD135 and CD11c, showing clonal heterogeneity for commitment to the cDC1 or cDC2 lineage, or both[49]. They arise directly from CDPs in the BM[50], and were shown to migrate from BM *in vivo* and complete maturation in the periphery giving rise to lymphoid CD8α^+ cDC1s or CD11b$^+$ cDC2s, and non-lymphoid CD103$^+$ cDCs[22, 51].

Recent studies using single cell analysis provide evidence for early lineage commitment in myelopoiesis, somewhat contradicting the stepwise branching model, highlighting the intrinsic heterogeneity of the pre-DC population and the existence within it of individual precursors directly committed to either cDC1 (CD8α^+) or cDC2 (CD4$^+$ CD11b$^+$) subsets[52]. These 2 committed pre-cDC subsets lack Siglec-H expression, and can be discriminated by Ly6C surface staining (expressed only in pre-cDC2s).

pDCs on the other hand complete maturation in the bone marrow and then migrate to peripheral tissues. The recently identified CCR9$^-$ pDC-like cells already possess the ability to migrate from the bone marrow to different peripheral tissues, where they complete differentiation into pDCs and cDCs depending on the tissue microenvironment.[32, 53].

Only a minority of the DCs found in peripheral tissues are of lymphoid origin, with a prevalence of pDCs[46]. Nevertheless, pDC output from common lymphoid progenitors (CLPs) can be augmented by a type I IFN-Flt3L signaling axis[54].

1.2.1 Cytokines and growth factors in DC differentiation

The development of pDCs and cDCs is dependent on several cytokines, among which Flt3L has an essential role. Its receptor, Flt3, also known as CD135, is expressed early on in hematopoietic development, on many hematopoietic stem cells (HSCs), on progenitors such as CLPs, CMPs, and on CDPs and MDPs[55]. Indeed, a lineage tracing mouse model, that enabled direct assessment of differentiation pathways *in vivo*, has shown that a Flt3$^+$ stage marks non-self-renewing HSCs that initiate all hematopoietic lineages[56]. CD135 expression is also maintained downstream of DC precursors, and on all DC subpopulations, while it is absent from

other circulating and tissue-resident leucocytes[57]. Production of Flt3L has been observed from endothelial cells, BM stromal cells and activated T cells[58].

Many studies have provided evidence for the pivotal role of Flt3 signaling in DC development: mice lacking Flt3 or its ligand showed defective hematopoiesis, with reduced numbers of HSCs, pDCs and cDCs[59]. On the other hand, in both mice and humans, overexpression or *in vivo* treatment with Flt3L leads to increased numbers of pDCs and cDCs in tissues as well as in the blood stream[60, 61]. In addition to its role in DC differentiation, Flt3L is an important regulator of homeostatic DC division in the periphery *in vivo*[62].

Another important cytokine for DC differentiation is granulocyte-macrophage colony-stimulating factor (GM-CSF), which has different effects at different levels of DC differentiation[63, 64]. Lack of GM-CSF in mice had little impact on the number of DC precursors, but compound deficiency of GM-CSF and Flt3L caused a further reduction of DC precursors, compared to Flt3L deficiency alone, suggesting that GM-CSF is not totally redundant in regulating DC precursors[65]. It is the main cytokine used for cDC generation *in vitro*, as it promotes differentiation of total bone marrow cultures into myeloid subsets, including large numbers of DCs that resemble splenic cDCs[66]. In concert with Flt3L, GM-CSF plays a critical role in the differentiation of both DC subsets, but it generally favors cDC development rather than pDCs, which are instead tightly regulated by Flt3L both *in vivo* and *in vitro*[67].

Csf-1, also known as macrophage colony-stimulating factor (M-CSF), regulates the survival and proliferation of macrophages. Its receptor (CD115) is expressed on MDPs, monocytes and macrophages as well as on CDPs. A model of early progenitor development hypothesizes that the balance of Flt3 versus M-CSF signaling might drive the diversion of MDPs towards CDPs rather than monocyte-macrophages, respectively[68]. CD115 is also expressed on precursor cells with clonogenic potential to both the cDC and pDC fate[48], and its downregulation defines commitment to the pDC lineage[69] Indeed, M-CSF can drive pDC and cDC development *in vitro* from BM precursors independently of Flt3L. Administrated *in vivo*, it is able to increase DC numbers in mice[70].

The observation of a different immune regulation in men and women has prompted

investigation of the role of hormones in immune cells development and function, and especially the role of estrogens in DC biology[71]. Estrogen receptor (ER)α is expressed in murine BM progenitor cells, including CLPs and myeloid progenitors, as well as on most mature immune cells[72]. Estrogen-mediated ERα activation is necessary for DC development *in vitro* in GM-CSF cultures, and its absence results in reduced total numbers of DCs, mainly pDCs, in Flt3L-driven *in vitro* differentiation[72, 73].

1.2.2 Transcriptional regulation of DC differentiation

Figure1.1 [25] details the main transcription factors (TFs) that regulate DC development within the myeloid lineage. At earlier levels, the divergence between lymphoid and myeloid lineages relies mainly on the protein levels of the TF spleen focus forming virus (SFFV) proviral integration oncogene (PU.1), which are strictly connected with the rate of cell cycling: fast dividing progenitors fail to accumulate PU.1, and this leads towards the lymphoid (B cell) lineage. On the other hand, a slower cell cycle leads to increased PU.1 protein, which drives myeloid lineage differentiation[74].

The development of $Flt3^+$ DC progenitors is dependent on runt related transcription factor (Runx)1 activity and especially on its cofactor core binding factor (Cbf)β: its deletion in hematopoietic lineages causes loss of DC progenitors and erythroid progenitors, with increased granulocyte-macrophage progenitors (GMPs) and a myeloproliferative disorder[75].

Interferon regulatory factor (Irf)8 is one of the major actors in the DC lineage: its expression is initiated by PU.1 as early as in the CMPs, and it is necessary to exclude granulocyte potential. Further downstream, it not only regulates MDP-CDP transition and monocyte differentiation, but also controls survival of $CD8^+$ cDCs and function of pDCs[76]. It is necessary for specification of the pre-cDC1 clonogenic progenitor, after which its autoactivation becomes dependent on the basic leucine zipper transcription factor, ATF-like (Batf)3, thus allowing completion of cDC1 development[50]. Irf8 downregulation is nevertheless necessary for terminal differentiation of cDC2, which become dependent on Irf4.

The TF zinc finger and BTB domain containing (Zbtb)46, expressed in all

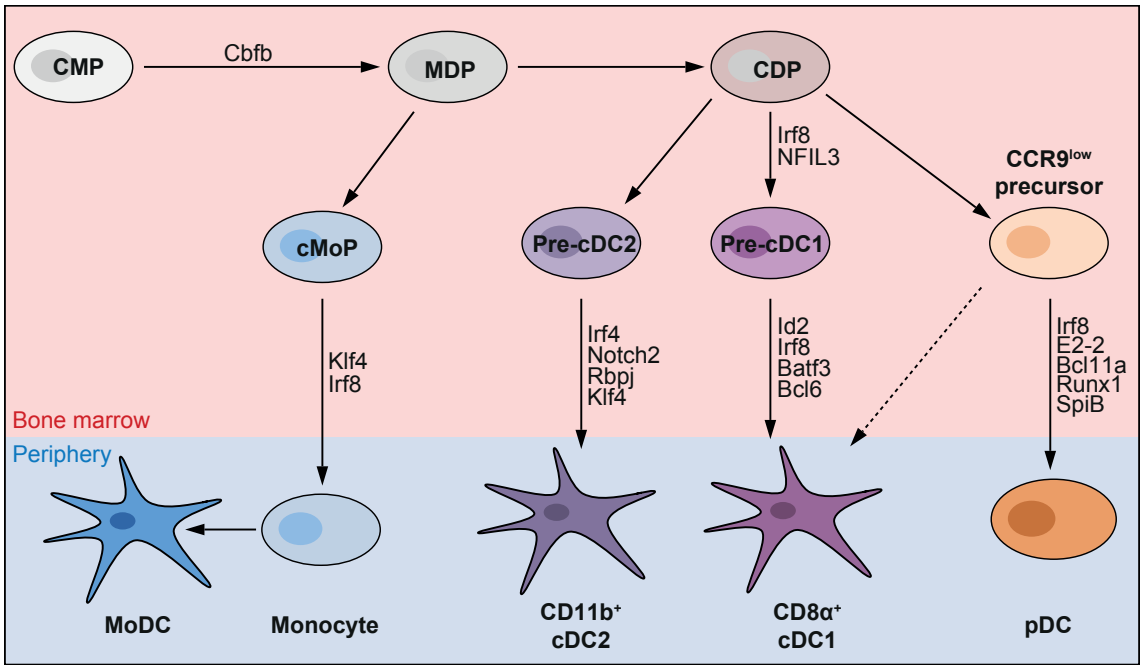


Figure 1.1: Stages and transcription factors of DC development. A scheme showing myeloid lineage development from the CMP, indicating transcription factors required for particular transitions between stages. This scheme shows DC lineage divergence from the CDP. Commitment to cDC1 and cDC2 branches of cDCs can occur in the bone marrow. Adapted from Murphy et al., 2016[25]

pre-cDCs, is a marker of cDC lineage commitment, as highlighted by Zbtb46-GFP reporter mouse studies. Nevertheless, its expression is not required for cDC differentiation, and its ablation causes only minor alterations[77, 78].

$CD8\alpha^+$ DCs require expression of nuclear factor, interleukin 3, regulated (Nfil3), a transcriptional repressor which controls expression of Batf3 and inhibitor of DNA binding (Id)2[79]. The latter forms inactive heterodimers with target E proteins, preventing their binding to the DNA[80]. One of such targets is transcription factor (Tcf)4, also known as E2-2, an essential transcription factor for the pDC lineage[81].

E2-2 controls transcription of B cell CLL/lymphoma (zinc finger protein) (Bcl) 11A and Irf8, supporting pDC differentiation, and pDC-specific genes such as TLR7, TLR9 and BDCA2 (in human). E2-2 is required for maintaining the cell fate in mature pDCs[82]. The competition of E2-2 with Id2 takes part in the pDC/cDC lineage divergence as early as at the CDP stage, although the complete mechanism is not fully understood at present. For example, this competition does not explain divergence of pDCs from cDC2s, as these cells do not require Id2 for development, even though they express it at later stages of maturation.

Recently, 2 other factors have been discovered supporting pDC differentiation, by actively suppressing Id2 expression: the ETO family transcriptional cofactor Mtg16[83], and the zinc finger E-box binding homeobox (Zeb)2 TF[84, 85]

1.2.3 pDC-specific transcription factor network

Although many TFs have been identified that specifically regulate pDC differentiation distinguishing it from the cDC lineage, little is known about the fine tuning of cell fate decisions in the intermediate steps that take a CDP to the mature plasmacytoid DC.

While the identification of E2-2 expressing CCR9^{-} pDC-like cells has added a step to this process, suggesting heterogeneity of cell fate potential until later steps of development, little is known about the factors that give identity to each and every cell, and whether this "conversion" potential observed is due to cell-intrinsic plasticity, or cell to cell variations that imprint small subpopulations with different lineage potentials.

Recent work from our lab defined a series of sequential steps in the *in vitro* differentiation of pDCs from CDPs by means of continuous single cell imaging [86], but this analysis is limited to surface markers and could not clarify the involvement of TFs in this process.

In addition, very little is known about the effects that inflammation and infection have on these processes, and whether they might actively shape the differentiation potential of the committed DC precursors. Several reports have shown expression of TLRs on many hematopoietic progenitors, and that their direct activation influences myelopoiesis (reviewed in Yáñez et al., 2013[87]). HSCs, as well as lineage-restricted progenitors such as CLPs, CMPs and GMPs, express TLR4 and TLR2, and *in vitro* stimulation with their ligands induces cell cycling and acquisition of myeloid markers[88]. CLPs also express TLR9, and its signaling primes these cells to become DCs *in vitro* and *in vivo*[89]. Some TLRs are also expressed at the CDP stage, and their stimulation leads to mobilization from the BM and migration towards draining lymph nodes, where they give rise to mature DCs[90].

2 Aims of the study

This project aimed at defining the transcriptional regulation of pDC differentiation in the steady state, as well as the alterations that systemic TLR stimulation induce on this pathway.

Recent research identified a series of step-wise phenotypic changes in Flt3L-dependent DC-lineage bone marrow precursors that mark the stages of differentiation of pDCs. These different stages can be discriminated by surface markers expression, and define a discrete number of cell types that are progressively committed to a mature pDC fate. In addition, single cell RNA sequencing has highlighted the heterogeneity of precursor populations, that is undetectable at the surface level.

The first aim of this project was therefore to define a strict and clear discrimination of DC precursors in order to isolate and analyze the different stages of DC subtype development. Moreover, these isolated populations could be challenged with TLR ligands, to assess their responsiveness to TLR stimulation and the resulting changes in cell type specific gene expression signatures.

The second aim was to define the differentially expressed genes in the different stages of steady state differentiation, and identify the network of factors that regulates this transition. This could lead to the identification of specific targets that have a pivotal role in cell fate determination. To do so, total mRNA sequencing was performed on the populations freshly isolated from murine bone marrow.

The third aim was to assess the role of TLR stimulation in shaping the differentiation process. I wanted to evaluate whether different stages of precursors are able to directly respond to TLR ligands, and whether this response influences the cell fate decisions. In addition, I wanted to identify the networks of TLR response, comparing various differentiation stages.

3 Materials and Methods

3.1 Materials

3.1.1 Reagents

Reagents

Reagent	Provider
β-estradiol	Sigma-Aldrich (Seelze, Germany)
β-mercaptoethanol	Sigma-Aldrich (Seelze, Germany)
Biocoll	Merck (Darmstadt, Germany)
CpG-A (ODN 2216)	Eurofins Genomics (Ebersberg, Germany)
CpG-B (ODN 1826)	Eurofins Genomics (Ebersberg, Germany)
CpG-C (ODN 2395)	Eurofins Genomics (Ebersberg, Germany)
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich (Seelze, Germany)
dNTP mix (each 10mM)	Promega (Mannheim, Germany)
Dulbecco's Modified Eagle's Medium (DMEM)	Invitrogen (Karlsruhe, Germany)
EDTA (0.5M, pH 8.0)	Invitrogen (Karlsruhe, Germany)
Fetal calf serum (FCS)	PAA (Pasching, Austria)
Glutamax™-I (100x)	Invitrogen (Karlsruhe, Germany)
Hexadimethrine bromide (Polybrene®)	Sigma-Aldrich (Seelze, Germany)
Lipofectamine® 2000	Invitrogen (Karlsruhe, Germany)
Non-essential aminoacids (NEAA)	Invitrogen (Karlsruhe, Germany)
Penicillin/Streptomicin (100x)	Invitrogen (Karlsruhe, Germany)
Phosphate buffered saline (PBS)	Invitrogen (Karlsruhe, Germany)
Phire Green Hot Start II PCR Master Mix	Thermo Scientific (Karlsruhe, Germany)

Table of reagents (continued)

Reagent	Provider
R848 (Resiquimod)	Sigma-Aldrich (Seelze, Germany)
Red blood cell lysis buffer	Sigma-Aldrich (Seelze, Germany)
RNAse Inhibitor (NxGen® RI, 40U/ml)	Lucigen (Middleton, WI, USA)
RPMI 1640	Biochrom (Berlin, Germany)
Sodium pyruvate (100x)	Invitrogen (Karlsruhe, Germany)
SuperScript™ III Reverse Transcriptase	Invitrogen (Karlsruhe, Germany)

3.1.2 Enzymes and recombinant cytokines

Recombinant proteins

Reagent	Provider
Collagenase D	Sigma-Aldrich (Seelze, Germany)
DNAse I	Sigma-Aldrich (Seelze, Germany)
Flt3L	Produced in house (as supernatant)
GM-CSF	Produced in house (as supernatant)
murine IL-6	PeproTech (Hamburg, Germany)
murine IL-3	PeproTech (Hamburg, Germany)
murine SCF	PeproTech (Hamburg, Germany)

3.1.3 Antibodies

Antibodies for flow cytometry

Antigen	Clone	Conjugate	Manufacturer
B220	RA3-6B2	BrilliantViolet 605™	BioLegend
CCR9	CW-1.2	eFluor™ 450	eBioscience
CD3	145-2C11	FITC	BioLegend
		APC-eFluor™ 780	eBioscience
CD8α	53-6.7	PE	BD Pharmigen
CD11b	M1/70	PerCP-Cy5.5	eBioscience

Antibodies for flow cytometry (continued)

Antigen	Clone	Conjugate	Manufacturer
CD11c	N418	PE-Cy7	eBioscience
CD19	1D3	FITC	BD Pharmigen
		APC-eFluor™ 780	eBioscience
CD86	GL1	PE	eBioscience
		BrilliantViolet 650™	BioLegend
CD90.1 (Thy1.1)	OX-7	AlexaFluor® 700	BioLegend
CD135 (Flt3)	A2F10	PE	eBioscience
Foxp1	Rabbit polyclonal	purified	Cell Signaling
Ly6G	1A8	FITC	BioLegend
		APC-Cy7	
MHCII (I-A/I-E)	M5/114.15.2	BrilliantViolet 650™	BioLegend
		APC-eFluor™ 780	eBioscience
NK1.1	PK136	FITC	BioLegend
		APC-eFluor™ 780	eBioscience
Siglec-H	440c	AlexaFluor® 647	Produced in house
Sirpα (CD172a)	P84	PerCP-Cy5.5	BioLegend
Goat α-rabbit	Poly4064	DyLight™ 649	BioLegend

Table 3.1: Table of Flow Cytometry Antibodies

3.1.4 Kits

MACS® cell isolation kits	Miltenyi Biotech (Bergisch Gladbach, Germany)
RNeasy Plus Mini kit	QIAGEN (Hilden, Germany)
Quick-RNA™ MicroPrep	Zymo Research (Freiburg, Germany)

3.1.5 Taqman™ probes for qPCR

Gene	Assay number	Transcripts detected[1]
Foxo1	Mm00490671_m1	NM_019739.3
Foxo4	Mm00840140_g1	NM_018789.2
Foxp1	Mm00474848_m1	NM_001197321.1, NM_001197322.1, NM_053202.2
Foxr1	Mm02600883_m1	NM_001033469.2
Foxr1	Mm02600884_g1	NM_001033469.2
Hprt (housekeeping)	Mm03024075_m1	NM_013556.2

3.1.6 Media and buffers

Cell culture media

DC medium	RPMI 1640
	10% FCS
	1% NEAA
	1% Glutamax-I
	1% Sodium Pyruvate
	1% Pen/Strep
	50µM β-mercaptoethanol
DMEM complete medium for HEK293T cell culture and virus production	DMEM
	10% FCS
	1% NEAA
	1% Glutamax
	1% Sodium Pyruvate
	1% Pen/Strep
	50µM β-mercaptoethanol

[1] RefSeq transcript identifiers. Multiple values indicate isoforms.

Growth medium	DC medium
for Hoxb8 cell line generation	10ng/ml IL-3
	20ng/ml IL-6
	250ng/ml SCF
Progenitor outgrowth medium (POM)	DC medium
for Hoxb8 stem cell line maintenance	1µM β-estradiol
	7% Flt3L containing supernatant
Freezing medium	90% FCS
	10% DMSO

Buffers

MACS buffer	PBS (w/o Ca^{2+} or Mg^{2+})
	2% FCS
	2mM EDTA
Sorting buffer (for RNA extraction)	PBS (w/o Ca^{2+} or Mg^{2+})
	1% RNAse Inhibitor

3.1.7 Mice

All mice were bred under SPF conditions in our animal house. Mice were used at age 6 to 15 weeks.

Strain	Original source	Application
C57BL/6J	Harlan, Paderborn	BM isolation for sorting *In vitro* culture experiments
Id2$^{eGFP/eGFP}$	Gabrielle T. Belz, The Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia[91]	*In vitro* culture experiments *In vivo* CpG challenge

Strain	Original source	Application
Foxp1$^{flox/flox}$	Jurgen Ruland, TUM, Munich, Germany	*In vitro* culture experiments Breeding
CD11c-Cre	Boris Reizis, NYU Langone Medical Center, New York, USA	Breeding with Foxp1$^{flox/flox}$

3.2 Methods

3.2.1 Cell culture

All cell cultures were maintained at 37°C with 5% CO_2 in humidified incubator with the appropriate media. FCS was heat inactivated at 56°C for 45 minutes prior to use in media or buffers.

3.2.2 Cell isolation from primary tissues

For cell isolation, mice were sacrificed by cervical dislocation. Bone marrow cells were isolated from the hind legs and hip bones; under a sterile hood, bone extremities were cut and the BM flushed out with DC medium, using a 24G needle and a 10 ml syringe. After flushing, the suspension was passed through the needle 2 or 3 times to disrupt clumps. The cell suspension was then passed through a 100 µm cell strainer and centrifuged for 5 minutes at 450 x *g*. The pellet was resuspended in 1 ml of red blood cell (RBC) lysis buffer, incubated for 5 minutes at room temperature and then washed with 20 ml of DC medium. After again centrifuging 5 minutes at 450 x *g*, cells were resuspended in MACS buffer and counted using Türk's solution to exclude left over RBCs and dead cells.

Splenocytes were isolated from freshly harvested spleens, that were injected with 5 ml of DC medium containing DNAse I (100 µg/ml) and Collagenase D (500 µg/ml) using a syringe with a 24G needle, then cut in small pieces and incubated for

30 minutes at 37°C. At the end of the incubation, the digested spleens were collected and forced through a 100 µm cell strainer using a syringe plunger, then washed with 10 ml of DC medium. After centrifuging for 5 minutes at 450 × *g*, red blood cell lysis and subsequent steps were performed as described above.

3.2.3 Lineage depletion

For sorting or *in vitro* stimulation experiments, BM cells were depleted of Lineage positive cells using MACS microbeads following the manufacturer's instructions. Briefly, the cell suspension was stained with FITC-conjugated lineage antibodies (CD3, CD19, NK1.1 and Ly6G) at a 1:200 dilution each, incubated for 15 minutes on ice, then washed with 10 ml cold MACS buffer. After centrifugation (5 minutes at 450 × *g*), cells were resuspended in MACS buffer containing 1:10 diluted anti-FITC microbeads and incubated again 15 minutes on ice. They were again washed and centrifuged, and then resuspended in 1 ml MACS buffer and loaded on a LS magnetic column supplied with 30 µm pre-filter, and passed through the column by gravity. The column was then washed 3 times with 3 ml MACS buffer, and the flow through collected, centrifuged and counted.

3.2.4 *In vitro* TLR stimulation

After lineage depletion cells were plated in 96 well plates at a concentration of 2×10^5 cells per well in 100 µl of DC medium. 100 µl of DC medium containing 2× concentration of the appropriate stimulus was added to the respective wells. Cells were incubated at 37°C with 5% CO_2 in humidified incubator for up to 24 hours.

At each time point (0, 2, 4, 6, 12 and 24 hours), the cells were transferred to FACS tubes, washed once with PBS and stained with the antibodies indicated in table 3.2.

3.2.5 Flow cytometry and sorting

Surface staining

Single cell suspensions were stained with surface antibodies by incubating 20 minutes at 4°C in buffer containing the appropriate dilution of each antibody; the

Antigen	Clone	Conjugate	Dilution
CD3	145-2C11	FITC	1:200
CD19	1D3	FITC	1:200
NK1.1	PK136	FITC	1:200
Ly6G	1A8	FITC	1:200
CD86	GL1	PE	1:200
CD11c	N418	PE-Cy7	1:200
Sirpα	P84	PerCP-Cy5.5	1:100
Siglec-H	440c	AlexaFluor647	1:200
MHCII	M5/114.15.2	APC-eFluor780	1:200
B220	RA3-6B2	BV 605	1:200
CCR9	CW-1.2	eFluor450	1:200

Table 3.2: Antibodies used to stain *in vitro* stimulated BM

buffer consists of a supernatant from the ATCC® HB-19™ hybridoma producing the 2.4G2 antibody, a CD16/32 (FCγR III/II) blocking antibody, to prevent unspecific staining. After incubation, cells were washed and resuspended in appropriate volume of either MACS buffer (for Flow Cytometry) or DC medium (for sorting).

Intracellular staining

For intracellular staining of Foxp1 I used the eBioscience™Foxp3 / Transcription Factor Staining Buffer Set according to manufacturer's instructions. Briefly, after staining of surface antigens (table 3.3), cells were washed and resuspended in Fixation/Permeabilization working solution, mixed well, and incubated for 30 minutes at 4°C in the dark. They were then washed with Permeabilization buffer, centrifuged and resuspended in 100 µl of 1X Permeabilization buffer containing the appropriate dilution (1:50) of rabbit anti-Foxp1 primary antibody, and incubated for 30 minutes at 4°C. Cells were then washed with Permeabilization buffer, centrifuged and resuspended as before in buffer containing 1:100 of secondary antibody (DyLight™ 649 conjugated Goat α-rabbit). Finally, cells were washed and centrifuged once with Permeabilization buffer and once with MACS buffer. They were resuspended in MACS buffer for FACS analysis.

Sorting

Cells were sorted on a FACS Aria III (Becton Dickinson) in the Flow Cytometry unit (Dr. Matthias Schiemann) of the Institute for Medical Microbiology, Immunol-

Bone Marrow			
Antigen	Clone	Conjugate	Dilution
CD3	145-2C11	APC-eFluor780	1:200
CD19	1D3	APC-eFluor780	1:200
NK1.1	PK136	APC-eFluor780	1:200
Ly6G	1A8	APC-Cy7	1:200
CD135	A2F10	PE	1:100
CD11c	N418	PE-Cy7	1:200
Sirpα	P84	PerCP-Cy5.5	1:100
Siglec-H	440c	AlexaFluor488	1:200
MHCII	M5/114.15.2	BV 650	1:200
B220	RA3-6B2	BV 605	1:200
CCR9	CW-1.2	eFluor450	1:200
Spleen			
Antigen	Clone	Conjugate	Dilution
CD3	145-2C11	APC-eFluor780	1:200
CD8α	53-6.7	PE	1:200
CD11c	N418	PE-Cy7	1:200
CD11b	M1/70	PerCP-Cy5.5	1:200
Siglec-H	440c	AlexaFluor488	1:200
MHCII	M5/114.15.2	BV 650	1:200
B220	RA3-6B2	BV 605	1:200
CCR9	CW-1.2	eFluor450	1:200

Table 3.3: Antibodies used for surface staining of BM and spleen.

ogy and Hygene of the Technical University Munich (TUM). The staining panel is indicated in table 3.4; sorting strategy is described in figure 4.1. Briefly, Lin$^-$ CD135$^+$ CD11c$^+$ cells were discriminated by expression of Siglec-H, B220, CCR9, MHCII and Sirpα into pre-DCs (Siglec-H$^-$ MHCII$^-$ Sirpα$^-$), CCR9^{low} cells (Siglec-H$^+$ CCR9^{low} B220$^{low/int}$) and pDCs (Siglec-H$^+$ CCR9^{hi} B220^{hi}). Cells were sorted to high purity and deposited in DC medium. The purity of each sample was verified at the end of each run.

FACS analysis

FACS analysis was performed with a Gallios (3 lasers, 10 colors) or a Cytoflex S (4 lasers, 13 colors) flow cytometers (Beckman Coulter, Krefeld, Germany). Data were analysed with FlowJo® Single Cell Analysis Software v10 (FlowJo LLC, Ashland, USA).

Antigen	Clone	Conjugate	Dilution
CD3	145-2C11	FITC	1:200
CD19	1D3	FITC	1:200
NK1.1	PK136	FITC	1:200
Ly6G	1A8	FITC	1:200
CD135	A2F10	PE	1:100
CD11c	N418	PE-Cy7	1:200
Sirpα	P84	PerCP-Cy5.5	1:100
Siglec-H	440c	AlexaFluor647	1:200
MHCII	M5/114.15.2	APC-eFluor780	1:200
B220	RA3-6B2	BV 605	1:200
CCR9	CW-1.2	eFluor450	1:200

Table 3.4: Antibodies used to stain murine BM for sorting

3.2.6 Stimulation of sorted cell populations

After sorting pre-DCs, CCR9^{low} precursors and pDCs into DC medium, volume was adjusted to 1 ml, and each population was divided into 10 equal aliquots (100µl) in microcentrifuge tubes. One was left untreated, the rest were stimulated with CpG-A or -C (0.5 µM) or R848 (3 µM) (3 each) by adding 100 µl of DC medium containing 2x concentration of the respective ligand. Samples were incubated at 37°C with 5% CO_2 in humidified incubator for up to 6 hours.

At each time point (0 hours for untreated control, 2, 4 or 6 hours for stimulated samples), tubes were directly centrifuged at 800 × g for 5 minutes and the supernatant removed, and RNA was isolated using the QIAGEN RNeasy Mini kit according to the manufacturer's instructions.

3.2.7 mRNA sequencing

mRNA sequencing libraries were prepared according to the Smart-seq2 protocol[92, 93] by Christoph Ziegenhein (Prof. Dr. Wolfgang Enard, Anthropology and Human Genomics, LMU Munich, Planegg-Martinsried).

Briefly, purified RNA (maximum 9ng) was reversed transcribed using SuperScript II (Invitrogen) with template switching oligonucleotides (TSOs), and the cDNA pre-amplified with KAPA HiFi HotStart DNA polymerase (KAPA Biosystems). The PCR was the purified with Agencourt Ampure XP beads (Beckman Coulter), quantified and quality checked on a 2100 Bioanalyzer Instrument with

Agilent high-sensitivity DNA kit (Agilent Technologies). Tagmentation (maximum 1ng cDNA) was performed with the Nextera XT DNA sample preparation kit (Illumina), the final product again purified with Ampure XP beads and quality verified with a Bioanalyzer, high-sensitivity DNA kit.

Sequencing was performed on 4 lanes of an Illumina High output flow cell, for a total of approximately 8×10^8 single end reads (50 bases long), corresponding to approximately 9×10^6 reads per sample. After pre-processing and quality control (figure 3.1), the reads were aligned with the UCSC mouse genome build GRCm38 (mm10), duplicates removed and the total raw counts normalized to transcripts per million (TPM). Depending on the analysis software, either the TPM or the total counts were used as input for the analyses.

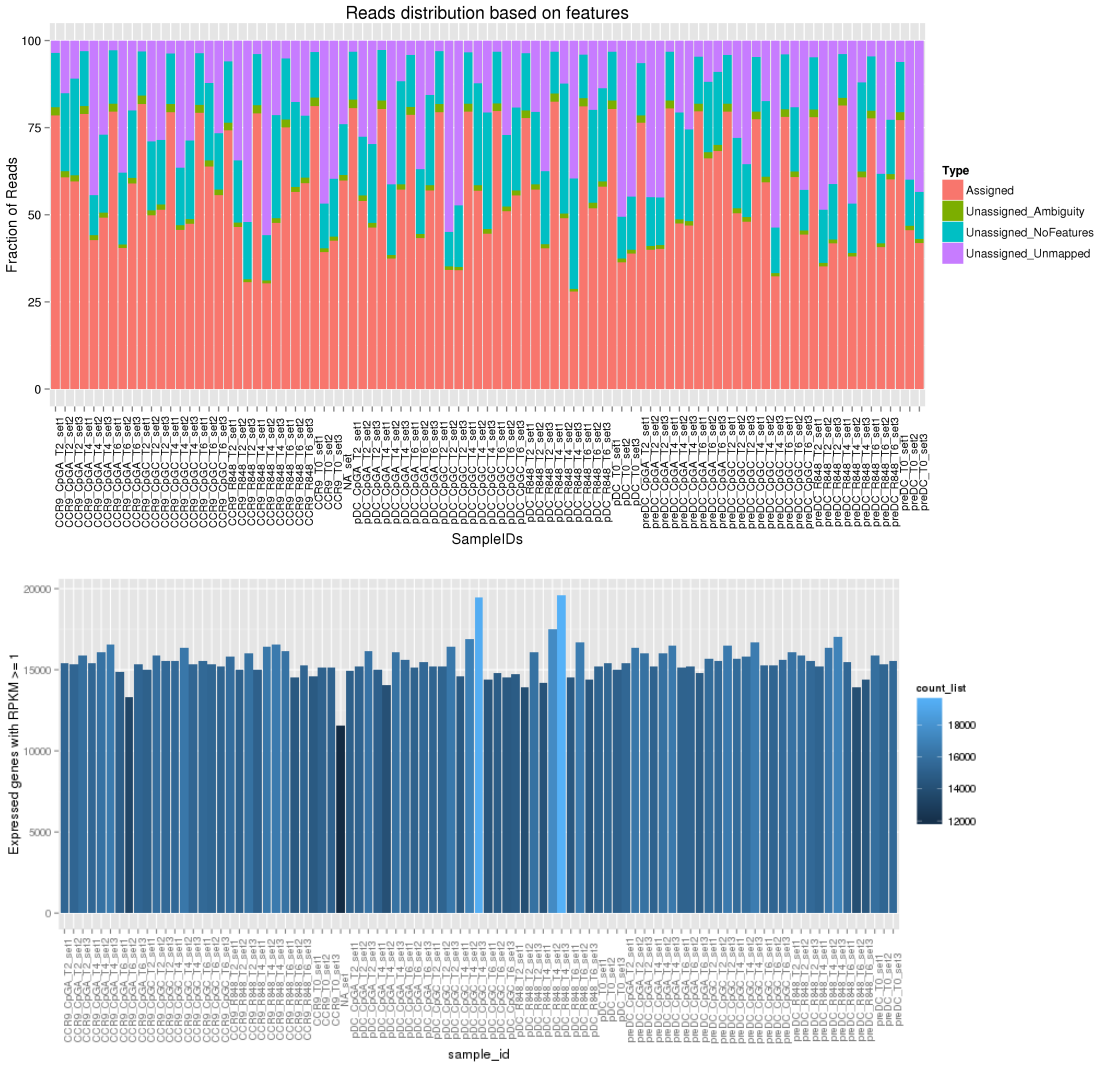


Figure 3.1: Quality control of the sequencing results for all samples. Top: proportion of reads correctly assigned to genes. Bottom: number of genes detected.

3.2.8 Reverse transcription-quantitative PCR

For target validation on unstimulated cells, sorting was performed again as described, adding a fourth population of Lin^+ $CD135^-$ cells as control, and cells deposited in PBS containing 1% RNAse Inhibitor (Sorting buffer).

In addition, four cell populations were sorted from the spleen (stained with the panel in table 3.5): after gating out Lineage positive cells, and selecting all $CD11c^+$ cells, pDCs were identified as Siglec-H^+ $B220^{high}$ $CCR9^{high}$. Within the Siglec-H^- population, cDC1s and cDC2s were discriminated as $CD8\alpha^+$ $CD11b^-$ and $CD8\alpha^-$ $CD11b^+$, respectively. Lineage positive cells were sorted as positive control (mostly T cells).

Antigen	Clone	Conjugate	Dilution
CD3	145-2C11	FITC	1:200
CD19	1D3	FITC	1:200
NK1.1	PK136	FITC	1:200
Ly6G	1A8	FITC	1:200
CD8α	53-6.7	PE	1:200
CD11c	N418	PE-Cy7	1:200
CD11b	M1/70	PerCP-Cy5.5	1:200
Siglec-H	440c	AlexaFluor647	1:200
MHCII	M5/114.15.2	APC-eFluor780	1:200
B220	RA3-6B2	BV 605	1:200
CCR9	CW-1.2	eFluor450	1:200

Table 3.5: Antibodies used to stain murine spleen for sorting

After checking the purity (> 95%), sorted cells were centrifuged and the supernatant discarded, and RNA extracted using Quick-RNA™ MicroPrep (Zymo Research).

RNA was quantified using a SimpliNano™ spectrophotometer (Biochrom, Harvard Bioscience, USA). Complementary DNA (cDNA) was produced with the SuperScript™ III reverse transcriptase (Invitrogen), according to the manufacturer's instructions. The equivalent of 20ng of RNA were used for each quantitative PCR (qPCR) reaction, which was performed with commercially available Taqman™ probes (section 3.1.5) according to the manufacturer's protocol, on a LightCycler® 480 Instrument II (Roche).

Data were analyzed using the $2^{-\Delta Ct}$ method. This method is equal to the better

known $2^{-\Delta\Delta Ct}$ method, with the exception that it does not normalize expression on one of the samples, allowing comparison of expression levels in the absence of a specific control sample to use as normalizer. Briefly, for each sample Ct values of the housekeeping gene (Hprt) are subtracted from the Ct value of each target gene, thus normalizing the latter for the sample's baseline gene expression. The ΔCt value obtained is then transformed to a negative power of two, obtaining values directly correlated to gene abundance relative to the housekeeping gene.

3.2.9 Id2$^{eGFP/eGFP}$ mice and *in vivo* experiments

In vivo TLR stimulation was performed on female Id2$^{eGFP/eGFP}$ mice of 8 weeks of age, which were injected subcutaneously in the right flank with 50 µg CpG-A diluted in 200 µl of sterile PBS. Controls received PBS only. Mice were sacrificed at 16 or 72 hours post injection by cervical dislocation, and BM and spleen collected for FACS analysis (staining panel in table 3.6).

Antigen	Clone	Conjugate	Dilution
CD3	145-2C11	APC-eFluor780	1:200
CD19	1D3	APC-eFluor780	1:200
NK1.1	PK136	APC-eFluor780	1:200
Ly6G	1A8	APC-Cy7	1:200
CD135	A2F10	PE	1:100
CD11c	N418	PE-Cy7	1:200
Sirpα	P84	PerCP-Cy5.5	1:100
Siglec-H	440c	AlexaFluor647	1:200
MHCII	M5/114.15.2	BV 650	1:200
B220	RA3-6B2	BV 605	1:200
CCR9	CW-1.2	eFluor450	1:200

Table 3.6: Antibodies used for surface staining of Id2$^{eGFP/eGFP}$ BM and spleen.

In vitro BM stimulation

BM was isolated from 2 female Id2-$^{eGFP/eGFP}$ mice of 8 weeks of age, as described above. Cells were plated in 6 well plates, 1.5×10^6 cells/ml in 3 ml of DC medium, and stimulated with 0.5 µM CpG-A or -C, or 3 µM R848, or left untreated, and incubated for 4 or 16 hours at 37°C with 5% CO_2 in a humidified incubator. At the end of the incubation period the cells were collected by gentle pipetting, washed once with PBS and stained as described above.

3.2.10 Hoxb8 progenitor cell lines

Hoxb8 progenitor cell lines were generated as described in Redecke et al., 2013[94], from BM isolated from wt mice and from Foxp1$^{fl/fl}$ mice.

Virus production

The plasmid MSCV–ERHBD–HOXB8 (kindly provided by Dr. Hans Häcker, St. Jude Children's research Hospital, Memphis, USA) was co–transfected together with the ecotropic packaging vector pCL–Eco (Addgene) into HEK293T cells using Lipofectamine 2000 (Invitrogen). 18 hours after transfection, the supernatant was replaced by fresh DMEM complete medium. After 24 hours virus–containing supernatant was collected and stored at 4°C, and fresh medium was added to the cells. After another 24 hours the supernatant was collected, pooled with the previous collection, filtered (0.45 µm), aliquoted and stored at -80°C.

The same method was used for the plasmids pSuper-Cre-Thy1.1 (Cre-RV) and pSuper-Thy1.1 (Mock-RV), both kindly provided by Prof. Vigo Heissmeyer, Institute for Immunology, LMU Munich, Germany.

Generation of the progenitor cell lines

Freshly isolated BM was washed once with DC medium, resuspended in 4 ml of DC medium and loaded on 3 ml of Biocoll (Merck), then separated by centrifugation for 30 minutes at 450 x *g*. The entire supernatant was collected and diluted with 45 ml PBS containing 1% FCS (final volume 50 ml), pelleted at 800 x *g* for 10 minutes, then resuspended in 10 ml DC medium, centrifuged at 450 x *g* for 5 minutes and finally resuspended at a concentration of 5×10^5 cells/ml in growth medium. After two days of cell culture, cells were collected and resuspended in progenitor outgrowth medium (POM). 2×10^5 cells were dispensed in 1 ml per well in a 12–well plate and infected with MSCV vectors (diluted 1:2 with POM) by spin inoculation at 1500 x *g* for 60 minutes in the presence of Lipofectamine (0.1%). After infection, cells were diluted by adding 1.5 ml POM for 24 hours, followed by removal and replacement of 2 ml of the cell culture medium. During the following cell culture period, cells were dispensed every 3–4 days in fresh POM and transferred into new wells, until

stably expanding.

Retroviral transduction of the Hoxb8 cell lines

The Foxp1$^{fl/fl}$-Hoxb8 cell line was used for transfection with the Cre-RV, to generate a Foxp1$^{-/-}$ line, and with the Mock-RV as control.

2.5×10^5 were seeded in each well of a 12-well plate, and infected with 1 ml of RV-containing supernatant supplemented with 5 µM β-estradiol and 7% Flt3L-containing supernatant, by spin inoculation at 1500 x g for 60 minutes in the presence of 8 µg/ml polybrene (Sigma), followed by incubation at 37°C, 5% CO_2. 2 hours after spinoculation, supernatant was removed and replaced with fresh POM, and cells were cultured normally in POM.

Cells were assessed for transduction efficiency on day 7 post infection, by surface staining for CD90.1 (Thy1.1, reporter). Positive and negative cells (that have not integrated the transcript, to use as controls) were FAC-sorted to over 99% purity (figure 3.2).

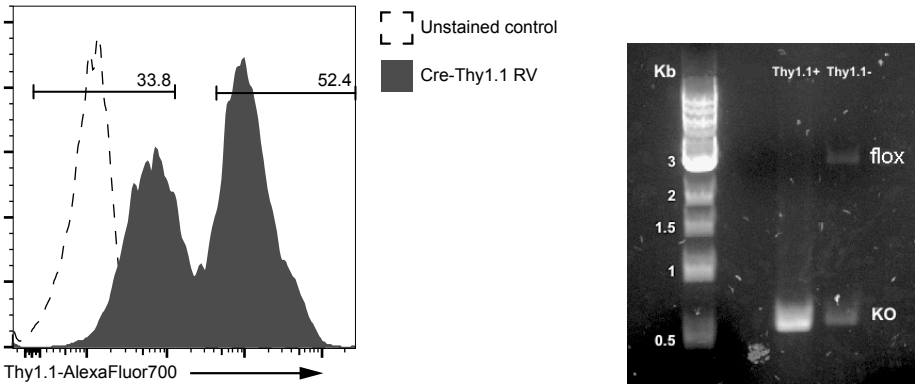


Figure 3.2: Sorting of Cre-transduced Foxp1$^{flox/flox}$-Hoxb8 cell line and verification PCR. At day 7 after retroviral transduction cells were stained for Thy1.1 (reporter gene) and a positive and negative fraction were sorted to high purity (left panel). Cre-mediated recombination was verified by PCR in both populations (right panel).

Effective recombination of the Foxp1 gene was verified by PCR (figure 3.2). The negative fraction showed heterozygous genotype, indicating recombinatory events happening without integration of the retrovirus, probably due to transient expression after infection. This fraction was therefore excluded, and the Mock-RV transduced line (also sorted for Thy1.1 positive cells) was used as control. This RV showed less transduction efficiency (12% positive cells at day 7), but the sorted cells showed

no difference in morphology or growth compared to the Cre-RV transduced or the untransduced parent line.

PCR protocol for the verification of Foxp1 recombination

Reaction mix:

H_2O	6µl
2× Phire PCR Master Mix	10µl
Forward primer	1µl
Reverse primer	1µl
DNA template	2µl
total	20µl

Primers:

Forward primer:	CTG CAC AGC AGG GTA GTT AGT G
Reverse primer:	ATG CTA GGC GGT ACT AAA TAG AAC

Protocol:

1	30” 98°C
2	5” 98°C
3	5” 65°C
4	15” 72°C
5	back to 2 (35 cycles)
6	1’ 72°C

hold 8°C

Amplicon sizes:

flox:	3135 bp
KO:	623 bp

3.2.11 Hoxb8 lines differentiation assay

For assessing differentiation, 3 Foxp1$^{flox/flox}$-Hoxb8 cell lines (the parental line without RV infection, Mock-RV transduced and Cre-RV transduced) were washed twice in DC medium to remove all traces of β-estradiol, and seeded in 24 well plates,

10^4 cells/well in 250 µl of DC medium, containing either: (A) 7% Flt3L containing supernatant; (B) 15% Flt3L supernatant (high dose); (C) 7% Flt3L and 50 ng/ml M-CSF; (D) 7% FLt3L and 1% GM-CSF containing supernatant. On day 3, 6 and 8, the volume was doubled with DC medium containing double amounts of the respective cytokines. On day 1, 2, 4, 6, 8 and 10 cells were collected and analyzed by flow cytometry (staining panel in table 3.7).

Antigen	Clone	Conjugate	Dilution
Siglec-H	440c	AlexaFluor488	1:200
CD11b	M1/70	PerCP-Cy5.5	1:200
MHCII	M5/114.15.2	APC-eFluor780	1:200
CCR9	CW-1.2	eFluor450	1:200
B220	RA3-6B2	BV 605	1:200
CD86	GL1	BV 650	1:200
CD135	A2F10	PE	1:100
CD11c	N418	PE-Cy7	1:200

Table 3.7: Antibodies used for surface staining of Hoxb8 cell lines.

3.3 Statistical analysis

Standard statistical analysis was performed using GraphPad Prism (GraphPad Software Inc., La Jolla, CA, USA). For pairwise comparisons, unpaired Student's t-test was used, with $\alpha = 0.05$ (unless stated otherwise). For comparisons between multiple samples, one-way ANOVA with Tukey's multiple comparisons test between all samples was used, unless stated otherwise.

RNA-seq data were analysed using the R software version 3.3.3[95]. The packages used for specific analyses are mentioned in the respective sections below.

qPCR data were analyzed using the $2^{-\Delta Ct}$ method.

3.3.1 Exploratory analysis and data mining of the complete data set

For the exploratory analysis of the complete data set, without selection of differentially expressed genes, I acquired or generated lists of genes of interest, used them to filter the data set and then generated clustered heatmaps to visualize the data. Hierarchical clustering was performed using the function `hclust` with the

option `dist = euclidean`, meaning that euclidean distance was used as measure of similarity to order and cluster the genes.

For each heatmap, the TPM values for each gene are scaled with the `scale` function, which first centers the values by subtracting the mean, then scales it dividing by the standard deviation. This standardizes the data in a normal distribution centered on 0, thus allowing better visualization and reducing the impact of outliers.

Principal component analysis (PCA) was performed with the `prcomp` function, with no additional parameters.

3.3.2 DESeq2 analysis on steady state populations

To analyze the steady state dataset I used the `DESeq2`[96] package, which takes as input the un-normalized read counts, models them as a negative binomial distribution, a widely accepted modeling method for RNA-seq data that accommodates the overdispersion among biological replicate count data, and analyzes it by means of a generalized linear model (GLM). It tests significance by means of Wald test for pairwise comparisons, or in our case (3 samples) by likelihood ratio test (LRT), which is conceptually similar to an analysis of variance (ANOVA) calculation in linear regression, except that in the case of the negative binomial GLM, it uses an analysis of deviance (ANODEV), where the deviance captures the difference in likelihood between a *full* model, where all the variable elements are included, and a *reduced* model, where some elements are removed. Threshold was set at $\alpha < 0.01$.

The genes thus selected as differentially expressed were visualized on a heatmap and hierarchically clustered (see previous section). Six clusters were assigned with the `cutree` function. Each cluster was then functionally analyzed using GeneOverlap and Cytoscape.

The full script can be found in Appendix A, script 1.

3.3.3 Weighted gene co-expression network analysis

The Weighted Gene Co-expression Network Analysis (WGCNA) algorithm[97, 98] is very effective for deriving groups of highly co-expressed genes, referred to as co-expression *modules*, from large gene expression data sets. I used the `WGCNA` R implementation[99, 100] to define modules in each cell type independently, as the

large differences inherent in the different populations and independent of stimulation would strongly skew the analysis without being of major interest (as they are already analyzed independently, as described in the previous section).

A thorough description of the algorithm can found in Pandey et al. 2013[101]. Briefly, it starts with a matrix of the absolute values of Pearson correlation coefficients between each gene pair (calculated from the log2 normalized and batch-corrected TPM expression values), and converts this into an adjacency matrix using a simple power function $f(x) = x^{\beta}$. The parameter β is determined in a way that the resulting adjacency matrix is approximately scale-free, a widely accepted property of biological networks (network topology is reviewed in Hu et al., 2016[102]). The `WGCNA` package in R offers a function to facilitate choice of β (also called soft thresholding power) by calculating, for a series of values of β, the fitting index R^2 of the linear model that regresses $log(p(k))$ on $log(k)$, with k being the connectivity and $p(k)$ the frequency distribution of connectivity. A perfect scale-free network will have fitting index equal to 1. For network construction the smallest value of β that gives the highest R^2 can be chosen, usually above 0.8, corresponding to an approximate scale-free network.

This way, I chose 5 for the pDC data set and 4 for the $CCR9^{low}$ cells. For the pre-DC data set however, the fit index failed to reach values above 0.8 for reasonable powers (less than 15). This is usually due to a strong driver in the data that makes a subset of the samples globally different from the rest. The difference causes high correlation among large groups of genes which invalidates the assumption of the scale-free topology approximation.

The cause for this behavior in this experiment is likely due to the strong changes between conditions caused by the time series of stimulated samples. Being this a variable of interest that cannot be removed (i.e. adjust the data for it), an appropriate value of β is chosen based on the number of samples, which in this case (30 samples) is 8 (for details refer to the FAQ section of the WGCNA manual).

The adjacency matrix thus generated is converted into a Topological Overlap Matrix (TOM) by the following formula:

$$TOM(i,j) = \frac{l_{ij} + a_{ij}}{min(k_i, k_j) + 1 - a_{ij}}$$

where a_{ij} is the adjacency score between genes i and j (calculated as described above), $k_i = \sum_u a_{iu}$ (sum of adjacencies with neighboring genes) and $l_{ij} = \sum_u a_{iu}a_{uj}$ (sum of the product of the adjacencies involving all common neighbors). The TOM measures the strength of the association between two genes based on the ratio of the similarity of their common neighborhood to the smaller of the individual neighborhoods of the two genes. Using adjacency as input, this also gives higher weight to genes that are already strongly associated.

The modules are then found by using average linkage hierarchical clustering on the TOM, and applying a dynamic tree-cutting algorithm on the resulting dendrogram, that merges modules with more than 25% similarity (option `mergeCutHeight`, default = 0.25).

I applied the WGCNA algorithm with a blockwise approach, meaning that the analysis was performed individually on blocks of maximum 5000 genes (`maxBlockSize` = 5000), and the modules found were merged at the end; minimum module size (`minModuleSize`) was set to 30. All other parameters were used with default values. See Langfelder and Horvath, 2008, for details on the parameters.

The complete script can be found in appendix A, script 2.

3.3.4 GeneOverlap: functional analysis of clusters and modules

To perform functional analysis on the clusters and modules I took advantage of the Molecular Signature Database (MSigDB), a collection of annotated gene sets developed by the Broad Institute for use with GSEA software[103]. The gene sets are organized in 8 major collections and several sub-collections: for this project I used the sub-collection `c2_kegg`, which contains 186 curated gene sets derived from the KEGG pathway database; and the sub-collection `c3_tft`, containing 615 motif gene sets, that include genes that share upstream cis-regulatory motifs which can function as potential transcription factor binding sites[104].

Each cluster and module was compared to each gene set in the collections using the `GeneOverlap` package[105], which uses Fisher's exact test to calculate p value, odds ratio and Jaccard index for the overlap of two lists over the background of the complete library (in this case the 29353 genes detected). P values < 0.05 were

considered significant.

I used the `GeneOverlap` package also in the signature-module analysis (section 5.5.2), where I evaluated how many of the detected modules of co-expressed genes were enriched for each cell type specific signature (CDP, cDC, pDC), thus identifying the most informative set of modules that indicate cell fate following stimulation.

The complete script can be found in Appendix A, script 3.

3.3.5 Cytoscape: network construction and visualization

Significant gene sets were visualized using Cytoscape[106–108], an open source software platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other state data. Within Cytoscape I used the app EnrichmentMap[109], which allows to visualize gene set enrichment results as a network: nodes represent gene-sets and edges represent mutual overlap. In this way, highly redundant gene-sets are grouped together in clusters, dramatically improving the capability to navigate and interpret enrichment results. The software was used with the default settings; given the redundancy of transcription factor motif gene sets, those matching the same TF were merged in a single node.

4 Results

4.1 Definition of DC precursor and pDC populations in murine bone marrow

The main aim of this project was to define the transcriptional regulation of pDC differentiation from pluripotent progenitors, through the intermediate steps of pre-DCs and CCR9^{low} pDC-like cells. For this purpose, it was important to be able to easily and clearly discriminate the populations of interest in the normal mouse BM, in order to sort them to high purity and perform mRNA-sequencing (RNA-seq). In addition, I wanted to evaluate the impact of direct TLR activation on differentiation: it was therefore important to choose TLR ligands able to elicit responses in the populations of interest at low doses and at early time points, in order to evaluate primary responses and direct effects on the transcriptome.

4.1.1 Pre-DCs, CCR9^{low} pDC-like cells and mature pDCs are discrete and well defined populations in the murine bone marrow

CCR9^{low} pDC-like precursors were initially identified by gating BST2^{+} Siglec-H^{+} CD11c^{+} cells in the bone marrow[32]. For this project, the aim was to isolate CCR9^{high} pDCs, CCR9^{low} pDC-like precursors and pre-DCs to high purity, while excluding all contaminant cells expressing similar markers, such as cDCs, B cells and other myeloid cells. To this aim, I tested several combinations of markers (not shown), and chose to use Siglec-H but not BST2, and added CD135 (Flt3), B220, MHCII and Sirpα. To exclude other cell types, I used a lineage cocktail containing

CD3, CD19, NK1.1 and Ly-6G (the full panel is presented in table 3.4). Exclusion of CD11b^{+} cells was avoided as CD11b is also expressed at low levels on pDCs.

By staining lineage depleted BM cells (see methods section 4.2.3) with this panel, it is possible to identify pre-DCs, CCR9^{low} cells and pDCs as subsets of the Lineage negative, CD135 (Flt3) and CD11c positive population (Fig. 4.1, P4), which essentially contains all DC-lineage cells (expressing Flt3) downstream of the CDPs (which do not yet express CD11c).

Pre-DCs, also called pre-cDCs, are a rare population of precursors with a prominent commitment to the cDC fate[51]. As was recently demonstrated in our lab by means of continuous live cell imaging[86], a pre-DC population arises at early time points *in vitro* directly from CDPs following upregulation of CD11c but not Siglec-H, that can also transition to Siglec-H^{+} CCR9^{low} pDC-like precursors and then to pDCs, in the presence of Flt3 ligand. Two recent publications[50, 52] showed that a Siglec-H^{-} B220^{-} pre-DC subpopulation isolated *ex vivo* is committed to the cDC lineage, and contains cells which are committed to give rise to cDC1 and cDC2. With this panel, these cells can be found within the Siglec-H^{-} B220^{-} population (P5), and are further characterized by the lack of expression of the maturation markers MHC class II and Sirpα. Throughout this thesis I will refer to this population as pre-DCs. They constitute approximately 0.1% of the total BM cells.

Gating on Siglec-H positive cells (P6), mature pDCs and CCR9^{low} pDC-like cells can then be distinguished by plotting CCR9 against B220. These markers show a continuum rather than a clear separation of CCR9 negative (and B220 negative) cells from mature pDCs. While the latter are CCR9^{high} B220^{high}, the CCR9^{low} cells also express lower levels of B220, with no clear separation between the two, suggesting a sequence of differentiation from CCR9^{low} B220^{low} to CCR9^{low}B220^{int} cells, to CCR9^{high} B220^{high} mature pDCs. For clarity and to avoid cross-contamination during sorting, only the CCR9^{low} population was included (which contains B220$^{low/int}$ cells), and will be referred to simply as CCR9^{low} precursors. This population is approximately 0.5% of the total BM cells, while mature pDCs constitute 1.5-2%.

Expression of the maturation markers MHCII and Sirpα in these three populations also suggests different degrees of differentiation, with CCR9^{low} cells showing an intermediate phenotype between pre-DCs and pDCs (bottom right panel).

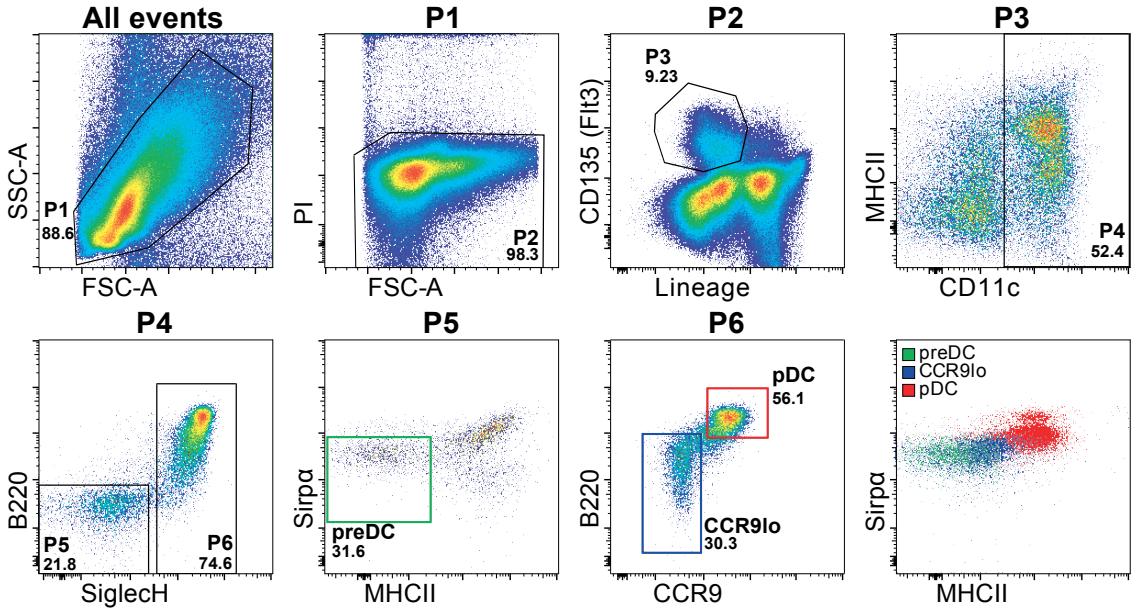


Figure 4.1: Example of staining and gating strategy to sort pre-DCs, CCR9^{low} precursors and pDCs from murine BM. Lineage depleted BM is used to sort to high purity the populations of interest, by first gating on lymphocytes (P1) and live cells (P2), and then on CD135 positive, Lineage negative cells (P3), followed by CD11c positive cells (P4). The three populations can be then discriminated by expression of the markers Siglec-H, B220, CCR9, Sirpα and MHC class II.

4.1.2 Pre-DCs, CCR9^{low} cells and pDCs show diverse responsiveness to different stimuli

Next, several TLR ligands were tested to evaluate the responsiveness of the different populations, and the kinetics of primary TLR activation in each of them, to select the appropriate time points for the subsequent transcriptome analysis. To this aim, I depleted freshly isolated BM of Lineage positive cells, and stimulated the remaining Lineage negative cells with optimal concentrations of ligands for TLR7 (R848, Resiquimod, 3µM), TLR9 (CpG-A, ODN 2216, 0.5µM), and TLR3 (poly I:C, 1µM), and live VSV virus (10^6 IU/ml), which was shown to activate TLR7 in pDCs and RIG-I in cDCs. Since in its wild-type form this virus suppresses type I IFN production, limiting DC activation and antiviral responses, I also used the mutated M51R variant (VSV-M) which lacks this protective mechanism[110]. Expression of the costimulatory molecule CD86 was then analyzed in the different cell populations by flow cytometry at 2, 4, 6, 12 and 24 hours.

As summarized in figure 4.2, TLR7 and -9 stimuli led to strong activation of

pDCs and CCR9^{low} precursors, and to a much lower extent of pre-DCs, as measured by upregulation of the activation marker CD86. Expression peaked at 6h after stimulation and then declined. VSV, in both WT and mutated forms, led to CD86 upregulation only at later time points, indicating delayed kinetics of activation, which can become confused with secondary activation of bystander cells by cytokines, such as Type I IFNs. Poly I:C had little to no effect on any cell type

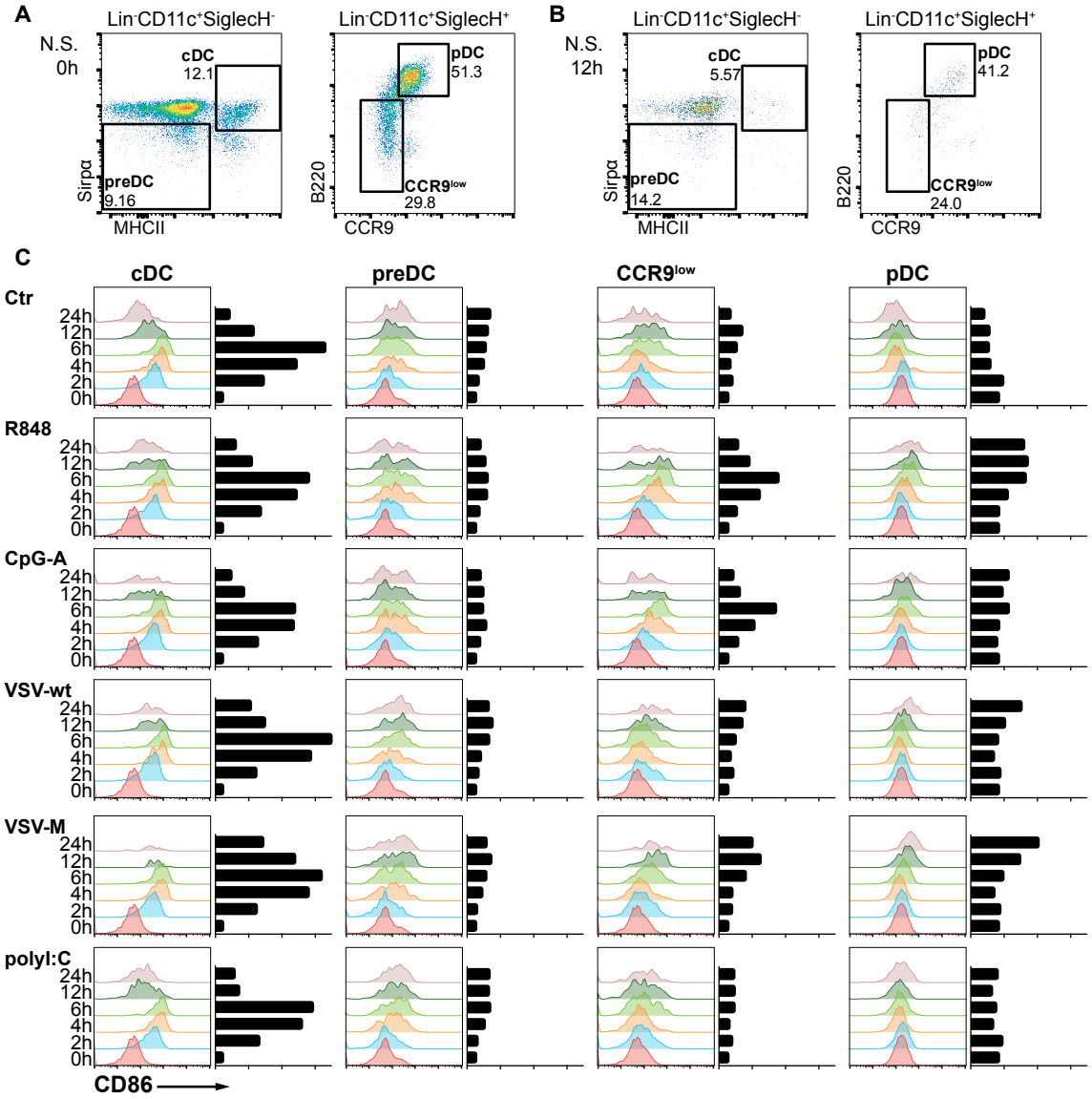


Figure 4.2: Evaluation of the responsiveness and kinetics of preDCs, CCR9^{low} precursors, pDCs and cDCs to different TLR ligands. Freshly isolated bone marrow cells were depleted of lineage positive cells and stimulated with different TLR ligands and VSV, and cell activation was evaluated at different time points. **A-B.** Gating example showing the untreated sample at time 0 (**A**) and time 12 hours (**B**). **C.** CD86 fluorescence of each population with the indicated treatments. Bar graphs represent mean fluorescence intensity (MFI) of CD86-PE staining. Representative results of 2 experiments are shown.

compared to the medium control, suggesting a lack of expression of TLR3 in these cell types. CDCs showed non-specific CD86 upregulation in all conditions, including the unstimulated control, indicating that none of these treatments directly activates them.

An important caveat of this analysis is that the activated cells might be lost to cell death or downregulation of markers early on in the experiment. As it is highlighted by panels A and B, already the untreated control shows a noticeable reduction of all $CD11c^+$ cells at 12h, with important changes also in the percentage of mature cDCs (while the relative frequency of pDCs and $CCR9^{low}$ precursors is unchanged).

Based on these results and observations, CpG-A and R848 were selected as stimuli for the RNA-seq experiments, as they both achieved strong direct activation at early time points. In addition, since CpG-A is a very strong IFN-α inducer but a relatively poor inducer of IL-6 expression, a second TLR9 ligand, CpG-C (ODN 2395), was added, which has been described as a good activator of both pathways, with similar kinetics to CpG-A[111]. Moreover, to limit the loss of cells of interest and confounding factors such as indirect stimulation by cytokines, pre-DCs, $CCR9^{low}$ precursors and pDCs were sorted previous to stimulation.

4.2 Transcriptome analysis of DC precursor and pDC populations

Figure 4.3 summarizes the experiment layout that was chosen for the analysis of the transcriptome of pre-DCs, $CCR9^{low}$ precursors and pDCs, in the steady state and after TLR-7 and -9 stimulation. In each of three independent experiments, each population was split immediately after sorting into 10 samples which were incubated with the appropriate stimulus for the time indicated. The control sample was directly lysed and the RNA isolated. At the end of the incubation time, total RNA was isolated, and cDNA libraries were prepared and sequenced.

After alignment and batch correction, the data were analyzed using R statistical software[95] and dedicated packages as described in the Methods.

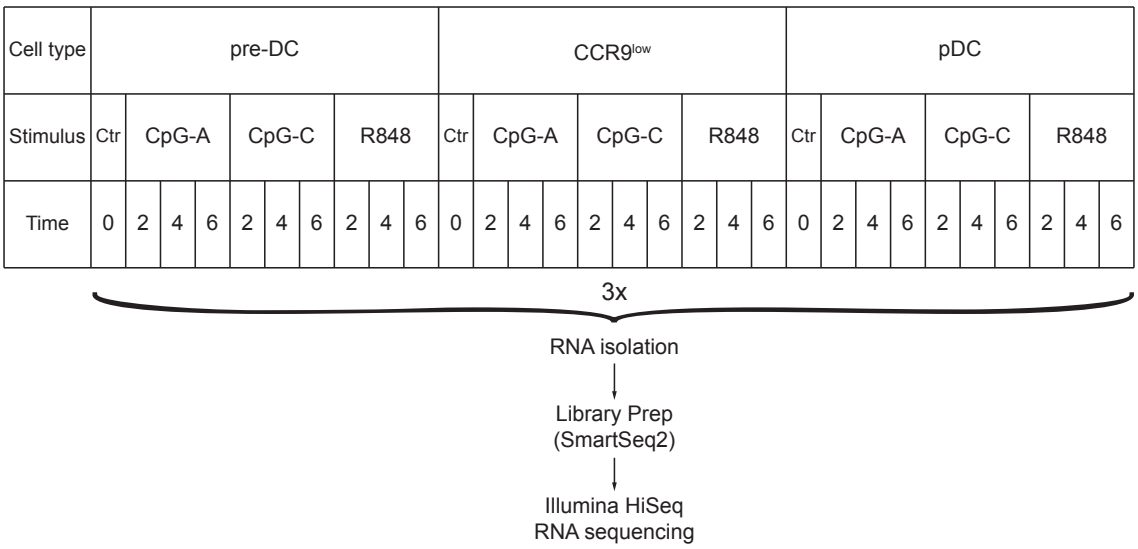


Figure 4.3: Scheme of the RNA sequencing experiment setup. Each cell population was sorted to high purity from primary BM cells using a BD FACS Aria III (Becton Dickinson), then divided in 10 experimental conditions and, at the indicated time points, the total RNA was isolated and cDNA libraries were prepared as described in the methods.

4.2.1 Type I IFN pathway genes are upregulated in response to all stimuli in CCR9^{low} precursors and pDCs but not in pre-DCs

As a first step in the analysis of the RNA-seq data, I used the whole data set, without selecting differentially expressed genes, to perform exploratory analysis using known gene lists and correlated pathways to evaluate the intrinsic characteristics of each cell type, their relations to each other and their changes following TLR stimulation. A list of genes of interest was generated by searching the Gene Ontology (GO) database (by means of the AmiGO tool v2.3[112]) for the string "Type I interferon pathway" (restricted to the organism *Mus Musculus*). This generated a list of 104 genes that are annotated to be involved in the IFN-I signaling pathway, including molecules involved in IFN induction and PRR signaling pathways, which was then used to generate an expression heatmap of all conditions with hierarchical clustering of the genes (by means of euclidean distance), to highlight clusters of co-regulation (figure 4.4). This heatmap gives an immediate visual summary of the responses to TLR stimulation in the three different populations.

In CCR9^{low} cells and pDCs, all Ifna (IFNα) genes, together with many others in

this pathway, are readily upregulated upon both TLR7 and -9 stimulation, already at the 2 hours time point. Type I IFN induced genes followed at 4 and 6 hours after stimulation in pDCs and CCR9^{low} precursors. In contrast, pre-DCs do not upregulate any Type I IFN specific genes, and the expression of other less specific genes, e.g. Tlr4 and -8, is unchanged following stimulation. At later time points, increased expression of a set of inflammatory genes (bottom part of the heatmap) was observed in pre-DCs, including mostly type I IFN induced genes, such as interferon regulatory factors (Irfs), Ddx58 (RIG-I) and Myd88; this could be due either to

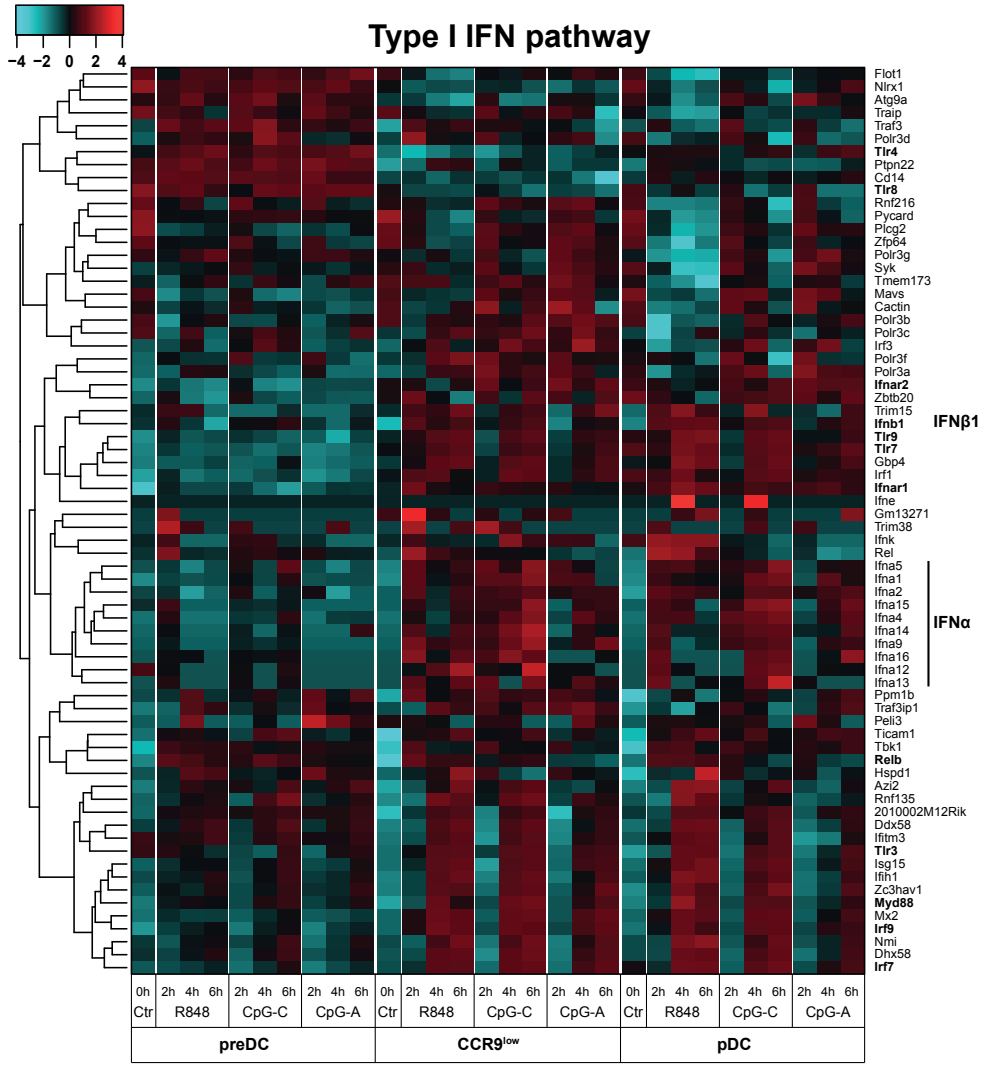


Figure 4.4: Analysis of the Type I IFN pathway genes reveals differences and similarities in the steady state and following stimulation. Genes annotated with the GO term "Type I interferon pathway", clustered by means of euclidean distance. Expression values are scaled for this specific selection, centered on the average value (=0, black) and lowest and highest values as extremes (cyan and red, respectively), to highlight differences within the selected set of genes.

spontaneous activation in the culture, or to activation of a very small subset of cells within the population, whose contribution is hardly detectable in the bulk analysis.

Taken together, these results suggest that most pre-DCs do not mount a Type I IFN response to TLR7 and -9 stimuli, while CCR9^{low} precursors and pDCs show a very similar behavior and phenotype, indicating high functional similarity between these two populations. In addition, in this analysis significant differences in the responses to the TLR7 and TLR9 ligands were not observed, showing that R848, CpG-A and CpG-C all activate strongly the Type I IFN pathway.

4.2.2 Pre-DCs express a different TLR repertoire

The reason for the lower responsiveness of pre-DCs to TLR7 and -9 stimulation is to be found in their characteristic "commitment" to the cDC lineage. This is exemplified in the pattern of expression of the Tlr genes (figure 4.5): in the steady state, pre-DCs express higher levels of Tlr1, -2, -3, -4, -6, -8, -11 and -13 compared to the other populations, while expression of Tlr7 and -9 is lower (p value = 0.0003 and 0.0503, respectively). On the other hand, CCR9^{low} precursors and pDCs share a very similar pattern of expression, with Tlr7, -9 and -12 expressed at higher levels than in pre-DCs, and all the other Tlrs expressed at markedly lower levels than in pre-DCs. Upon stimulation, the expression of Tlrs in pre-DCs remained unchanged, while CCR9^{low} precursors upregulated Tlr7, -9, -3 and -12, thereby acquiring the TLR expression pattern of pDCs. In pDCs the same Tlrs were upregulated, to a lesser extent, suggesting a less mature phenotype in CCR9^{low} precursors that could be driven to maturation by the stimuli themselves.

Although there are no evident changes in pre-DCs following stimulation, it

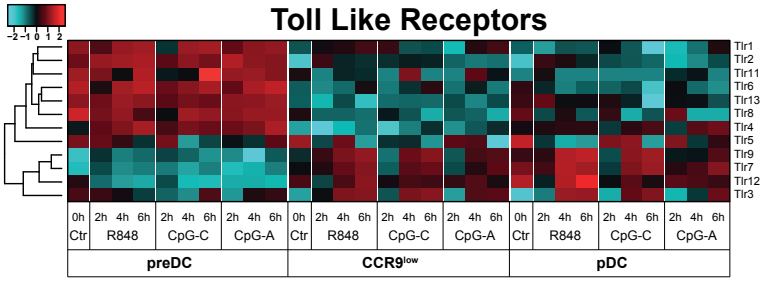


Figure 4.5: Toll-like receptors are differentially expressed in pre-DCs compared to CCR9^{low} precursors and pDCs. Expression of TLR genes in the data set. Expression values are scaled for this selection (see caption of figure 4.4).

cannot be excluded that a small subset of them does respond specifically to the TLR ligands used, but their contribution is masked in the bulk population, becoming undetectable with this type of analysis.

4.2.3 Principal component analysis highlights distinctions between populations in the steady state, as well as functional similarities following stimulation

Analysis of the whole data set generated by mRNA sequencing, without selection of differentially expressed genes, reveals intrinsic characteristics of each cell population, and how they compare to each other. Indeed, when analyzing steady state populations alone by principal component analysis (PCA), pre-DCs and pDCs are clustered at opposite sides of the plot, indicating major differences between the two populations (figure 4.6A). $CCR9^{low}$ precursors are found relatively equidistant from pre-DCs and pDCs, suggesting an intermediate phenotype with characteristics of mature pDCs mixed with precursor-like features.

Including the stimulated samples in the analysis (Figure 4.6B-D) highlights the functional differences between the populations: steady state pre-DCs are clearly separated from $CCR9^{low}$ cells and pDCs, which now overlap in a single cluster. Following stimulation, pre-DCs do separate from the steady state, indicating changes in their transcriptome, and remain separate from the other stimulated populations. Nevertheless, the progression of changes is consistent with that of the more differentiated cells, although not as pronounced, suggesting a specific activation pattern. Indeed, with all stimuli both $CCR9^{low}$ precursors and pDCs are found at progressively greater distance from their respective steady state counterparts with increasing time, in line with a regulated process of activation with sequential changes in gene expression.

Differences between the populations and the various stimuli will be discussed in details in the WGCNA analysis section (Section 4.5).

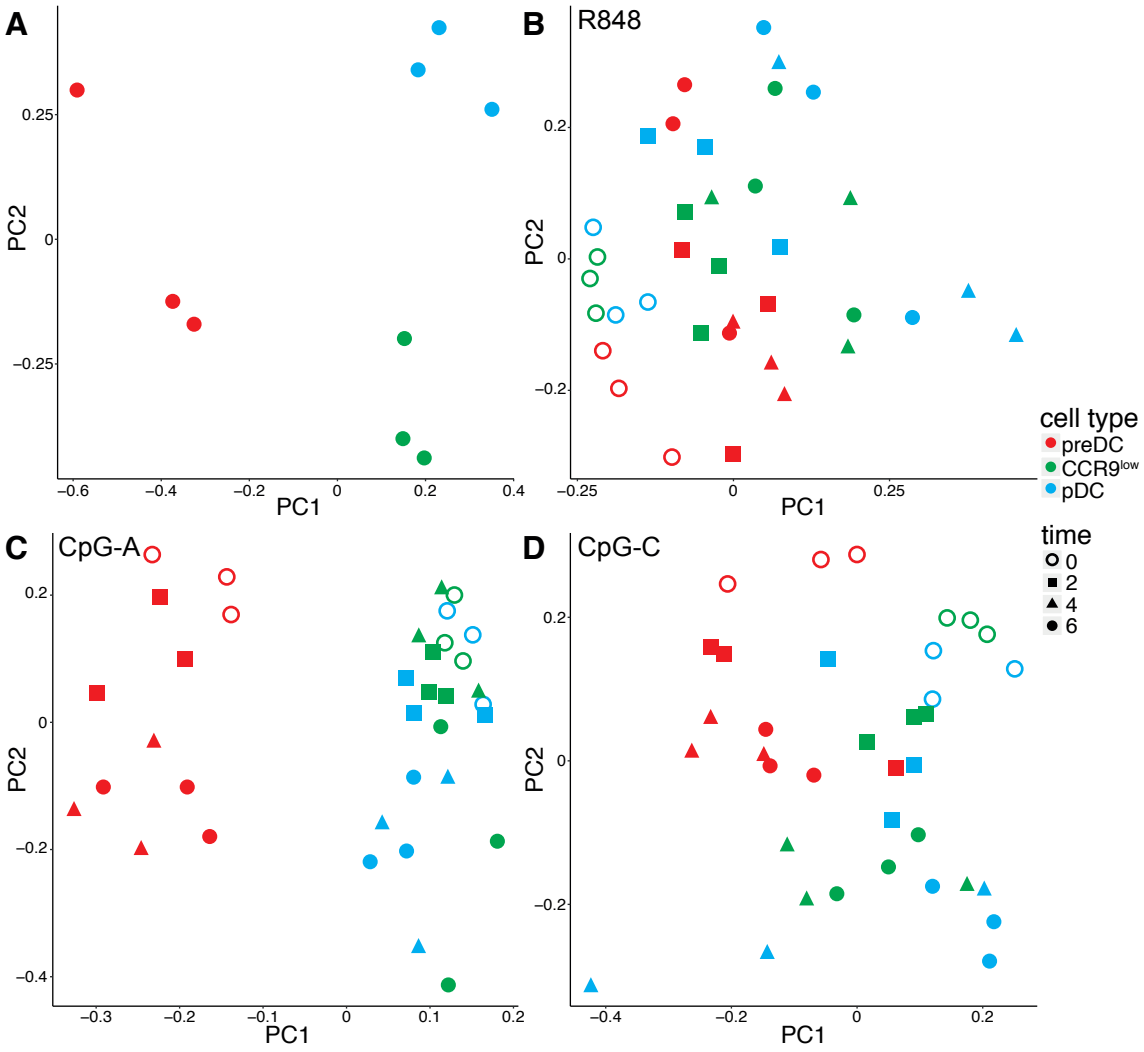


Figure 4.6: Principal component analysis of the steady state and stimulated populations. The whole unfiltered data set was used for PCA. **A.** Considering only the steady state populations, CCR9^{low} cells (green) are positioned at relatively equal distances from pre-DCs (red) and pDCs (blue), suggesting an intermediate phenotype, with mixed characteristics of precursor cells as well as mature pDCs. **B-D.** Including the stimulated samples in the analysis, CCR9^{low} cells and pDCs appear to be more closely related and more distant from pre-DCs already at steady state, and with similar behaviors after stimulation.

4.2.4 Independently generated cell-type specific signatures characterize the different populations and are regulated upon TLR activation

To categorize these cell types based on previous knowledge and population definitions, I took advantage of cell-type specific gene expression signatures generated independently from the Immgen database. I used signatures for BM CDPs and

splenic (mature) pDCs previously generated in our lab[53], and for splenic cDCs from Miller et al.[113]. All these signatures were generated by comparing the different DC subsets in the steady state and extracting sets of population-specific genes, upregulated only in the selected cell type relative to all other subsets.

After generating heatmaps for each of them from all of our populations, I compared expression of these signature genes in the steady state as well as following stimulation (figure 4.7).

As expected, steady state pDCs specifically express higher levels of all genes of the pDC signature compared to the other populations, while they show lower expression of most of the other two signatures. Pre-DCs on the other hand, express much lower levels of pDC signature genes, but have higher expression of both CDP and cDC signature genes, indicating their precursor nature as well as their commitment to the cDC lineage. Interestingly, CCR9^{low} precursors express almost all genes of the pDC signature, although to a lower extent than pDCs, while also showing high expression of most of the CDP signature genes, similar to pre-DCs, and a small subset of the cDCs genes. This is in line with previous findings that these cells, while phenotypically and functionally very similar to mature cells, still retain differentiation potential and can switch to cDC phenotype under certain conditions[32, 53].

Following activation, pre-DCs and pDCs show very little changes in their signatures, indicating lower responsiveness of the former and final commitment of pDCs, while CCR9^{low} cells readily downregulate both CDP and cDC specific genes, and upregulate their pDC signature, suggesting a stimulus-induced differentiation. Surprisingly, a small subset of cDC specific genes (top of the cDC signature heatmap) are quickly upregulated upon stimulation in both CCR9^{low} precursors and pDCs. These genes include Cd83 and Cd86, which are important for antigen presentation and are normally expressed at higher levels in cDCs than pDCs. Furthermore, Id2, a lineage defining factor for cDCs, is upregulated by stimulation in CCR9^{low} precursors and in pDCs.

Focusing on TFs that are known to be differentially expressed in the different DC subtypes (figure 4.8), in the steady pre-DCs showed higher expression of many myeloid and cDC specific factors (top of the heatmap) and not of the factors highly

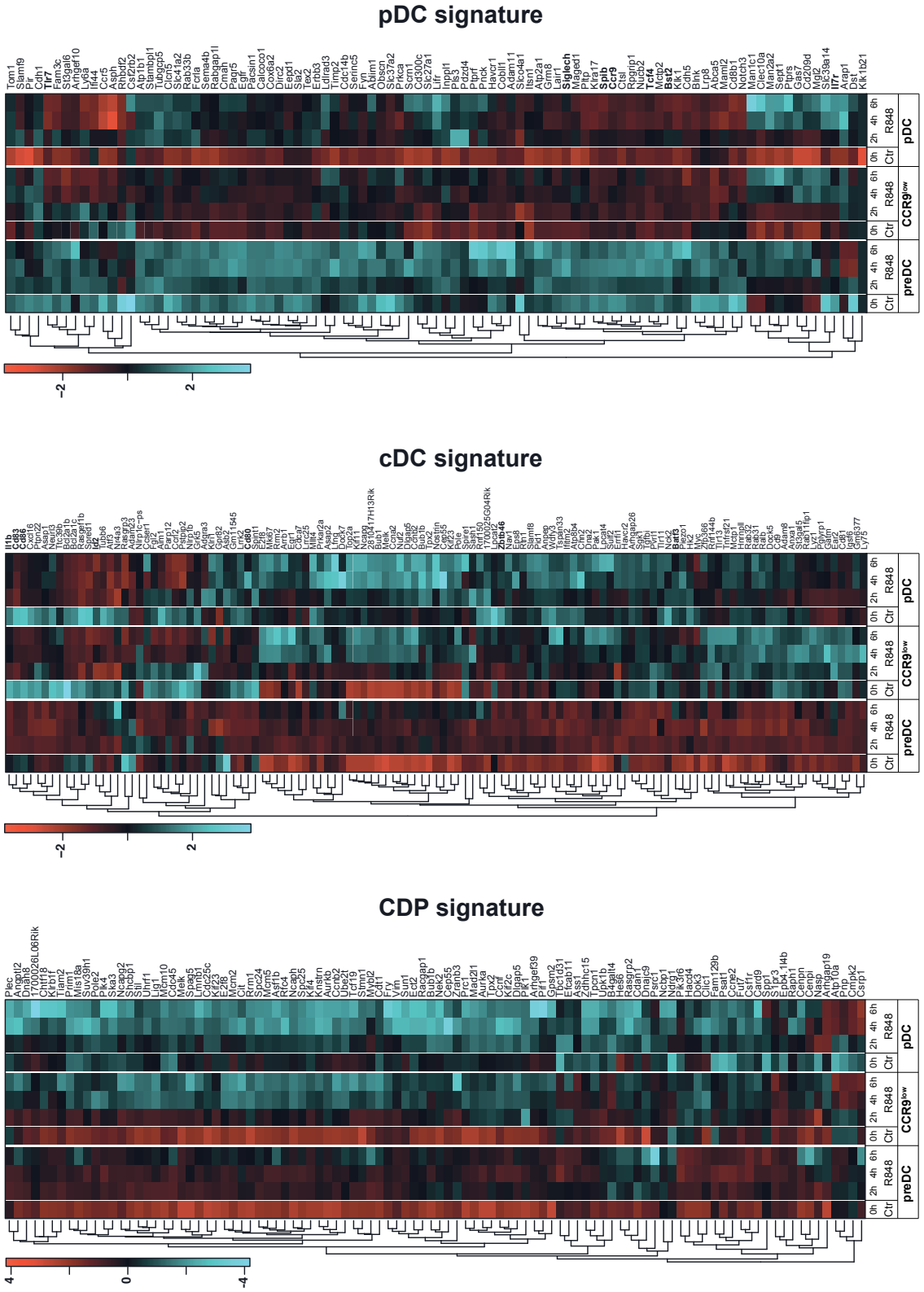


Figure 4.7: Expression of DC subtype-specific signatures. Signatures were generated from data retrieved from the Immgen database in our lab (CDP and pDC)[53] or by Miller et al. (cDC)[113]. Values are scaled independently for each heatmap (see caption of figure 4.4). For clarity, only R848 is shown for the stimulated samples.

expressed in pDCs (bottom). Interestingly, CCR9^{low} precursors expressed the same levels of pDC-specific factors (Tcf4, Bcl11a, Irf8 and others) than pDCs, as well as some pre-DC-expressed genes (Klf4, Cbfb, Runx1). Upon stimulation, pre-DCs showed little change in TFs expression, with limited upregulation of Stat3 and Relb. Both pDCs and CCR9^{low} cells however, upregulated a subset of cDC specific TFs, including Id2, Bcl6 and Nfil3, as well as inflammatory TFs such as Nfkb1 and Relb, and at later time points Irf2 and Notch2. Interestingly, pDCs showed more downregulation of Runx2, Bcl11a and Irf4 compared to CCR9^{low} precursors.

These results suggest the activation of an inflammatory program in response to TLR7 and -9 stimuli that involves cDC-specific transcription factors.

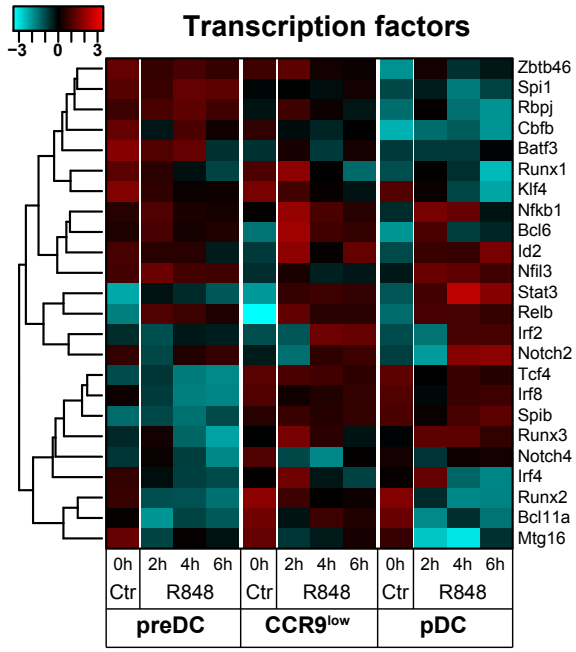


Figure 4.8: Expression of DC subtype-specific transcription factors. CDC and pDC specific transcription factors are differently expressed in the different cell populations, and are regulated upon stimulation.

4.2.5 Expression of Id2 is upregulated in CCR9^{low} precursors and pDCs by TLR stimulation

I was able to verify this finding by using an Id2 reporter mouse that was available in our lab, which has an IRES-eGFP reporter cassette within the endogenous Id2 locus, resulting in a reliable reporter for Id2 expression, without loss of function [91].

Id2$^{eGFP/eGFP}$ mice were subcutaneously injected with CpG-A and Id2-eGFP

expression was detected by flow cytometry in BM cells and splenocytes after 16 or 72 hours. Furthermore, BM cells from these mice were isolated and treated *in vitro* with TLR ligands for 4 and 16 hours. Cells were gated as previously described, and a cDC gate (Lin$^-$ CD135$^+$ CD11c$^+$ Siglec-H$^-$ MHCII$^+$ Sirp$\alpha^{+/-}$) was included as positive control.

In vivo (figure 4.9), unstimulated BM pDCs and CCR9^{low} cells expressed very low levels of Id2, while pre-DCs showed higher expression, although lower than cDCs, as expected. Upon stimulation, both pre-DCs and cDCs transiently upregulated Id2 expression, which was back to normal expression levels 72 hours post injection. Interestingly, CCR9^{low} precursors showed upregulation already at 16h, and maintained it at the later time point, while mature pDCs upregulated Id2 only later, at 72h. Interestingly, in the unstimulated spleen expression of Id2 was higher on both cDCs and CCR9^{low} precursors. These two cell types also showed a similar behavior upon stimulation, with strong transient upregulation of the protein, clearly detectable at 16h. Pre-DCs, while expressing similar Id2 levels in the spleen as in the BM at the steady state and at 16h, showed stronger upregulation at 72h. Moreover, pDCs showed faster upregulation in the spleen than in the BM, with expression of Id2-eGFP increasing already at 16h.

These results confirm at the protein level the upregulation of Id2 gene expression which was detected by RNA sequencing.

Systemic TLR stimulation however causes a strong activation of many types of cells, and may trigger migration of responding cells from hematopoietic tissues to lymphoid organs, making the analysis of rare populations difficult.

Therefore, I isolated BM cells from untreated Id2-GFP mice and stimulated them *in vitro*, in order to better follow activation of specific populations without the risk of losing them to migration. Moreover, I also stimulated with CpG-C and R848, in addition to CpG-A, to better compare to the RNA-seq results.

In this experiment, no effect was detectable after 4h, while at the 16h time point all stimuli had induced upregulation of Id2 in all cell types, including cDCs and pre-DCs (figure 4.10), confirming the results of the in vivo experiment.

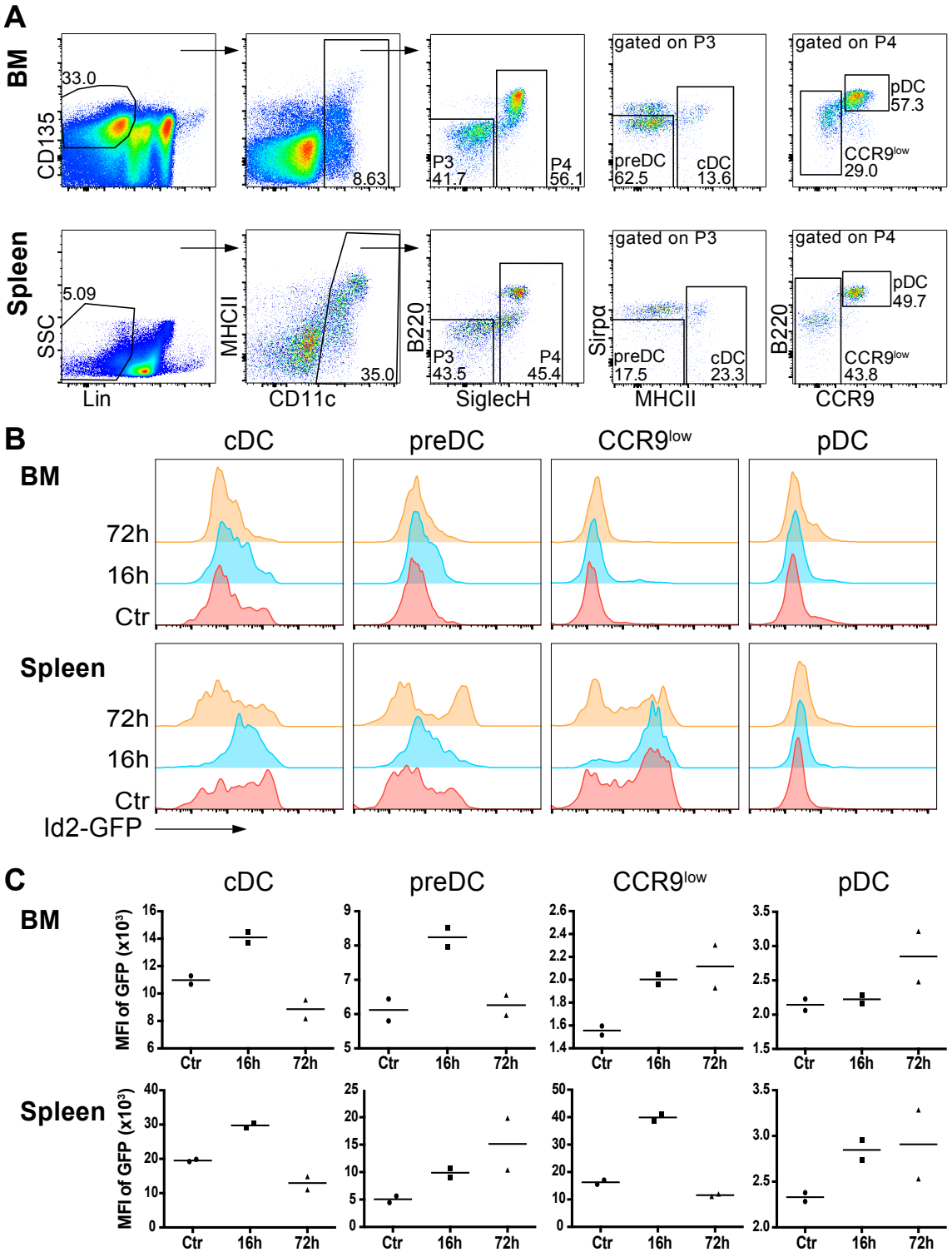


Figure 4.9: Id2 expression following stimulation *in vivo*. Id2-GFP reporter mice were injected subcutaneously with CpG-A. At the indicated time points, mice were sacrificed and spleen and BM were analyzed by flow cytometry. **A.** Example of gating in BM (top) and spleen (bottom). **B.** Representative histograms and **C.** quantification of GFP fluorescence in the indicated populations. n = 2 mice per group.

Taken together, these results validate the RNA-seq findings that Id2, a cDC-specific transcriptional regulator, is upregulated upon TLR7 and -9 stimulation in pDCs and pDC precursors, both at the mRNA and at the protein level.

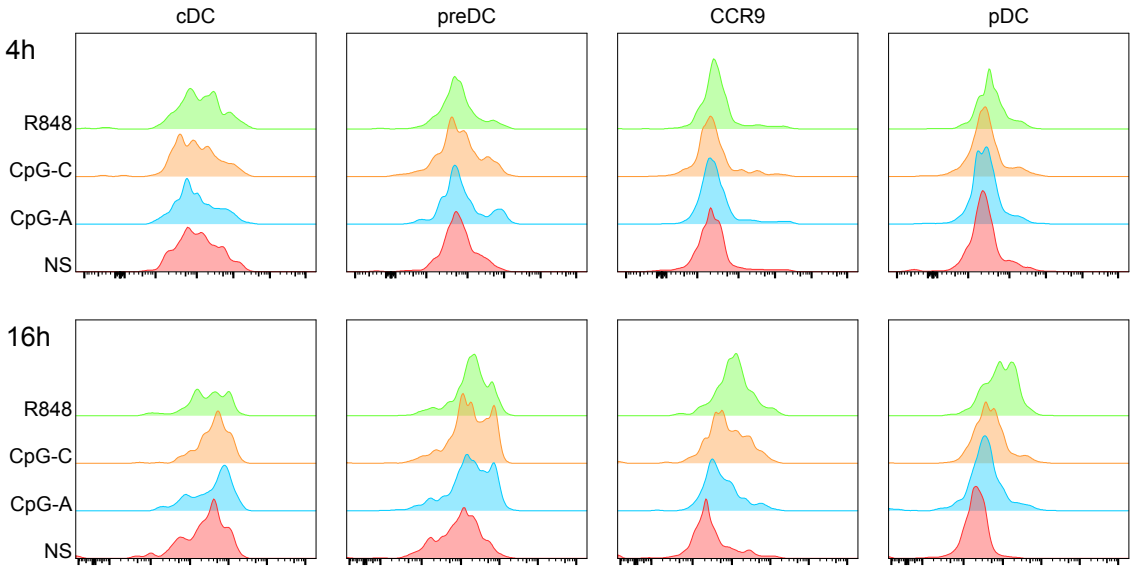


Figure 4.10: Id2 expression following BM stimulation *in vitro*. BM cells were isolated from Id2-eGFP reporter mice, and treated *in vitro* with TLR7 or -9 ligands (3µM R848, 0.5µM CpG-C and 0.5µM CpG-A) or cultured with medium alone (NS, non-stimulated), for 4 or 16 hours. Cell populations were gated as indicated in figure 4.9A. Results of one representative experiment are shown (two experiments with similar results were performed).

4.3 Analysis of differentially expressed genes in steady state pre-DCs, CCR9^{low} precursors and pDCs

For statistical analysis of the RNA-seq data, separate analyses of the gene expression patterns of the unstimulated and the stimulated populations were performed (figure 4.11). This way, I could focus first on normal differentiation, investigating differential expression across untreated populations and searching for regulators of pDC differentiation, and investigate the responses to TLR activation as a second step.

The steady state data set is composed of 3 conditions, with 3 replicate samples each. To identify clusters of co-regulated genes, I selected differentially expressed

(DE) genes by means of an ANOVA-like test (likelihood ratio test, LRT), and performed hierarchical clustering. The design of the TLR stimulation experiment was more complex, with 3 cell types divided in 10 conditions each (for a total of 90 samples), with one independent variable (treatment) and one ordered variable (time). To simplify the analysis and limit confounding factors, I divided this data set by cell type and analyzed each independently using the Weighted Gene Co-expression Network Analysis (WGCNA) algorithm. This analysis defines modules (equivalents to clusters in LRT analysis) of co-expressed genes, which can then be analyzed for functional significance. Comparisons of the WGCNA results for each population highlights differences and similarities between the cell types.

For functional analysis, I calculated the overlap with gene sets from the MSigDB collection 3, "transcription factor targets" (`c3_tft`) and collection 2, "KEGG path-

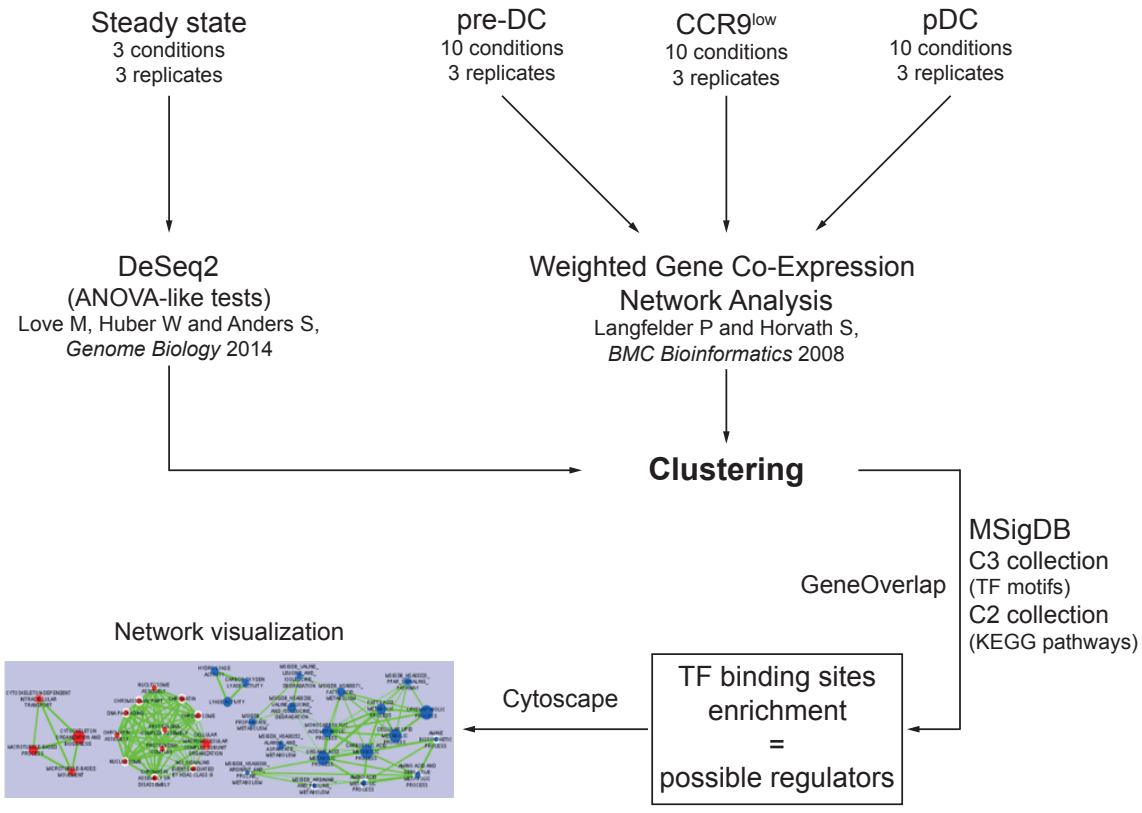


Figure 4.11: Analysis workflow for the RNA sequencing data. Given the complexity of the experimental design, the steady state and the stimulated data sets were analyzed separately, the latter separated into the three cell types. After the differentially expressed genes were calculated and clustered as indicated, each co-regulated cluster (or module) was screened for functional significance as well as for transcription factor binding sites enrichment, in order to identify putative regulators. Finally, Cytoscape was used to visualize the network of regulatory factors.

ways" (`c2_kegg`, biological pathways from the KEGG repository), for each cluster and module. The results of this analysis were merged and visualized using Cytoscape to build network maps of regulatory factors and regulated pathways. The complete analysis strategy is summarized in figure 4.11.

4.3.1 Genes differentially expressed during differentiation are co-regulated in clusters and belong to functional pathways for DCs

The likelihood ratio test (LRT) on the steady state data set identifies 2274 genes differentially expressed (DE) with p value < 0.01. These can be clustered by means of euclidean distance into six main clusters of co-expression, of varying size and expression patterns (figure 4.12). To understand the functional meaning of each cluster, I used the R package GeneOverlap[105] to compare each one with the MSigDB collection `c2_kegg`, which contains gene sets corresponding to KEGG biological pathways. GeneOverlap works by comparing two lists of genes against a genome background (the total of the expressed genes) in the form of an hypergeometric distribution, using the Fisher's exact test to calculate the probability of the two lists overlapping by chance, and returning for each comparison a p value and odds ratio, as well as a Jaccard index for similarity of the lists. For simplicity, here I consider only the p value as a measure of significance of the overlap between each cluster and the KEGG pathway gene sets.

Table 4.1 summarizes the results of the functional enrichment, clarifying for each enriched KEGG pathway the most significantly DE genes in the cluster. Disease pathways (e.g. *Leishmania infection*) are often highly redundant with physiological pathways.

Cluster 1 contains 255 genes that are expressed at lower levels in CCR9^{low} precursors only, including many adhesion molecules (Itga8, Itgb2) and complement proteins (C3, Cd55), enriched in pathways related to infection as well as innate immunity and cell adhesion.

Cluster 2 contains 263 genes that are expressed at higher levels in pre-DCs and at lower levels in the other 2 populations. These include several cytokines

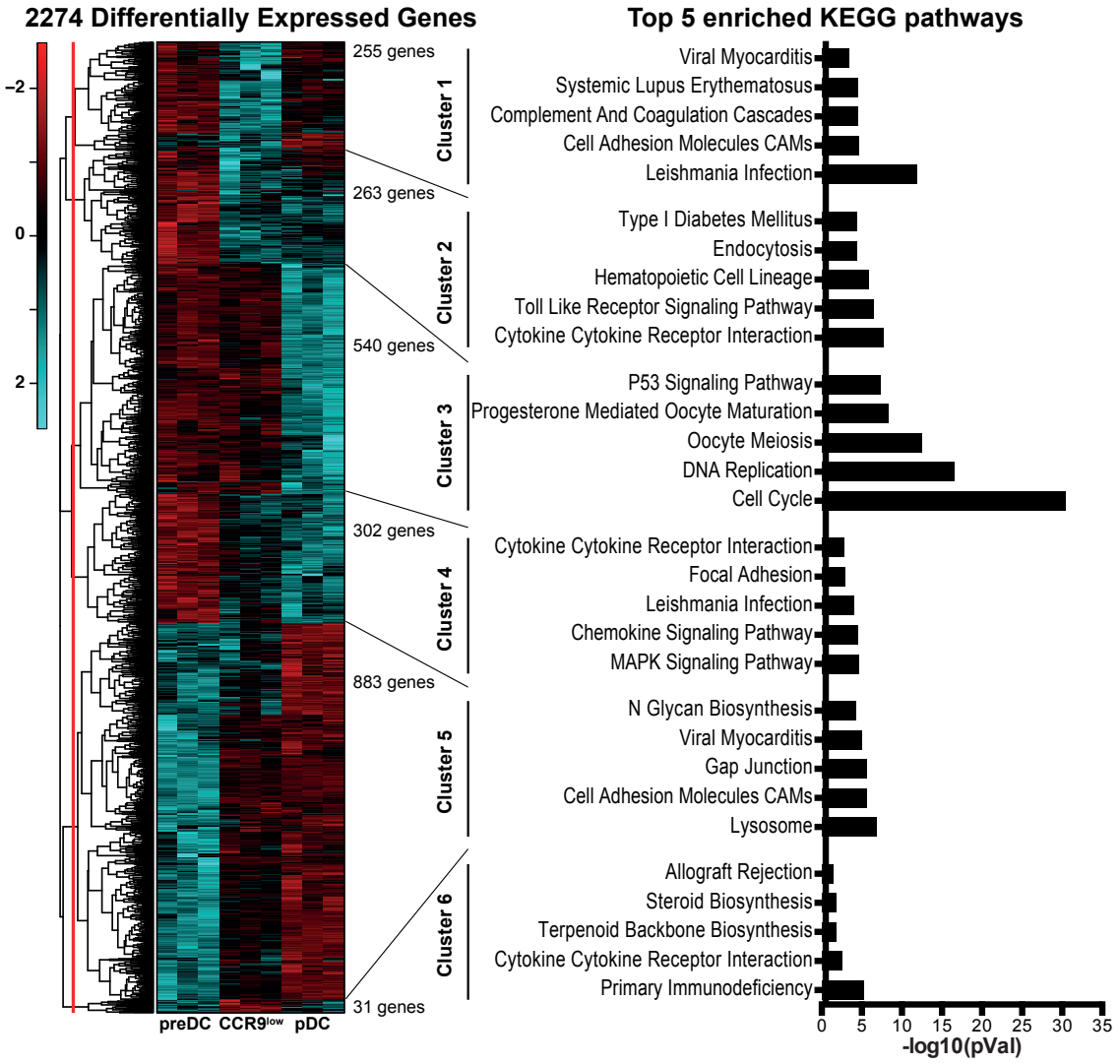


Figure 4.12: Differentially expressed genes in the steady state, and enrichment for KEGG pathways. The 2274 differentially expressed genes identified by likelihood ratio test (LRT) with p value < 0.01 were hierarchically clustered by means of euclidean distance, and the 6 main clusters were compared with MSigDB collection `c2_kegg`, identifying several pathways significantly enriched in each cluster. The red line on the dendrogram indicates cutoff height for cluster assignment. The bar graph represents the 5 most significantly enriched KEGG pathways in each cluster (-log10[p value]).

(Il1b, Tnf) and chemokines (Ccl3, Cxcl16) and their receptors (Csf3r, Cxcr2, Ifngr1 and 2), as well as costimulatory molecules (Cd40) and activation markers (Cd80, Cd86), belonging to inflammatory pathways. These genes are known to be expressed at higher levels in conventional DCs in the steady state compared to pDCs, and this is once again in line with previous results showing that Siglec-H^{-} pre-DCs are committed to the cDC lineage.

Cluster 3 contains 540 genes that are expressed at lower levels in pDCs only.

These belong to pathways controlling DNA replication and cell cycle (cyclins and Cdks such as Ccnb1 and 2, Cdc45, Cdk1 and more): indeed, mature pDCs in the BM are slowly dividing cells[53], which is congruent with the downregulation of proliferation genes, whereas pre-DCs and CCR9^{low} cells are precursor populations with higher proliferative capacity.

Cluster 4 contains 302 genes that are strongly expressed in pre-DCs, low in CCR9^{low} precursors and lowest in pDCs, including Id2, Itgam (CD11b), Spi1 (PU.1) and Tlr2 and 3, and other cDC specific genes belonging to immune and inflammatory pathways, that are indeed still expressed at low levels in CCR9^{low} precursors.

Cluster 5 contains 883 genes that are low in pre-DCs, higher in CCR9^{low} precursors and highest in pDCs, including genes that are necessary for antigen uptake and processing (in the lysosomal pathway), adhesion and cell-cell communication; not surprisingly, this cluster also contains most pDC-specific genes and transcription factors, such as Ccr9, Siglech, Tcf4 (E2-2), Bcl11a and many others.

Cluster 6 contains only 31 genes that are higher in CCR9^{low} cells compared to the other 2 populations, including Cd27, Itgax (CD11c) and Il12a.

These results confirm the nature of pre-DCs as cDC-imprinted proliferating precursors. They also highlight the functional similarity between CCR9^{low} precursors and pDCs (Cluster 5) as well as their difference in terms of precursor potential, e.g. considering proliferation (Cluster 3) and cDC potential (Cluster 4).

Cluster	Pathway	Enrichment p value	Pathway genes in cluster[a]
1	Leishmania infection	1.38×10^{-12}	*C3, Ncf4, Irak4, Mapk13, Jun* +1
	Cell adhesion molecules (CAMs)	2.36×10^{-05}	*Itga8, Vcan, Itgb2, Ncam1, Cldn15*
	Complement and coagulation cascade	3.07×10^{-05}	*C3, F10, C5Ar1, F5, Cd55* +1
	Systemic lupus erythematosus	3.23×10^{-05}	*C3, Elane*
	Viral myocarditis	4.42×10^{-04}	*Cd55, Itgb2*
2	Cytokine/Cytokine receptor interaction	2.22×10^{-08}	*Il1b, Tnf, Ccr1, Tnfrsf21, Csf3r* +10
	Toll-like receptor signaling pathway	3.60×10^{-07}	*Il1b, Tnf, Cd14, Ccl3, Cd86* +5
	Hematopoietic cell lineage	1.29×10^{-06}	*Il1b, Itga1, Tnf, Csf3r, Cd14* +3

[a] Genes overlapping in the pathway gene set and the cluster. The 5 most significant for differential expression are shown, numbers indicate remaining genes in the overlap.

Cluster	Pathway	Enrichment p value	Pathway genes in cluster[a]
	Endocytosis	4.35×10^{-05}	*Rab31, Asap1, Hspa1a, Rab11fip1, Hspa1b* +5
	Type I diabetes mellitus	4.54×10^{-05}	*Il1b, Tnf, Cd86, Cd80, Ica1*
3	Cell cycle	3.60×10^{-31}	*Bub1b, Bub1, Cdk1, Cdc45, Mcm3* +30
	DNA replication	2.85×10^{-17}	*Lig1, Pola1, Pole, Mcm3, Mcm5* +10
	Oocyte meiosis	3.24×10^{-12}	*Bub1, Cdk1, Sgol1, Ccnb2, Ccnb1* +13
	Progesterone mediated oocyte maturation	5.86×10^{-09}	*Bub1, Cdk1, Ccnb2, Cdc25b, Ccna2* +8
	p53 signaling pathway	4.95×10^{-08}	*Cdk1, Ccnb2, Ccnb1, Ccne1, Pmaip1* +7
4	MAPK signaling pathway	2.35×10^{-05}	*Pak1, Dusp3, Dusp2, Flna, Arrb1* +7
	Chemokine signaling pathway	2.98×10^{-05}	*Pak1, Hck, Ccr2, Arrb1, Elmo1* +5
	Leishmania infection	9.97×10^{-05}	*Ptgs2, Itgam, Tlr2, Marcksl1, Mapk3* +1
	Focal adhesion	0.001	*Pak1, Met, Itga5, Flna, Dock1* +3
	Cytokine/Cytokine receptor interaction	0.002	*Met, Csf1r, Ccr2, Il13ra1, Acvrl1* +4
5	Lysosome	1.33×10^{-07}	*Gns, Ctso, Sort1, Npc1, Ap1m2* +12
	Cell adhesion molecules (CAMs)	2.83×10^{-06}	*Cd4, Sdc4, Sdc2, Cdh5, Cd8a* +7
	Gap junction	2.86×10^{-06}	*Egfr, Prkcg, Prkcb, Tubb2b, Adcy5* +8
	Viral myocarditis	1.09×10^{-05}	*Ccnd1, Cd28, Fyn, Myh13, Myh14* +3
	N-glycan biosynthesis	6.06×10^{-05}	*Man2a2, Alg2, Man1a2, St6gal1, Mgat1* +3
6	Primary immunodeficiency	6.82×10^{-06}	*Tnfrsf13c, Igll1, Lck*
	Cytokine/Cytokine receptor interaction	0.003	*Il12a, Tnfrsf13c, Cd27*
	Terpenoid backbone biosynthesis	0.016	*Idi1*
	Steroid biosynthesis	0.018	*Cyp51*
	Allograft rejection	0.039	*Il12a*

Table 4.1: KEGG Pathways enriched in the clusters

4.3.2 Enrichment for transcription factor target genes provides putative transcriptional regulators

One aim of this project was to identify regulatory networks of pDC differentiation in the steady state, and find the TFs that contribute to this regulation. To do so, I again used GeneOverlap to compare each cluster with the MSigDB `c3_tft` collection, where each gene set is composed of genes with promoter regions ([-2kb,2kb] around transcription start site) containing a conserved motif sequence, which is annotated as binding site for known or putative TFs (based on work by Xie et al. 2005[104]). The TFs matching the enriched sets therefore represent possible transcriptional regulators of the specific cluster.

The factors thus identified were visualized using the Cytoscape app EnrichmentMap[109], which generates a network of nodes (the enriched gene sets) and edges (the overlap between gene sets) to highlight connections between regulators. In addition, in post-analysis the generated network was overlaid with the enriched KEGG pathways, thus visualizing networks of regulated pathways with their putative regulators. Given that an individual TF is often represented by several redundant motifs in MSigDB, those mostly cluster together. For easier visualization, these have been merged when possible, and only the individual TFs are shown.

The data were analyzed with the EnrichmentMap default settings (threshold p value < 0.005).

Cluster 1 is enriched for 6 motifs, corresponding to only 3 different transcription factors (Gata1, Jun and Cebpb) and a motif (GGGNNTTTCC) not matching any known TF. These have no common targets, and therefore fail to form a network.

In cluster 2 (figure 4.13) 32 gene sets were found, corresponding to 21 TFs. As already observed in the previous section, genes in this cluster are typical cDC expressed genes, and this is reflected in the functional analysis: KEGG pathways related to inflammation and innate immunity are regulated by canonical myeloid transcription factors such as Cebpb and PU.1, or inflammatory TFs such as NFAT or NF-$\varkappa$B.

Interestingly, in cluster 3 (figure 4.14), where 41 significant gene sets corresponding to 10 known TFs are found, the most significant overlap is seen with

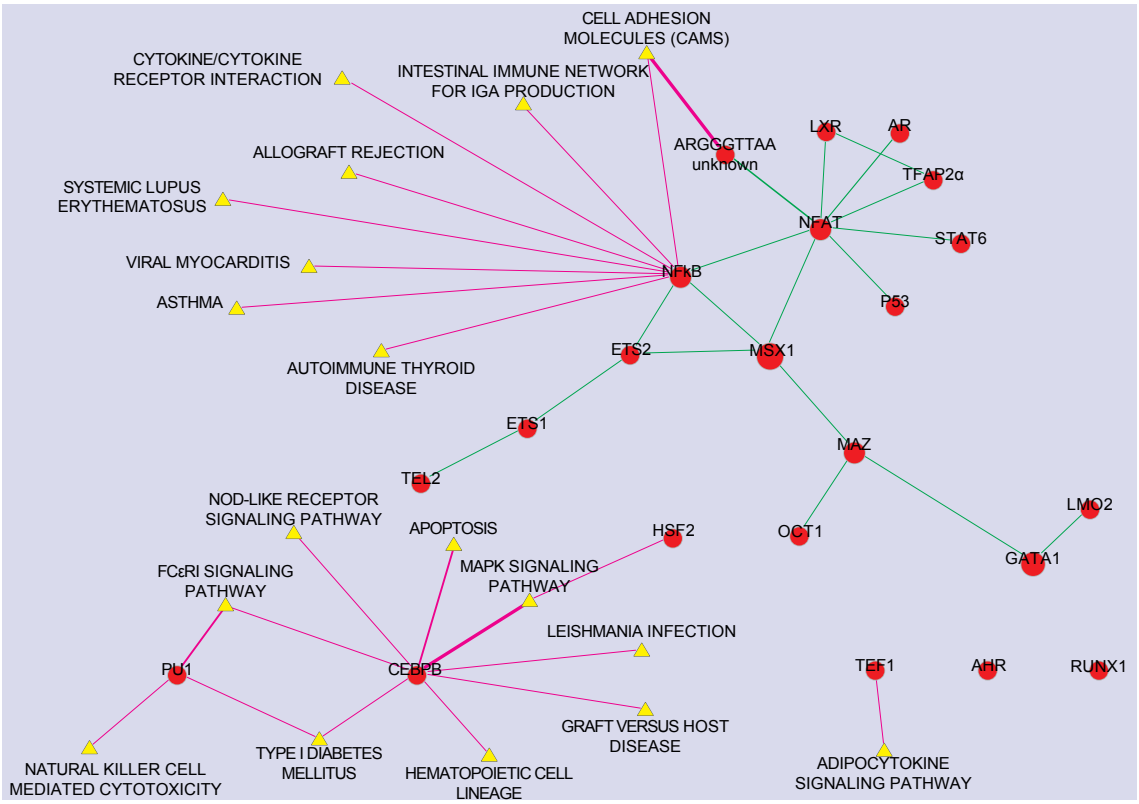


Figure 4.13: Regulatory network of cluster 2. Transcription factors and KEGG pathways significantly enriched in cluster 2. Circles represent TFs, triangles KEGG pathways (nodes). Connecting lines (edges) indicate overlap between the gene sets. Circle size is proportional to the overlap with the cluster. Thickness of the edges is proportional to the overlap between nodes. Color scale of the nodes indicates p values (blue = 0.005, red < 0.0001).

members of the E2 transcription factors family (21 gene sets), which are known regulators of DNA replication and cell cycle. Indeed, this cluster is enriched with genes involved in cell cycle and DNA replication, which are highly expressed in pre-DCs and downregulated in pDCs, but still expressed in CCR9^{low} cells, in line with the reduced proliferative capacity of mature cells compared to precursors. The TF genes E2f7 and E2f8 themselves are found in this cluster, while E2f2 is found in cluster 4.

Cluster 4 (figure 4.15) is enriched for a high number of motifs (67 gene sets), many of which are binding sites for inflammatory TFs (Stat3, Stat6, NFAT) or TFs found in the myeloid lineage (PU.1, Cebpb), that indeed control expression of genes expressed in the cDC lineage, and are here downregulated. Of interest in this cluster are the Foxo4 and Foxf2 gene sets (right insert): these TFs are members of a family of transcriptional repressors, and could play a role in specifically downregulating cDC genes to favor pDC differentiation.

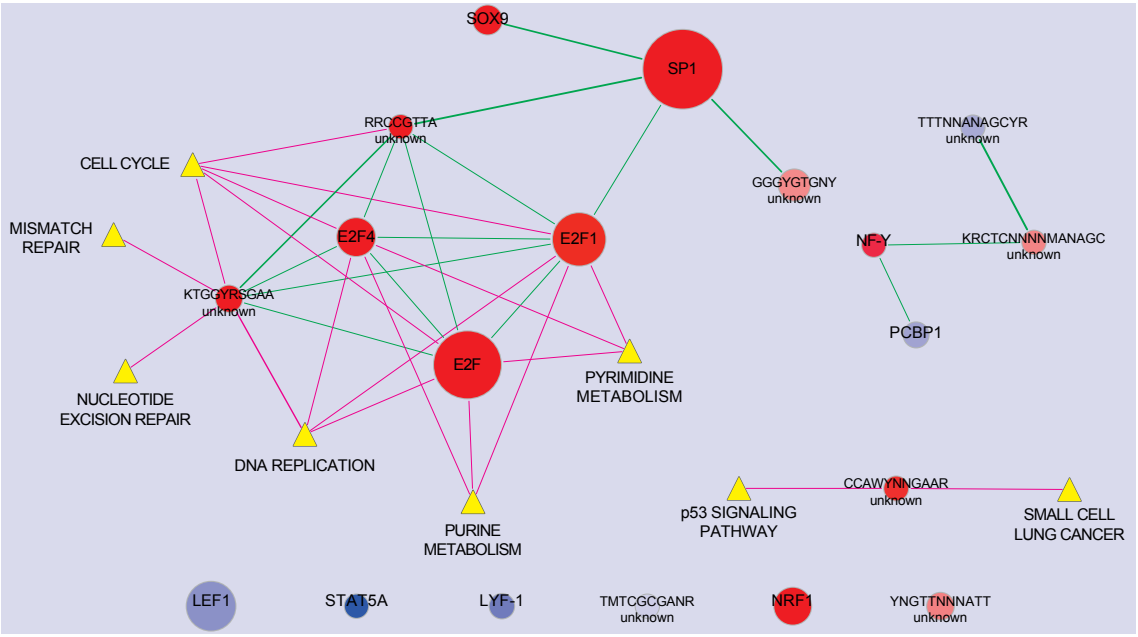


Figure 4.14: Regulatory network of cluster 3. Transcription factors and KEGG pathways significantly enriched in cluster 3. Circles represent TFs, triangles KEGG pathways (nodes). Connecting lines (edges) indicate overlap between the gene sets. Circle size is proportional to the overlap with the cluster. Thickness of the edges is proportional to the overlap between nodes. Color scale of the nodes indicates p values (blue = 0.005, red < 0.0001).

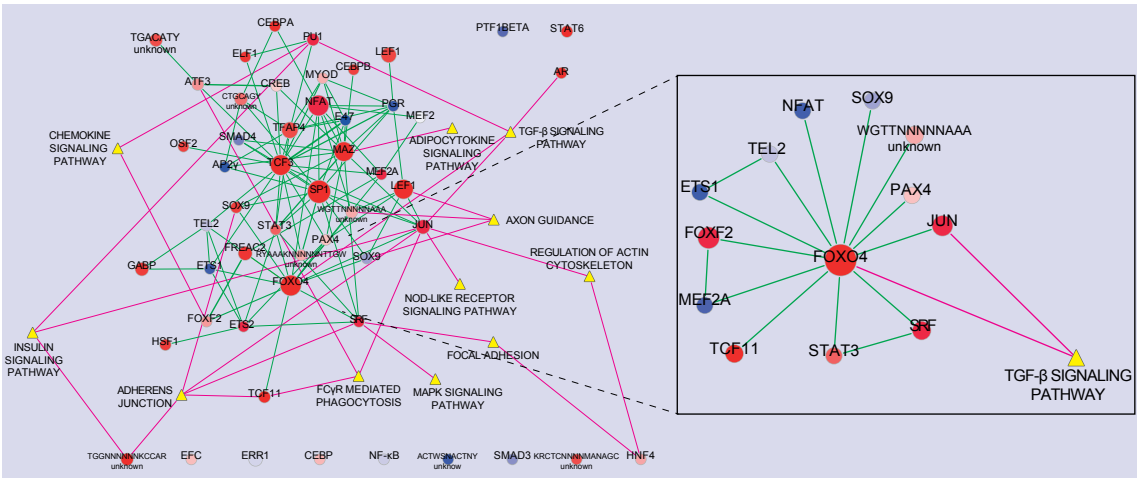


Figure 4.15: Regulatory network of cluster 4. Transcription factors and KEGG pathways significantly enriched in cluster 4. Circles represent TFs, triangles KEGG pathways (nodes). Connecting lines (edges) indicate overlap between the gene sets. Circle size is proportional to the overlap with the cluster. Thickness of the edges is proportional to the overlap between nodes. Color scale of the nodes indicates p values (blue = 0.005, red < 0.0001).

In cluster 5 (figure 4.16) 93 gene sets were enriched, with high redundancy of targets, creating a complex network of regulation and cooperation of several factors. The pDC-specific genes that are contained in this cluster (Siglech, Ccr9, Tlr7 and

others) are known to be regulated by factors such as Tcf4, Bcl11a, Spib and others. These are not represented in the collection, and therefore not enriched in the network as expected. However, other E-box binding proteins are highly enriched, such as Tcf3 and Tfap4: these are basic helix-loop-helix (bHLH) factors belonging to the same family as Tcf4, and binding motifs based on the E-box consensus sequence (CANNTG). This is indicative of an enrichment in genes that can be regulated by E-box binding proteins, such as Tcf4. Interestingly, several members of the Forkhead box (Fox) family of transcription factors are central in this network (right insert), suggesting a role for members of this family also in suppressing pDC specific genes in the pre-DCs and to a lower extent in CCR9^{low} precursors to maintain precursor state and plasticity for differentiation into cDCs.

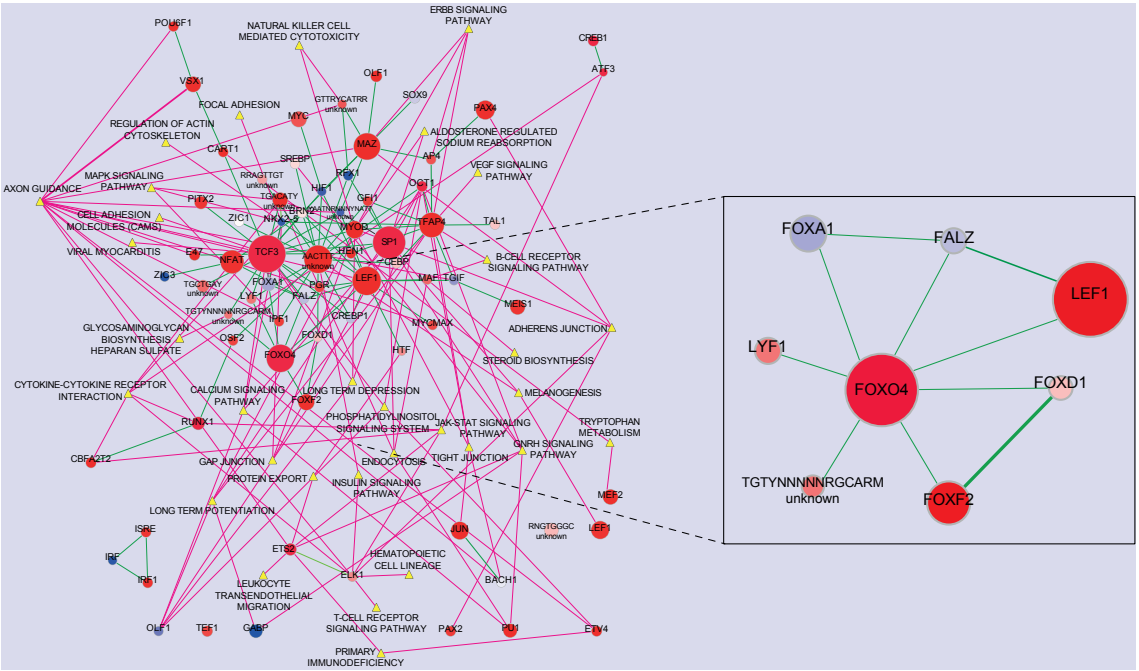


Figure 4.16: Regulatory network of cluster 5. Transcription factors and KEGG pathways significantly enriched in cluster 5. Circles represent TFs, triangles KEGG pathways (nodes). Connecting lines (edges) indicate overlap between the gene sets. Circle size is proportional to the overlap with the cluster. Thickness of the edges is proportional to the overlap between nodes. Color scale of the nodes indicates p values (blue = 0.005, red < 0.0001).

Cluster 6 is enriched for 7 gene sets corresponding to 5 known TFs (Pax4, Myc, Usf2, Myb, FXR). Similarly to cluster 1, these are not related and fail to form a network.

These results suggest a complex network of regulation of differentiation, with

reduction of cDC specific genes and proliferative capacity and induction of "plasmacytoid" pathways. The prominent enrichment for members of the Fox family of TFs was intriguing, as it is a large family with members playing different roles in the immune system. Many proteins belonging to the Fox family have been identified as regulators of T and B cell differentiation and function, for example Foxp1, which is required for the maintenance of $CD8^+$ T cell quiescence. Recent publications have also highlighted their role in DC development, for example for Foxo1[114] (which unfortunately is not directly represented in the MSigDB), but the importance in DC development has not been studied systematically.

4.4 The role of Fox transcription factors for DC differentiation

The murine family of Forkhead box (Fox) TFs consists of 43 members, most of which have not been characterized with regard to their function and biological significance. This limited knowledge is also reflected in the MSigDB, where only 8 of these TFs are represented, for which conserved binding motifs are known. These motifs, although mapping to different Fox proteins, are similar and partly overlapping, based on the core consensus sequence GTAAACA: it is therefore reasonable to hypothesize that other members of the family, sharing the Forkhead DNA binding domain, bind to the same or similar target sequences. Therefore, I explored the whole family to look for factors that are differentially expressed in pDCs, in order to find specific regulators of differentiation towards the pDC lineage.

Fox proteins which were found to be significantly upregulated with differentiation in the RNA-seq data set were chosen as candidates for validation by qPCR and other techniques.

4.4.1 Members of the Fox family of transcription factors are differentially expressed during differentiation

Out of the 43 members of the Fox family of transcription factors, 14 were found in our steady state data set with an expression value (TPM) > 0; of these, 5 were

differentially expressed during differentiation (figure 4.17).

Many targets of Fox family proteins are found in cluster 5, upregulated with differentiation, including several pDC specific genes (Tcf4, Bcl11a, Siglec-H); this suggests that some Fox proteins might actively repress pDC-lineage specific genes, to drive cDC differentiation or to maintain pluripotency/precursor potential. This might be the case for proteins such as Foxm1 and Foxn2, which are downregulated in pDCs and found in cluster 3 and cluster 4, respectively.

Within this framework, it is reasonable to speculate that Fox proteins that show high expression specifically in pDCs define this cell fate by suppressing other lineage programs (exemplified by the Fox targets found in cluster 4, including among others Id2 and Bcl6).

Therefore, I focused on those factors that are upregulated with progression to mature pDCs, for these are more likely to actively and specifically shape the plasmacytoid identity. These are Foxp1 and Foxr1, which also belong in cluster 5, with

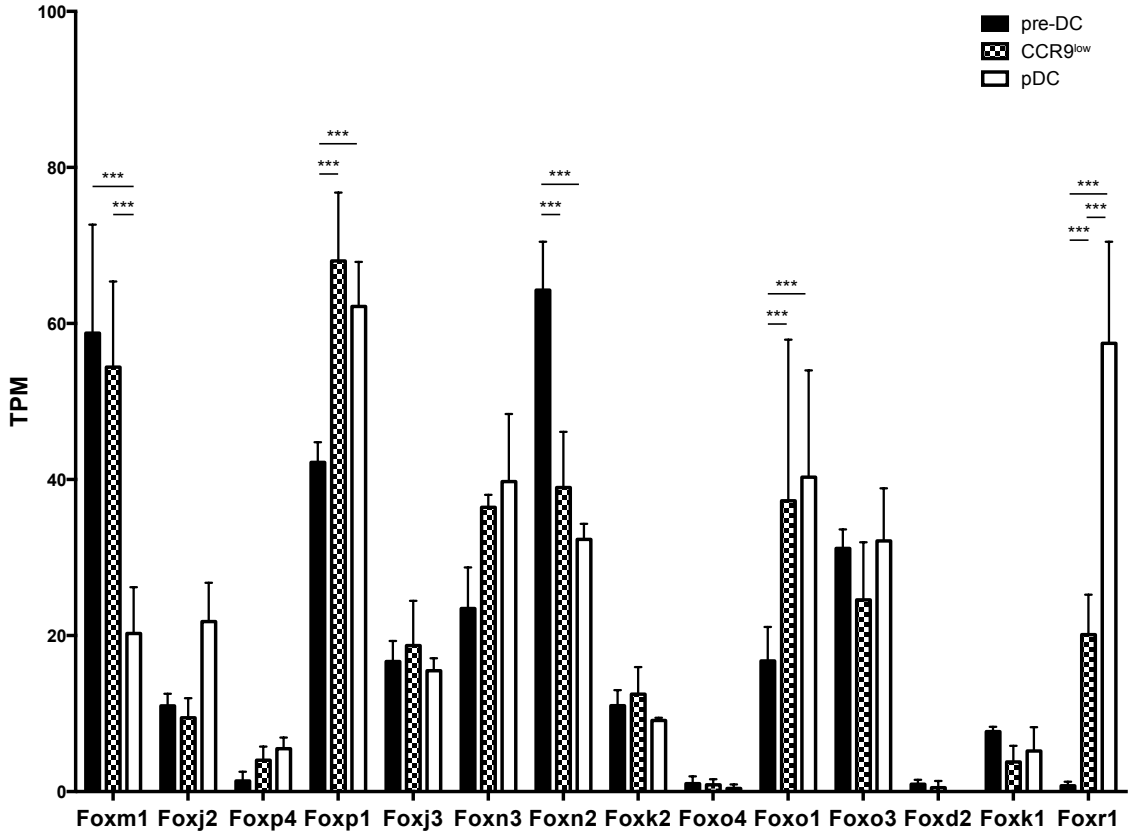


Figure 4.17: Expression of Fox proteins in the RNA-seq dataset. 14 out of 43 members of the Fox family of transcription factors are expressed in the RNA-seq data set. *** = p value < 0.005

the pDC-upregulated genes, and Foxo1, which is the only factor of this family that was shown to regulate DC function. I also included Foxo4 in the analysis, although its expression is very low and not significantly changed, as it is a central node in the statistical analysis, and its known binding motif has a highly significant overlap with both cluster 4 (p value = 5.85×10^{-07}) and cluster 5 (p value = 7.77×10^{-07}).

Foxo1 is a pro-apoptotic transcription factor controlled by the PI3K-Akt signaling pathway, which regulates several aspects of T cell biology, such as activation[115, 116], homing and survival[117] and differentiation into memory cells[118]; it is also an important factor in DC differentiation and function[114, 119].

Foxo4 is known as an anti-oncogenic factor, and has similar functions to Foxo1 in T cells[116, 117], but has not been extensively studied.

Foxp1 regulates early B cells development[113] and human plasma cells differentiation and survival[120, 121], and counteracts Foxo1 proapoptotic activity in T cells to promote survival and quiescence[122, 123]. It also plays a role in Tfh cells differentiation[124].

Foxr1 has only been identified in lymphomas and other tumors as a pro-oncogenic factor being activated following chromosomal aberrations[125, 126].

4.4.2 Fox family transcription factors are differentially expressed in pDC and cDC subpopulations

Taking advantage of the Immgen database, I looked for expression of these targets among key immune populations (figure 4.18) and the different DC subtypes (figure 4.19). As previously described, Foxo1 and Foxp1 are highly expressed in T cells and B cells, where they contribute to maintenance of the quiescent state in naive cells and control activation and function. Among DC subsets, Foxo1 and Foxp1 both showed higher expression in pDCs compared to all other DC subsets. Interestingly, Foxr1 was the most specific for pDCs, with higher expression when compared to all other DC subtypes as well as to the key populations. However, it must be noted that its expression in the database is very low in all samples, very near the background (generally considered around 100 relative units).

These data show that Foxo1, Foxp1 and possibly Foxr1 are expressed in DCs, and are higher in the plasmacytoid subset, making them optimal targets for valida-

tion as potential regulators of pDC differentiation.

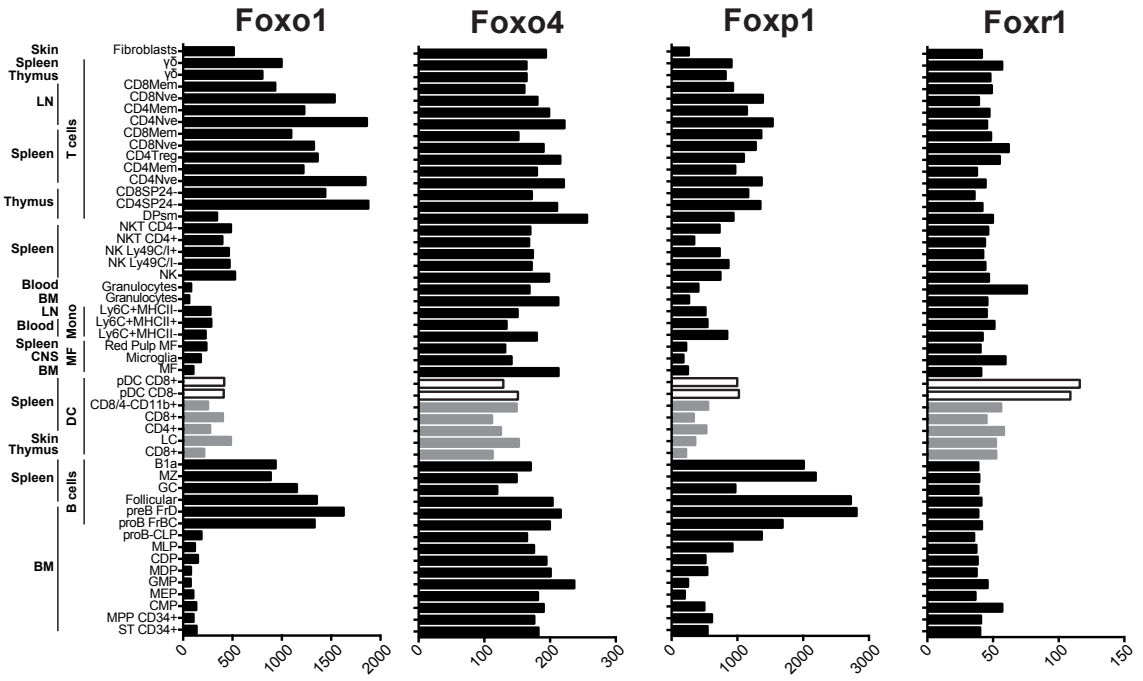


Figure 4.18: Expression of selected Fox proteins in the Immgen database (key populations). Grey bars highlight DC populations; empty bars highlight pDCs.

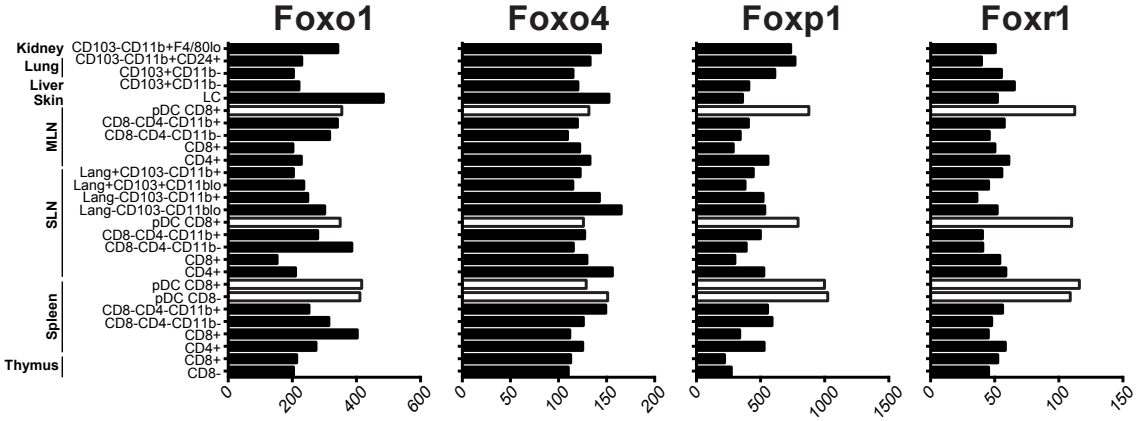


Figure 4.19: Expression of selected Fox proteins in the Immgen database (DC subsets). Empty bars highlight pDC populations.

4.4.3 qPCR analysis confirms differential expression of Foxo1 and Foxp1 in pDCs and CCR9^{low} precursors

To validate the results of the RNA-seq analysis, all three populations (pre-cDCs, CCR9^{low} precursors and pDCs) were again sorted from murine BM cells

and qPCR was performed for Foxo1, Foxo4, Foxp1 and Foxr1, using commercially available Taqman probes. As controls, Lineage positive cells were sorted from BM cells (containing T and B cells), and pDCs, cDC1, cDC2 and Lineage positive cells (comprising mostly T cells) were sorted from splenocytes.

Foxo1 was highly expressed in splenic Lin^+ cells, as expected (figure 4.20), but also relatively higher in pDCs in both organs when compared with the other cell populations. Interestingly, Foxp1 was expressed at higher levels in pDCs of both BM and spleen, compared to all other cell types, indicating a specific expression of Foxp1 mRNA in differentiated pDCs. As expected, Foxo4 showed very low expression in all samples, just above the limit of detection. Unfortunately, Foxr1 mRNA was not detectable in any sample, using two different primer/probe sets.

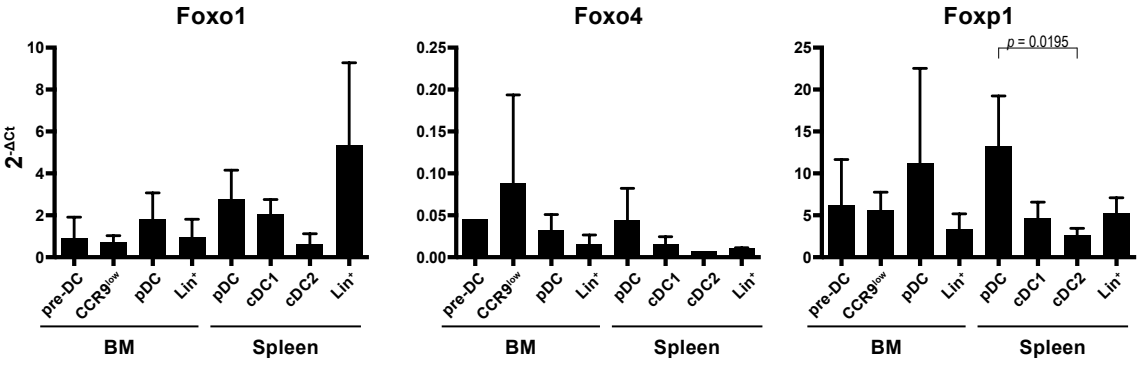


Figure 4.20: Expression of Fox genes mRNA in DC precursors and subpopulations measured by qPCR. Foxo1, Foxo4, Foxp1 and Foxr1 mRNA expression was quantified by RT-qPCR. Foxo1 and Foxp1 showed pDC specific expression when compared to other DC subsets. Foxr1 was not detected in any sample. Mean values and standard deviation of 3 independent experiments are shown.

Taken together, the RNA sequencing and qPCR results indicate a specific expression of the Foxp1 gene in mature pDCs. The expression pattern of Foxp1, together with its known function as a transcriptional repressor and cell cycle regulator, and the availability of validated antibodies and of a conditional knock-out mouse strain ($Foxp1^{flox/flox}$), allowed further phenotypical validation and functional investigation.

Foxp1 is a transcriptional repressor that acts as an oncogene in several tumors by downregulating the expression of proapoptotic genes. In the hematopoietic system, it controls survival of B cells[120, 122] by the same mechanism, and quiescence of T cells[123, 127]

4.4.4 Staining with Foxp1 antibody confirms expression in pDCs at the protein level

Using a Foxp1 specific antibody, I evaluated its expression in the spleen and BM of wild type mice. BM populations were gated as described above (figure 4.1) with the addition of a Siglec-H$^-$ MHCII$^+$ cDC population as control. Splenic DCs were gated within CD3$^-$ CD11c$^+$ cells as Siglec-H$^+$ B220^{high} pDCs and Siglec-H$^-$ B220$^-$ cDCs. T and B cells were also included, the former as CD3$^+$ cells, B cells gated as CD3$^-$ CD11c$^-$ MHCII$^+$ CD11b$^-$ and B220$^+$.

Notably, in the spleen the protein was clearly detectable in B cells, and at lower levels in T cells. In splenic pDCs Foxp1 was expressed at similarly high levels as in B cells, while it was found at significantly lower levels in cDCs (figure 4.21). cDC subsets (CD8α^+ cDC1 and CD11b$^+$ cDC2) showed no difference in Foxp1 expression, and are therefore shown together.

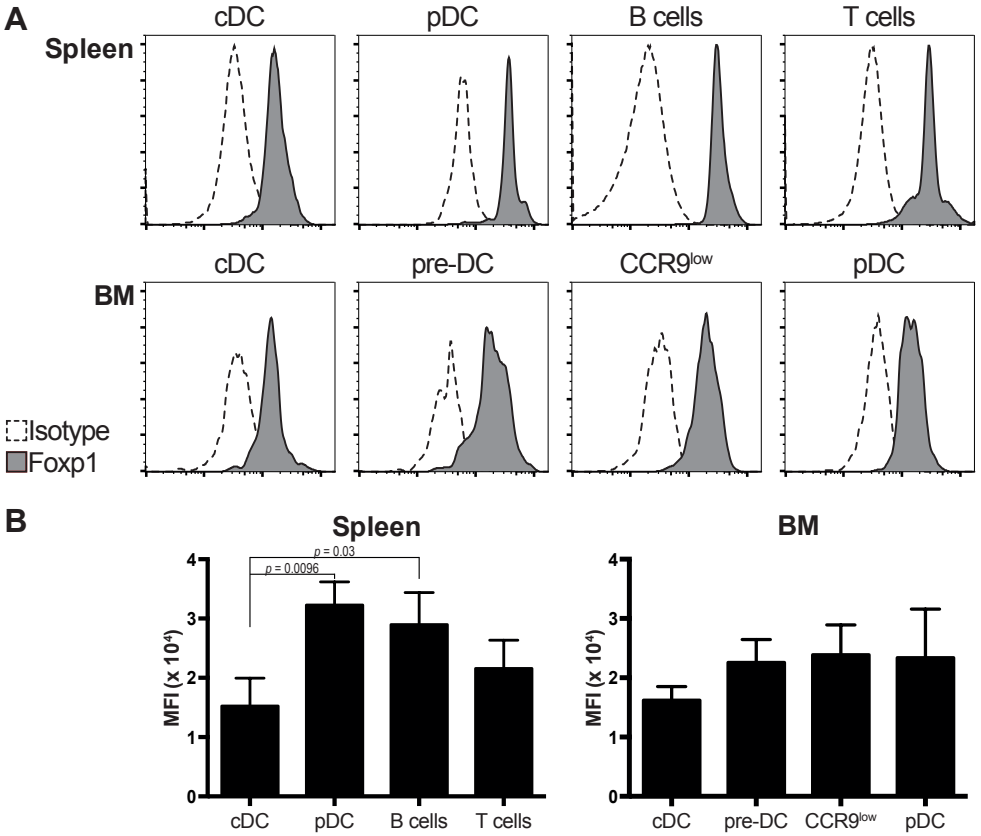


Figure 4.21: Mean Fluorescence Intensity (MFI) of Foxp1 staining in WT spleen and BM. Foxp1 was stained intracellularly with a specific antibody. **A.** exemplary histograms of the staining in spleen and BM. **B.** MFI is compared for the populations indicated. Mean and SD, n = 3 mice.

In the BM, Foxp1 protein was expressed at lower levels than in the spleen, but was still detectable in pre-DCs, CCR9^{low} precursors and pDCs at higher levels than in cDCs.

These data confirm presence of the Foxp1 protein in pDCs and precursors, while it is significantly lower in cDCs, in line with the expression pattern observed at the mRNA level.

4.5 Analysis of the response to stimuli with WGCNA

As briefly explained in section 4.3, the dataset of stimulated populations was analysed using the Weighted Gene Co-expression Network Analysis (WGCNA) algorithm[98, 99], which uses a Topological Overlap Matrix to find modules of co-expressed genes by average linkage hierarchical clustering, leading to more cohesive and biologically meaningful modules from large gene expression data sets than other clustering methods (see methods section 3.3.3 for detailed explanation of the algorithm). Module eigengenes were then used to find correlations with traits of interest (in this case Time and Treatment), and the significant modules were functionally analyzed.

4.5.1 WGCNA highlights similarities between CCR9^{low} cells and pDCs in response to TLR ligands

For each population, I first analyzed the modules significantly correlated with the traits (treatment and/or time; p value < 0.01), evaluating enrichment for KEGG functional pathways using GeneOverlap. Correlation depends on the order of the samples: with time (being naturally ordered), positive correlation indicates upregulation, negative indicates downregulation of the module's genes. Treatments are arbitrarily ordered as R848, CpG-C and CpG-A. Therefore, positive correlation indicates expression in R848 $<$ CpG-C $<$ CpG-A, and negative correlation the opposite. A correlation p value is also calculated. The genes in the significant modules are therefore differentially expressed due to the correlated variable(s), and not by random, unspecific variations in the expression.

In line with the observations made earlier by principal component analysis,

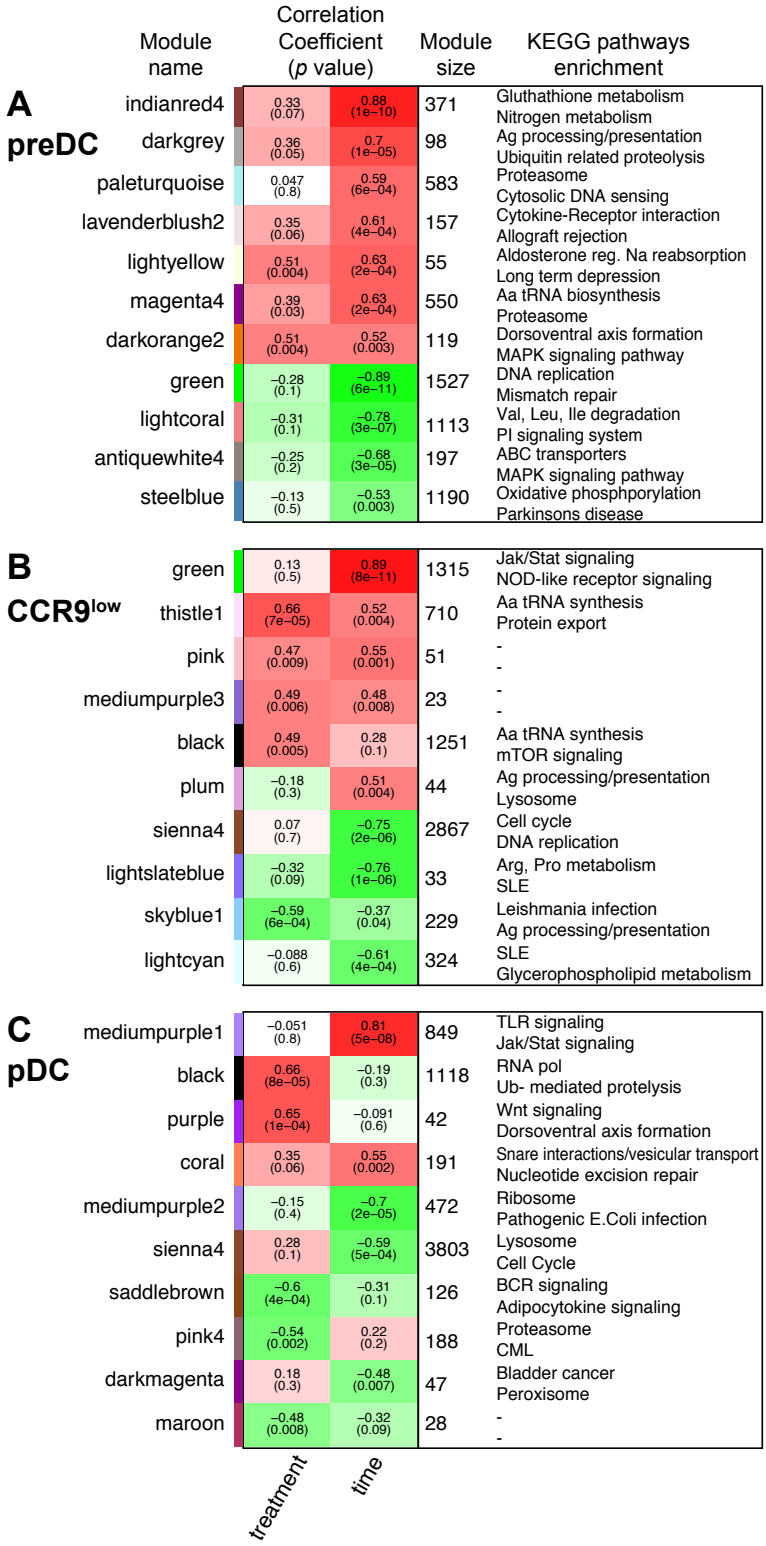


Figure 4.22: Modules significantly correlated with treatment and/or time. Module eigengenes are correlated with the traits. Treatments are arbitrarily ordered as R848, CpG-C, CpG-A. Therefore, positive correlation (red) indicates expression in R848 < CpG-C < CpG-A, negative correlation (green) R848 > CpG-C > CpG-A. Being time a naturally ordered variable, positive correlation means upregulation over time, negative correlation means downregulation. The two most significantly enriched KEGG pathways are shown for each module.

CCR9^{low} precursors and pDCs show high functional similarity (Figure 4.22B-C), with upregulation of genes involved in TLR signaling (modules CCR9^{low} `green` and pDC `mediumpurple1`) and downregulation of metabolic and cell-cycle related genes (modules CCR9^{low} `sienna4` and pDC `sienna4`) with all three stimuli. In both populations genes related to RNA synthesis show positive correlation with the treatments (modules CCR9^{low} `thistle1` and pDC `black`), suggesting that CpG treatments elicit more de novo transcription compared with R848. On the other hand, genes involved in antigen processing and proteasome (modules CCR9^{low} `skyblue1` and pDC `pink4`), as well as BCR and cytokine signaling (pDC module `saddlebrown`) were expressed at higher levels in R848 stimulated samples than in CpG-A stimulated samples, and did not change with time, suggesting a stronger effect of the TLR7 ligand on these pathways.

Pre-DCs on the other hand (figure 4.22A), while differing from the other cell types, show a differential expression pattern in line with previous data suggesting that they are indeed responsive to the given stimuli: they upregulate metabolic pathways (module `indianred4`), antigen processing and proteasomal genes (modules `darkgrey`, `paleturquoise` and `magenta4`) and inflammatory genes (modules `paleturquoise` and `lavenderblush2`). At the same time, they downregulate DNA replication and cell cycle genes (module `green`), similarly to the other cell types, suggesting also the onset of a differentiation program.

4.5.2 Modules enriched for DC specific signatures reveal stimulus-induced cell fate decisions

To better evaluate the impact that TLR stimulation has on the cell fate of each population, I calculated for each cell type which modules are enriched for the specific CDP, cDC and pDC signatures used for the exploratory analysis (Section 4.2.4). Again, the R package GeneOverlap was used to calculate enrichment with Fisher's exact test. Table 4.2 shows the numbers of modules enriched for each signature.

In pDCs (figure 4.23C), signature genes can be found in a total of 8 modules, with one containing both CDP and cDC signature genes (`coral2`) and one containing genes from all 3 signatures (`sienna4`); the only module with differential expression

		# Enriched modules			
Cell Type	# Modules	CDP	cDC	pDC	Any
pre-DC	95	4	6	4	10
CCR9^{low}	101	2	7	4	8
pDC	101	4	4	3	8

Table 4.2: Number of modules enriched for the specific signatures. "Any" indicates the total number of modules enriched for any signature (as they are not mutually exclusive).

(excluding `sienna4`, discussed later) is `cyan` (p value = 0.01, coefficient = 0.45), enriched for cDC genes, such as Cd83 and Cd86, Il1b and Zbtb46: this confirms the observations of the exploratory analysis (Section 4.2.4), where upregulation of the same genes associated with DC maturation and cDC differentiation was observed in pDCs and CCR9^{low} cells following stimulation (Figure 4.7).

In pre-DCs (figure 4.23A), pDC signature genes are found in 4 modules, 2 of which are upregulated (`magenta4` and `paleturquoise`, including mainly genes involved in RNA synthesis and the proteasome, figure 4.22A), one downregulated (`lightcoral`, including metabolic and PI signaling pathway genes, figure 4.22A) and one unchanged (`darkgreen`). In the upregulated modules several inflammatory genes (Ccr5, Ifi44, Tlr7) can be found. In the downregulated module, genes such as Cd209d (DC-SIGN, known to be downregulated after DC stimulation), Erbb3 (activator of PI3K/Akt pathway) and Cdc14b (involved in DNA replication) can be found, regulating cell adhesion and cell cycle/proliferation. At the same time, CDP and cDC signature genes are found together in modules that are downregulated (`green` and `antiquewhite4`, including mainly genes involved in DNA replication and cell cycle, figure 4.22A) or unchanged (`yellow`, `plum`), and some DC maturation genes (including Cd80, Cd83 and Cd86) are upregulated in module `lavenderblush2`. This suggests that stimulation induces a DC activation and maturation program in pre-DCs, at least in a small proportion of cells within the population, whose contribution is now visible thanks to the high sensitivity and resistance to noise of the WGCNA analysis.

In CCR9^{low} cells (figure 4.23B), genes contained in cDC and CDP signatures are downregulated (modules `lighslateblue` and `sienna4`, including mainly genes involved in the cell cycle, figure 4.22B), while pDC genes are strongly upregulated (module `green`), suggesting that indeed these cells respond strongly to TLR7 and

-9 stimuli, and are driven to differentiate into pDCs.

The two `sienna4` modules (in CCR9^{low} cells and pDCs) are very similar, large modules with thousands of genes, strongly downregulated, and containing many cell cycle, DNA replication and metabolic genes, such as Cdc25c, Ccnb2 and Cdc45 (CDP signature), E2f8, Ccna2 (cDC signature), as well as Cd209d, Il7r and Clec10a (pDC signature) in both cell types. These genes are repressed upon stimulation. Interestingly, they also include Cbfa2t3 (Mtg16), which is known to directly inhibit Id2 expression, providing a mechanism for the upregulation of Id2 observed following stimulation.

Functional analysis and TF binding sites enrichment performed on selected modules confirms these observations (see Appendix B for figures and detailed explanation): modules downregulated in pre-DCs are enriched for DNA-replication and cell cycle pathways, and binding sites for E2 family members (module `green` figure B.2). Interestingly, the module `paleturquoise`, which is upregulated, is enriched for binding sites for many Interferon regulatory factors (Irfs), centered on Irf8 and regulating pathways of cytosolic DNA sensing and RLR signaling, suggesting a specific Type I IFN pathway activation following TLR7 and 9 stimulation (figure B.1B).

In CCR9^{low} cells, regulation of upregulated modules (`green`, figure B.4A) is also dependent on a network of Irfs and inflammatory pathways, confirming the specific activation. Downregulation of the `sienna4` module genes is dependent on a complex network of TFs, among which interestingly are also members of the Fox family (Foxo4 and Foxf2, figure B.3, blue arrowheads), suggesting the activation of a differentiation program.

The pDC module `sienna4` also shows a similarly complex network as the CCR9^{low} precursors one, with many common regulators, but interestingly does not contain targets for Fox family members (figure B.5), suggesting a final commitment not influenced by stimulation.

In conclusion, these data show that the stimulation with TLR7 and -9 ligands elicits a specific response in all 3 cell types, or at least in subsets of those, and strongly suggest that TLR7 and -9 stimulation promotes differentiation of CCR9^{low} precursors towards a pDC fate.

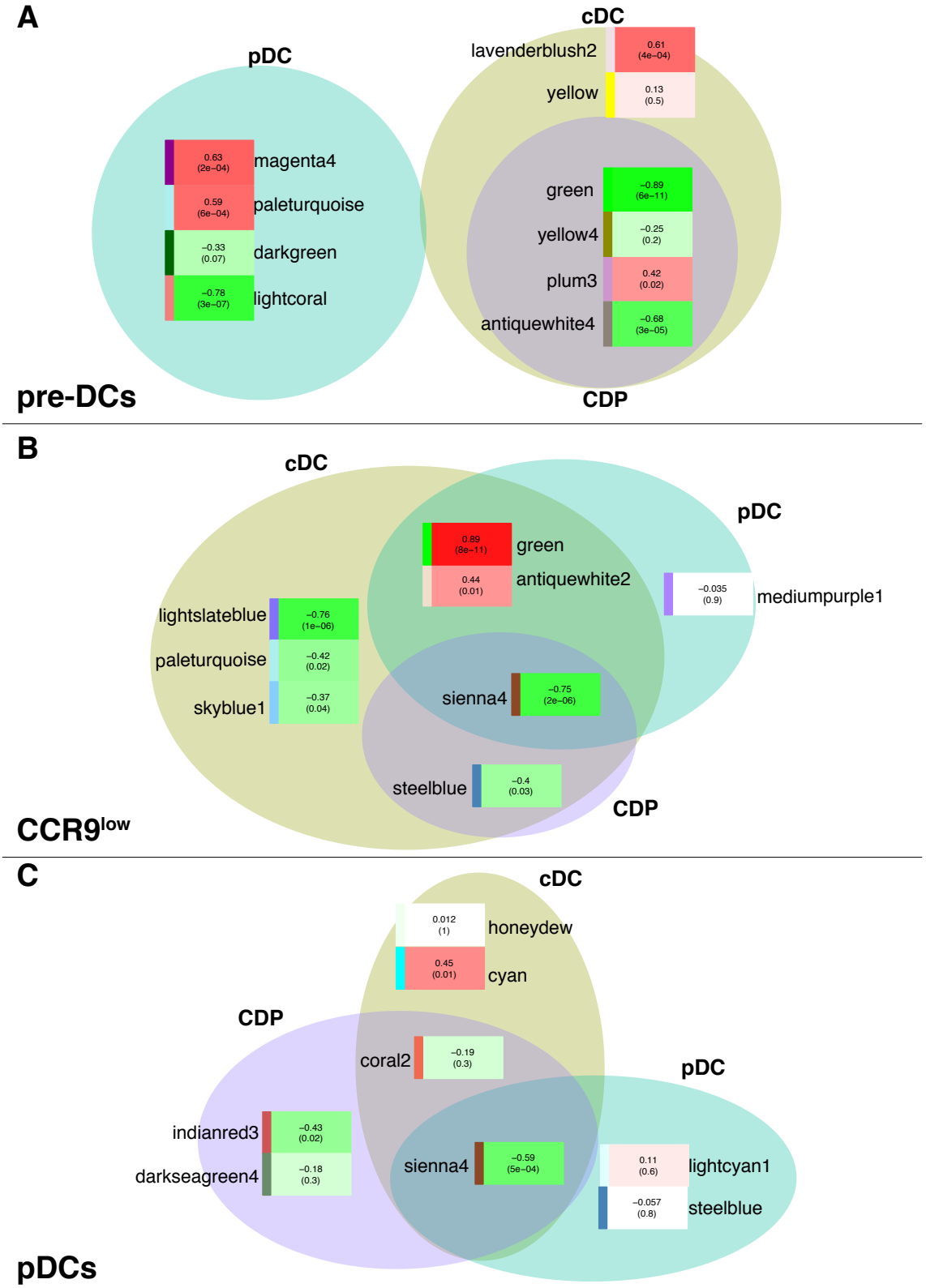


Figure 4.23: Signatures-modules overlap. Modules that are significantly enriched for the cell-type specific signatures are displayed. Circles indicate the signatures (purple = CDP; beige = cDC; blue = pDC) and contain the enriched modules, squares indicate correlation of each module with time [correlation factor and (p value)].

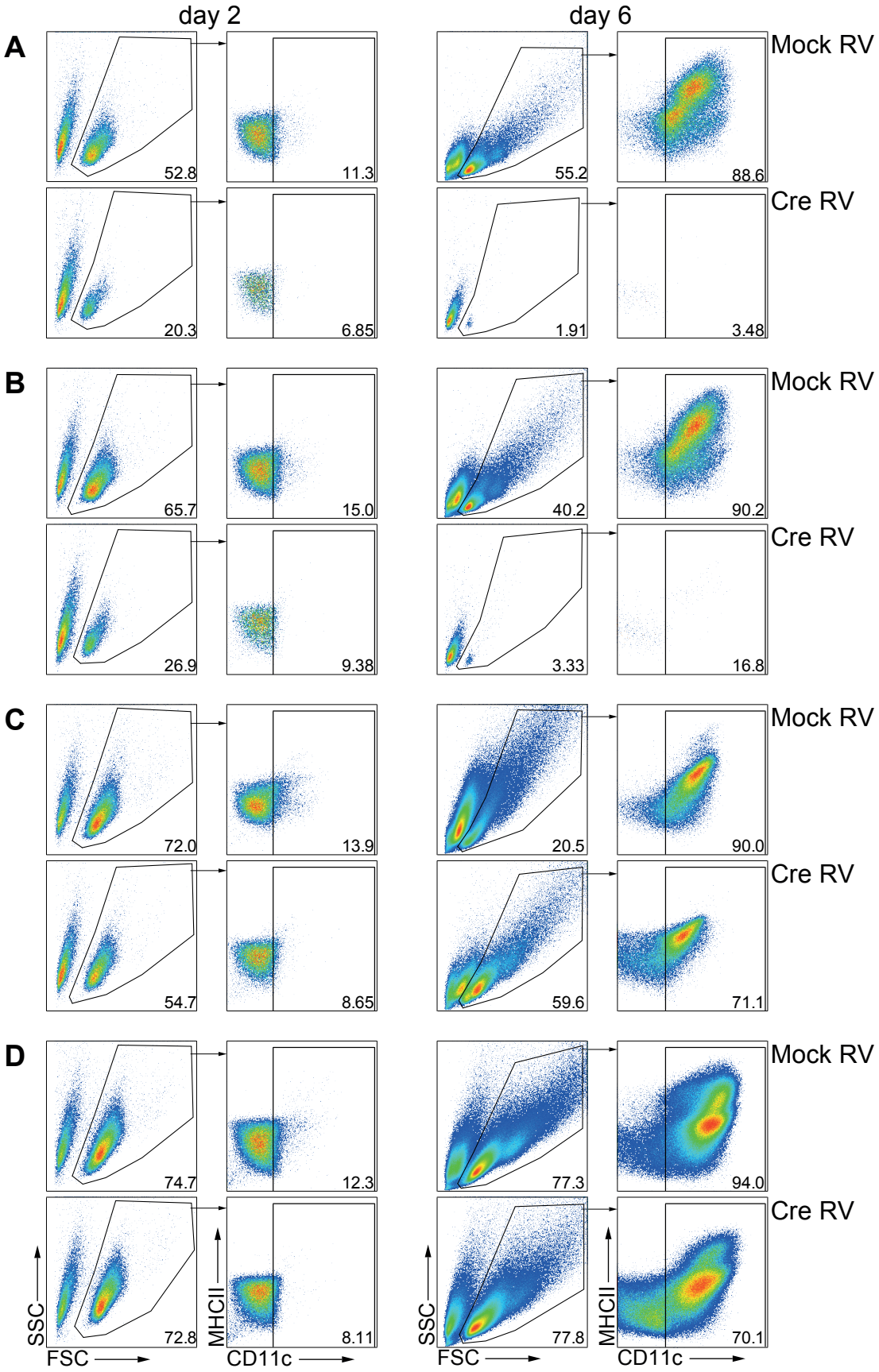
4.6 Foxp1 deficiency influences DC development from progenitor cells

Given the prominent role of Fox family members in pDC differentiation highlighted in the steady state analysis, and observed as well in the activation-induced differentiation of CCR9^{low} precursors, I wanted to evaluate the functional relevance of Foxp1 (the most prominent target, validated at the protein level). To assess this, I generated progenitor cell lines from Foxp1$^{flox/flox}$ BM cells, using the Hoxb8 system[94]. The cell lines thus generated are self renewing myeloid progenitors that are held in an undifferentiated state while in the presence of β-estradiol and Flt3L, and can be differentiated into different DC subsets upon withdrawal of β-estradiol, using different growth factors.

I generated Foxp1-KO progenitors by transducing the Foxp1$^{flox/flox}$ line with a Cre-encoding retrovirus. The progenitor cell lines generated were viable and showed no difference compared to the parent cell line or the cell line which was transduced with a control retrovirus, not encoding Cre recombinase.

The cell lines were differentiated in the presence of Flt3L only, in standard concentration (7% of supernatant from a Flt3L-producing CHO cell line, roughly equivalent to 70ng/ml of recombinant Flt3L) or high concentration (15%), or with the addition of 50ng/ml recombinant M-CSF or 1% GM-CSF-containing supernatant to 7% Flt3L.

Interestingly, in the presence of Flt3L alone, Foxp1-KO cells failed to survive and neither expanded nor upregulated CD11c expression, independently of the Flt3L concentration (Figures 4.24 and 4.25, panels A and B), whereas the parental cell line and the cell line transduced with control retrovirus strongly expanded (10-fold after 6 days of culture) and maintained survival until at least day 6 of the culture. The addition of M-CSF or GM-CSF was able to rescue the survival of Foxp1-KO cells (panels C and D), and their differentiation into MHCII^{low} DCs, to similar extents compared to the Foxp1-sufficient cell lines. In figure 4.26, it is clear that at day 6 after estrogen withdrawal most Foxp1-KO cells fail to survive in the presence of Flt3L only, dying at very early stages, before upregulation of CD11c. Interestingly, even in the presence of Flt3L and GM-CSF, when Foxp1-KO cells expanded and survived


day 2
day 6
A
B
C
D
Mock RV
Cre RV
52.8
11.3
55.2
88.6
20.3
6.85
1.91
3.48
65.7
15.0
40.2
90.2
26.9
9.38
3.33
16.8
72.0
13.9
20.5
90.0
54.7
8.65
59.6
71.1
74.7
12.3
77.3
94.0
72.8
8.11
77.8
70.1
SSC
FSC
MHCII
CD11c

normally, upregulation of CD11c expression was delayed (91.5% [SD 3.6%]in Ctr-RV versus 59.9% [SD 11.2%] in KO on day 6).

These results suggest a pivotal role for Foxp1 in the Flt3 signaling axis, specifically controlling survival and proliferation of Flt3L-depending cells, and their differentiation into DCs.

Figure 4.24 *(preceding page)*: Foxp1-Hoxb8 cell lines differentiation (Part 1). Foxp1-Hoxb8 cell lines were differentiated for 10 days in the presence of standard (7%) or high (15%) concentrations of Flt3L only (**A-B**), 7% Flt3L + M-CSF (**C**) or 7% Flt3L + GM-CSF (**D**). Representative plots of day 2 and day 6. In each panel, only the lines transduced with Mock-RV (top) and Cre-RV (Foxp1-KO, bottom) are shown. With Flt3L only, Foxp1-KO cells are clearly declining already at day 2 of differentiation. Exemplary results of one out of three experiments.

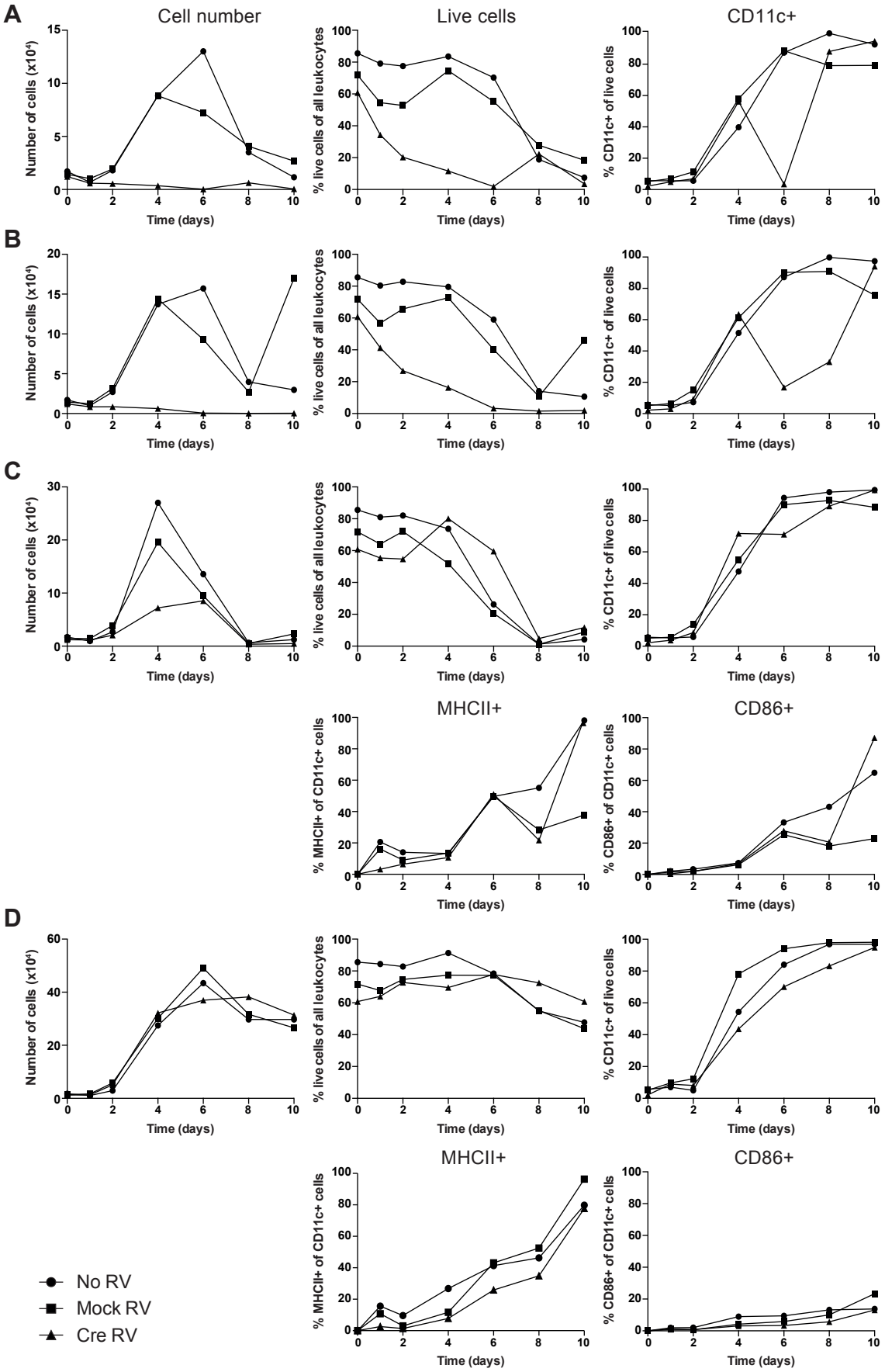

A
Cell number
Live cells
CD11c+
B
C
MHCII+
CD86+
D
Number of cells (x10⁴)
% live cells of all leukocytes
% CD11c+ of live cells
% MHCII+ of CD11c+ cells
% CD86+ of CD11c+ cells
Time (days)
No RV
Mock RV
Cre RV

Figure 4.25 *(preceding page)*: Foxp1-Hoxb8 cell lines differentiation (Part 2). Panels represent differentiation conditions (**A.** 7% Flt3L, **B.** 15% Flt3L, **C.** Flt3L + M-CSF, **D.** Flt3L + GM-CSF). Graphs indicate the total cell counts, percentage of live cells and percentage of CD11c$^+$ cells. For treatments with M-CSF and GM-CSF, where enough cells can be evaluated, also percentage of MHCII$^+$ and MHCII$^+$ CD86$^+$ (within the CD11c$^+$) are shown. Exemplary results of one out of three experiments.

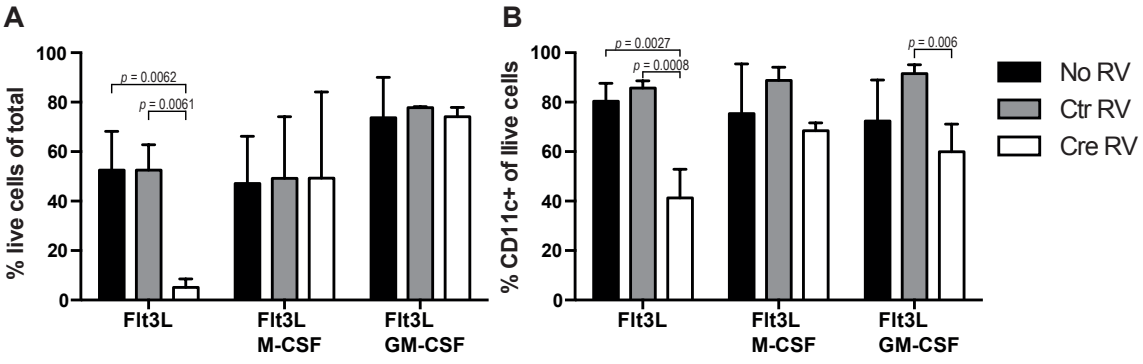


Figure 4.26: Foxp1-Hoxb8 cell lines survival and CD11c$^+$ output at day 6 of differentiation. Percentage of live cells of total lymphocytes (**A**) and CD11c positive cells of live cells (**B**) 6 days after β-estrogen removal. Mean values and standard deviation are shown for 3 independent experiments. Statistical analysis performed with two-way ANOVA and Tukey's multiple comparisons test.

5 Discussion

The differentiation of DC subpopulations from progenitors in the BM has been object of studies for decades, and several aspects and mechanisms have been clarified by numerous pivotal studies. Nonetheless, many details regarding the regulation of the process are still elusive, especially about the functionally distinct pDCs.

In this thesis project, I used RNA sequencing and computational methods to analyze the global gene expression pattern of pDCs and different precursor populations in the murine BM, and the influence that TLR stimulation, mimicking systemic viral infection, has on the transcriptional program of the different populations.

This approach allowed me to assess pDC differentiation at the transcriptional level, and to define the changes that occur within progressively lineage-restricted populations. Clusters of co-regulated genes offered a view on the functional changes of differentiation, as well as the basis for the identification of regulatory factors.

Indeed, TF networks could be defined in the unstimulated populations that are with high probability regulators of the differentiation process. Within these large and complex networks, I could identify a specific transcription factor, Foxp1, that is highly likely to have a central role in regulating steady state differentiation.

In addition, analysis of the effects of TLR7 and 9 stimulation on precursor cells highlighted the functional differences between mature pDCs and different precursor populations, suggesting different levels of cell fate commitment as well as responsiveness to stimuli. The data also showed that precursor cell activation could go hand in hand with differentiation into mature pDCs.

5.1 pDC progenitors in the murine bone marrow

Previous studies have identified several different steps of DC differentiation in the healthy mammalian BM. Early CD11c^{+} MHCII^{-} precursors were identified in the murine BM, which give rise *in vitro* to all DC subsets including pDCs[49]. More recently, the CD11c^{+} pre-DC population was characterized at the single cell level, and a subpopulation of Siglec-H^{-} migrating pre-cDCs was identified, as a heterogeneous collection of cells showing early commitment to the different cDC subsets[52]. A Siglec-H^{+} B220^{-} Ly6C^{-} subpopulation was also described, able to generate all DC subtypes including pDCs, which could therefore be an earlier precursor. Recent research in our lab highlighted heterogeneity and commitment of very early progenitors, the common DC progenitors, by means of single cell continuous imaging and tracking *in vitro*, consistent with the model that pDC commitment can occur at different levels before or after the CDP stage[86].

I designed a panel of surface markers that allows to clearly discriminate several different steps of pDC differentiation within the CD11c^{+} MHCII$^{-/low}$ precursor compartment in the murine BM, downstream of the CDP stage. By focusing the analysis on cells which are negative for lineage markers (CD3, CD19, NK1.1 and Ly6G) and express CD135 (Flt3) and CD11c, precursors of cDCs, pDCs and differentiated pDCs could be discriminated by the expression pattern of Siglec-H, B220 and CCR9.

Mature pDCs express high levels of all three markers, and are the most abundant population, amounting to approximately 2% of the total BM cells. Lower expression of the chemokine receptor CCR9 was previously found to characterize a population of pDC-like cells that are already able to produce vast amounts of type I interferon in response to TLR7 and -9 stimulation, but retain plasticity to convert to the cDC lineage in specific conditions (e.g. tissue microenvironment)[53]. Analysis of this population showed that CCR9 expression directly correlates with expression of B220, resulting in a continuous transition from CCR9^{low} B220^{low} cells to CCR9^{high} B220^{high} mature pDCs, through intermediate, overlapping steps of CCR9^{low} B220$^{int/high}$ cells. The B220 expressing cells had been excluded from prior analyses of pre-DCs[32, 53]. The whole CCR9^{low} population, containing both B220^{low} and B220$^{int/high}$ cells, was

used in the present analysis. These cell make up approximately 0.5% of the total BM.

Very little is known about commitment to the pDC lineage at earlier stages. Recent studies have shown that a population of CD11c^{+} Siglec-H^{-} B220^{-} precursors is made up of cells committed to either of the cDC subpopulations, but not to pDCs[52]. Given the plasticity of the system, and the observation that CCR9^{low} cells still retain cDC potential, I included this pre-DC population in the analysis, with the aim to evaluate not only its possible, albeit limited, contribution to the pDC pool in the steady state, but also the potential of influencing their commitment by means of direct TLR stimulation. These cells are extremely rare in normal BM, amounting to no more than 0.1% of the total cells.

In the study by Schlitzer et al.[52] the CD11c^{+} MHCII^{-} Siglec-H^{+} B220^{-} BM fraction contained Ly6C^{+} cells which give rise to both cDC subsets, and a fraction of Ly6C^{-} cells which give rise to pDCs and both cDC subsets. This less committed population is partially overlapping with the Siglec-H^{+} CCR9^{low} precursor population studied here, but lacks expression of B220. By inclusion of the cells with progressively increased B220 expression the full spectrum of Siglec-H^{+} precursors was included in the analysis.

5.2 Analysis of the RNA sequencing data

The large amount of data generated by mRNA sequencing allows for very broad and in-depth analyses to be performed, while also posing challenges in terms of statistical significance and interpretation of the results, especially for complex experimental designs with different combinations of conditions and treatments.

Several tools and softwares are available today for analyzing RNA-seq data[128], allowing a wide variety of information to be extracted. It was of paramount importance therefore, to choose the best strategy in order to obtain statistically relevant results that are meaningful and resistant to the high level of noise inherent in these large data sets. To better answer the questions of this project, the analysis was divided in 3 parts: (a) an exploratory analysis, a form of data mining that uses previously known information to categorize and characterize the RNA-seq data, (b)

a statistical analysis on the unstimulated data set, to define differentially expressed genes and compare the different populations based on statistically significant changes in the transcriptome, and (c) a statistical analysis on the stimulated data set, to define differential expression and statistically analyze the effects of stimulation on the different cell types, and compare the results.

The implementation of this analysis workflow allowed the characterization of mature pDCs and CCR9^{low} precursors as closely related, but still partially distinct populations, separated from pre-DCs both phenotypically and functionally. The results of the statistical analysis allowed the definition of a model of pDC differentiation based on a complex network of interconnecting factors, among which a pivotal role is played by the transcriptional repressor Foxp1. Comparison of the responses to stimuli then confirmed the functional differences suggested by the exploratory analysis, but also highlighted a response-driven differentiation in CCR9^{low} precursors, as well as a limited, but significant specific response of pre-DCs.

5.3 Exploratory analysis defines BM populations and responses to TLR ligands

Exploratory analysis was used to evaluate the data in their entirety, without selection for differential expression: this type of data mining is based on previous knowledge and independently generated data, such as specific pathways of interest (e.g. from the KEGG repository) and cell type-specific gene signatures (e.g. those generated for pDCs and cDCs by Miller et al.[113]). These can be used as descriptive measures, to evaluate the intrinsic characteristics of each sample relative to cell types of interest, as well as to evaluate responsiveness and changes in lineage features when TLR stimulation is applied.

The stimuli used for this experiment were chosen in order to induce strong type I IFN responses, a typical feature of pDCs. Analyzing the expression of genes involved in this pathway across all samples clearly showed that most transcripts were readily upregulated upon stimulation both in mature pDCs and in CCR9^{low} precursors, confirming that TLR responsiveness is a feature already fully developed in precursors of pDCs[32], whereas antigen presentation capacity is confined to differ-

entiated DCs[129]. The contribution of CCR9^{low} precursors to the IFN-I response may enhance the antiviral defense both in the BM and the periphery: Runx2 expression (necessary for pDC egression from the BM[130, 131]) is maintained in these precursors following activation, while it is downregulated in mature pDCs, suggesting a higher migratory capacity that can potentiate systemic antiviral responses, as well as replenish the peripheral pDC pool. This mechanism might also influence hematopoiesis and contribute to the type I IFN-mediated emergency myelopiesis observed with chronic TLR7 signaling[132] as well as the type I IFN-Flt3 axis augmenting pDC differentiation from CLPs[54].

On the other hand, pre-DCs showed only limited changes in expression of type I IFN pathway genes, consistent with their commitment and similarity to cDCs, which are less responsive to viral challenges. Nonetheless, a limited increase in some IFN-I induced genes (Relb, Myd88) at later time points could be observed, which could be due to non-specific IFN production induced by culture conditions. However, it cannot be excluded that a small proportion of TLR7 and -9 ligand-responsive cells exist within this population, whose limited contribution to IFN-I production is masked in the bulk RNA-seq data, but sufficient to cause limited activation of inflammatory response genes.

Evaluation of the expression of Tlr genes confirms the inherent unresponsiveness of pre-DCs as a population, given that Tlr7 and 9 genes are expressed at very low levels, and in general the pattern of Tlr expression does not change upon stimulation. CCR9^{low} precursors express a similar repertoire of Tlrs compared to mature pDCs, and at similar levels. Moreover, both populations readily upregulate the receptors upon TLR7 or 9 activation, in line with a specific response.

The results of principal component analysis interestingly showed clearly that CCR9^{low} cells are a population of precursors with intermediate characteristics of mature pDCs and pre-DCs in the steady state. Stimulated, they almost fully overlap mature cells, with sequential changes increasing with time, further confirming previous observations of functional similarity[32, 53]. Pre-DCs are clearly different cells, as expected. However, in PCA unresponsiveness would correspond to a clustering of all samples (unstimulated and stimulated) in a restricted area, or scattering in unspecific directions in case of non-specific activation due to culture conditions.

Interestingly, this population also showed sequential changes with all 3 stimuli, consistent with that of responsive populations, albeit of reduced magnitude. This could be interpreted as a limited but specific response to TLR7 and -9 stimuli.

Indeed, in the study previously cited that established the commitment of pre-DCs to cDC subsets at the single cell level[52], a very small percentage of pDCs could be detected after differentiating Siglec-H$^-$ precursors. These results suggest the presence of a small fraction of cells with surface characteristics of committed pre-cDCs, but possessing functional characteristics of the pDC lineage. It must be noted that the cell sorting technology used here, while being very precise and reliable (quality check after sorting always showing more than 95% purity), leaves a margin for few contaminating cells to be included in the sorted populations. Only single cell RNA sequencing analysis could resolve whether cells expressing the pre-cDC gene signature contain cells which express TLR7 and TLR9, or if few CCR9^{low} precursors or pDCs were included in the sorted pre-cDC population.

To better define these populations in terms of DC lineage or precursor similarities, and evaluate wether stimulation has an effect on cell identity, I used independently generated gene signatures (i.e. sets of genes specifically expressed in one cell type and not others) of the CDP and pDC[32] and cDC lineages[113].

In this analysis, pDCs were clearly identified by the expression of the pDC signature only, while pre-DCs expressed CDP and cDC genes, in line with their precursor state and commitment. Most striking in this analysis was the expression pattern of CCR9^{low} cells: in the steady state they expressed most of the pDC signature genes, although to a lower extent than the mature cells. But they also expressed most CDP genes, to levels comparable to those of pre-DCs, indicating an intermediate phenotype, in line with the observations made with PCA. In addition, they also showed expression of a small subset of cDC specific genes, which are a confirmation of their plasticity to become cDCs in specific conditions[32, 53].

After TLR7 or -9 stimulation, pre-DCs showed very little changes in gene expression, with limited downregulation of CDP and cDC signature genes and no significant upregulation of any interesting feature, consistent with their general unresponsiveness to these TLR ligands, previously discussed. pDCs on the other hand, responded to stimuli by upregulating the type I IFN pathway, described earlier, and

appeared to slightly downregulate many of their signature genes (Siglech, Ptprc [B220], Tcf4, Spib and others). This is consistent with published results showing downregulation of Tcf4 and SpiB mRNA expression in pDCs derived from Flt3L BM culture after stimulation with CpG-B for 48 hours and in the human pDC cell line CAL-1[81]. No major alterations were observed in their signature, consistent with the terminal differentiation of these cells. CCR9^{low} precursors were again the most striking: in response to stimuli, they strongly downregulated their CDP signature and the cDC genes expressed in the steady state, and in general acquired an expression pattern highly similar to that of pDCs. These results suggest that TLR7 or 9 stimulation might act on these precursor cells as a signal to complete maturation. TLR activation-induced DC differentiation of earlier progenitors was previously observed in models of systemic infection and chronic TLR stimulation[88, 90, 132, 133]. However, these *in vivo* studies describe a secondary effect of systemic type I IFN signaling rather than direct TLR-induced differentiation. The RNA-seq data suggest a direct effect on CCR9^{low} precursors, given that changes in signature genes could be observed as early as 2 hours after stimulation.

Another interesting observation of this analysis is that upon activation both pDCs and CCR9^{low} precursors show significant upregulation of a small subset of cDC signature genes: among these are canonical DC activation genes, such as CD83 and CD86 (normally expressed at higher level in steady state cDCs), but also the transcription factors Id2, Nfil3 and Bcl6, which are master determinants of cDC identity[79, 80, 134].

Consistent with previous observations that Tcf4 (E2-2) and its targets are downregulated in pDCs following activation[81] mentioned above, it also reflects the observations of Ghosh et al.[82] in pDCs with conditional deletion of E2-2, where Id2 and other cDC genes were upregulated. An important speculation of this work is that this upregulation might be a consequence of the reduced expression of targets of E2-2 that are transcriptional repressors, such as Bcl11A. Therefore, the upregulation of Id2 mRNA observed here could be the consequence of downregulation of factors such as Zeb2 and Mtg16, which are known to directly inhibit its expression[83, 85]. Indeed, the Mtg16 gene (Cbfa2t3) is significantly downregulated following stimulation (by WGCNA analysis). This observation could be validated using an Id2-eGFP

reporter mouse[91], confirming that upregulation of Id2 happens at the protein level both in pDCs and CCR9^{low} precursors, upon TLR7 and 9 stimulation. Further research is required to dissect the functional relevance of this interesting finding.

Taken together, the results of exploratory analysis show that pre-DCs, CCR9^{low} precursors and pDCs are distinct populations, and that CCR9^{low} precursors are an intermediate stage of pDC differentiation. An alternative explanation for the gene expression pattern found in CCR9^{low} cells is that they are a heterogeneous population containing different subsets of precursors, which are committed to pDCs or cDC subpopulations. Functionally, pre-DCs are clearly separated from the other two populations, which are highly overlapping, but they might also harbor a small subpopulation of TLR7 and -9 responsive cells. The core signatures of pDCs and pre-DCs are not significantly changed by stimulation, suggesting a terminal differentiation of CCR9^{high} pDCs, with changes in function but not identity, and a stable commitment for pre-DCs, which are not influenced by TLR7 or 9 stimulation (at the population level). CCR9^{low} precursors on the other hand not only show significant changes in functional pathways, consistent with high responsiveness to TLR stimuli, but also rapidly acquire a fully differentiated pDC phenotype, suggesting a stimulus-induced differentiation. This phenomenon, called emergency myelopoiesis, has been observed in hematopoietic stem and progenitor cells (HSPCs) with different types of infections[87]. This could be a mechanism necessary to maintain cell balance and immune protection by quickly replacing the cells lost to activation-induced cell death. Finally, TLR7 and 9 stimulation induces cDC-specific proteins, such as Id2, to be expressed in pDCs and pDC-committed precursors.

5.4 Steady state differentiation: differential gene expression and regulatory network

To address the second aim of this project, the definition of the regulatory network of pDC differentiation, I performed statistical analysis on the RNA-seq data from the steady state samples.

The more than 2000 differentially expressed genes could be hierarchically clustered into 6 clusters of co-expression, with different patterns of up- or down-regulation.

The functional analysis of each cluster revealed the distinguishing features between the 3 cell types. Mature pDCs downregulate genes involved in cell proliferation and cell cycle, that are expressed in both precursor cell types, in line with the lower proliferative capacity of the differentiated cells. On the other hand, pDC specific genes, including their canonical markers (Siglech, Ccr9, Tcf4 and others) cluster together with numerous cell adhesion molecules and genes necessary for antigen processing, and are strongly upregulated in CCR9^{low} cells compared to pre-DCs, and are obviously even higher in pDCs.

To define the regulatory factors of each module, analysis of enrichment for transcription factor binding sites was performed, taking advantage of the MSigDB[103], a large publicly available database (similarly to the work by Pandey et al., 2013[101]). Within this database, a large, curated collection of TF targets gene sets[104] can be used for enrichment analysis. This enrichment analysis is a powerful bioinformatics tool to visualize possible regulatory networks and formulate hypothesis about the factors that are involved. Genes in each MSigDB gene set are validated for containing a specific motif sequence in their promoter region (4kb around transcription start site). When interpreting the results of the TF binding site enrichment analysis we need to consider that: (i) shorter, highly conserved motifs might be over-represented; (ii) TFs annotated with multiple binding sites might also be over-represented; (iii) co-factors, additional sequences and distant enhancers are not considered. Therefore, the collection has an inherent bias for TFs with multiple binding sequences, as well as with short and highly conserved motifs. In addition, some sequences do not match annotation for known TFs (they are validated only as binding sites), and many known TFs still have unknown or not conserved DNA binding sites.

Analysis of enrichment for transcription factor binding sites revealed complex networks of regulation: targets for TFs of the E2-family (important regulators of cell cycle and DNA replication) are enriched in clusters downregulated with differentiation: indeed, it has been demonstrated that differentiated pDCs do not proliferate[135]. Proliferating precursors of pDCs were only detected in the BM, whereas proliferating pre-cDCs can also be found in the spleen. BrdU labeling experiments showed that splenic pDC have a slower BrdU incorporation than splenic cDCs, which is not due to longer lifespan of pDCs, but to the fact that they proliferate only in

the BM and then migrate to the spleen.

E2-2 is considered a master regulator of the pDC lineage, and a positive feedback loop that promotes E2-2 isoform expression and pDC differentiation has been recently described[136]. Unfortunately, a gene set of E2-2 targets is not available in the MSigDB, thus lacking an important node in the regulatory network of upregulated clusters. However, other E-box binding proteins, with target sequences closely related to E2-2, are found highly enriched in upregulated clusters and not others; this is indicative of a regulation by structurally related proteins, which can include E2-2.

Many binding motifs for multifunctional, redundant TFs (e.g. Sp1, AP1 and Myc) were found enriched in both upregulated and downregulated clusters. Among these non specific factors, a noticeable enrichment for targets of the Forkhead box (Fox) family of transcription factors could be observed. These transcriptional repressors are known to regulate differentiation as well as function of T and B cells[115, 117, 118, 120–123, 127], and one of them, Foxo1, has also been characterized in DCs[114, 116, 119]. Considering that their targets were both upregulated and downregulated, it could be hypothesized that different members of this family define different DC lineages, by suppressing opposing transcriptional programs to determine a specific cell fate.

Building on this hypothesis, I analyzed expression of this family of TFs in the data set and selected those significantly upregulated for validation.

5.5 The Fox family of transcription factors: target discovery and validation

Several members of the Fox family of TFs were found to be expressed in the RNA-seq data set. Interestingly, 3 of them are also significantly upregulated with differentiation. These are Foxo1, Foxp1 and Foxr1.

Taking advantage of the Immgen database, I verified that these factors are indeed also expressed in pDCs, higher than in other DC subtypes; this is especially evident for Foxp1 and Foxr1.

This observation was validated by qPCR, where Foxp1 was found to be ex-

pressed higher in pDCs compared to all other cell types, both in BM and spleen. Unfortunately, Foxr1 was not detectable in any sample, by any of the probe sets tested.

This result, together with the availability of a Foxp1$^{flox/flox}$ mouse strain, led me to focus on this factor to perform functional validation as a potential regulator of pDC differentiation.

Indeed, its specific expression was further validated by means of FACS staining with a specific antibody, that showed presence of the protein in mature pDCs both in BM and in the spleen.

In addition, BM cells from a Foxp1$^{flox/flox}$ mouse were used to generate a Foxp1-KO progenitor cell line. This progenitor cell line showed no significant alteration compared to its parental line or to the same cell line transduced with control retrovirus and treated in the same manner, both Foxp1 sufficient.

But when differentiated in the presence of Flt3L only, in order to generate pDCs, the Foxp1-KO cells were unable to survive and died quickly, independently of the Flt3L concentration. Their survival, expansion and differentiation into pDCs was rescued by the addition of M-CSF or GM-CSF.

As Flt3 expression was unchanged in Foxp1-KO cells, this result suggests an important role of Foxp1 downstream of Flt3 signaling, in regulating the survival of Flt3 dependent DCs, already at very early stages of differentiation.

It has been shown in T cells that Foxp1 is regulated by the PI3K/Akt/mTOR pathway, and it provides a negative feedback loop to maintain quiescence[123] as well as promote survival[122]. Flt3 signaling is mainly dependent on Stat3 activity, but it has also been shown to activate the PI3K/mTOR pathway in DCs[137] and in hematological malignancies, where it promotes cell growth and survival[138]. An organ-specific role has been described for PI3Kγ (specifically the catalitic subunit p110γ) in mediating pDC and cDC responsiveness to Flt3L in the lung[139]. In addition, pDCs showed higher mTOR expression compared to other DC subsets, and are therefore less sensitive to rapamycin, expanding in response to the combination of rapamycin and Flt3L, contrary to cDCs[140]. This is in contrast with the negative feedback loop of PI3K that Foxp1 activates in T cells, mentioned above, suggesting a different, pDC specific pathway. It can be hypothesized that in pDCs

PI3K/Akt/mTOR induced Foxp1 is required for Flt3-mediated survival, and in its absence the apoptotic pathway activated by Akt-induced Foxo proteins prevails and leads to rapid cell death (as observed in T cells[122]).

Interestingly, both M-CSF and GM-CSF activate the PI3K/Akt/mTOR pathway to promote survival[64, 141]: as these growth factors are able to rescue the viability of Foxp1-KO precursors, it is clear that Foxp1 function is restricted to the Flt3 pathway. Indeed, DC subsets that depend on GM-CSF for differentiation express lower levels of Foxp1, confirming the specificity of this TF for Flt3L-dependent pDC differentiation.

More research is necessary to dissect the precise mechanisms by which Foxp1 regulates survival of Flt3-dependent cells, as well as its pDC-specific role in differentiation.

5.6 TLR stimulation and cell fate decisions

In order to address the third aim of this project, to dissect the effects of TLR stimulation on the transcriptional networks of the different cell types, a powerful analysis tool was required to extract valuable information from a highly complex data set, i.e. 3 cell types treated with 3 different ligands and 3 time points.

This was provided by the Weighted Gene Co-expression Network Analysis. This algorithm has been successfully employed in the analysis of large-scale data sets in genomics and proteomics, given its high resistance to noise and power in identifying groups of strictly co-regulated genes (called modules). Applications include oncology[142], plant biology[143], neuroimmunology[144] and many others (a Medline search for "WGCNA" retrieves 243 publications to date), but it has not been employed to analyze the response of different immune cells so far.

The analysis was applied to each cell type separately, to highlight the different networks of response to stimuli, and the results compared. This is the first comprehensive RNA seq analysis showing the response of murine pDCs to TLR7 and TLR9 stimulation. Previous studies using microarrays have shown that differential response programs are activated by different stimuli (TLR9 ligand vs influenza virus[145]). Here, limited differences could be observed between TLR7 and

9 stimulation in the induction of *de novo* RNA synthesis, that do not highlight any significant functional difference, whereas a core inflammatory response was induced by all stimuli in both pDCs and CCR9^{low} cells, highly enriched for targets of Irf proteins, in line with observations made with microarray analysis of a human pDC line, CAL-1, stimulated with TLR9 ligands[146].

As already seen in the exploratory analysis, pDCs and CCR9^{low} precursors show high functional similarity, and this is once again confirmed by the WGCNA results: comparing the two populations, modules with similar patterns of expression also share similar functional enrichment, indicating a common pathway of response to TLR stimuli. This responsiveness and full functionality of CCR9^{low} cells was already shown using type I IFN protein detection by ELISA[32]. The transcriptome analysis suggests that the CCR9^{low} population could be more heterogeneous, and contain an immature subpopulation of pDCs expressing type I IFNs as strongly or even more strongly than CCR9^{high} pDCs, which is mixed with pre-DCs. Siglec-H was shown to downmodulate the type I IFN response[147]: higher expression of Siglec-H in CCR9^{high} pDCs could explain why they produce less type I IFN than CCR9^{low} cells.

On the other hand, pre-DCs show a different behavior, with fewer genes being co-regulated in smaller modules and differentially expressed after treatment. Interestingly, some of the upregulated pre-DC modules show functional enrichment for activation pathways typical of pDCs, and genes involved in cell cycle and DNA replication, a hallmark of progenitor cells, are significantly co-repressed. This could suggest that although these cells are committed to the cDC lineage and seemingly not responsive to TLR7 and 9 stimuli, a small portion of them is able to specifically mount a response, with the consequence that they lose their progenitor phenotype.

This observation is further supported by analysis of signature enrichment in the modules: indeed, in pre-DCs modules that are enriched for the pDC signature are also upregulated upon stimulation, as well as modules containing inflammatory genes. This stimulus induced differentiation is even more evident in CCR9^{low} precursors, which strongly upregulated modules with inflammatory genes belonging to the pDC signature, and downregulated cDC and CDP genes. Mature pDCs instead, only upregulated a set of inflammatory genes, that belongs to the cDC signature (as previously observed), but were otherwise unchanged.

Regulatory network analysis of these modules also highlighted the involvement of specific inflammatory TFs in pre-DCs, consistent with a specific response to the stimuli, as well as downregulation of E2 family target genes, which are involved in cell cycle and replication, an hallmark of pDC differentiation.

Interestingly, CCR9^{low} progenitors downregulated a vast set of genes including proliferation genes, but were also enriched for targets of members of the Fox family, once again suggesting an involvement of these factors not only in steady state, but also in stimulus-induced differentiation.

5.7 Concluding remarks and future perspectives

This project, through the use of RNA sequencing and statistical analysis, identified networks of TFs regulating pDC differentiation in the BM, and highlighted a central role for the TF Foxp1 in maintaining viability of Flt3L-dependent DCs, to allow differentiation. More research in this direction is required to dissect the mechanisms by which Foxp1 regulates DC differentiation, for example by using a DC specific Foxp1 KO mouse model (CD11c-Cre x Foxp1$^{flox/flox}$). Mice lacking Foxp1 in DCs will allow the evaluation of cDC and pDC differentiation dynamics *in vivo*, as well as of the effects of systemic TLR activation.

By analyzing DC precursor populations stimulated with TLR7 and 9 ligands, I found evidence for stimulus-induced differentiation of CCR9^{low} precursors into pDCs, as well as an unexpected responsiveness of pre-DCs, which could be explained by the heterogeneity of this population, that could include cells not completely committed to the cDC lineage. To shed light upon the heterogeneity of these populations, single-cell RNA-seq can be used, which allows observation of the transcriptional dynamics of individual cells, thus permitting the identification of early lineage- and function-defining events, as well as the presence of small subsets of differently behaving cells, that are not possible to discriminate by means of surface markers and are obviously masked in bulk analyses.

Single cell transcriptome analysis is also required to resolve if the Siglec-H^{+} CCR9^{low} populations contains separate subsets of pDC-committed and cDC-committed cells, or whether they are precursors with dual potential and gradual commitment

to pDCs along with the upregulation of markers of mature pDCs (B220, CCR9). In addition, the technique can also be applied to stimulated cells, and their transcriptomes compared to steady state cells to evaluate whether responses can drive specific differentiation, and whether this happens by reprogramming of differently committed cells, or by specific selection of cells already meant to respond specifically at early stages.

In conclusion, this work used powerful statistical tools to analyze the vast amount of data generated by total RNA sequencing, extracting valuable and statistically relevant information that led to the identification of a plausible target that regulates DC differentiation with subset-specific functions, and of the functional connections between different stages of pDC differentiation.

6 Summary

Plasmacytoid dendritic cells (pDCs) are fundamental players in antiviral immune responses, readily producing large amounts of Type I IFN upon encounter with TLR7 and 9 ligands.

They originate in the BM, differentiating from common DC progenitors, in the myeloid lineage, although a contribution of lymphoid progenitors has also been identified. A late pDC precursor retaining cDC differentiation potential was recently discovered, characterized by lower expression of the chemokine receptor CCR9, but already possessing the ability to respond to viral challenge like fully differentiated pDCs.

The molecular mechanisms that drive cell fate decisions and lineage determination in the BM are still unclear, and although several critical TFs have been identified that shape DC subtype identity and function, very little is known about how fate-defining events are organized, and whether and by which mechanisms pathogens themselves can alter this process.

In this project, mature pDCs and precursor populations were discriminated in healthy murine BM by expression of cell surface markers and were sorted to high purity. The mature, Siglec-H^+ $CCR9^{high}$ $B220^{high}$ pDCs were compared by transcriptome analysis (total mRNA sequencing) to their immediate precursors, Siglec-H^+ $CCR9^{low}$ cells, and to a population of Siglec-H^- $MHCII^-$ $Sirp\alpha^-$ pre-DCs, commonly known for their commitment to the cDC lineage. In addition their response to TLR7 and 9 ligands was investigated to evaluate the effects of activation on the differentiation process.

An exploratory analysis of the unfiltered RNA seq data set emphasized the high similarity between pDCs and $CCR9^{low}$ precursors, phenotypically and functionally, and the clear separation of these populations from the Siglec-H negative pre-DC pop-

ulation. TLR stimulation also showed varying effects on the cell populations, further confirming the similarities between mature cells and CCR9^{low} precursors, but also suggesting a process of activation-induced differentiation in the latter population.

Clusters of genes differentially expressed at the steady state highlighted the transitory processes from the CCR9^{low} precursor to the pDC, which involve down-regulation of cell cycle and replication genes, and upregulation of inflammatory and pDC specific transcripts. Enrichment analysis for transcription factor (TF) binding sites showed a complex network of regulatory factors which are involved in pDC differentiation and function. The binding sites of the family of Forkhead box (Fox) transcription factors, were strongly enriched in differentially expressed gene clusters and the expression pattern of several Fox family members correlated with pDC differentiation leading to the hypothesis that they might be involved in allowing pDC differentiation and defining functional identity.

Validation experiments highlighted Foxp1 as a promising candidate with specificity for pDCs. Evaluation of DC generation from a Foxp1-knockout Hoxb8- immortalized progenitor cell line revealed a central role of this TF in DC development

References

1. Beutler, B. Innate immunity: an overview. *Mol Immunol* **40,** 845–859 (2004).
2. Murray, P. J. *et al.* Macrophage activation and polarization: nomenclature and experimental guidelines. *Immunity* **41,** 14–20 (2014).
3. Lavin, Y. & Merad, M. Macrophages: gatekeepers of tissue integrity. *Cancer Immunol Res* **1,** 201–209 (2013).
4. Mayadas, T. N., Cullere, X. & Lowell, C. A. The multifaceted functions of neutrophils. *Annu Rev Pathol* **9,** 181–218 (2014).
5. Zipfel, P. F. & Skerka, C. Complement regulators and inhibitory proteins. *Nat Rev Immunol* **9,** 729–740 (2009).
6. Garred, P. *et al.* A journey through the lectin pathway of complement-MBL and beyond. *Immunol Rev* **274,** 74–97 (2016).
7. Bottazzi, B., Doni, A., Garlanda, C. & Mantovani, A. An integrated view of humoral innate immunity: pentraxins as a paradigm. *Annu Rev Immunol* **28,** 157–183 (2010).
8. Xu, Y. *et al.* Macrophages transfer antigens to dendritic cells by releasing exosomes containing dead-cell-associated antigens partially through a ceramide-dependent pathway to enhance CD4(+) T-cell responses. *Immunology* **149,** 157–171 (2016).
9. Steinman, R. M. & Cohn, Z. A. Identification of a novel cell type in peripheral lymphoid organs of mice. I. Morphology, quantitation, tissue distribution. *J Exp Med* **137,** 1142–1162 (1973).
10. Steinman, R. M. & Cohn, Z. A. Identification of a novel cell type in peripheral lymphoid organs of mice. II. Functional properties in vitro. *J Exp Med* **139,** 380–397 (1974).
11. Steinman, R. M., Lustig, D. S. & Cohn, Z. A. Identification of a novel cell type in peripheral lymphoid organs of mice. III. Functional properties in vivo. *J Exp Med* **139,** 1431–1445 (1974).
12. Steinman, R. M., Adams, J. C. & Cohn, Z. A. Identification of a novel cell type in peripheral lymphoid organs of mice. IV. Identification and distribution in mouse spleen. *J Exp Med* **141,** 804–820 (1975).
13. Idoyaga, J., Suda, N., Suda, K., Park, C. G. & Steinman, R. M. Antibody to Langerin/CD207 localizes large numbers of CD8alpha+ dendritic cells to the marginal zone of mouse spleen. *Proc Natl Acad Sci USA* **106,** 1524–1529 (2009).

14. Dudziak, D. *et al.* Differential antigen processing by dendritic cell subsets in vivo. *Science* **315,** 107–111 (2007).

15. Shortman, K. & Liu, Y.-J. Mouse and human dendritic cell subtypes. *Nat Rev Immunol* **2,** 151–161 (2002).

16. Allan, R. S. *et al.* Migratory dendritic cells transfer antigen to a lymph node-resident dendritic cell population for efficient CTL priming. *Immunity* **25,** 153–162 (2006).

17. Schuler, G., Romani, N. & Steinman, R. M. A comparison of murine epidermal Langerhans cells with spleen dendritic cells. *J Invest Dermatol* **85,** 99s–106s (1985).

18. Merad, M., Ginhoux, F. & Collin, M. Origin, homeostasis and function of Langerhans cells and other langerin-expressing dendritic cells. *Nat Rev Immunol* **8,** 935–947 (2008).

19. Bedoui, S. *et al.* Cross-presentation of viral and self antigens by skin-derived CD103+ dendritic cells. *Nat Immunol* **10,** 488–495 (2009).

20. Bogunovic, M. *et al.* Origin of the lamina propria dendritic cell network. *Immunity* **31,** 513–525 (2009).

21. Varol, C. *et al.* Intestinal lamina propria dendritic cell subsets have different origin and functions. *Immunity* **31,** 502–512 (2009).

22. Ginhoux, F. *et al.* The origin and development of nonlymphoid tissue CD103+ DCs. *J Exp Med* **206,** 3115–3130 (2009).

23. Becker, M. *et al.* Ontogenic, Phenotypic, and Functional Characterization of XCR1(+) Dendritic Cells Leads to a Consistent Classification of Intestinal Dendritic Cells Based on the Expression of XCR1 and SIRPα. *Front Immunol* **5,** 326 (2014).

24. Gurka, S., Hartung, E., Becker, M. & Kroczek, R. A. Mouse Conventional Dendritic Cells Can be Universally Classified Based on the Mutually Exclusive Expression of XCR1 and SIRPα. *Front Immunol* **6,** 35 (2015).

25. Murphy, T. L. *et al.* Transcriptional Control of Dendritic Cell Development. *Annu Rev Immunol* **34,** 93–119 (2016).

26. Facchetti, F. *et al.* Plasmacytoid monocytes (so-called plasmacytoid T cells) in granulomatous lymphadenitis. *Hum Pathol* **20,** 588–593 (1989).

27. Fitzgerald-Bocarsly, P., Feldman, M., Mendelsohn, M., Curl, S. & Lopez, C. Human mononuclear cells which produce interferon-alpha during NK(HSV-FS) assays are HLA-DR positive cells distinct from cytolytic natural killer effectors. *J Leukoc Biol* **43,** 323–334 (1988).

28. Liu, Y.-J. IPC: professional type 1 interferon-producing cells and plasmacytoid dendritic cell precursors. *Annu Rev Immunol* **23,** 275–306 (2005).

29. Swiecki, M. & Colonna, M. The multifaceted biology of plasmacytoid dendritic cells. *Nat Rev Immunol* **15,** 471–485 (2015).

30. Hochrein, H. *et al.* Differential production of IL-12, IFN-alpha, and IFN-gamma by mouse dendritic cell subsets. *J Immunol* **166,** 5448–5455 (2001).

31. Villadangos, J. A. & Young, L. Antigen-presentation properties of plasmacytoid dendritic cells. *Immunity* **29,** 352–361 (2008).

32. Schlitzer, A. *et al.* Identification of CCR9- murine plasmacytoid DC precursors with plasticity to differentiate into conventional DCs. *Blood* **117,** 6562–6570 (2011).

33. Björck, P., Leong, H. X. & Engleman, E. G. Plasmacytoid dendritic cell dichotomy: identification of IFN-α producing cells as a phenotypically and functionally distinct subset. *J Immunol* **186,** 1477–1485 (2011).

34. Niederquell, M. *et al.* Sca-1 expression defines developmental stages of mouse pDCs that show functional heterogeneity in the endosomal but not lysosomal TLR9 response. *Eur J Immunol* **43,** 2993–3005 (2013).

35. Kamogawa-Schifter, Y. *et al.* Ly49Q defines 2 pDC subsets in mice. *Blood* **105,** 2787–2792 (2005).

36. Fang, W.-n., Shi, M., Meng, C.-y., Li, D.-d. & Peng, J.-p. The Balance between Conventional DCs and Plasmacytoid DCs Is Pivotal for Immunological Tolerance during Pregnancy in the Mouse. *Sci Rep* **6,** 26984 (2016).

37. Ouabed, A. *et al.* Differential control of T regulatory cell proliferation and suppressive activity by mature plasmacytoid versus conventional spleen dendritic cells. *J Immunol* **180,** 5862–5870 (2008).

38. Sharma, M. D. *et al.* Plasmacytoid dendritic cells from mouse tumor-draining lymph nodes directly activate mature Tregs via indoleamine 2,3-dioxygenase. *J Clin Invest* **117,** 2570–2582 (2007).

39. Irla, M. *et al.* MHC class II-restricted antigen presentation by plasmacytoid dendritic cells inhibits T cell-mediated autoimmunity. *J Exp Med* **207,** 1891–1905 (2010).

40. Nestle, F. O. *et al.* Plasmacytoid predendritic cells initiate psoriasis through interferon-alpha production. *J Exp Med* **202,** 135–143 (2005).

41. Webster, B., Assil, S. & Dreux, M. Cell-Cell Sensing of Viral Infection by Plasmacytoid Dendritic Cells. *J Virol* **90,** 10050–10053 (2016).

42. Lund, J. M., Linehan, M. M., Iijima, N. & Iwasaki, A. Cutting Edge: Plasmacytoid dendritic cells provide innate immune protection against mucosal viral infection in situ. *J Immunol* **177,** 7510–7514 (2006).

43. Swiecki, M., Gilfillan, S., Vermi, W., Wang, Y. & Colonna, M. Plasmacytoid dendritic cell ablation impacts early interferon responses and antiviral NK and CD8(+) T cell accrual. *Immunity* **33,** 955–966 (2010).

44. Miller, E. & Bhardwaj, N. Dendritic cell dysregulation during HIV-1 infection. *Immunol Rev* **254,** 170–189 (2013).

45. Collin, M. & Milne, P. Langerhans cell origin and regulation. *Curr Opin Hematol* **23,** 28–35 (2016).

46. Manz, M. G., Traver, D., Miyamoto, T., Weissman, I. L. & Akashi, K. Dendritic cell potentials of early lymphoid and myeloid progenitors. *Blood* **97,** 3333–3341 (2001).

47. Fogg, D. K. *et al.* A clonogenic bone marrow progenitor specific for macrophages and dendritic cells. *Science* **311,** 83–87 (2006).

48. Onai, N. *et al.* Identification of clonogenic common Flt3+M-CSFR+ plasmacytoid and conventional dendritic cell progenitors in mouse bone marrow. *Nat Immunol* **8,** 1207–1216 (2007).

49. Naik, S. H. *et al.* Development of plasmacytoid and conventional dendritic cell subtypes from single precursor cells derived in vitro and in vivo. *Nat Immunol* **8,** 1217–1226 (2007).

50. Grajales-Reyes, G. E. *et al.* Batf3 maintains autoactivation of Irf8 for commitment of a CD8α(+) conventional DC clonogenic progenitor. *Nat Immunol* **16,** 708–717 (2015).

51. Liu, K. *et al.* In vivo analysis of dendritic cell development and homeostasis. *Science* **324,** 392–397 (2009).

52. Schlitzer, A. *et al.* Identification of cDC1- and cDC2-committed DC progenitors reveals early lineage priming at the common DC progenitor stage in the bone marrow. *Nat Immunol* **16,** 718–728 (2015).

53. Schlitzer, A. *et al.* Tissue-specific differentiation of a circulating CCR9- pDC-like common dendritic cell precursor. *Blood* **119,** 6063–6071 (2012).

54. Chen, Y.-L. *et al.* A type I IFN-Flt3 ligand axis augments plasmacytoid dendritic cell development from common lymphoid progenitors. *J Exp Med* **210,** 2515–2522 (2013).

55. Adolfsson, J. *et al.* Upregulation of Flt3 expression within the bone marrow Lin(-)Sca1(+)c-kit(+) stem cell compartment is accompanied by loss of self-renewal capacity. *Immunity* **15,** 659–669 (2001).

56. Boyer, S. W., Schroeder, A. V., Smith-Berdan, S. & Forsberg, E. C. All hematopoietic cells develop from hematopoietic stem cells through Flk2/Flt3-positive progenitor cells. *Cell Stem Cell* **9,** 64–73 (2011).

57. Karsunky, H., Merad, M., Cozzio, A., Weissman, I. L. & Manz, M. G. Flt3 ligand regulates dendritic cell development from Flt3+ lymphoid and myeloid-committed progenitors to Flt3+ dendritic cells in vivo. *J Exp Med* **198,** 305–313 (2003).

58. 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* **31,** 563–604 (2013).

59. McKenna, H. J. *et al.* Mice lacking flt3 ligand have deficient hematopoiesis affecting hematopoietic progenitor cells, dendritic cells, and natural killer cells. *Blood* **95,** 3489–3497 (2000).

60. Maraskovsky, E. *et al.* In vivo generation of human dendritic cell subsets by Flt3 ligand. *Blood* **96,** 878–884 (2000).

61. Manfra, D. J. *et al.* Conditional expression of murine Flt3 ligand leads to expansion of multiple dendritic cell subsets in peripheral blood and tissues of transgenic mice. *J Immunol* **170,** 2843–2852 (2003).

62. Waskow, C. *et al.* The receptor tyrosine kinase Flt3 is required for dendritic cell development in peripheral lymphoid tissues. *Nat Immunol* **9,** 676–683 (2008).

63. Zhan, Y., Xu, Y. & Lew, A. M. The regulation of the development and function of dendritic cell subsets by GM-CSF: more than a hematopoietic growth factor. *Mol Immunol* **52,** 30–37 (2012).

64. Van de Laar, L., Coffer, P. J. & Woltman, A. M. Regulation of dendritic cell development by GM-CSF: molecular control and implications for immune homeostasis and therapy. *Blood* **119,** 3383–3393 (2012).

65. Kingston, D. *et al.* The concerted action of GM-CSF and Flt3-ligand on in vivo dendritic cell homeostasis. *Blood* **114,** 835–843 (2009).

66. Inaba, K. *et al.* Generation of large numbers of dendritic cells from mouse bone marrow cultures supplemented with granulocyte/macrophage colony-stimulating factor. *J Exp Med* **176,** 1693–1702 (1992).

67. Greter, M. *et al.* GM-CSF controls nonlymphoid tissue dendritic cell homeostasis but is dispensable for the differentiation of inflammatory dendritic cells. *Immunity* **36,** 1031–1046 (2012).

68. Schmid, M. A., Kingston, D., Boddupalli, S. & Manz, M. G. Instructive cytokine signals in dendritic cell lineage commitment. *Immunol Rev* **234,** 32–44 (2010).

69. Onai, N. *et al.* A clonogenic progenitor with prominent plasmacytoid dendritic cell developmental potential. *Immunity* **38,** 943–957 (2013).

70. Fancke, B., Suter, M., Hochrein, H. & O'Keeffe, M. M-CSF: a novel plasmacytoid and conventional dendritic cell poietin. *Blood* **111,** 150–159 (2008).

71. Laffont, S., Seillet, C. & Guéry, J.-C. Estrogen Receptor-Dependent Regulation of Dendritic Cell Development and Function. *Front Immunol* **8,** 108 (2017).

72. Carreras, E. *et al.* Estradiol acts directly on bone marrow myeloid progenitors to differentially regulate GM-CSF or Flt3 ligand-mediated dendritic cell differentiation. *J Immunol* **180,** 727–738 (2008).

73. Seillet, C. *et al.* Estradiol promotes functional responses in inflammatory and steady-state dendritic cells through differential requirement for activation function-1 of estrogen receptor α. *J Immunol* **190,** 5459–5470 (2013).

74. Kueh, H. Y. *et al.* Positive feedback between PU.1 and the cell cycle controls myeloid differentiation. *Science* **341,** 670–673 (2013).

75. Satpathy, A. T. *et al.* Runx1 and Cbfβ regulate the development of Flt3+ dendritic cell progenitors and restrict myeloproliferative disorder. *Blood* **123,** 2968–2977 (2014).

76. Sichien, D. *et al.* IRF8 Transcription Factor Controls Survival and Function of Terminally Differentiated Conventional and Plasmacytoid Dendritic Cells, Respectively. *Immunity* **45,** 626–640 (2016).

77. Satpathy, A. T. *et al.* Zbtb46 expression distinguishes classical dendritic cells and their committed progenitors from other immune lineages. *J Exp Med* **209,** 1135–1152 (2012).

78. Meredith, M. M. *et al.* Expression of the zinc finger transcription factor zDC (Zbtb46, Btbd4) defines the classical dendritic cell lineage. *J Exp Med* **209,** 1153–1165 (2012).

79. Kashiwada, M., Pham, N.-L. L., Pewe, L. L., Harty, J. T. & Rothman, P. B. NFIL3/E4BP4 is a key transcription factor for CD8α+ dendritic cell development. *Blood* **117,** 6193–6197 (2011).

80. Hacker, C. *et al.* Transcriptional profiling identifies Id2 function in dendritic cell development. *Nat Immunol* **4,** 380–386 (2003).

81. Cisse, B. *et al.* Transcription factor E2-2 is an essential and specific regulator of plasmacytoid dendritic cell development. *Cell* **135,** 37–48 (2008).

82. Ghosh, H. S., Cisse, B., Bunin, A., Lewis, K. L. & Reizis, B. Continuous expression of the transcription factor e2-2 maintains the cell fate of mature plasmacytoid dendritic cells. *Immunity* **33,** 905–916 (2010).

83. Ghosh, H. S. *et al.* ETO family protein Mtg16 regulates the balance of dendritic cell subsets by repressing Id2. *J Exp Med* **211,** 1623–1635 (2014).

84. Wu, X. *et al.* Transcription factor Zeb2 regulates commitment to plasmacytoid dendritic cell and monocyte fate. *Proc Natl Acad Sci USA* **113,** 14775–14780 (2016).

85. Scott, C. L. *et al.* The transcription factor Zeb2 regulates development of conventional and plasmacytoid DCs by repressing Id2. *J Exp Med* **213,** 897–911 (2016).

86. Dursun, E. *et al.* Continuous single cell imaging reveals sequential steps of plasmacytoid dendritic cell development from common dendritic cell progenitors. *Sci Rep* **6,** 37462 (2016).

87. Yáñez, A., Goodridge, H. S., Gozalbo, D. & Gil, M. L. TLRs control hematopoiesis during infection. *Eur J Immunol* **43,** 2526–2533 (2013).

88. Nagai, Y. *et al.* Toll-like receptors on hematopoietic progenitor cells stimulate innate immune system replenishment. *Immunity* **24,** 801–812 (2006).

89. Welner, R. S. *et al.* Lymphoid precursors are directed to produce dendritic cells as a result of TLR9 ligation during herpes infection. *Blood* **112,** 3753–3761 (2008).

90. Schmid, M. A., Takizawa, H., Baumjohann, D. R., Saito, Y. & Manz, M. G. Bone marrow dendritic cell progenitors sense pathogens via Toll-like receptors and subsequently migrate to inflamed lymph nodes. *Blood* **118,** 4829–4840 (2011).

91. Jackson, J. T. *et al.* Id2 expression delineates differential checkpoints in the genetic program of CD8α+ and CD103+ dendritic cell lineages. *EMBO J* **30,** 2690–2704 (2011).

92. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods* **10,** 1096–1098 (2013).

93. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9,** 171–181 (2014).

94. Redecke, V. *et al.* Hematopoietic progenitor cell lines with myeloid and lymphoid potential. *Nat Methods* **10,** 795–803 (2013).

95. R Core Team. *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing (Vienna, Austria, 2016). `https://www.R-project.org`.

96. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).

97. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4,** Article17 (2005).

98. Langfelder, P. & Horvath, S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol* **1,** 54 (2007).

99. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9,** 559 (2008).

100. Langfelder, P. & Horvath, S. Fast R Functions for Robust Correlations and Hierarchical Clustering. *J Stat Softw* **46** (2012).

101. Pandey, G., Cohain, A., Miller, J. & Merad, M. Decoding dendritic cell function through module and network analysis. *Journal of Immunol Methods* **387,** 71–80 (2013).

102. Hu, J. X., Thomas, C. E. & Brunak, S. Network biology concepts in complex disease comorbidities. *Nat Rev Genet* **17,** 615–629 (2016).

103. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102,** 15545–15550 (2005).

104. Xie, X. *et al.* Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434,** 338–345 (2005).

105. Shen, L. & Sinai, M. *GeneOverlap: Test and visualize gene overlaps* R package version 1.10.0 (2013). `http://shenlab-sinai.github.io/shenlab-sinai/`.

106. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13,** 2498–2504 (2003).

107. Christmas, R. *et al.* in *Am Assoc Cancer Res Education Book 2005* 12–16 (2005).

108. Cline, M. S. *et al.* Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* **2,** 2366–2382 (2007).

109. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS ONE* **5,** e13984 (2010).

110. Ahmed, M. *et al.* Ability of the matrix protein of vesicular stomatitis virus to suppress beta interferon gene expression is genetically correlated with the inhibition of host RNA and protein synthesis. *J Virol* **77,** 4646–4657 (2003).

111. Vollmer, J. *et al.* Characterization of three CpG oligodeoxynucleotide classes with distinct immunostimulatory activities. *Eur J Immunol* **34,** 251–262 (2004).

112. Carbon, S. *et al.* AmiGO: online access to ontology and annotation data. *Bioinformatics* **25,** 288–289 (2009).

113. Miller, J. C. *et al.* Deciphering the transcriptional network of the dendritic cell lineage. *Nat Immunol* **13,** 888–899 (2012).

114. Dong, G. *et al.* FOXO1 regulates dendritic cell activity through ICAM-1 and CCR7. *J Immunol* **194,** 3745–3755 (2015).

115. Fabre, S. *et al.* Stable activation of phosphatidylinositol 3-kinase in the T cell immunological synapse stimulates Akt signaling to FoxO1 nuclear exclusion and cell growth control. *J Immunol* **174,** 4161–4171 (2005).

116. Riol-Blanco, L. *et al.* Immunological synapse formation inhibits, via NF-kappaB and FOXO1, the apoptosis of dendritic cells. *Nat Immunol* **10,** 753–760 (2009).

117. Kerdiles, Y. M. *et al.* Foxo1 links homing and survival of naive T cells by regulating L-selectin, CCR7 and interleukin 7 receptor. *Nat Immunol* **10,** 176–184 (2009).

118. Hess Michelini, R., Doedens, A. L., Goldrath, A. W. & Hedrick, S. M. Differentiation of CD8 memory T cells depends on Foxo1. *J Exp Med* **210,** 1189–1200 (2013).

119. Sukhbaatar, N., Hengstschläger, M. & Weichhart, T. mTOR-Mediated Regulation of Dendritic Cell Differentiation and Function. *Trends Immunol* **37,** 778–789 (2016).

120. Van Keimpema, M. *et al.* FOXP1 directly represses transcription of proapoptotic genes and cooperates with NF-κB to promote survival of human B cells. *Blood* **124,** 3431–3440 (2014).

121. Van Keimpema, M. *et al.* The forkhead transcription factor FOXP1 represses human plasma cell differentiation. *Blood* **126,** 2098–2109 (2015).

122. Van Boxtel, R. *et al.* FOXP1 acts through a negative feedback loop to suppress FOXO-induced apoptosis. *Cell Death Differ* **20,** 1219–1229 (2013).

123. Wei, H. *et al.* Cutting Edge: Foxp1 Controls Naive CD8+ T Cell Quiescence by Simultaneously Repressing Key Pathways in Cellular Metabolism and Cell Cycle Progression. *J Immunol* **196,** 3537–3541 (2016).

124. Miyazaki, M. *et al.* The E-Id protein axis modulates the activities of the PI3K-AKT-mTORC1-Hif1a and c-myc/p19Arf pathways to suppress innate variant TFH cell development, thymocyte expansion, and lymphomagenesis. *Genes Dev* **29,** 409–425 (2015).

125. Santo, E. E. *et al.* Oncogenic activation of FOXR1 by 11q23 intrachromosomal deletion-fusions in neuroblastoma. *Oncogene* **31,** 1571–1581 (2012).

126. Pommerenke, C. *et al.* Chromosome 11q23 aberrations activating FOXR1 in B-cell lymphoma. *Blood Cancer J* **6,** e433 (2016).

127. Durek, P. *et al.* Epigenomic Profiling of Human CD4(+) T Cells Supports a Linear Differentiation Model and Highlights Molecular Regulators of Memory Development. *Immunity* **45,** 1148–1161 (2016).

128. Conesa, A. *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17,** 13 (2016).

129. Loschko, J. *et al.* Antigen targeting to plasmacytoid dendritic cells via Siglec-H inhibits Th cell-dependent autoimmunity. *J Immunol* **187,** 6346–6356 (2011).

130. Chopin, M. *et al.* RUNX2 Mediates Plasmacytoid Dendritic Cell Egress from the Bone Marrow and Controls Viral Immunity. *Cell Rep* **15,** 866–878 (2016).

131. Sawai, C. M. *et al.* Transcription factor Runx2 controls the development and migration of plasmacytoid dendritic cells. *J Exp Med* **210,** 2151–2159 (2013).

132. Buechler, M. B., Teal, T. H., Elkon, K. B. & Hamerman, J. A. Cutting edge: Type I IFN drives emergency myelopoiesis and peripheral myeloid expansion during chronic TLR7 signaling. *J Immunol* **190,** 886–891 (2013).

133. Buechler, M. B., Gessay, G. M., Srivastava, S., Campbell, D. J. & Hamerman, J. A. Hematopoietic and nonhematopoietic cells promote Type I interferon- and TLR7-dependent monocytosis during low-dose LCMV infection. *Eur J Immunol* **45,** 3064–3072 (2015).

134. Watchmaker, P. B. *et al.* Comparative transcriptional and functional profiling defines conserved programs of intestinal DC differentiation in humans and mice. *Nat Immunol* **15,** 98–108 (2014).

135. Zhan, Y. *et al.* Plasmacytoid dendritic cells are short-lived: reappraising the influence of migration, genetic factors and activation on estimation of lifespan. *Sci Rep* **6,** 25060 (2016).

136. Grajkowska, L. T. *et al.* Isoform-Specific Expression and Feedback Regulation of E Protein TCF4 Control Dendritic Cell Lineage Specification. *Immunity* **46,** 65–77 (2017).

137. Sathaliyawala, T. *et al.* Mammalian target of rapamycin controls dendritic cell development downstream of Flt3 ligand signaling. *Immunity* **33,** 597–606 (2010).

138. Lagunas-Rangel, F. A. & Chávez-Valencia, V. FLT3-ITD and its current role in acute myeloid leukaemia. *Med Oncol* **34,** 114 (2017).

139. Nobs, S. P. *et al.* PI3-Kinase-γ Has a Distinct and Essential Role in Lung-Specific Dendritic Cell Development. *Immunity* **43,** 674–689 (2015).

140. Biswas, M. *et al.* Synergy between rapamycin and FLT3 ligand enhances plasmacytoid dendritic cell-dependent induction of CD4+CD25+FoxP3+ Treg. *Blood* **125,** 2937–2947 (2015).

141. Bradley, E. W., Ruan, M. M., Vrable, A. & Oursler, M. J. Pathway crosstalk between Ras/Raf and PI3K in promotion of M-CSF-induced MEK/ERK-mediated osteoclast survival. *J Cell Biochem* **104,** 1439–1451 (2008).

142. Zhang, C. *et al.* The identification of key genes and pathways in hepatocellular carcinoma by bioinformatics analysis of high-throughput data. *Med Oncol* **34,** 101 (2017).

143. Fu, F., Zhang, W., Li, Y.-Y. & Wang, H. L. Establishment of the model system between phytochemicals and gene expression profiles in Macrosclereid cells of Medicago truncatula. *Sci Rep* **7,** 2580 (2017).

144. Gao, C. *et al.* Inflammatory and apoptotic remodeling in autonomic nervous system following myocardial infarction. *PLoS ONE* **12,** e0177750 (2017).

145. Iparraguirre, A. *et al.* Two distinct activation states of plasmacytoid dendritic cells induced by influenza virus and CpG 1826 oligonucleotide. *J Leukoc Biol* **83,** 610–620 (2008).

146. Steinhagen, F. *et al.* IRF5 and IRF8 modulate the CAL-1 human plasmacytoid dendritic cell line response following TLR9 ligation. *Eur J Immunol* **46,** 647–655 (2016).

147. Blasius, A. L., Cella, M., Maldonado, J., Takai, T. & Colonna, M. Siglec-H is an IPC-specific receptor that modulates type I IFN secretion through DAP12. *Blood* **107,** 2474–2476 (2006).

Acknowledgments

A first, sincere thank you to Prof. Dr. Anne Krug, a brilliant supervisor who inspired me and gave me the invaluable opportunity to learn a great deal of new things and carry out this ambitious project, through its highs and lows, with unrelenting patience and decisive support over the last 4 years.

Thank you to our collaborators, Lynette Henkel and Dr. Matthias Schiemann for the sorting, Christoph Ziegenhein and Prof. Wolfgang Enard for performing the library preparations and being of constant support with the sequencing experiments, Prof. Helmut Blum at LAFUGA, Gene Center, for performing the sequencing, and Dr. Tobias Straub, for the helpful discussions of the bioinformatics analyses and all statistics-related issues.

I also thank all the current and former members of the AG Krug, Alexander, Katharina, Mona, Yvonne, Mahulena, for their help and useful criticism, discussions and suggestions during the endless labseminars; a special mention to Ana-Marija, Anamarija, Lisa, Rebecca and Elena, for being friends as well as colleagues.

A special heartfelt thank you to Ezgi Dursun-Rüttgers, a great teacher and mentor, but most of all a friend, who always supported and advised me during these years.

I am immensely grateful to all the friends and colleagues that I have met in these 4 years as a PhD student, inside and outside of the lab.

Un grazie enorme ovviamente va alla famiglia, perché la mamma è sempre la mamma, ma anche il papà è sempre il papà e la sorella è sempre la sorella, quindi non potrei essere dove sono senza un po' di ciascuno.

Curriculum Vitae

Personal details

	Andrea Musumeci
Date of birth	April 22nd, 1987
Place of birth	Sassuolo (MO), Italy
Current address	Steinhauserstraße 15, 81677 Munich

Scientific education

from 08.2013	Ph.D. thesis in the laboratory of Prof. Dr. Anne Krug, initially II. Med. Klinik, Klinikum rechts der Isar, TUM, then Institut für Immunologie, Ludwig Maximilians Universität Munich
12.2011–07.2013	Research assistant in the laboratory of Dr. Vincenzo Russo, Cancer Gene Therapy unit, Università Vita-Salute San Raffaele, Milano, Italy
10.2009–11.2011	Master of Science in Medical, Molecular and Cellular Biotechnology at Università Vita-Salute San Raffaele, Milano, Italy. Final grade 110/110 *cum laude*.
10.2006–10.2009	Bachelor of Science in Medical and Pharmaceutical Biotechnology at Università Vita-Salute San Raffaele, Milano, Italy. Final grade 110/110.

Education

2001–2006	High School, Liceo Classico Statale L.A. Muratori, Modena, Italy
1998–2001	Middle School, Scuole Medie Lanfranco, Modena, Italy
1993–1998	Elementary School, Scuola Elementare Buon Pastore, Modena, Italy

Original publications

Schuh E., **Musumeci A.**, Thaler FS., Laurent S., Ellwart JW., Hohlfeld R., Krug A., Meinl E., Human Plasmacytoid Dendritic Cells Display and Shed B Cell Maturation Antigen upon TLR Engagement. *J Immunol* **8**, 3081-3088 (2017), doi: 10.4049/jimmunol.1601746

Dursun E., Endele M., **Musumeci A.**, Failmezger H., Wang SH., Tresch A., Schroeder T., Krug AB.; Continuous single cell imaging reveals sequential steps of plasmacytoid dendritic cell development from common dendritic cell progenitors. *Sci Rep* **6**, 37462 (2016), doi: 10.1038/srep37462

Lanterna C., **Musumeci A.**, Raccosta L., Corna G., Moresco M., Maggioni D., Fontana R., Doglioni C., Bordignon C., Traversari C., Russo V.; The administration of drugs inhibiting cholesterol/oxysterol synthesis is safe and increases the efficacy of immunotherapeutic regimens in tumor-bearing mice. *Cancer Immunol Immunother* **11**, 1303-1315 (2016), doi: 10.1007/s00262-016-1884-8

Raccosta L., Fontana R., Maggioni D., Lanterna C., Villablanca EJ., Paniccia, A., **Musumeci A.**, Chiricozzi E., Trincavelli ML., Daniele S., Martini C., Gustafsson JA., Doglioni C., Feo SG., Leiva A., Ciampa MG., Mauri L., Sensi C., Prinetti A., Eberini I., Mora JR., Bordignon C., Steffensen KR., Sonnino S., Sozzani S., Traversari C., Russo V.; The oxysterol-CXCR2 axis plays a key role in the recruitment of tumor-promoting neutrophils. *J Exp Med* **9**, 1711-1728 (2013), doi: 10.1084/jem.20130440

Congress Abstracts

11th European Network of Immunology Institutes (ENII) Summer School of Advanced Immunology, Porto Cervo, Italy (07.05.2016-14.05.2016)
Andrea Musumeci, Ezgi Dursun, Anne Krug. The transcriptional landscape of plasmacytoid dendritic cells differentiation. (poster)

Club Francophone des Cellules Dendritiques (CFCD) annual meeting: "DC ontology, functional specialization and targeting", Paris, France (07.12.2015-08.12.2015)
Andrea Musumeci*, Ezgi Dursun*, Max Endele, Henrik Failmezger, Achim Tresch, Gabrielle Belz, Timm Schroeder, Anne Krug. DC progenitor cells in murine bone marrow are heterogeneous and precommitted to generate pDCs or cDCs. (oral presentation) - *Best oral presentation award*

European Molecular Biology Organization (EMBO) Workshop: Unravelling Biological Secrets by Single-Cell Expression Profiling, Heidelberg, Germany (25.09.2014-26.09.2014)
Andrea Musumeci, Ezgi Dursun, Max Endele, Timm Schroeder, Anne Krug. Unraveling DC precursor cells heterogeneity and early commitment at the single cell level. (poster)

München, 17.10.2017

Appendix A

R scripts for RNA-seq data analysis

List of R scripts

Script 1: DESeq2

Functions for the analysis of differential expression on steady state samples, using the R package `DESeq2`. Refers to methods section 3.3.2.

Script 1: `DESeq2`: differential expression analysis

```
library(DESeq2)

## total read counts as input
reads=read.delim("AGKrug_counts_rmdup.txt")

## select only steady state samples
reads=reads[, grep("T0", colnames(reads))]
reads=reads[-grep("gERCC", rownames(reads)),]
reads=reads[, c(7:9, 1:3, 4:6)]            # reorder samples

## experiment design
colData <- data.frame(cell.type= rep(c("preDC", "CCR9", "pDC"),
                                     each=3),
                      row.names = colnames(reads))

## build dataset
dataset <- DESeqDataSetFromMatrix(countData = reads,
                                  colData = colData,
                                  design = ~ cell.type)

datasetFULL <- DESeq(dataset)              # GLM model of the data

## LRT for multiple comparisons
datasetLRT <- DESeq(dataset, test = "LRT", reduced = ~1)

resLRT <- results(datasetLRT)

## select differentially expressed genes (p value < 0.01)
DEgenes <- resLRT[!is.na(resLRT$padj),]
idx <- DEgenes$padj < 0.01
DEgenes <- DEgenes[idx,]

## heatmap of DE genes
library(Biobase)
load("WGCNA (final analysis)/steady state (DESeq2)/eset_mod_T0.rda")
                    # load the expression set
DEnames = rownames(DEgenes)
flt=featureData(eset_T0)$ensemblID %in% DEnames   #filter for DE genes
eset.f=eset_T0[flt,]
n=length(DEnames)
source("scripts/heatmap.R")                # load the heatmap function
pdf("DESeq2.DEgenes.pdf", 4, n/10)         # save to file
heatmap.ts(eset.f)
dev.off()
```

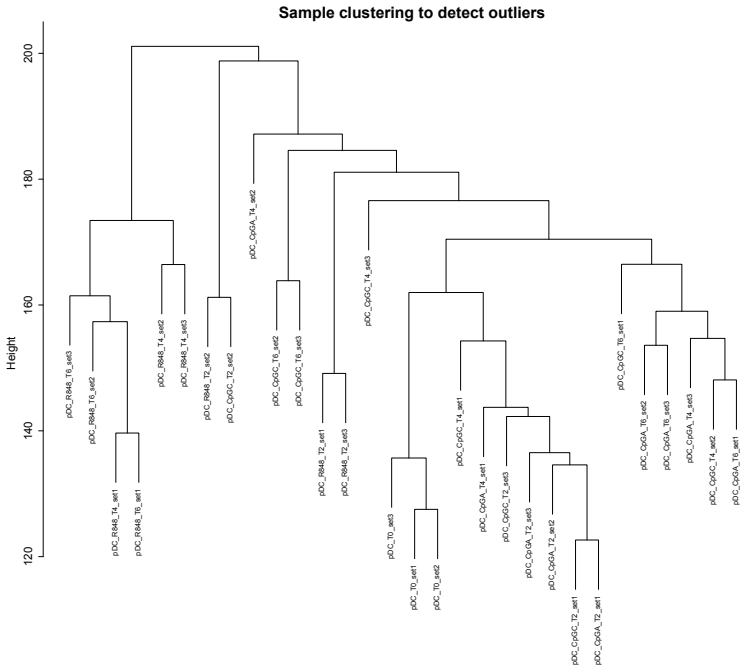
The resulting heatmap is presented in figure 4.12.

Script 2: WGCNA

Functions for the co-expression analysis on stimulated samples using the package `WGCNA`, by cell type. Refers to methods section 3.3.3. The example shown and the relative images are from the pDC data set.

Script 2: `WGCNA`: co-expression analysis

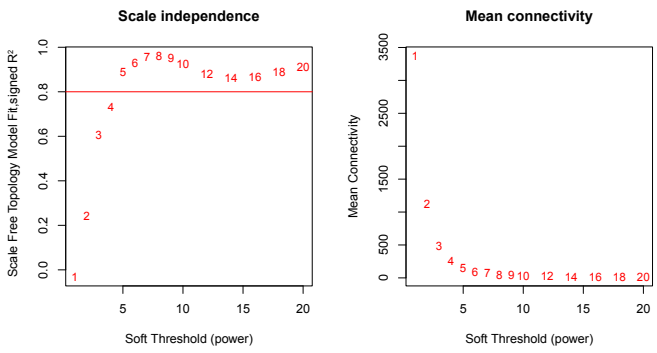
```
library(WGCNA)
library(Biobase)
options(stringsAsFactors=F)

load("eset.rda")  #load complete expression set

cell <- "pDC"      #define cell type to analyze (preDC, CCR) or pDC)
eset_0 <- eset[, phenoData(eset)$cell.type == cell]

datExpr0 <- t(exprs(eset_0))

## check for genes and samples with too many missing values
gsg <- goodSamplesGenes(datExpr0, verbose=3)
gsg$allOK          #if TRUE, proceed without filtering

## cluster samples to detect outliers
sampleTree <- hclust(dist(datExpr0), method="average")
pdf(file = paste("sampleTree_", cell, ".pdf", sep = ""), 12,9)
par(cex=0.6)
par(mar=c(0,4,2,0))
plot(sampleTree, main="Sample clustering to detect outliers",
     sub="", xlab="", cex.lab=1.5, cex.axis=1.5, cex.main=2)
dev.off()
```



Sample clustering to detect outliers.

The clustering does not detect outliers. We proceed without filtering.

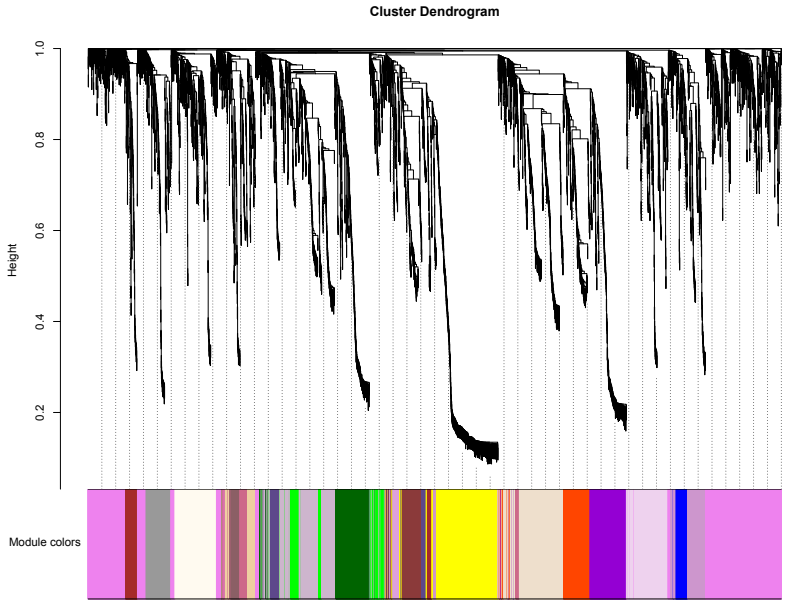
```
## if no outliers detected, proceed without filtering
datExpr <- datExpr0

## load the traits to correlate with the modules, Treatment and Time
datTraits <- data.frame(pData(eset_0)[,2:3])
rownames(datTraits) <- rownames(datExpr)

## Automatic blockwise network construction
enableWGCNAThreads()  #multithreading for faster computing

networkType <- "signed hybrid" # the type of network to be generated

## pick soft thersholding power
powers <- c(c(1:10), seq(from=12, to=20, by=2))
sft <- pickSoftThreshold(datExpr, powerVector=powers,
                         verbose=5, networkType = networkType)

pdf(file = paste("SoftThreshold.", cell, ".pdf", sep = ""), 9,5)
par(mfrow=c(1,2))
cex1 <- 0.9
plot(sft$fitIndices[,1],
     -sign(sft$fitIndices[,3])*sft$fitIndices[,2],
     xlab="SoftThreshold (power)",
     ylab="Scale Free Topology Model Fit, signed R^2",
     type="n",
     main=paste("Scale independence"))
text(sft$fitIndices[,1],
     -sign(sft$fitIndices[,3])*sft$fitIndices[,2],
     labels=powers,
     cex=cex1,
     col="red")
abline(h=0.9, col="red")
plot(sft$fitIndices[,1], sft$fitIndices[,5],
     xlab="SoftThreshold (power)",
     ylab="Mean Connectivity",
     type="n",
     main=paste("Mean Connectivity"))
text(sft$fitIndices[,1], sft$fitIndices[,5],
     labels=powers, cex=cex1, col="red")
dev.off()
```



Selection of the appropriate β value.

The fit indices indicate that an appropriate soft threshold power for this data set is 5. It is selected independently for each data set.

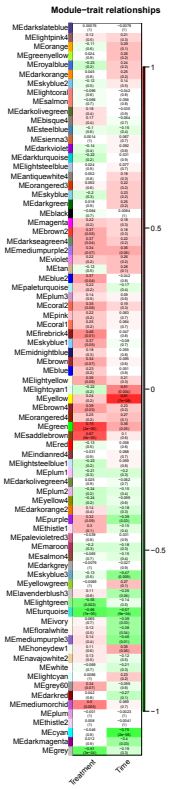
```
## choose appropriate power based on fit indices
softPower <- 5

## blockwise network construction
net <- blockwiseModules(datExpr, maxBlockSize=5000,
                        power=softPower,
                        TOMType = "unsigned",
                        minModuleSize=30,
                        reassignThreshold=0,
                        mergeCutHeight=0.25,
                        networkType = "signed hybrid",
                        numericalLabels=T,
                        pamRespectDendro=F,
                        saveTOMs=T,
                        saveTOMFileBase="WGCNA_TOM",
                        verbose=3)

## plot resulting dendrogram and modules
pdf(file = paste("DendroColors.", cell, ".pdf", sep = ""), 12, 9)
mergedColors <- labels2colors(net$colors)
plotDendroAndColors(net$dendrograms[[1]],
                    mergedColors[net$blockGenes[[1]]],
                    "Module colors",
                    dendroLabels = F, hang=0.03,
                    addGuide=T, guideHang=0.05)
dev.off()
```



Cluster dendrogram and module assignment.

Modules are assigned dynamically. Those with more than 25% similarity are merged. Next the modules are related to traits.

```
## relate module eigengenes to traits
moduleLabels <- net$colors
moduleColors <- labels2colors(net$colors)
nGenes <- ncol(datExpr)      #define number of genes
nSamples <- nrow(datExpr)    #define number of samples

##calculate MEs with color labels
MEs0 <- moduleEigengenes(datExpr, moduleColors)$eigengenes
MEs <- orderMEs(MEs0)
datTraits$treatment <- as.integer(as.factor(datTraits$treatment))

## calculate correlation and p value
moduleTraitCor = cor(MEs, datTraits, use = "p")
moduleTraitPvalue = corPvalueStudent(moduleTraitCor, nSamples);

##visualize module-trait relationship
textMatrix <- paste(signif(moduleTraitCor,2),
                    "\n(", signif(moduleTraitPvalue,1), ")",
                    sep="")
dim(textMatrix) <- dim(moduleTraitCor)

pdf(file = paste("heatmap.", cell, ".pdf", sep = ""), 5, 40)
par(mar=c(6,14,5,3))
labeledHeatmap(Matrix=moduleTraitCor,
               xLabels=names(datTraits), yLabels=names(MEs),
               ySymbols=names(MEs), colorLabels=F,
               colors=greenWhiteRed(50),
               textMatrix=textMatrix, setStdMargins=F,
               cex.text=0.5, zlim=c(-1,1),
               main=paste("Module-trait relationships"))
dev.off()
```



Final result of WGCNA analysis showing all modules and their correlation to traits

Script 3: GeneOverlap

Functions for the functional analysis of clusters and modules, calculating enrichment of gene sets from the MSigDB. Refers to methods section 3.3.4. The example given here refers to cluster analysis of steady state.

Script 3: `GeneOverlap`: functional enrichment analysis

```
library(xlsx)
library(GeneOverlap)

db <- "c3.tft"   #collection to interrogate (c2.kegg or c3.tft)

## select the cluster or module names to analyze
mods <- c("Clust1", "Clust2", "Clust3", "Clust4", "Clust5", "Clust6")

## load collection
load(paste("../", db, ".Mm.R", sep = ""))
x <- get(db)
gs <- 29353 #total number of genes in the expression set, as background
```

A first `for` loop loads each cluster or module from the specified excel file as a list of member genes and sends it to the second nested `for` loop, which calculates overlap with each gene set in the collection and saves the result in a data frame. This is filtered for significance and finally saved to an excel file, ending the loop.

```
## first main loop
for(mod in mods){
list <- read.xlsx(file = "newclusterlist_T0.xlsx",
                  sheetName = mod, header = T)
list <- as.character(na.omit(list[, "name"]))

result <- data.frame(row.names = mod)

## second nested loop
for (i in 1:length(x)) {
  go.obj <- newGeneOverlap(list, as.character(x[[i]]@geneIds),
                           genome.size = gs)
  go.obj <- testGeneOverlap(go.obj)
  set <- as.character(x[[i]]@setName)
  result[, set] <- getPval(go.obj)
  }   # end nested loop

## select significant results and organize output
result <- data.frame(t(result[,result < 0.05]))
result$x <- NA
result$y <- rownames(result)
result <- result[,c(3,2,1)]
colnames(result) <- c("Name", "Description", "pVal")

write.xlsx(result, file = paste("Overlap.result.", db, ".xlsx",
                                sep = ""),
           sheetName = mod, append = T, row.names = F)
}   # end main loop
```

The final output is an excel file with sheets named after each cluster or module, which contain a list of significantly enriched gene sets and their respective p value. These can be then directly used in Cytoscape for network visualization.

Appendix B

Regulatory networks of the significant modules from WGCNA

Pre-DCs

Module	Signature	Enrichment p value	*Overlapping genes*
antiquewhite4	CDP	5.68×10^{-04}	*1700026L06Rik Angptl2 Fry Klrb1f Tpcn1*
	cDC	3.83×10^{-11}	*Arrb1 Pak1 Lrrc25 1700025G04Rik Nostrin Rab11fip1 Mctp1 Gm6377 Eps8 St3gal5 Tspan33 Wdfy3*
darkgreen	pDC	5.92×10^{-47}	*Mctp2 Cmah Slc41a2 Ablim1 Ctsl Siglech Ccr9 Pltp Rpgrip1 Pacsin1 Klra17 Gas7 Blnk Fcrla Tcf4 Rabgap1l Sla2 Cox6a2 Serinc5 Sema4b Paqr5 Lrp8 Eepd1 Maged1 Timp2 Abca5 Dirc2 Scrn1 Spib Slc37a2 Tom1 Cd300c Fyn Tex2 Nucb2 Ldlrad3 Calcoco1 Maml2*
green	CDP	2.37×10^{-18}	*Uhrf1 Mcm10 Arhgap19 Lig1 Shcbp1 Hes6 Cenpn Stmn1 Mcm2 Cdc25c Pole2 E2f8 Rfc4 Spag5 Plec Prim1 Sun1 Tcf19 Nasp Racgap1 Rasgrp2 Mybl2 Ube2t Spc24 Cdc45 Cdt1 Dnajc9 Kif23 Mcm5 Suv39h1 Ncapg2 Stil Lmnb1*
	cDC	0.010	*Lpcat4 Pglyrp1 Cdca7 E2f8 Egr1 Gatm Id2 Kif23 Lfng Mllt4 Myc Pole Rrm2*
lavenderblush2	cDC	3.06×10^{-07}	*Asb2 Bcl2a1b Cd80 Cd83 Cd86 Cxcl16 Il1b Spred1*
lightcoral	pDC	9.55×10^{-05}	*Grm8 Cdc14b Erbb3 Cd209d Lair1 Clec10a Adam11 Lifr Man2a2 Slc27a1 Fam3c Inppl1 Man1c1*
magenta4	pDC	0.011	*Cd8b1 Slco4a1 Stambpl1 Dst Slc39a14 Tubgcp5*
paleturquoise	pDC	0.014	*Slamf9 Pir Ly6a Ccr5 Ifi44 Tlr7*
plum3	CDP	0.016	*Atp10a Spp1 Csf1r Pik3r6*
	cDC	8.03×10^{-18}	*Adam23 Aim1 Arhgap26 Asap1 Ccrl2 Chn2 Asap2 Dock5 Piezo1 Gm11545 Grk5 Ifitm2 Igsf6 Lpcat2 Ptpn22 Rab32 Rnf150 Sgk1 Tgfbi*
yellow	cDC	0.008	*Pid1 Lrrk2 Lrrk2 Lyz1 Nav1 Nlrp1c-ps Rtn1 Slamf8 Tlr13 Zbtb46 Zfp366*

Module	Signature	Enrichment p value	*Overlapping genes*
yellow4	CDP	1.87×10^{-27}	*Fut7 Dok3 Bub1b Dlgap5 Kif2c Ska3 Ncaph Plk1 Ect2 Psat1 Knstrn Cit Nek2 Aurkb Ccnb2 Plk4 Cep55 Spc25 Mad2l1 Rrm1 Aurka Ccnf Tbc1d31 Prc1 Hacd4 Gpsm2 Efcab11 Pif1 Zranb3 Kif4 Melk Clic1 Card9 Tpx2 Arhgef39*
	cDC	5.39×10^{-13}	*2810417H13Rik Bub1 Cep55 Dlgap5 Hnrnpll Melk Ncapg Anpep Atp8b4 Bub1b Ccna2 Ccnb2 Dna2 Hk2 Kif11 Mki67 Nuf2 Rab31 Ralb Rnf144b Sulf2 Tnfrsf21 Top2a Tpx2*

Table B.1: Genes found in signature-enriched modules of pre-DCs, grouped by signature

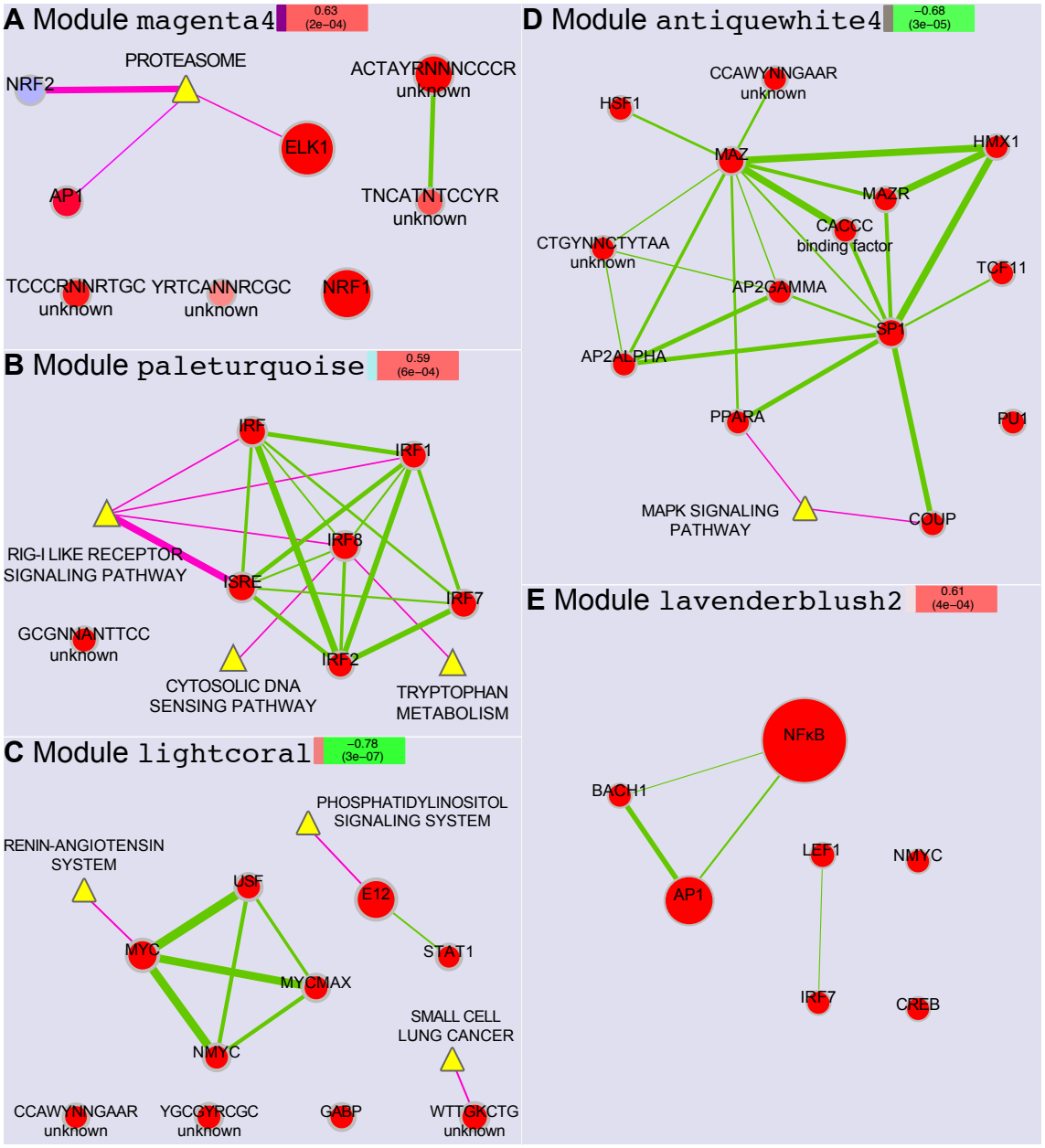


Figure B.1: Regulatory networks of the significant modules of pre-DCs. Transcription factors and KEGG pathways significantly enriched in the indicated modules.

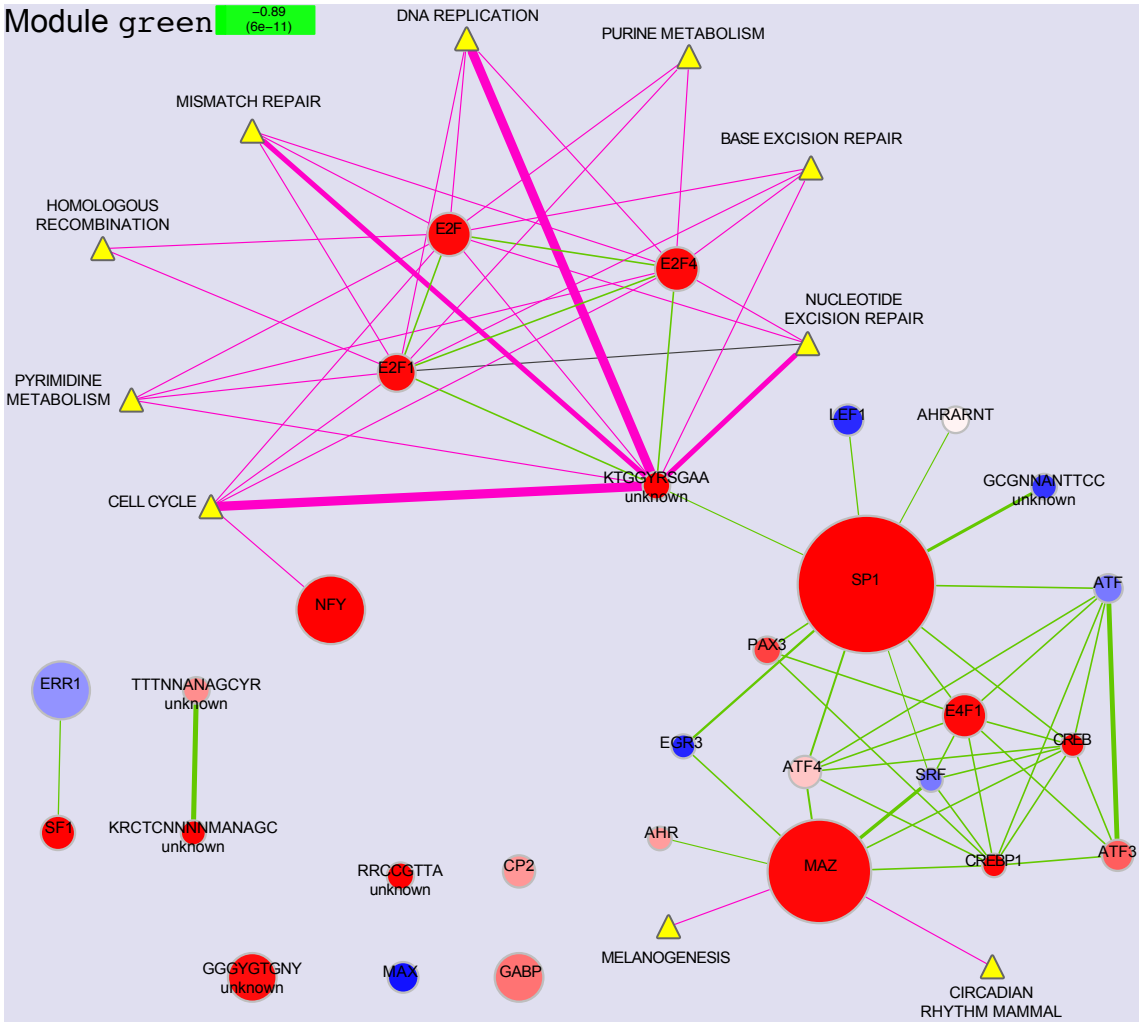


Figure B.2: Regulatory network of the pre-DC module `green`. Transcription factors and KEGG pathways significantly enriched in the `green` module of preDCs.

CCR9^{low} precursors

Module	Signature	Enrichment p value	*Overlapping genes*
antiquewhite2	cDC	9.07×10^{-05}	*Arhgap26 Asap1 Atf3 Bcl2a1c Cd83 Cxcl16 Il1b Neurl3 Rasgef1b Spred1*
	pDC	0.038	*Abca5 Rpgrip1 Blnk Dst Asph*
green	cDC	0.047	*Aim1 Ccrl2 Fgl2 Grk5 Lrrk2 Lrrk2 Nlrp1b Parp12 Pstpip2 Ptpn22 Sgk1*
	pDC	0.014	*Lrp8 Ifi44 Cd8b1 Ly6a Klra17 St3gal6 Tlr7 Fyn Arhgef10 Fam3c*
lightslateblue	cDC	0.008	*Asap2 Gatm*
mediumpurple1	pDC	7.89×10^{-05}	*Slc41a2 Ablim1 Atp1b1 Pacsin1 Cox6a2 Stambpl1 Dirc2 Arap1 Pdzd4 Pls3 Tex2 Calcoco1 Tubgcp5 Inppl1*
paleturquoise	cDC	0.030	*Errfi1 Gpr82 Id2 Lpcat2 Nr4a3 Tubb6*
sienna4	CDP	4.10×10^{-40}	*Fut7 Upk1b Tpcn1 Dnah8 Bub1b Uhrf1 Arhgap19 Zdhhc15 Dlgap5 Lig1 Kif2c Shcbp1 Hes6 Asf1b Ncaph Ect2 Cenpn Knstrn Nek2 Stmn1 Mcm2 Aurkb Cdc25c Ccnb2 Plk4 Cep55 Spc25 Pole2 E2f8 Mad2l1 Fry Rrm1 Cenpi Aurka Ccnf Rfc4 Spag5 Angptl2 Plec Prim1 Sun1 Tcf19 Prc1 Racgap1 Csf1r Rasgrp2 Mybl2 Ube2t Spc24 Gpsm2 Cdc45 Kif4 Melk Dnajc9 Kif23 Card9 Tpx2 Mis18a Arhgef39 Mcm5 Suv39h1 Ncapg2 Stil Lmnb1*
	cDC	2.59×10^{-09}	*2810417H13Rik Arrb1 Atp8b4 Cep55 Dlgap5 Dock5 E2f8 Egr1 Eps8 Kif11 Kif23 Melk Nostrin Pole Prkar2a Rab31 Rab32 Ralb Bub1 Bub1b Ccna2 Ccnb2 Cdca7 Ear2 Mki67 Mllt4 Ncapg Nuf2 Pak1 Rab11fip1 Rnf150 Rrm2 Top2a Tpx2 Wdfy3*
	pDC	2.64×10^{-09}	*Il7r Rabgap1l Clcn5 Cd209d Tom1 Nucb2 Ptprs Sep-01 Mctp2 Cmah Atp2a1 Gas7 Lair1 Mgl2 Paqr5 Clec10a Adam11 Timp2 Lifr Man2a2 Pnck Rab33b Ldlrad3 Man1c1 Grm8 Siglech Ptprf Fcrla Slc27a1 Cd300c Slc39a14*
skyblue1	cDC	0.015	*Adam8 Havcr2 Myc Zbtb46*
steelblue	CDP	0.030	*S1pr3 Mcm10 Ccne2 Klrb1f Ska3 Chtf18 Raph1 1700026L06Rik*

Module	Signature	Enrichment p value	*Overlapping genes*
	cDC	0.002	*1700025G04Rik Asb2 Chn2 Dock7 Piezo1 Gm6377 Hnrnpll Lfng Lyz1 Pglyrp1 Rtn1 St3gal5*

Table B.2: Genes found in signature-enriched modules of CCR9^{low} precursors, grouped by signature

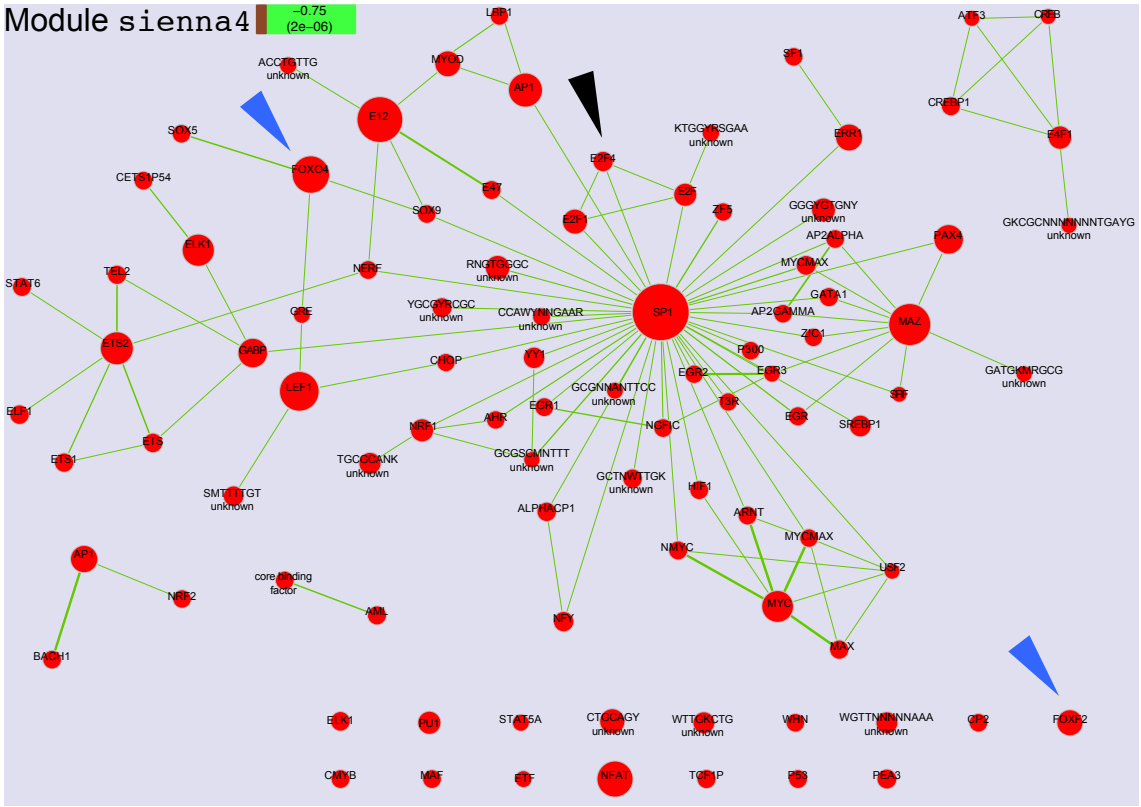


Figure B.3: Regulatory network of the CCR9^{low} module `sienna4`. Transcription factors significantly enriched in the `sienna4` module. Blue arrows highlight FOX family TFs, the black arrow highlights a E2F subnetwork.

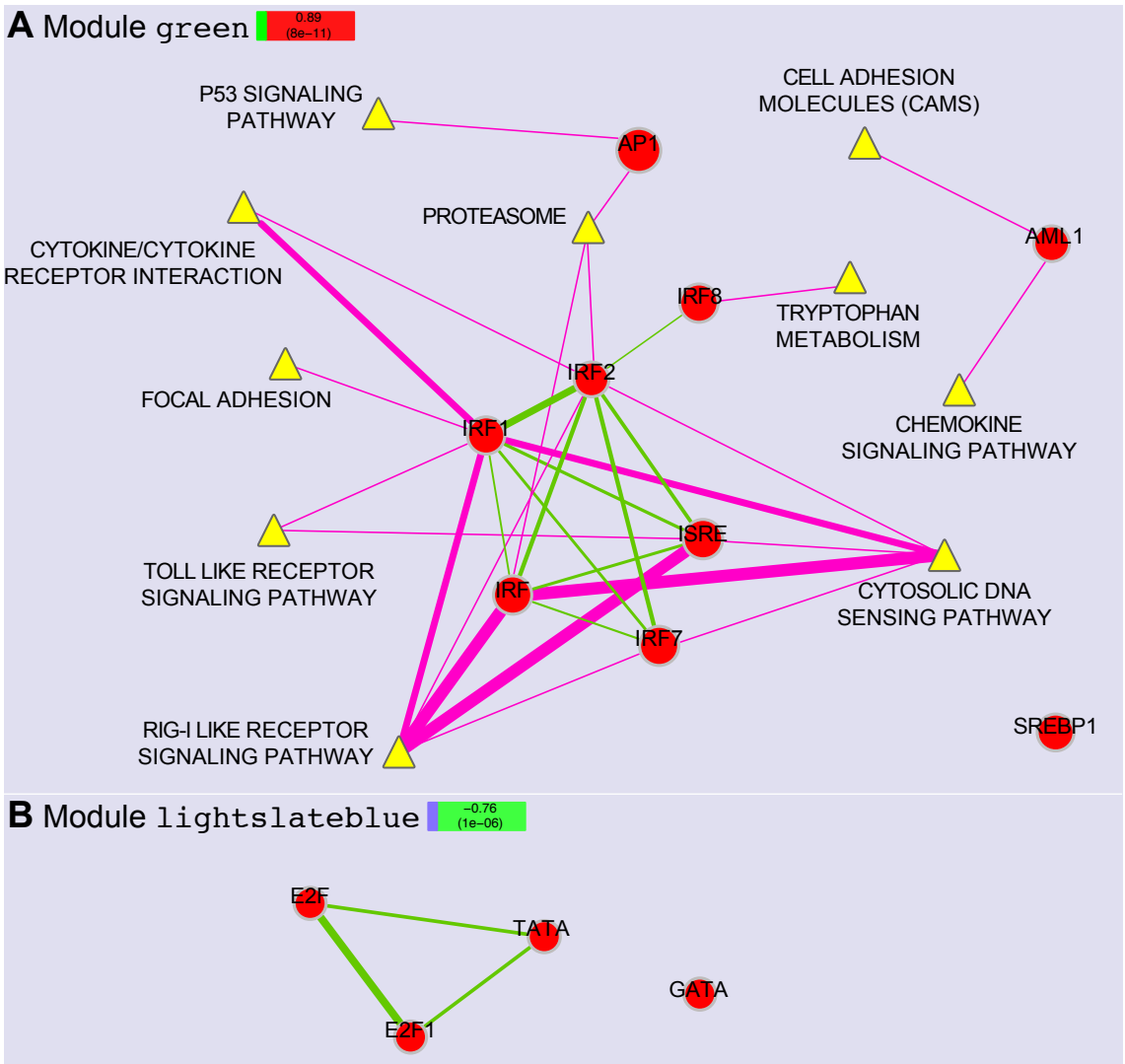


Figure B.4: Regulatory network of the significant modules of CCR9^{low} cells. Transcription factors and KEGG pathways significantly enriched in the indicated modules.

pDCs

Module	Signature	Enrichment p value	*Overlapping genes*
coral2	CDP	1.55×10^{-05}	*Upk1b Zdhhc15 Psat1 Cenpn Csf1r Fam129b Kif23*
	cDC	8.92×10^{-21}	*Anpep Arhgap26 Atp8b4 Cd9 Cdca7 Errfi1 Piezo1 Ifitm2 Kif23 Lrrk2 Lrrk2 Mctp1 Neurl3 Pak1 Prkar2a Ptpn22 Rab31 Rab32 Ralb Sgk1 Tlr11*
cyan	cDC	4.04×10^{-09}	*Asap1 Atf3 Bcl2a1b Bcl2a1c Cd83 Cd86 Cxcl16 Il1b Rab11fip1 Rasgef1b Spred1 Tubb6 Zbtb46*
darkseagreen4	CDP	0.032	*Dnah8 S1pr3 Hacd4 Efcab11*
honeydew	cDC	0.004	*Hnrnpll Nlrp1b St3gal5 Tspan33*
indianred3	CDP	3.05×10^{-04}	*Chtf18 Pif1 Card9 Arhgef39 Ncapg2*
lightcyan1	pDC	5.21×10^{-33}	*Havcr1 Mctp2 Slamf9 Grm8 Ctsl Siglech Ccr9 Pltp Cd8b1 Rpgrip1 Ly6a Klra17 Prkca Blnk Tcf4 Lair1 Lrp8 Cobll1 Bst2 Maged1 Abca5 Spib Slc37a2 Tlr7 Notch3 Fyn Nucb2 Maml2 Fam3c*
sienna4	CDP	1.68×10^{-27}	*Tpcn1 Bub1b Uhrf1 Mcm10 Dlgap5 Lig1 Kif2c Shcbp1 Klrb1f Hes6 Ska3 Asf1b Tiam2 Ncaph Plk1 Ect2 Knstrn Cit Nek2 Stmn1 Mcm2 Aurkb Cdc25c Ccnb2 Vim Plk4 Cep55 Spc25 Pole2 E2f8 Fry Rrm1 Aurka Ccnf Rfc4 Spag5 Angptl2 Plec Prim1 Sun1 Tcf19 Prc1 Racgap1 Rasgrp2 Mybl2 Ube2t Spc24 Gpsm2 Cdc45 Kif4 Melk Clic1 B4galt4 Dnajc9 Tpx2 Mis18a Mcm5 Stil Lmnb1*
	cDC	0.003	*Anxa1 Bub1 Ccna2 Cd81 Cep55 Dlgap5 E2f8 Egr1 Eps8 Gm6377 Kif11 Melk 2810417H13Rik Bub1b Ccnb2 Lfng Mki67 Ncapg Nuf2 Pglyrp1 Pole Rrm2 Spire1 Sulf2 Tnfrsf21 Top2a Tpx2*
	pDC	1.16×10^{-04}	*Cmah Cd209d Il7r Rabgap1l Clcn5 Sema4b Egfr Slc41a2 Ptprf Mgl2 Clec10a Adam11 Timp2 Lifr Man2a2 Rab33b Ptprs Ldlrad3 Man1c1 Gas7 Arap1 Scrn1 Slc27a1 Cd300c Slc39a14 Sep-01 Inppl1*

Module	Signature	Enrichment p value	*Overlapping genes*
steelblue	pDC	2.87×10^{-18}	*Klk1 Obscn Cdh1 Ablim1 Atp1b1 Cdc14b Pacsin1 Erbb3 Atp2a1 Fcrla Sla2 Cox6a2 Serinc5 Paqr5 Eepd1 Stambpl1 Dirc2 Pdzd4 Tom1 Pnck Pls3 Tex2 Calcoco1 Tubgcp5*

Table B.3: Genes found in signature-enriched modules of pDCs, grouped by signature

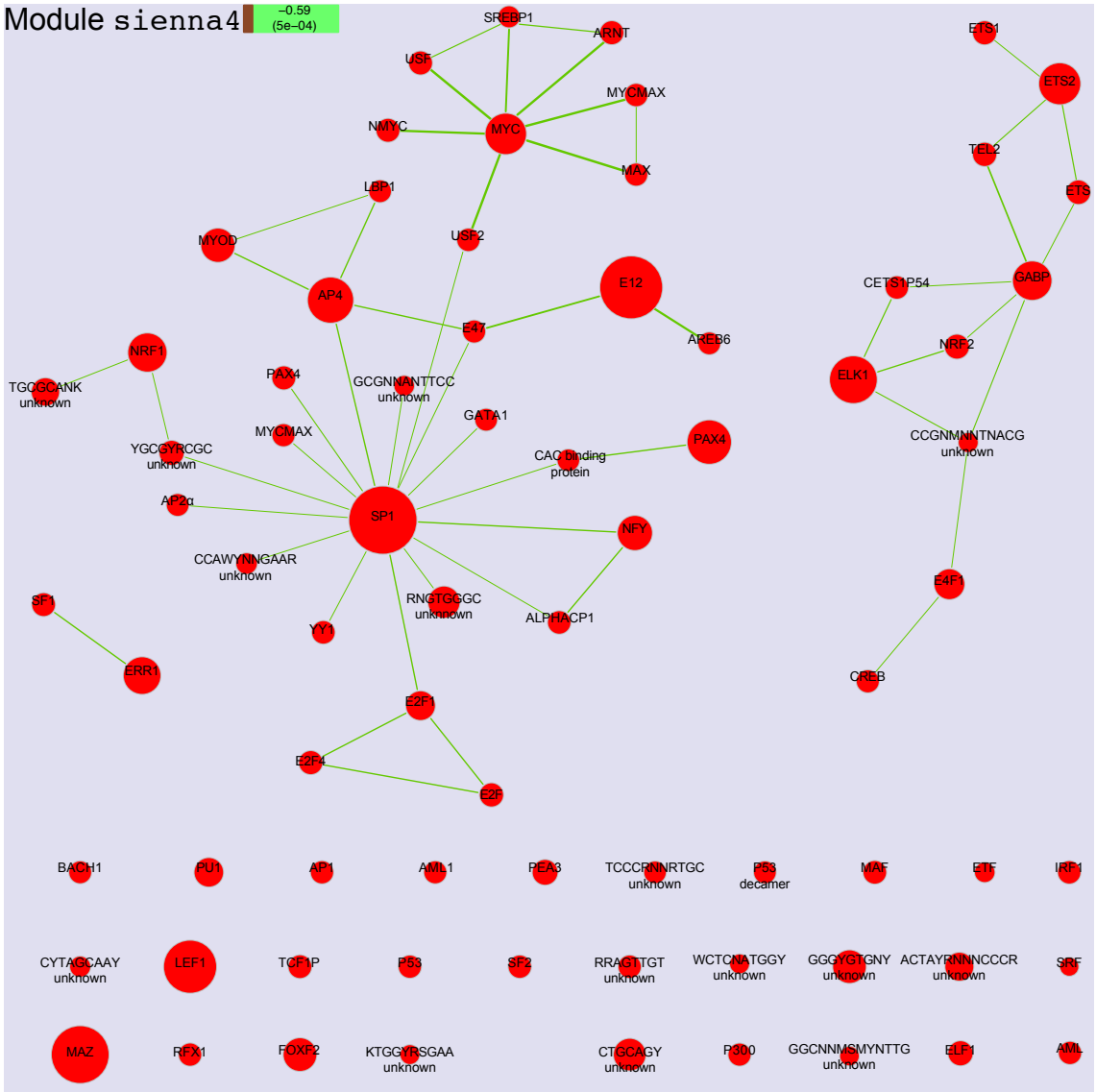


Figure B.5: Regulatory network of the pDC module `sienna4`. Transcription factors significantly enriched in the module.